A metabolomics and transcriptomics comparison of *Narcissus pseudonarcissus* cv. Carlton field and *in vitro* tissues in relation to alkaloid production

Aleya Ferdausi

The University of Liverpool

April 2017

UNIVERSITY OF LIVERPOOL

Thesis submitted in accordance with the requirements of the University of Liverpool for the degree of Doctor in Philosophy by Aleya Ferdausi

Acknowledgements

First, I would like to express my profound gratitude to my supervisor Dr Meriel Jones for her continuous support, motivation, suggestions, and guidelines to continue my project work smoothly. I am also grateful to her for giving critical evaluation on my thesis. I would also like to thank my secondary supervisor Professor Anthony Hall for his support and guidelines to lead me on the bioinformatics discovery. I would like to thank my PhD assessors Professor Martin Mortimer and Dr James Hartwell for their valuable suggestions and feedback on my annual project progress.

My sincere acknowledgements also goes to Dr Xianmin Chang for his entire help throughout the major part of my research such as providing *Narcissus* bulbs, tissue culture and alkaloid analysis method development, calculations and data interpretations. Mark Preston, Centre of Proteome analysis for helping with GC-MS analysis and Dr Phelan Marie, NMR Centre, University of Liverpool for helping with NMR analysis. Dr Ryan Joynson, for his helps regarding transcript annotation. Centre of Genomic Research, University of Liverpool, for RNA-sequencing. Dr Jane Pulman for her helps to learn basic molecular biology techniques.

My sincere thanks are given to Jean Wood, Senior Technician, Lab G, Institute of Integrative Biology. I am also grateful to all other staff, research groups, post-graduate students and all members of the Biosciences building, University of Liverpool for their friendly and humble attitude, which made me feel my work place like home.

I also wish to express special thanks to Commonwealth Scholarship Commission, UK, and University of Liverpool, UK, for funding this PhD project. I am also thankful to Bangladesh Agriculture University, Bangladesh and Government, Peoples Republic of Bangladesh for allowing me the 4-years study-leave for pursuing my PhD.

I would like to thank my friends Wazeera, Majida and Nirja. I am also grateful to my lovely pet birds Red, Blue and Chris. Special thanks to go Shams Khan for the daffodil plant drawing.

Finally, those people who are everything to me, my parents, my two little sisters, Sharmin and Azmiri, my nephew, Saifan and my brother, Shuvo; without their support and love I would not be what I am today.

With love, I dedicate my PhD thesis to my parents.

Table of contents: **Page No.**

LIST OF FIGURES

Figure Number ..... Figure legend

LIST OF TABLES

Table Number ..... Table title

ABSTRACT

The Amaryllidaceae alkaloids e.g. galanthamine (Gal), lycorine, narciclasine and pretazettine are noted for their pharmaceutical properties. The biosynthesis of alkaloids by plants using *in vitro* systems has been considered as a tool for drug discovery and production since total chemical synthesis is not economic. The biosynthetic pathways, especially for Gal, are starting to be understood, but still far from complete. This study focused on understanding biosynthesis in whole plants and developing cell culture systems that could be optimised for alkaloid production. Metabolite profiling and knowledge about probable genes involved in secondary metabolism and their regulation in different tissues will provide insight into the *Narcissus* plant biology related to alkaloid production.

In vitro cultures of *N. pseudonarcissus* cv. Carlton were initiated from twin-scale explants, cultured on Murashige and Skoog agar medium (MS) fortified with different concentrations of growth regulators. Callus was obtained mainly from MS medium containing high concentration of auxin, while media with low auxin and MS basal medium gave bulblets with both white and green shoots. Regenerated bulblets developed from callus developed in both high and low auxin MS media. These tissue culture derived materials and field-grown samples were analysed using GC-MS for Gal content. The highest amount of Gal was obtained from basal plate tissue followed by bulb, leaves, and *in vitro* tissues. A trace amount of Gal was found in callus along with possibly other alkaloids.

An NMR-based metabolomic analysis showed that the relative concentrations of compounds involved in phenylalanine and tyrosine metabolism, the initial stage biosynthetic pathways that yield the Amaryllidaceae alkaloids, were higher in field samples than *in vitro* samples. These results support the GC-MS findings of high Gal production in field samples.

Initial RT-PCR analysis using gene specific transcripts possibly involved in Gal biosynthesis i.e. P450s. *PAL*, *TYDC* and *OMT* showed their expression in *in vitro* as well in field tissues, which is intriguing. A transcriptome analysis was performed comparing expression in the highest (basal plate) versus lowest (callus) Gal containing tissues. The transcriptome analysis results were in accordance with the previous findings, as it was observed that the transcripts involved in the initial biosynthetic pathways leading to precursors (tyrosine, phenylalanine) were up regulated in callus while the transcripts involved in later pathways leading to alkaloids were up regulated in basal plate.

Abbreviations

^{o}C	Degree Centigrade	CTAB	Hexadecyltrimethylammonium bromide
%	Percentage	cv./cvs.	Cultivated variety/varieties
≥	Greater than or equal to	CYP	Cytochrome P450
>	Greater than	C3H	Coumarate-3-hydroxylase
≤	Less than or equal to	C4H	Cinnamate-4-Hydroxylase
<	Less than	D	Day
~	Equivalent to	Da	Dalton
δ	Delta	DB	Dormant bulb
β	Beta	DBS	Dormant basal plate
γ	Gamma	DBL	Direct bulblets
µg	Microgram	DE	Discovery Environment
g	Gram	DNA	Deoxyribonucleic acid
µM	Micro molar	DW	Dry Weight
µm	Micrometre	DWS	Direct white shoot
µl	Micro litre	EC	Enzyme class
µs	Micro second	e.g.	As for example
1D	One-dimensional	ERF	Ethylene response factor
2D	Two-dimensional	eV	Electron Volt
2, 4-D	2,4-Dichlorophenoxyacetic acid	ES	External standard
$^{2}H_2O$	Deuterated H_2O	FAO	Food and Agricultural Organisation
^{1}H	Protium	FPKM	Fragments mapped per kilobase of exon per million reads mapped
^{13}C	Carbon-13	FU	Fluorescence unit
^{15}N	Nitrogen-15	FW	Fresh Weight
^{31}P	Phosphorus-32	*g*	Gravitational force
AB	Andrew's Choice bulb	Gal	Galanthamine
ABS	Andrew's Choice basal plate	GAPDH	Glyceraldehydes-3-phosphate dehydrogenase
ABA	Abscisic acid	GC-MS	Gas Chromatography-Mass Spectrometry
ACT	Actin	GO	Gene ontology
AChE	Acetylcholinesterase	GS	Green shoot
AKR	Aldo-keto reductase	GTF	Gene Transfer Format
ANOVA	Analysis of Variance	GTP	Guanosine triphosphate
AP2	APETALA 2	h	Hour
ATP	Adenosine triphosphate	HCL	Hydrochloric acid
AUX	Auxin	HMDB	Human Metabolome Database
B1, B2....	Carlton bulb numbers	HPLC	High Performance-Liquid Chromatography
BA	Benzyladenine	HSD	Honest Significant Difference
BAP	Benzylaminopurine	HSQC	Heteronuclear Single Quantum Correlation
BLAST	Basic Local Alignment Search Tool	$HgCl_2$	Mercuric chloride
bp	Base pair	Hz	Hertz
BWA	Burrows-Wheeler transform	IAEA	International Atomic Energy Agency
C	Carbon	i.e.	That is
CAL	Callus	ID	Identification
CB	Carlton bulb	IAA	Indoleacetic Acid
CBS	Carlton basal plate	IBA	Indole-3-butyric acid
cDNA	Complementary DNA	Inc.	Incorporation
CGR	Centre of Genomic Research	IS	Internal standard
CH	Chitosan	KEGG	Kyoto Encyclopedia of Genes and Genomes
CoA	Coenzyme A	kg	Kilogram
Co.	Company	kHz	Kilohertz
Cod	Codeine	KN	Kinetin
cm	Centimetre	KO	KEGG Orthology
CPMG	Meiboom-Gill modification of the Carr-Purcell	kPa	Kilo Pascal

L/l	Litre	PC	Principal component
LC-MS	Liquid Chromatography-Mass Spectrometry	PCA	Principal Component Analysis
LSD	Least Significant Difference	PMT	Putrescine *N*-methyltransferase
Ltd.	Limited	POP	Polyphenol oxidase
mg	Milligram	PVP	Polyvinylpyrrolidone
min	Minute	PW	Pulse width
m	Metre	RB	Regenerated bulblet
ml	Millilitre	RD	Relaxation delay
mm	Millimetre	RIN	RNA integrity number
M	Molar	rRNA	Ribosomal RNA
MAQ	Mapping and Assembly with Quality	RNA	Ribonucleic acid
mM	Millimolar	RNA-seq	RNA sequencing
MetPA	Metabolomics Pathway Analysis	RT or rt	Retention time
MHz	Megahertz	RT-PCR	Reverse Transcriptase PCR
mRNA	Messenger RNA	RWS	Regenerated white shoot
miRNA	Micro RNA	SAM	*S*-adenosylmethionine
MS	Murashige and Skoog	SD	Standard deviation
MJ	Methyl jasmonate	SE	Standard error
ml	Millilitre	SDR	Short-chain dehydrogenase/reductase
MW	Molecular weight	s	second
m/z	Mass-to-charge ratio	SMPDB	Small Molecule Pathway Database
n	Number of observation	SNP	Single-nucleotide polymorphism
nM	Nanomolar	SOAP	Short Oligonucleotide Analysis Package
nt	Nucleotide	SOLID	Sequencing by Oligonucleotide Ligation and Detection
NAA	Naphthaleneacetic acid	sp./spp.	Species
NADP/ NADPH	Nicotinamide adenine dinucleotide phosphate-oxidase	sRNA	Small RNA
NGS	Next Generation Sequencing	T	Treatment
ng	Nanogram	TAIR	The Arabidopsis Information Resource
NCBI	National Center for Biotechnology Information	TCIN	*Trans*-cinnamic acid
NIST	National Institute of Standards and Technology	TCA cycle	Citrate cycle
NaOCl	Sodium hypochlorite	TIC	Total ion current
NaOH	Sodium hydroxide	tRNA	Transfer RNA
NaN_3	Sodium azide	TSP/TMS	Tri-methylsilyl propanoic acid
NpN4OMT/ N4OMT	Norbelladine 4'-*O*-methyltransferase	TYDC	Tyrosine decarboxylase
NMR	Nuclear magnetic resonance spectroscopy	UN	Unknown
NMT	*N*-methyltransferase	UK	United Kingdom
NOE	Nuclear Overhauser Effect	USA	United States
NHS	National Health Services	UV-B	Ultraviolet B
OH	Hydroxide	UDP	Uridine phosphorylase
OMT	*O*-methyltransferase	VC	Vitamin C
PacBio	Pacific Biosciences	www	World Wide Web
PCR	Polymerase Chain Reaction	WHO	World Health Organisation
pg	Pico gram	w/v	Weight/volume
ppm	Parts per million	YE	Yeast extract
psi	Pounds per square inch	WHO	World Health Organisation
pH	Negative logarithm of hydrogen ion	YE	Yeast extract
PAL	Phenylalanine ammonia lyase		

1 Chapter One: Background

Plants and plant cell cultures provide a broad range of commercial and industrial applications serving as resources for secondary compounds, bio-based fuel, plastics, preservatives, enzymes, aromas, flavours, fragrances and cosmetics (Anand, 2010). Due to the complex chemical structures and limited knowledge on the biosynthesis of plant secondary metabolites, their nature of production and accumulation in plants is unclear (Verpoorte *et al.,* 2002). However, new technologies such as metabolomics and transcriptomics are starting to give new insights on plant secondary metabolism (Lubbe *et al.,* 2013; Berkov *et al.,* 2014a). This project is about the application of these emerging technologies to give insight into secondary metabolism in *N. pseudonarcissus* cv. Carlton, a plant with pharmacologically interesting secondary metabolites.

1.1 The genus *Narcissus*

Narcissus is one of the 75 genera within the monocotyledon family Amaryllidaceae that includes 1100 plant species (He *et al.,* 2015). The *Narcissus* genus originates from the Mediterranean area with the centre of diversity in the Iberian Peninsula. It comprises around one hundred wild species mainly distributed in South-Western Europe and North Africa, with some populations in the Balkans, France and Italy (Ronsted *et al.,* 2008).

The UK is the world's largest producer of daffodils and *Narcissus* cut flowers. Growing of daffodils adds around £23 M/year to the economy in the UK, which also grows about half of the world's daffodil bulbs (Jones *et al*., 2011). *Narcissus* cultivars have been classified in 13 divisions with more than 27,000 distinct varieties (Royal Horticultural Society, www.rhs.org.uk).

In addition to their ornamental properties, *Narcissus* plants have been found to contain many alkaloids with pharmaceutical properties including galanthamine, (Torras-Claveria *et al.,* 2013) approved in UK to relieve some symptoms of Alzheimer's disease (NHS, 2016; www.nhs.uk).

A typical *Narcissus* plant includes an underground bulb, where leaves arise from the base of a flower stem or stalk with rounded tips with a single or multiple flowers on

the flower stem (www.kew.org) (Figure 1.1). The bulbs consist of a disc-shaped basal plate with adventitious roots (Hanks, 2002).

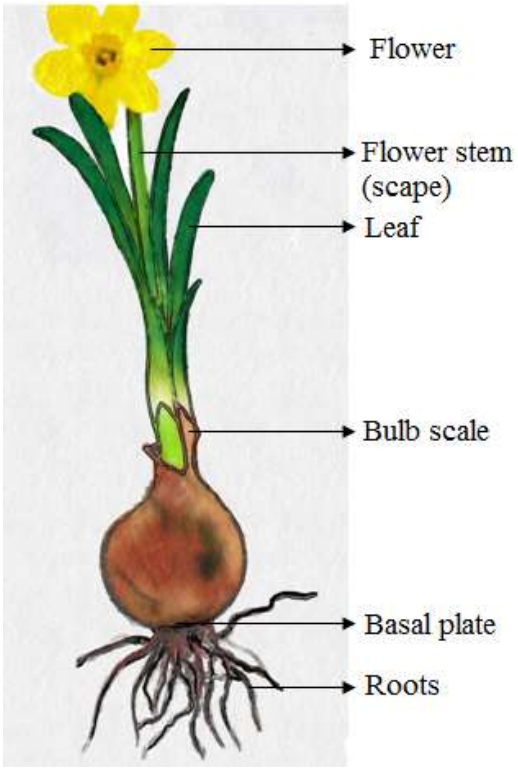

Fig. 1.1 A typical *Narcissus* (daffodil) plant showing the main plant parts (twitter.com/mindartrandom).

N. pseudonarcissus cv. Carlton is currently used for commercial production of galanthamine because of its wide commercial availability and vigour (Takos and Rook, 2013) as well as the fact that cultivation practices are well established (Lubbe *et al.,* 2011).

1.1.1 Growth cycle of *Narcissus*

Narcissus are predominantly spring-flowering perennial plants containing an underground storage organ (bulb) with a life cycle that includes a summer dormancy period for bulbs and seeds; breaking of this dormancy often requires a cold period (Takos and Rook, 2013). A cold period (4 ^{o}C to 9 ^{o}C) in autumn and winter breaks bulb dormancy and growth begins. Leaves emerge above ground in spring (Herranz *et al.,* 2013). The flower stem (scape) emerges rapidly and single or multiple flowers open. Rapid leaf senescence occurs immediately after flowering (Hanks, 2002). New bulbs form from the leaf bases and bulb scales containing reserves of carbohydrate and protective secondary metabolites. Dormancy over the summer allows the bulbs to conserve moisture and avoid predators (Hanks, 2002; Graham and Barrett, 2004).

1.1.2 *Narcissus* morphology and development

Daffodils comprise branching systems of bulb units (terminal and lateral) with adventitious roots growing from the basal plate (Figure 1.2) (Hanks, 2002).

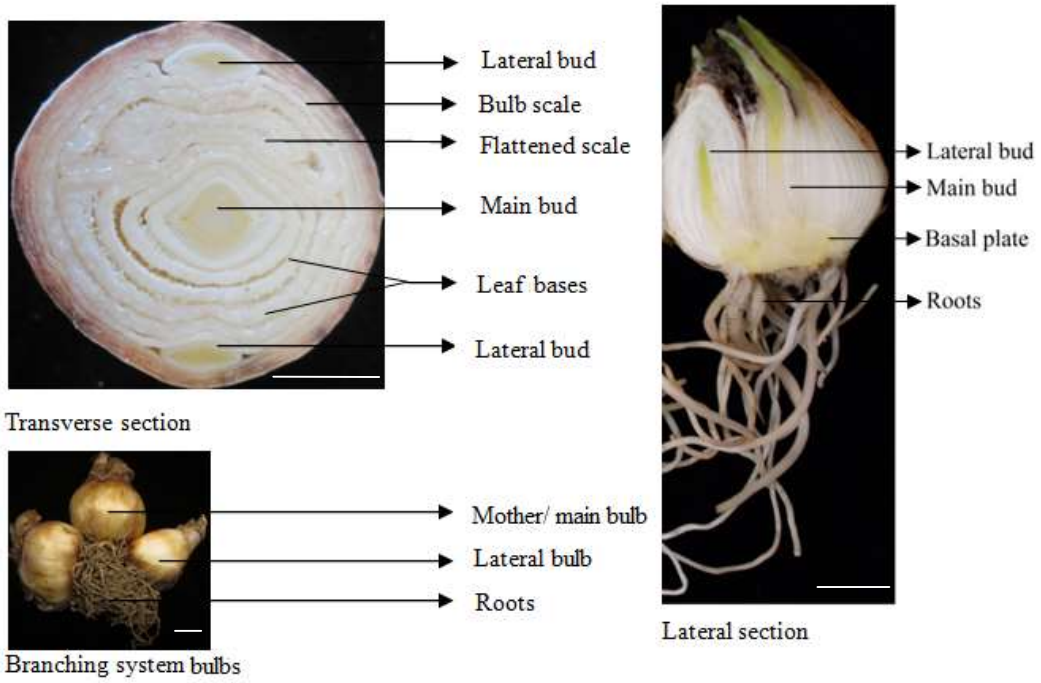

Fig. 1.2 Transverse and lateral sections of *Narcissus* bulb showing main and lateral buds and different bulb scales as well as branching systems bulbs with roots. Scale bars: 1.5 cm.

The main or terminal bud flowers in the first year, while lateral units are subject to apical dominance by the terminal units and develop only once the terminal bud has become floral, losing its dominance (Hanks, 2002). The flattened scale consists of true scales entirely within the bulb; after anthesis the base of the flower stalk becomes flattened. The two or three layers of scales surrounding the flattened-scale are leaf bases which are distinguished from the bulb scales by the thicker tip and a scar. The bulb scales constitute the outer layers of bulbs which are eventually shed as dry tunic (Hanks, 2002). Indeed new bulb units can be initiated within the mother bulb (grown from terminal bud) either in May or December (Branch, 2013).

1.2 Plant tissue culture

Plant cells are distinguished for their totipotency and possess the ability to differentiate, proliferate, and subsequently regenerate into mature plants under appropriate culture conditions (Neelakandan and Wang, 2012). Plant tissue culture is

the aseptic manipulation of plant cells, tissues or organs under defined physical and chemical conditions *in vitro* to grow in a disorganised state (Chaturvedi *et al.,* 2007).

1.2.1 Aspects of plant cell or tissue culture

Plant tissue culture is typically considered to involve three steps. The first step is to obtain a suitable plant part from its usual environment, which is often called explant in tissue culture studies. The next step is to ensure the use of strict aseptic techniques throughout the whole process from explant preparation to explant culture on an appropriate culture medium. The third step is to maintain the culture under a defined physical and chemical environment (Molnár *et al.,* 2011).

Culturing of explants in an appropriate medium often results in an unorganised, growing, and dividing mass of cells called callus. Differentiation takes place during callus formation in both morphology and metabolism, and one of the major consequences of such differentiation is the lack of ability to achieve photosynthesis (Bhatia and Dahiya, 2015).

Shoot and root cultures can be obtained through *in vitro* regeneration of plantlets from explant without an intervening callus phase. This technique is widely used for mass propagation. The plantlets can be isolated and transplanted into soil. In addition, the shoots or roots can be used separately for suspension cultures (Molnár *et al.,* 2011). Root and shoot cultures have emerged as potential methods to study the biochemistry and molecular biology of secondary metabolite biosynthetic pathways due to their organised structure, as production of secondary metabolites often requires specialised tissue types (Loyola-Vargas *et al.,* 2008).

Plant tissue culture has evolved as an experimental model to study cellular physiology, cell division, differentiation, metabolism changes and morphogenesis, which are important for key developmental processes, i.e. embryogenesis, meristem formation and stress related genome plasticity (Neelakandan and Wang, 2012).

1.2.2 Growth requirements for plant tissue culture

The requirements of plant cells or tissues grown *in vitro* are similar to plants growing in nature, as they also require proper light, temperature, humidity and nutrient supply (Molnár *et al.,* 2011).

The growth of cells or tissues in culture condition is hugely affected through maintaining aseptic conditions. Aseptic condition mainly comprises with the sterilisation techniques used (Molnár *et al.,* 2011). Plants grown in the natural environment are habitually contaminated with microorganisms (George *et al.,* 2008). Hence, to maintain an aseptic condition they need to be free from all contaminants before use as an explant. The most commonly used chemical solutions for explant sterilisation are calcium hypochlorite, sodium hypochlorite, bleach solution and ethanol (Vasil and Thorpe, 2013). Several studies have been reported on surface sterilisation combined with hot water treatments, which considerably reduced contamination in plant tissue culture (Abu Zahra and Oran, 2007; Ptak *et al.,* 2008; El Tahchy *et al.,* 2011a).

An artificial growth medium is required for explant culture. Media are made up from solutions of macronutrients, micronutrients, sugar, vitamins, a solidifying agent (agar, agarose or a gellan gum) and growth regulators (mainly auxins and cytokinins) (George *et al.,* 2008). Usually autoclaving (121 ^{o}C, 15 psi for 30-40 min) is used for media sterilisation. Heat sensitive liquids such as growth regulators that are not soluble in water need to be dissolved in a suitable solvent (e.g. ethanol) first and then sterilised using membrane filters of 0.45 or 0.2 µm pore size before adding into the media (Vasil and Thorpe, 2013).

Working surface (laminar air-flow cabinet) of tissue transfer is best to treat with disinfectant and preferably with 70% ethanol. All the metal instruments are best to sterilize by immersion in absolute ethanol and then faming (Vasil and Thorpe, 2013). Finally, a well-maintained growth room is mandatory for successful growth of cultures. The typical plant tissue culture growth room should maintain a consistent temperature range of 22-28 ^{o}C and 16h photoperiod (Vasil and Thorpe, 2013).

1.3 Plant secondary metabolism

1.3.1 Secondary metabolites

An enormous diversity of natural products or metabolites are produced by plants. A small portion of these is indispensable for plant growth and development and classified as primary metabolites (Zhao *et al.,* 2013). Primary metabolites such as amino acids, sugars and fatty acids, can also be channelled to the synthesis of low

molecular weight secondary compounds like alkaloids, flavonoids, terpenes, phenolic compounds, volatiles, tannins, resins and glycosides (Namdeo, 2007).

Primary metabolism is defined as uniform, universal and essential for plant growth. Secondary metabolites are defined as diverse, dispensable for plant growth and development but indispensable for survival (Hartmann, 1996). Collectively, plants can biosynthesize a huge diversity of secondary metabolites where more than 200,000 have been identified (Mizutani and Sato, 2011).

1.3.2 Role of secondary metabolism in plants

In all natural habitats, plants may experience a diversity of potential biotic and abiotic stresses. These come from the interactions between the host plant and the bacteria, viruses, fungi, nematodes, mites, insects, mammals and other herbivorous animals, that rely on them (Mazid *et al.,* 2011). Secondary metabolism plays a major role in plant survival through mechanisms such as attraction of pollinators, defence against herbivores and microbes and abiotic stress including UV-B exposure as well as communication between plants and other organisms (Schäfer and Wink, 2009). Secondary metabolites are also involved in signalling and regulation of primary metabolic pathways. Some secondary metabolites can function as plant growth hormones, which are often used to regulate the metabolic activity within cells and can influence the overall development of the plant (Christophersen, 1995).

1.3.3 Types of secondary metabolites

Plant secondary metabolites can be broadly classified in three major groups i.e. terpenes, phenolic compounds and alkaloids as exemplified in Figure 1.3.

1.3.3.1 Terpenes

Terpenes are composed of 5-C isopentanoid units (Mazid *et al.,* 2011). The largest class of secondary metabolites is terpenes derived from acetyl-CoA are the important agents of insect toxicity and have strong feeding repellence to many herbivorous insects and mammals. As an example, it has been reported that *Chrysanthemum* species contains pyrethroids (monoterpene esters) that showed strong toxicity to insects including moths, beetles and wasps (Mazid *et al.,* 2011).

Abscisic acid (ABA) is a sesquiterpene hormone, which regulates the initiation and maintenance of seed and bud dormancy, plant response to water stress and acts as a transcriptional activator (Mazid *et al.,* 2011). ABA is also involved in the activation of defence system in plants against UV-B (Berli *et al.,* 2010). The diterpene gibberellins are another group of plant hormones that function in plant developmental processes including leaf expansion, flower, and fruit set, seed germination, biomass production, and assimilate translocation (Gupta *et al.,* 2001; Ouzounidou and Ilias, 2005). Terpenes also play vital role on plant physiology such as increase the efficiency of chlorophyll during photosynthesis and serve as component of cell membrane to maintain permeability of small molecules (Mazid *et al.,* 2011).

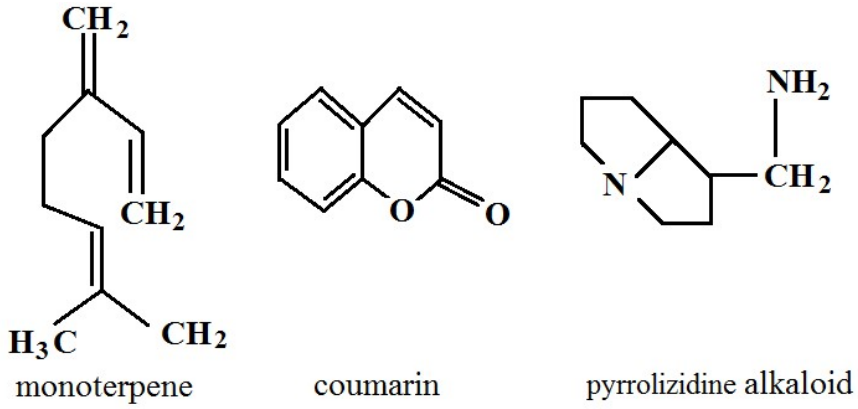

Fig. 1.3 Representative structures of secondary metabolites; terpenes (monoterpene) redrawn from Mazid *et al.,* 2011, phenolic compounds (coumarin) redrawn from Stewart, 2007 and nitrogen containing alkaloids (pyrrolizidine alkaloid) redrawn from Stahl, 2003.

1.3.3.2 Phenolic compounds

Phenolic compounds are synthesized from the shikimic acid pathway (Mazid *et al.,* 2011). They have a wide range of anti-microbial activities, including anti-fungal properties against soil borne plant pathogenic fungi (Brooker *et al.,* 2007). Coumarin, lignin, flavonoids, isoflavonoids and tannins are well-studied phenolic compounds produced by plants. Lignin is responsible for physical toughness that deters feeding by herbivorous animals and its chemical durability makes plants indigestible to insects and most fungi (Mazid *et al.,* 2011). Flavonoids play a tremendous role in plant pigmentation, defence and protection from UV-B radiation (Lake *et al.,* 2009).

1.3.3.3 Alkaloids

Alkaloids contain nitrogen and are synthesized from amino acids, particularly tyrosine, aspartic acid, tryptophan and lysine (Pearce *et al.,* 1991). They mainly serve in defence against microbial infection and herbivore attack and on a cellular level show significant effects on membrane transport, protein synthesis and enzyme activities (Creelman and Mullet, 1997).

1.3.4 Production of secondary metabolites in plant cell or tissue culture

Plant secondary metabolites are important sources of bioactive compounds such as pharmaceuticals, flavours and food additives indicating that their biological activity extends from the native plant to humans (Zhao, 2005; Karuppusamy, 2009). Plants with bioactive compounds can be difficult to cultivate and their genetic diversity may be threatened due to over-harvesting, forest degradation, agricultural encroachment, urbanisation and other factors (Nalawade and Tsay, 2004). Therefore, the biotechnological production of secondary metabolites in plant cell or organ cultures has become a tempting alternative to the extraction of whole plant material (Oksman-Caldentey and Inze, 2004) and gives some advantages. This includes a controlled physical and chemical environment for the synthesis of bioactive compounds independent of climatic and soil factors, and source from uniform, sterile plant material free from natural contaminants. Purification of the metabolite may be easier from this uniform material than from whole plants, thus reducing production costs (Nalawade and Tsay, 2004). In addition, it may be possible to select cultivars able to produce higher amounts in culture. Cell or tissue culture materials may also be 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).

1.3.5 Plant and plant tissue culture as a source of commercially important secondary metabolites

Plant derived compounds usually have highly complex structures and specific stereochemistry requirements that provide their valuable biological properties as well as barriers in commercial production (Verpoorte *et al.,* 2002).

1.3.5.1 Pharmaceuticals

Plants are an important source of pharmaceuticals and plant biomass is still the most cost-effective source for many medical drugs. The WHO estimates that about 80% people of developing countries rely on plants as a source of traditional medicines and plant extracts are applied in more than 85% of the treatments (Hadacek, 2002). In addition, about 119 chemical compounds extracted from higher plants are used in pharmaceuticals worldwide including developed countries (Farnsworth, 1988). Some of these 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).

In vitro methods for secondary metabolite production have very little commercial success (Karuppusamy, 2009). The few long-standing commercial examples of high-value secondary metabolites production from cell culture are shikonin (anti-bacterial) from *Lithospermum erythrorhizon*, sanguinarine (anti-plaque) from *Papaver somniferum* and berberine (against intestinal ailments) from *Coptis japonica* cell cultures (Rao and Ravishankar, 2002; Karuppusamy, 2009). However, research has made very significant advances with some secondary metabolites derived from cell cultures such as taxol and vanillin, produced to a semi-commercial status (Karuppusamy, 2009). Moreover, it has been reported that the Indian Research Centre: Bio-organic Division is carrying out research on mass cultivation of selected cell lines of *Rauvolfia serpentina* (ajmaline, reserpine), *P. somniferum* (thebaine, codeine, and morphine), *Artemisia annua* (artemisinin) and other plant species (Smetanska, 2008).

1.3.5.2 Commercial alkaloids

The commercial importance of some alkaloids has focused attention on developing production methods from both intact plants and cell cultures (Takos and Rook, 2013). The pharmaceutical company GlaxoSmithKline supplies 25% of the world's medicinal opiates and the vast majority of morphine legally derived from *P. somniferum* used to mitigate severe pain (www.gsk.com).

Both intact and cell cultures of *Catharanthus roseus* has been reported to contain about 130 different terpene indole alkaloids (Jacobs *et al.,* 2004). Two, vinblastine

and vincristine, are produced in small quantities for use as anticancer drugs. Both are present as minor compounds when compared with their precursors (tryptamine and ajmalicine) they can be produced using semi-synthetic methods. This has been used by the pharmaceutical company Eli Lily (Eli Lily and Company Ltd. www.lilly.co.uk) to produce vincristine.

Among the Amaryllidaceae alkaloids, only galanthamine is produced on a commercial scale for the pharmaceutical industry (Takos and Rook, 2013). Lycorine and lycoramine from *Lycoris radiata* (red spider lily) are used in Chinese traditional medicine. A Chinese pharmaceutical company (Zhejiang Yixin Pharmaceutical Co. Ltd., www.herbs-tech.com) has tested commercial production of these alkaloids through HPLC methods.

1.3.5.3 Other commercial products

Plant cell and tissue culture techniques were approved as a method to produce natural compounds for food purposes by the United Nations Food and Agriculture Organisation (FAO) in 1994 (Dal Toso and Melandri, 2011). An important aspect is evaluation of risks to consumers from these novel food sources. A bio-safety evaluation of food ingredients derived from plant cell, tissue and organ cultures was elaborated with protocols to evaluate the toxicity of these products and also their potential bioactivity (Murthy *et al.,* 2015). A report was published in 2002 by the FAO and IAEA (International Atomic Energy Agency) on the production of bioactive compounds with increased value from tissue culture, focusing on how they can be processed in the most economical way by researchers and industry (Ahloowalia *et al.,* 2002).

A number of food additives have been obtained from plant cell cultures such as anthocyanins, carotenoids, vanillin, capsaicin, jasmine and aniseed oil (Rao and Ravishankar, 2002).

1.3.6 Problems associated with secondary metabolism in plant tissue culture

A specific secondary compound is often found in only a few species, or even within varieties within a species (Smetanska, 2008). The production of these compounds depends on the plant's physiological and developmental stage. Some of them

accumulate in the plant in specialised cells or organs and their production is often very low (less than 1% DW) (Namdeo, 2007).

Plant growth usually includes three growth phases, cell formation or formative phase, cell enlargement or elongation phase and cell differentiation or stationary phase (Beemster *et al.,* 2003). The growth of cultures *in vitro* typically increases due to cell division and elongation and it continues for short period than stops due to the exhaustion of some nutrient factor or accumulation of toxic compounds in the culture medium (Bhojwani and Razdan, 1986). Therefore, one predicted reason for low yield of secondary metabolites in plant cell or tissue culture is the brief stationary phase of *in vitro* cultured tissue, which inhibits the action of enzymes normally present or active in mature plant or in differentiated tissues (Muhitch and Fletcher, 1985; Dias *et al.,* 2016). Another reason behind this low yield is the lack of biosynthetic knowledge on how secondary metabolites are synthesized or how their synthesis is regulated (Verpoorte *et al.,* 2002).

1.3.7 Strategies for enhanced production of secondary metabolites

Several biotechnological strategies have been undertaken for the enhanced production of valuable secondary compounds such as media modification with growth regulators, optimizing culture condition, elicitation, screening for high yielding cell lines, hairy root culture, cell immobilization, use of bioreactors, biotransformation and metabolic engineering of biochemical pathways (Rao and Ravishankar, 2002; Namdeo, 2007).

1.3.7.1 Organ culture

Organ culture has been reported as a successful biotechnological tool, for example organ culture (small cuttings of bulbs) of *Fritillaria unibracteata* showed 30 to 50 times higher growth rate than wild plants as well as a higher amount of alkaloid than found in the wild bulb (Gao *et al.,* 1999). Shoot cultures of *Frangula rupestris* for the production an anthraquinone also has been reported (Kovačević and Grubišić, 2005) and *Gentianella austriaca* shoot culture showed production of the same amount of secondary metabolites as intact plants (Vinterhalter *et al.,* 2008).

1.3.7.2 Manipulation of nutrient or culture media

Manipulation of culture media including the levels of sugar, nitrogen, phosphate, growth regulators, precursors or intermediate compounds along with optimized culture environment (temperature, illumination, medium pH) is useful for increased secondary metabolite accumulation (Rao and Ravishankar, 2002). In addition, elicitors can be used that are substances, can trigger physiological and morphological defence responses (Namdeo, 2007) and can result in activation of secondary pathways (Anand, 2010). Methyl jasmonate (MJ), chitosan (a fungal elicitor) and yeast extracts are the most widely used elicitors in plant cell culture systems (Anand, 2010) as shown in Table 1.1.

Table 1.1 Production of secondary metabolites by elicitor treatments in plant culture systems.

Plant species	Elicitors	Secondary metabolites	References
Coleus blumei	Methyl jasmonate	Rosmarinic acid	Szabo *et al.,* 1999
Coleus forskolin	Methyl jasmonate	Forskolin	Babu, 2000
Cupressus lusitanica	Methyl jasmonate	β-thujaplicin	Zhao *et al.,* 2001a
Hyoscyamus albus	Methyl jasmonate	Phytoalexins	Kuroyanagi *et al.,* 1998
Silybum marianum	Methyl jasmonate	Silymarin	Sánchez-Sampedro *et al.,* 2005
T. canadensis, T. cuspidata	Methyl jasmonate	Taxoids	Ketchum *et al.,* 2003
T. wallichiana	Methyl jasmonate	Taxanes	Babu, 2000
Catharanthus roseus	Methyl jasmonate	Catharanthine	Zhao *et al.,* 2001b
Glycyrrhiza glabra	Methyl jasmonate	Soyasaponin	Hayashi *et al.,* 2003
Brugmansia suaveolens	Methyl jasmonate	Tropane alkaloids	Alves *et al.,* 2007
Medicago truncatula	Yeast extract	Beta-amyrin	Broeckling *et al.,* 2005
Silybum marianum	Yeast extract	Silymarin	Sánchez-Sampedro *et al.,*2005
Catharanthus roseus	Yeast extract	Indole alkaloids	Menke *et al.,* 1999
Nicotiana tabacum	Yeast extract	Phytoalexins	Wibberley *et al.,* 1994
Salvia miltiorrhiza	Yeast extract	Diterpenoid tanshinones	Yan *et al.,* 2005
Coleus blumei	Yeast extract	Rosmarinic acid	Szabo *et al.,* 1999
Ruta graveolens	Chitosan	Rutacridone epoxide	Eilert *et al.,* 1984

Adapted from Rao and Ravishankar, 2002 and Namdeo, 2007.

1.3.7.3 Metabolic engineering

In many cases, the production of secondary metabolites from plant cell cultures is too low for commercialisation; in such cases, metabolic engineering can provide strategies for enhanced production of secondary compounds (Wu *et al.,* 2006; Hasunuma *et al.,* 2008). However, to undertake metabolic engineering relevant knowledge of the pathways is necessary since it aims for increased productivity of a specific metabolite by increasing the carbon flux through a biosynthetic pathway such as by over expression of genes for rate limiting enzymes or blocking feedback inhibition and competitive pathways (Karuppusamy, 2009). The key goal of plant

metabolic engineering is to provide new plants with improved or new chemical traits (Sugiyama *et al.,* 2011, Zhao *et al.,* 2013).

The first attempt of metabolic engineering to produce a medicinally useful compound was reported in 1992; the cDNA encoding *H6H* (hyoscyamine 6 β-hydroxylase) from *Hyoscyamus niger* was introduced into *Atropa belladona* using *Agrobacterium tumefaciens* which resulted in a high level of scopolamine in the transgenic *Atropa* (Yun *et al.,* 1992).

The pathways of plant secondary metabolism contain multiple enzymes. Multiple genes can be delivered together e.g. transgenic tobacco with over expression of terpene synthase and avian farnesyl diphosphate synthase genes resulted in high-level terpene production (Wu *et al.,* 2006). Whole pathway genetic engineering also has been reported in many plants such as *Arabidopsis*, tobacco, rice and *Brassica* (Zhao *et al.,* 2013). 'Golden Rice' is an example of pathway engineering where the entire β-carotene biosynthetic pathway has been engineered several times to improve the nutritional quality of rice (Ye *et al.,* 2000); but is still only now approaching approval for commercial production.

Despite these successful examples, other studies of plant metabolic engineering did not achieve the desired or anticipated results (Lange and Ahkami, 2013) due to the complexity of plant secondary metabolism, underscoring the need to fully understand the biochemical pathways for the production of secondary metabolites of interest (Dudareva and Pichersky, 2008).

1.3.8 Molecular basis of plant secondary metabolism

The biochemical pathways for the production of secondary metabolites require primary metabolites to undergo a series of modifications catalysed by enzymes such as cytochrome P450s, methyltransferases, glycosyltransferases and acyltransferases. Knowledge provided by the functional annotation of genes from sequenced plant genomes is important for identifying genes and pathways involved in the production of secondary metabolites (Zhao *et al.,* 2014).

The first plant genome sequenced in 2000, was *Arabidopsis thaliana* and around 20% of its genes were annotated for the biosynthesis of secondary metabolites at the

time (Kaul *et al.,* 2000). Most of the putative genes of secondary metabolism exist in families with multiple members because of recurrent gene duplication (Ober, 2005). Some genes belong to small families (<10 members) involved in the production of a specific secondary product such as cinnamate 4-hydroxylase; while others present in mid-size families (10-100 members) such as cinnamoyl-CoA reductase and terpene synthase. However, some genes belong to very large families (>100 members) involved in the production of a plethora of plant secondary metabolites. Examples of these latter include the cytochrome P450s, UDP-dependent glycosyltransferases and methyltransferases (Zhao *et al.,* 2014).

1.4 Amaryllidaceae alkaloids

1.4.1 Types of alkaloids

The Amaryllidaceae benzylisoquinoline alkaloids are of great interest due to their wide range of biological activities and pharmaceutical properties (Bastida *et al.,* 2011). They are of nine distinct types as determined by skeletal characterisation (Bastida *et al.,* 2011) namely norbelladine, lycorine, homolycorine, haemanthamine, tazettine/ pretazettine, narciclasine, montanine, crinine and galanthamine as shown in Figure 1.4.

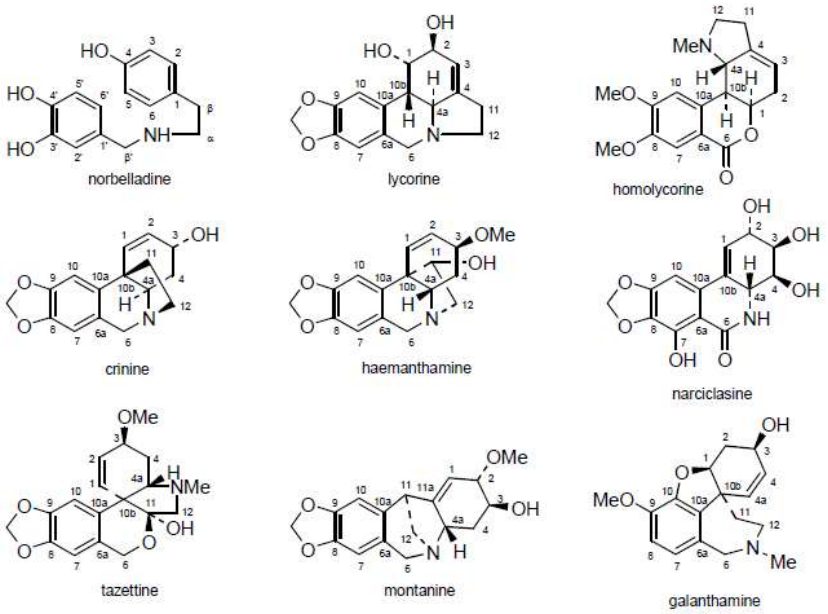

Fig. 1.4 Chemical structures of some Amaryllidaceae alkaloids (Bastida *et al.,* 2011).

Galanthamine (Gal) and lycorine are the two best studied Amaryllidaceae alkaloids (Mroczek and Mazurek, 2009). Galanthamine was first isolated from *Galanthus*

nivalis commonly known as snowdrops (Berkov *et al.,* 2013). *Leucojum aestivum* (snowflakes), *Narcissus* cultivar Carlton, *Lycoris radiata* and *Ungernia victoria* are current industrial source of galanthamine and several chemical synthesis routes have also been reported (Berkov *et al.,* 2014a).

1.4.2 Pharmaceutical properties

Amaryllidaceae species have been used for thousands of years in traditional herbal medicine and the initial evidence of their therapeutic application was recorded in the fourth century by the use of oil from *N. poeticus* L. for the treatment of uterine tumours (He *et al.,* 2015). Amaryllidaceae alkaloids possess a number of pharmacological properties including inhibition of acetylcholinesterase and ascorbic acid biosynthesis, cytotoxicity, anti-cancer, anti-tumour and anti-viral activities (Tahchy *et al.,* 2010).

Galanthamine (Gal) is a long acting, selective and reversible alkaloid used as an acetylcholinesterase (AChE) inhibitor for the treatment of early to mid-stage Alzheimer's disease (Berkov *et al.,* 2014a), which works by increasing the concentration of acetylcholine at sites of neurotransmission and also modulates activity at nicotinic receptors. It has marketing authorization in the UK for the symptomatic treatment of mild to moderately severe dementia of the Alzheimer's type (NHS, 2016; www.nhs.uk). The dual action mechanism of Gal makes it a more attractive drug for treatment in comparison to other AChE inhibitors such as donepezil and rivastigmine, which are known to have greater inhibitory activity than Gal but show no modulation of the nicotinic receptors (Berkov *et al.,* 2009).

Lycorine is a powerful inhibitor of cell growth and division, which has been demonstrated to have antitumor activity and also antiviral activities against poliovirus and measles (Szlavik *et al.,* 2004). Narciclasine is an isocarbostyril compound similar in structure to lycorine with highly promising anti-cancer properties (Takos and Rook, 2013; Bastida *et al.,* 2006). Haemanthamine also has biological properties, including as a protein synthesis inhibitor and has antiparasite and antiviral activity, as well as bioactivity to induce apoptosis in cancer cells (Osorio *et al.,* 2010). Jonquailine, a pretazettine type alkaloid, isolated from *Narcissus* showed significant antiproliferative effects against melanoma,

glioblastoma, and lung cancer, including resistance to apoptosis and multi-drug resistance (Masi *et al.,* 2015)

1.4.3 Biochemical understanding of biosynthesis of Amaryllidaceae alkaloids

If the biosynthetic pathway of plant secondary metabolites is understood and the starting substances are determined, it may be possible to use insights from enzymatic pathways for *in vitro* chemical synthesis (Leonard *et al*., 2009). Studies have been made into the biosynthesis of Amaryllidaceae alkaloids belonging to several different subgroups (Eichhorn *et al*., 1998; Herbert, 2001; Kornienko and Evidente, 2008).

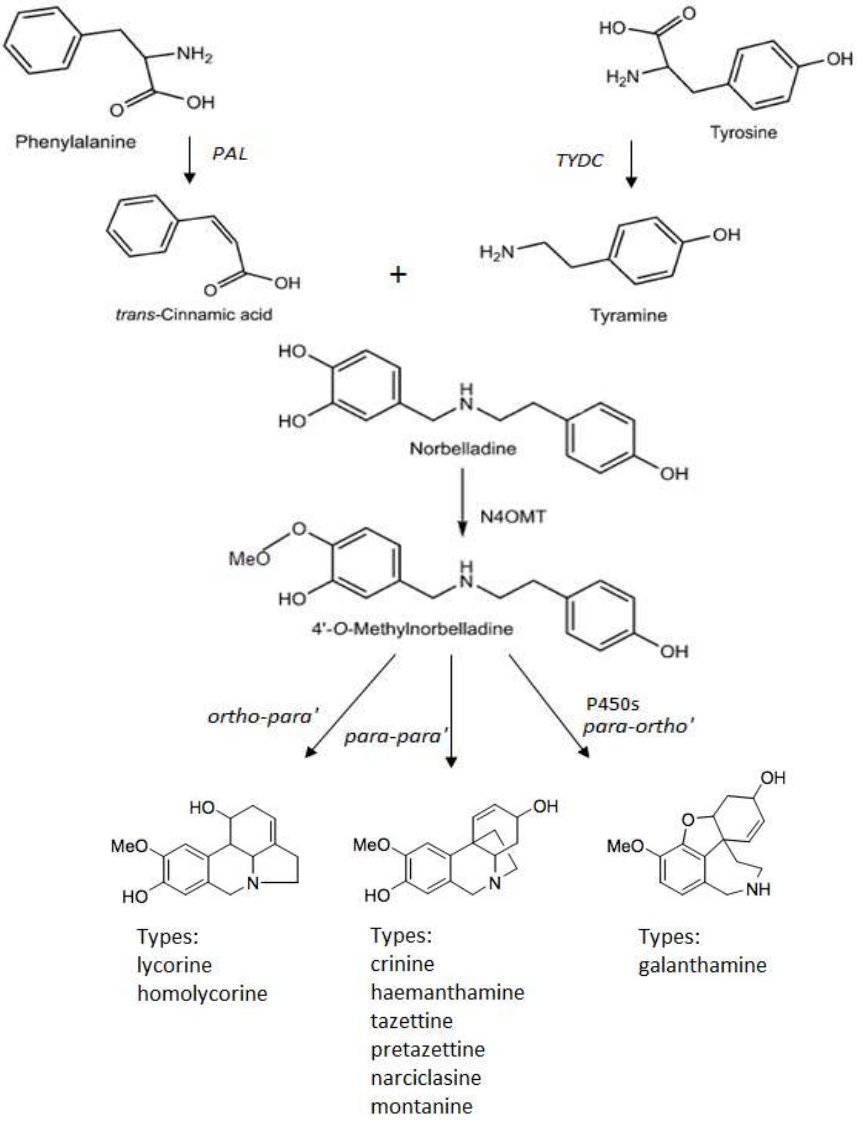

Fig. 1.5 Schematic overview of Amaryllidaceae alkaloid biosynthesis (The pathways shown may take place in one or more steps; more details about the pathways are available in Chapter 3). PAL= phenylalanine ammonia lyase, TYDC= tyrosine decarboxylase, N4OMT= norbelladine 4'-*O*-methyltransferase (redrawn from Takos and Rook, 2013).

All Amaryllidaceae alkaloids are derived from the aromatic amino acids phenylalanine and tyrosine, which produce the common pathway precursor 4′-*O*-methylnorbelladine as shown in Figure 1.5 (Takos and Rook, 2013).

The enzymatic conversion of phenylalanine to *trans*-cinnamic acid and ammonia by the enzyme phenylalanine ammonia lyase (PAL), and decarboxylation of tyrosine by tyrosine decarboxylase (TYDC) to yield tyramine (Figure 1.5), are the start points of Amaryllidaceae alkaloid biosynthesis (Eichhorn *et al.,* 1998). Further degradation of cinnamic acid produces protocatechuic aldehyde and its junction with tyramine results in a Schiff's base, which is converted to norbelladine and its derivatives by the enzyme norbelladine 4'-*O*-methyltransferase (N4OMT). Later studies indicated that norbelladine or related compounds (4'-*O*-methylnorbelladine) could undergo intramolecular oxidative phenol coupling leading to three different groups of Amaryllidaceae alkaloids namely crinine, lycorine and galanthamine (El Tahchy *et al.,* 2011b).

The *para-ortho′* oxidative coupling leads to a dienone which spontaneously cyclises to demethylnarwedine which upon stereoselective reduction leads to demethylgalanthamine. Subsequent *N*-methylation of this compound gives galanthamine. The *para-para′* coupling gives another dienone which depends on nitrogen nucleophilic addition to provide crinine derivatives, and finally *ortho-para′* coupling yields lycorine (Tahchy *et al.,* 2010).

1.4.4 Molecular biological understanding of biosynthesis of Amaryllidaceae alkaloids

Molecular biology knowledge of Amaryllidaceae alkaloid biosynthesis would enable rational approaches to the optimization of commercial alkaloid production (Takos and Rook, 2013). Transcriptome analysis combined with metabolic profiling can allow gene identification by correlating alkaloid production with gene expression profiles (Goossens and Rischer, 2007). Putative genes including phenylalanine ammonia lyase (*PAL*), *N*-methyltransferase (*NMT*), cytochrome P450s (*CYPs*), tyrosine decarboxylase (*TYDC*) and *O*-methyltransferase (*OMT*) and other potentially important candidate genes involved in Amaryllidaceae alkaloids synthesis have been identified in *Lycoris* (Wang *et al.,* 2013) and also in *Narcissus* (Pulman,

2015; Kilgore *et al.,* 2014, Kilgore *et al.,* 2016a). Cytochrome P450s are known enzymes catalysing intramolecular phenol coupling reactions leading to galanthamine type alkaloids (Ikezawa *et al.,* 2008).

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). A cytochrome P450, *CYP96T1* has also 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).

1.4.5 Alkaloids in *Narcissus*

Bastida *et al.,* (2006) reported the occurrence and distribution of alkaloids in approximately 100 varieties of *Narcissus* along with wild *Narcissus* species; the most abundant alkaloids were lycorine, haemanthamine and galanthamine. About 23 alkaloids of the galanthamine, lycorine, homolycorine and haemanthamine skeleton types were reported from the bulbs of *N. pseudonarcissus cv.* Carlton (Bastida *et al.,* 2006; Berkov *et al.,* 2009).

Bulbs and leaves are the major tissues assayed for alkaloids in *Narcissus*. For example, bulb and leaves of more than 100 different ornamental varieties of *Narcissus* were screened for their galanthamine content and their acetylcholinesterase inhibitory activity (Torras-Claveria *et al.,* 2013). The highest amount of galanthamine in leaves was reported as 0.46% and in bulb was 0.14% based on dry weight (DW). A total of 22 ornamental varieties exhibited high acetylcholinesterase inhibitory activity (Torras-Claveria *et al.,* 2013). In a previous study, bulbs of the *Narcissus* cv. 'Sir Winston Churchill' also showed high acetylcholinesterase inhibitory activity (Ingkaninan *et al.,* 2000).

A later study indicated lycorine-type alkaloids predominated in leaves and galanthamine-type in bulbs of *Narcissus* (Torras-Claveria *et al.,* 2014). Galanthamine was found as the major alkaloid in *N. pseudonarcissus* bulbs, followed

by haemanthamine, lycorine, lycoramine and *O*-methyllycorenine (Gotti *et al.,* 2006). Another study showed that the bulbs of *N. pseudonarcissus* contained galanthamine the major alkaloid followed by haemanthamine and narciclasine (Lubbe *et al.,* 2013). Along with galanthamine, the lycoramine, norgalanthamine, epi-norgalanthamine, vittaine, *O*-methyllycorenine, lycorenine, haemanthamine and homolycorine were detected in extracts of *N. pseudonarcissus* cv. Carlton bulbs by GC-FID (gas chromatography-flame ionization detection) (Lubbe *et al.,* 2011).

GC-MS screening of *in vivo N. pseudonarcissus* showed the presence of trispheridine, anhydrolycorine and crinine, whereas galanthamine, lycorine, trispheridine, anhydrolycorine, crinine, demethylmartidine and narwedine were observed *in vitro* (El Tahchy *et al.,* 2011a). Shoot clumps and callus obtained from *N. confusus* contained galanthamine, *N*-formylnorgalanthamine, haemanthamine and tazettine (Selles *et al.,* 1999). Wild plants of *N. confusus* have also been reported to have galanthamine, *N*-formylnorgalanthamine, haemanthamine, homolycorine and tazettine with all identified alkaloids in most of the plant parts (leaves, stem, flowers, roots and bulbs) but varying in content at different growth phases (leaf emerging to flower senescence) (López *et al.,* 2003).

1.4.6 Galanthamine accumulation in plants

1.4.6.1 *In vivo* (field) production

Plants are still considered as the main source of galanthamine production for pharmaceutical industries even if it can be produced synthetically (Lubbe *et al.,* 2011). The galanthamine content varies among the species within the Amaryllidaceae; leaves usually contain 0.05 to 0.15% (DW) while in bulbs it varies from 0.05 to 0.65% (DW) (López *et al.,* 2003). Examples of high galanthamine containing plants are *G. woronowii* (0.9% DW), *U. victoria* (0.6% DW), *L. aestivum* (0.3% DW), and several species of *Narcissus* (0.2% DW) (Cherkasov and Tolkachev, 2002; López *et al.,* 2003).

Seasonal accumulation of galanthamine in *Narcissus* and other Amaryllidaceae was reported in several studies (Kreh, 2002; López *et al.,* 2003; Lubbe *et al.,* 2013). The accumulation of galanthamine was higher in bulbs than in the other plant organs of *N. confusus* and the highest amount was observed in bulbs during leaf emergence

(2.5% DW) followed by flowering (1.6% DW) and the pre-flowering phase (1.1% DW) of plant development (López *et al.,* 2003). Another study on *N. pseudonarcissus* cv. Carlton showed increased galanthamine concentration in bulbs from the shoot emergence to a maximum before flowering and decreased during flowering but showed a slight increase again at senescence of the aerial parts (Lubbe *et al.,* 2013). Moreover, it has also been observed that galanthamine and narciclasine were present at the same levels in leaves, with steady levels until full flowering followed by a decrease after flowering (Lubbe *et al.,* 2013). *N. confusus* bulb has been reported to contain the highest amount (2.5% DW) of galanthamine in plants of this genus (López *et al.,* 2003). These patterns of galanthamine accumulation throughout the life cycle of *Narcissus* have led to the proposal that it might be located into the bulb during the vegetative period but progressively translocated to the aerial parts during flowering and senescence (López *et al.,* 2003).

1.4.6.2 *In vitro* (tissue culture) production

Increased demand of galanthamine has attracted the attention of researchers to *in vitro* culture as an alternative approach for its sustainable production (Pavlov *et al.,* 2007). During the last thirty years, plants including *Narcissus* species have been using to establish *in vitro* production of secondary metabolites (Berkov *et al.,* 2009). The first articles on galanthamine biosynthesis from *in vitro* shoot cultures of *N. confusus* and *L. aestivum* were published in the 1990s (Bergoñón *et al.,* 1992 and 1996; Selles *et al.,* 1997). Mainly undifferentiated callus or suspension cultures were investigated and were improved for this purpose, but very small amounts of the secondary metabolites were produced in undifferentiated cells and have not been commercially viable (Berkov *et al.,* 2009; Schumann *et al.,* 2012; Georgiev *et al.,* 2012).

The degree of cell and tissue differentiation strongly influences the galanthamine content in *in vitro* cultures (Berkov *et al.,* 2009). Cellular differentiation often affects the alkaloid biosynthetic pathways; in particular it enhances the *para-ortho'* oxidative coupling producing galanthamine precursors (Codina, 2002). Callus induction is a vital point for micro-propagation and alkaloid production; due to their simple structure, calli might be a good experimental system to study the biosynthesis of alkaloids, although not the best for alkaloid production (Selles *et al.,* 1999).

A trace amount of galanthamine was observed in undifferentiated cell cultures of *N. confusus* (0.03 µg/g DW) and *L. aestivum* (12 µg/g DW). However, differentiation to shoots (0.14 µg/g DW) or small plantlets (1.43 µg/g DW) resulted in an increased amount of galanthamine and related alkaloids (Pavlov *et al.,* 2007; Selles *et al.,* 1997). The biosynthesis of galanthamine and other Amaryllidaceae alkaloids by shoot cultures derived from different explants (i.e. bulbs, leaves, seeds, young fruits) of *L. aestivum and N. confusus* has been investigated and found to contain 25-454 µg/g DW Gal (Berkov *et al.,* 2014a). The *in vitro* shoot clumps from *L. aestivum* contained a higher proportion of Gal than in *N. confusus* but other major accompanying compounds (haemanthamine, pretazettine and lycorine) were only found in *N. confusus* (Berkov *et al.,* 2011).

1.5 Metabolomics in plant secondary metabolism

1.5.1 Metabolomics

The term 'metabolome' is applied to the low molecular weight compounds present in biological systems consisting both the primary and secondary metabolites (Verpoorte *et al.,* 2007). Metabolomics is generally defined as both the qualitative and quantitative study of metabolomes contained within cells, tissues, and organisms (Yang *et al.,* 2009) to ultimately provide a clear metabolic picture of a living organism (Kim *et al.,* 2010).

Metabolomics has become an important discipline for systems biology, functional genomics and biotechnology (Fiehn *et al.,* 2000; Saito and Matsuda, 2010). There are two approaches to metabolomics; targeted metabolic profiling, aimed at identification and quantification of specific metabolites for a specific reason and non-targeted metabolomics, which focuses on comprehensive metabolic information (Berkov *et al.,* 2011).

1.5.2 Plant metabolomics

Metabolomic studies have provided new knowledge about the dynamics of secondary metabolite production in plants. A metabolome dataset containing 36 distinct tissues of *Arabidopsis* has been reported where the results suggested tissue-specific production of secondary metabolites (Matsuda *et al.,* 2010). Some studies

have revealed the dynamics of secondary metabolite production in response to environmental stresses. An increased level of condensed tannins was observed in *Populus tremuloides* and *P. nigra* in response to drought, pathogen and herbivore attack (Osier and Lindroth, 2001; Hale *et al.,* 2005). Another study on *Arabidopsis* showed that short-term exposure to UV-B resulted in the increased accumulation of primary metabolites, which was suggested to have key function in production of UV-B absorbing secondary metabolites like flavonoids and sinapoyl malate (Kusano *et al.,* 2011). Metabolomics can also establish connection between biochemical pathways through identified metabolites; an *Arabidopsis* metabolic investigation indicated there was a direct or indirect connection between the phenylpropanoid/ flavonoid pathway and the glucosinolate pathway (Böttcher *et al.,* 2008).

About 3,000 metabolites have been detected in tobacco leaves (Wahlberg and Enzell, 1987). The presence of such a large number of compounds in a single organ makes plant metabolomic analysis more complex. Moreover, metabolites differ in their chemical structure, behaviour, polarity, stability, and concentration, which make the analysis extremely difficult if using only one analytical method (Kim *et al.,* 2010).

1.5.3 Plant metabolomic platforms

The metabolomic platforms can be grouped into three groups, based on their underlying principles namely, chromatographic methods (gas or liquid chromatography, capillary electrophoresis, thin layer chromatography), mass spectrometry, or nuclear magnetic resonance (NMR) spectroscopy (Kim *et al.,* 2006).

Mass spectrometry-based metabolomic analysis has been primarily chosen for plants due to its selectivity, sensitivity, comprehensiveness and relatively low cost (Zhao *et al.,* 2013). However, another important criterion for developing a metabolomics platform is the ease and time needed for sample preparation as well as metabolite identification and quantitation (Verpoorte *et al.,* 2007). NMR-based analyses, when compared with chromatographic techniques are highly reproducible, rapid and require very simple sample preparation (Van Der Kooy *et al.,* 2009). The spectrum can provide detailed structural information on metabolites, including chemical shifts and coupling constants (Zhi *et al.,* 2012). Secondary metabolites can be directly

observed in NMR spectra obtained from plant tissues (Schripsema *et al.,* 2007) including crude extracts of cell cultures (Schripsema, 2010). Several NMR-based analyses have been applied to the quantification of alkaloids (i.e. galanthamine, lycorine and haemanthamine) in *N. pseudonarcissus* cv. Carlton (Lubbe *et al.,* 2010; Lubbe *et al.,* 2011; Lubbe *et al.,* 2012; Lubbe *et al.,* 2013).

1.6 Transcriptomic study in the biosynthesis of plant secondary metabolites

1.6.1 Transcriptomic methodology

Omic technologies are aimed at the detection of genes (genomics), mRNA (transcriptomics), proteins (proteomics) and metabolites (metabolomics) in biological samples. The transcriptome is the total mRNA in a cell or organism and acts as a template for protein synthesis. The transcriptome reflects the genes that are actively expressed at a given condition (Horgan and Kenny, 2011). Transcriptome sequencing is a next-generation sequencing (NGS)-based approach, which delivers unbiased information to profiling and analysing RNA without the need for prior knowledge of the genome or transcriptome. It is the method of choice frequently used for the analysis of differentially expressed genes (Horgan and Kenny, 2011).

Transcriptome sequencing begins with isolating RNA and converting (total RNA or mRNA) to complementary DNA (cDNA). For library preparation, strand-specific or random primers can be used. The cDNA can undergo either single-read or paired-end sequencing once after adapters are ligated to the cDNA fragments. Either the sequencing reads can be assembled *de novo* or they can be aligned to a reference genome or transcriptome (GATC Biotech; www.gatc-biotech.com).

1.6.2 Transcriptomics in non-model plants

Transcriptomic studies have been applied to more than four hundred species of non-model plants including Amaryllidaceae (Johnson *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 for their production becomes a challenging task (Hall *et al.,* 2013). Furthermore, limited genomic resources are available for the most valuable specialised metabolites, and generally in non-model plant species (Strickler *et al.,*

2012). Therefore, comprehensive study of selected plant species is needed for better understanding of the biosynthesis of specific metabolites (Xiao *et al.,* 2013).

A transcriptome based project (PhytoMetaSyn; www.phytometasyn.ca) on 75 non-model plants that produce natural products (terpenoids, alkaloids and polyketides), has been undertaken to identify biosynthetic genes responsible for the accumulation of targeted metabolites of six subgroups, i.e. monoterpenoid indole alkaloids, benzylisoquinoline alkaloids, sesquiterpenes, diterpenes, triterpenes and polyketides (Xiao *et al.,* 2013). In this study two NGS sequencing platforms, 454 and Illumina GA were used to generate transcriptome data and the results showed that the number of reads and percentage of unigenes matching reference sequences were greater for the Illumina GA compared with the 454 platform.

A large-scale transcriptome sequence and gene expression profiles for *in vitro* cultures of three medicinally important plant species (*Camptotheca, Catharanthus* and *Rauvolfia*) were performed based on the Illumina platform (Góngora-Castillo *et al.,* 2012). This study represented the expression patterns of important monoterpenoid indole alkaloid biosynthetic pathway genes including methyltransferases, cytochrome P450s, *NADPH*-reductases and many more. Gene transcripts corresponding to known benzylisoquinoline alkaloids biosynthetic enzymes, i.e. *OMT*, *NMT*, *TYDC* and others were detected in different organs of *P. bracteatum* (Rezaei *et al.,* 2016) as well as in *P. somniferum* cell cultures (Desgagné-Penix *et al.,* 2010).

1.6.3 Transcriptomic studies of Amaryllidaceae

Recent transcriptomic study on the non-model plant *L. aurea* based on *de novo* sequence assembly discovered putative genes involved in the Amaryllidaceae alkaloid biosynthesis i.e. *PAL*, *TYDC*, *OMT*, *NMT*, P450 and others (Wang *et al.,* 2013). In another study Kilgore *et al.,* (2014) performed RNA-seq using Illumina in *N. pseudonarcissus* and released the *de novo* assembled transcriptome on the public database MedPlant RNA Seq Database (www.medplantrnaseq.org) that also includes transcriptome data for *Galanthus* species. The candidate genes (4-*O'*-methyltransferase) (Kilgore *et al.,* 2014) and *CYP96T1* (Kilgore *et al.,* 2016a), leading to important Amaryllidaceae alkaloids have also been identified.

1.7 Aims of the project

The overall goal of this research was to provide a comparison, using transcriptomics and metabolomics methods of differing *N. pseudonarcissus* cv. Carlton tissues containing substantially different levels of the alkaloid galanthamine. Tissue culture can provide a simple but intriguing technique to study secondary metabolism (alkaloid biosynthesis). The focus has therefore been to compare callus from *in vitro* culture with field-grown bulb tissues but with interest in other *in vitro* tissues and elicitor treatments as well. The initial step was to utilise a complete and reproducible method for the achievement of undifferentiated and differentiated or organised tissues from tissue culture of *N. pseudonarcissus* cv. Carlton. This was followed by analysis for galanthamine and other alkaloids with a simple and rapid GC-MS (Gas Chromatography-Mass Spectrometry) metabolic profiling and by NMR metabolomics for other small metabolites. Finally, a transcriptomic analysis was undertaken, comparing transcript expression in callus and basal plate tissue. These differential metabolomic and transcriptomic comparisons have been analysed to obtain knowledge about the differences in metabolic pathway activity between the two tissues.

2 Chapter Two: Tissue culture of *N. pseudonarcissus* cv. Carlton

2.1 Introduction

This chapter outlines culture protocol for *N. pseudonarcissus* cv. Carlton, from bulb explants (twin-scale) using MS culture media. The objective was to obtain tissues, particularly undifferentiated callus, to contrast with field samples (bulb and basal plate). These would be used in metabolomic and transcriptomic studies to investigate difference related to alkaloid production.

2.1.1 Difficulties with *Narcissus* propagation

Propagation of *Narcissus* through vegetative methods is not very efficient because of their low propagation speed; conventional horticultural methods (soil preparation, fertilizing, planting, mulching, harvesting and storage) are laborious (Hanks, 2002). Commercial *Narcissus* bulb production takes several years from seed to bulb (5-7 years) (Hanks, 2002). *Narcissus* bulbs show strong dormancy that is slow to break, without a low temperature treatment (4 to 9 ^{o}C) prior to field planting (Hanks, 2002). In addition, bulbs are frequently infected with soil-borne fungi and viruses due to asexual propagation (Chen *et al.,* 2005).

Narcissus are mainly propagated by dividing clumps as well as by seed propagation and commercially *Narcissus* are also propagated by tissue culture or twin-scaling (Royal Botanic Gardens, UK. www.kew.org). Several studies have been reported on efficient rapid micro-propagation of *Narcissus* species through callus induction, adventitious shoot initiation and proliferation (Santos *et al.,* 2002; Sage *et al.,* 2004; Chen *et al.,* 2005; Abu Zahra and Oran, 2007). Tissue culture systems, especially based on direct or indirect somatic embryogenesis can significantly increase the propagation rate (Malik, 2008).

Multiplication can be improved by techniques such as twin scaling, chipping and cutting the bulbs into leaf base and basal plate tissues to culture in a moist inert medium until bulblet formation occurs (Staikidou *et al.,* 1994). Bulbs can be cut into a number of longitudinal segments; then segments are separated into pairs of twin-scales joined by a portion of basal plate up to 60-100 per bulb. While in chipping, whole bulbs are partly divided longitudinally into several segments (8-16 per bulb); each segment with few scales (Hanks, 2002).

2.1.2 Plant parts (explants) for *Narcissus* culture

A variety of plant parts have been used for tissue culture in *Narcissus* for cell differentiation and regeneration (Sage, 2004). Some of them are described here.

2.1.2.1 Leaf explant culture

Leaves are easy to harvest, cut and prepare as explants but limited success has been reported and they are only available during spring (Herranz *et al.,* 2013). Leaves of *N. pseudonarcissus* have been used as explants for shoot culture (El Tahchy *et al.,* 2011a) and bulbils formation (Staikidou *et al.,* 1994). Leaf base and leaf lamina explants were used to produce somatic embryos in *N. pseudonarcissus* cvs. Golden Harvest and St. Keverne (Sage *et al.,* 2000).

2.1.2.2 Scape (flower stem) culture

Flower stem explants are easy to prepare as explants but like leaves are also available in spring (during flowering) (Herranz *et al.,* 2013). Flower stem (scapes) dissected from mature bulbs have been used to induce nodular callus (Sage, 2004) and flower stem explants produced somatic embryos in *N. pseudonarcissus* cvs. Golden Harvest and St. Keverne earlier than leaf or bulb scale explants (Sage *et al.,* 2000).

2.1.2.3 Anther and ovary culture

In vitro anther culture can provide an efficient micro-propagation technique for callus induction and plant regeneration (Chen *et al.,* 2005). Ovary culture has been successful for the production of somatic embryos in *N. pseudonarcissus* cv. Carlton (Malik, 2008). Selecting appropriate reproduction stage and preparing explants from anther or ovaries can be more difficult than other explants. In addition, *Narcissus* flower contains a tricarpellate inferior ovary and two whorls of three anthers (Hanks, 2002) providing a limited number of explants.

2.1.2.4 Chipping

Chips were used for shoot and root formation in *Narcissus* cvs. Carlton and Dutch Master (Sharma and Kanwar, 2002). Bulb scales derived from terminal bulbs of *N. papyraceus* cv. Shirazi were used as explants for somatic embryogenesis (Anbari *et al.,* 2007). Multiple bulb scale explants were used for shoot induction in *N. tazetta*

(Abu Zahra and Oran, 2007). Therefore, use of chipping technique is successful in *Narcissus* (Hanks, 2002) but can require a large volume of bulbs for commercial culture.

2.1.2.5 Twin-scaling

Twin-scales were used as primary explants cultured on a modified MS medium supplemented with growth regulators for shoot induction and proliferation in *N. asturiensis* (Santos *et al.,* 2002). Twin-scales along with a portion of basal plate were used for shoot and root formation in *Narcissus* Carlton and Dutch Master cultivars (Sharma and Kanwar, 2002). Twin-scales and immature scape explants regenerated to small plantlets (De Bruyn *et al.,* 1992) although twin-scaling was more successful for regeneration. The single-scale segments with basal plate showed a clear tendency to produce bulblets at a higher rate than the ones without basal plate (Yanagawa, 2004). Twin-scales from intermediate scales of bulbs previously stored at 15 ^{o}C or 30 ^{o}C produced the maximum number of bulbils (Hanks, 1985). Twin-scales more consistently gave shoots than single scales and stem scales tended to produce callus more readily on higher growth regulators levels in *Narcissus* (Hussey, 1982).

From the above studies, it is evident that twin-scale explants are the most suitable plant part for *in vitro* culture of *Narcissus*. Bulbs are available throughout the whole year while leaf, flower stem, anther and ovary explants are only available in a specific time of year and plant growth stage. Hence, twin-scale explants were chosen as explant for *N. pseudonarcissus* cv. Carlton tissue culture.

2.1.3 Pre-culture treatments to break dormancy and reduce contamination

The establishment of *in vitro* culture in *Narcissus* is often difficult due to a high rate of contamination in cultures initiated from explants originated from underground plant parts (Sochacki and Orlikowska, 2005). Fungal diseases including basal rot caused by *Fusarium oxysporum f. sp. narcissi*, the most serious disease affecting *Narcissus* bulbs in the UK can contaminate bulbs (Bowes *et al.,* 1992; Hanks, 2002).

2.1.3.1 Cold treatment of bulb prior to culture

It has been observed that a cold treatment of low temperature (3 to 9 ^{o}C) for a period (4 to 12 weeks) is mandatory to break *Narcissus* bulb dormancy prior to planting or

tissue culture. Cold treatment of bulbs before tissue culture is a widely used technique to break bulb dormancy (Hanks, 2002). The *N. tazetta* bulbs were subjected to a cold treatment of 4-8 ^{o}C for 4 weeks (Chen *et al.,* 2005) prior to *in vitro* culture. *N. pseudonarcissus* cv. Carlton bulbs were chilled for 12 weeks at 5 ^{o}C before initiating *in vitro* culture (Malik, 2008). *N. asturiensis* bulbs were stored at 9 ^{o}C for four weeks to overcome dormancy before cutting them into twin-scales for *in vitro* bulb formation (Santos *et al.,* 2002).

Based on the previous studies a cold storage of 4-5 weeks at 4 ^{o}C was selected for *N. pseudonarcissus* cv. Carlton bulb treatment prior to hot water treatment.

2.1.3.2 Hot water treatment to reduce contamination in culture

The application of a hot water treatment is necessary to kill nematodes and microorganisms present in bulbs (Hol and Van Der Linde, 1992; Hanks, 2002). This become even more essential while in light of the UK recommendation to treat all *Narcissus* planting stock at 44.4 ^{o}C for 3 h before field planting (Hanks, 2002).

Hot water treated bulbs from *N. tazetta* (1h at 54 ^{o}C) were subsequently sterilised using 70% ethanol for 3 min and 1% NaOCl for 20 min for adventitious shoot induction (Abu Zahra and Oran, 2007) to combat contamination. *N. pseudonarcissus* cv. Carlton bulb treatment at 44.4 ^{o}C for 3h prior to sterilisation in 0.1% mercuric chloride and 0.75% chloramine T resulted in the reduction of contamination by only 14% to 17% (Sochacki and Orlikowska, 2005).

It was also reported that a lower temperature (1h at 50 ^{o}C) was less effective at reducing contamination but a higher temperature (1h at 58 ^{o}C or 62 ^{o}C) resulted in either the twin-scales turning black or prevented the formation of new shoots in Golden Harvest. An intermediate hot water treatment (1h at 54 ^{o}C) resulted in reduced contamination without affecting regeneration and the same results were obtained for other *Narcissus* cultivars including Carlton (Hol and Van Der Linde, 1992).

Therefore, a lower temperature could be successful to kill bulb borne pathogens if the bulbs are treated for long duration, or an increased temperature with a reduced duration could also be a good choice. Consequently, a low temperature treatment for

long duration (45 ^{o}C, 2h) and higher temperature (52 ^{o}C or 54 ^{o}C) with reduced time duration (1 h) were chosen for *N. pseudonarcissus* cv. Carlton bulb treatment prior to surface sterilisation.

2.1.3.3 Surface sterilisation

Surface sterilisation of plant parts prior to tissue culture is also essential to reduce the contamination rate. Many chemicals have been reported all of which reduced the microbial load. Leaf explants of *N. pseudonarcissus* were sterilised with 70% ethanol for one minute and then shaken in 15% Domestos (with NaOCl and NaOH content below 5%) for another 15 min (El Tahchy *et al.,* 2011). Bulb sterilisation with commercial bleach solution (1.5%) for 30 min followed by shaking in 0.1% $HgCl_2$ solution for 4 min on a rotary shaker has been used in *N. papyraceus* cv. Shirazi (Anbari *et al.,* 2007). Unopened flower buds of *Narcissus* were immersed in 75% ethanol for 30 s and then in 2% NaOCl for 15 min for surface disinfection (Chen *et al.,* 2005). Initially bulb explants from *Narcissus* were washed with 0.1% Teepol surfactant under running water followed by a dip in 70% ethanol for one minute and a final sterilisation was done with 0.1% mercuric chloride for 10 min (Sharma and Kanwar, 2002). Bulb scale explants from terminal bulb units, derived from mature bulbs of *Narcissus,* were surface sterilised in a 10% aqueous solution of Domestos commercial bleach solution (Sage *et al.,* 2000).

A combination of hot water treatment and surface sterilisation of bulbs prior to *in vitro* culture is the best strategy to kill inner-borne fungus and other contaminants in *Narcissus*. Examples where this combination has been effective include twin-scale explants of Golden Harvest and Carlton that after disinfect with 1% sodium hypochlorite for 30 min, remained contaminated by *Fusarium* species. A hot water treatment (1h at 54 ^{o}C) prior to surface sterilisation showed reduced contamination to 5% (Hol and Van Der Linde, 1992). This indicates the contamination is located inside the tissues and surface disinfection alone is not suitable to kill inner borne contaminants.

Therefore, a surface sterilisation step using bleach solution followed by hot water treatment was applied for *N. pseudonarcissus* cv. Carlton bulb sterilisation prior to *in vitro* culture.

2.1.4 Suitable culture media for *Narcissus*

Success in plant tissue culture also depends on the selection of nutrient medium. The cells of most plant species can be grown on a completely defined medium. The Murashige and Skoog (MS) medium, or modifications thereof is widely used for plant tissue culture (Che *et al.,* 2006; Ptak *et al.,* 2010; El Tahchy *et al.,* 2011; Singh and Chaturvedi, 2012; Rezaei *et al.,* 2016) because it consists of mineral salts, a carbon source (generally sucrose), vitamins and growth regulators. In addition, MS medium contains the correct amounts of inorganic nutrients to satisfy the nutritional and the physiological needs of many plant cells in culture (Parimalan *et al.,* 2008; Safdari and Kazemitabar, 2010; Anjusha and Gangaprasad, 2016). A distinguishing feature of the MS medium is its high content of nitrate, potassium and ammonium relative to other nutrient media (Gamborg *et al.,* 1976). Successful use of MS medium has been reported for tissue differentiation, plantlet regeneration and multiplication in several *Narcissus* species (Selles *et al.,* 1999; Chen *et al.,* 2005; Taleb *et al.,* 2013).

2.1.5 Effect of growth regulators on culture

The presence of appropriate concentrations and combinations of plant growth regulators in the culture media are the basic requirements for both callus induction and plant regeneration from explants (Kaya and Aki, 2013). The auxins (NAA, 2, 4-D and IBA) and cytokinins (KN, BA and BAP) are the most widely used growth regulators and important for regulating growth and morphogenesis in plant tissue culture (George *et al.,* 2008).

Auxins are involved in cell division, cell elongation and root formation; on the other hand, cytokinins regulate protein synthesis and cell differentiation (Yadava and Chawla, 2002; George *et al.,* 2008). Growth regulators have a major effect and a regulatory role on the growth of callus and root cultures (Islam *et al.,* 2005). Auxins have a crucial role in callus induction and their action is facilitated by lower concentrations of cytokinins (Wiktorowska *et al.,* 2010). On the other hand, lower concentration of auxin combined with higher cytokinin stimulates callus proliferation (Verma *et al.,* 2012).

2.1.5.1 Callus induction and maintenance

Callus induction in *Narcissus* is mainly influenced by a higher level of auxin combined with a lower concentration of cytokinin. MS media containing 1.5 mg/l BA and 3mg/l NAA was the most suitable medium for callus induction and proliferation derived from bulb explants of *N. tazetta* var. *italicus* (Taleb *et al.,* 2013). Calli were successfully induced with low concentrations of auxin (1.0 to 2.5 mg/l 2, 4-D) while an increased concentration of auxin (3.8 to 4.4 mg/l 2, 4-D) decreased the survival rates of explants, indicating a toxic effect (El Tahchy *et al.,* 2011a). The same study also showed that the effects of auxin and cytokinin alone were not significant but the combination of both had a positive effect on callus induction, root and bulb formation.

A similar finding was reported in another study, i.e. only 2, 4-D (0.5 mg/l) was not sufficient for callus induction from anther culture in Chinese *Narcissus* while the highest callus induction efficiency was achieved on the medium supplemented with BA (0.5-2 mg/l) in addition to 2, 4-D (0.5 mg/l) (Chen *et al.,* 2005). It was also been reported from the same study that MS medium without any supplement of growth regulators did not show any callus formation.

MS media supplemented with 2 or 4 mg/l NAA with 4 or 8 mg/l BA produced callus like proliferation whereas lower concentration of both (0.12 or 0.25 mg/l NAA and 2 or 4 mg/l BA) gave the greatest number of shoots in *Narcissus* Golden Harvest and Fortune (Hussey, 1982). Auxin concentrations of 2 to 4 mg/l 2, 4-D were used for callus induction in *N. confusus* while a higher concentration (10 mg/l 2, 4-D) was used for their maintenance for more than two years (Selles *et al.,* 1999).

It is noted from the above studies that combination of auxins and cytokinins accelerate callus induction and proliferation in *Narcissus* and their action mostly effective when they are applied at low concentration. In addition, a higher level of auxin is required for callus maintenance (Selles *et al.,* 1999).

2.1.5.2 Shoot initiation or bulblet formation

The cytokinins play prominent role in cell differentiation (George *et al.,* 2008). The highest numbers of adventitious shoot formation was obtained by supplementing the MS culture medium with high a concentration of BAP (10 mg/l) and low NAA

concentration (0.5 mg/l) (Abu Zahra and Oran, 2007). That study is in accordance with the main function of cytokinin in cell differentiation as higher level of cytokinin resulted in better shoot formation.

MS media supplemented with lower levels of auxin initiated more shoots and the highest number of shoots was obtained from the medium containing 0.5 mg/l IAA in *Narcissus* (Sochacki and Orlikowska, 2005). The multiplication rates of leafy shoots were higher in MS medium with NAA (0.12 mg/l) and BA (5.99 mg/l) rather than IBA (1 mg/l) (Santos *et al.,* 2002), which indicates NAA could be a more appropriate auxin for shoot initiation in *Narcissus.* Somatic embryos obtained from *N. pseudonarcissus* cv. Carlton showed high regeneration to small plantlets when transferred in MS medium containing 1 mg/l NAA and 1 mg/l BAP (Malik, 2008). MS medium containing 1.0 mg/l NAA and 10 mg/l BA resulted in maximum bulblet formation in Carlton (Sharma and Kanwar, 2002).

2.1.5.3 Callus differentiation or formation of regenerated bulblets

Callus differentiation in plant tissue culture allows organised tissues to be obtained through various degrees of differentiation. Moreover, auxin and cytokinin mainly influence this differentiation (Ikeuchi *et al.,* 2013). Different combinations of NAA, BAP and KN were used for bulblet regeneration in *Narcissus* and the highest number of regenerated bulblets was obtained from medium containing NAA (0.93 mg/l) combined with BAP (2.2 mg/l) and 2, 4-D (1.1 mg/l) (Anbari *et al.,* 2007). The same study showed MS medium containing IBA (1 mg/l) as a rooting medium for initiation of roots in the bulblets.

The maximum number of bulblets was regenerated using MS medium containing 0.1 mg/l 2, 4-D or NAA and 0.1 mg/l BA in *Narcissus* (Sage, 2004). The auxins (NAA, IAA and IBA) stimulated bulblet development, compared to control cultures without auxin (Seabrook and Cumming, 1982; Chow *et al.,* 1992; Staikidou *et al.,* 2005), whereas this was inhibited by BA (Staikidou *et al.,* 1994) in the presence or absence of NAA (Chow *et al.,* 1992).

All of these studies suggested that choosing a combination of NAA and BAP, as growth regulators would be a good choice for *Narcissus*.

Therefore, it can be concluded that MS medium supplemented with growth regulators are inevitable for tissue culture i.e. callus induction, shoot initiation, plantlet regeneration and root formation in *Narcissus*. The selection of the proper combination and concentration of growth regulators are important for better growth and differentiation. In this study a lower (4 mg/l NAA) and a higher (20 mg/l NAA) auxin concentration were selected to combine with low concentrations of cytokinins (1.5 mg/l BAP and 0.5 mg/l KN) for generating callus culture of *N. pseudonarcissus* cv. Carlton. In addition, MS medium supplemented with high auxin (20 mg/l NAA; 10 mg/l 2, 4-D and 8 mg/l 2, 4 -D) was selected as trial treatments for callus maintenance based on Selles *et al.,* 1999.

2.1.6 Effect of sucrose concentration on culture

Sucrose is the carbon source in culture media and required for cell growth and development. High sucrose concentrations (6% or 9% or 12%) induce better organogenesis but decrease callus formation in *Narcissus* (El Tahchy *et al.,* 2011a). Increasing the sucrose concentration of MS medium from 3% to 6% or 9% showed the percentage of shoot formation in *Narcissus* was increased to 48% and 71%, respectively (Chow *et al.,* 1992). The *N. pseudonarcissus* cv. Carlton produced maximum shoots and bulblets in MS medium containing 1 mg/l NAA, 5-10 mg/l BA and only 3% sucrose (Sharma and Kanwar, 2002).

According to these findings, a sucrose concentration of 3% or 6% would be beneficial to obtain the maximum number of shoots and bulblets but 6% might not be the best one for callus formation. Therefore, a sucrose concentration of 5% was selected for this study.

2.2 Materials and methods

2.2.1 Plant materials

Narcissus bulbs from variety 'Carlton' supplied by New Generation Daffodil Ltd. UK, were used for tissue culture. Bulbs were maintained under cold treatment (4 ^{o}C) for at least one month to break the bulb dormancy. Then cold treated bulbs were kept at room temperature for at least 24 h before hot water treatment.

2.2.1.1 Hot water treatment

Fresh and healthy looking bulbs, each weighing around 40 to 45 g, were selected for hot water treatment. Outer dead scales were removed, top and bottom portions were cut. Then the bulbs were kept in a hot water bath (Thermo Fisher Scientific, UK) maintaining three different temperatures (45 ^{o}C, 52 ^{o}C and 54 ^{o}C) for 1 to 2 h. The bulbs were then dried with tissue paper and for complete drying were kept at room temperature for another 24 h before surface sterilisation. Following trials, a routine hot water treatment (52 ^{o}C, 1h) was used for all bulbs in preparation for tissue culture.

2.2.1.2 Surface sterilisation

After hot water treatment, the dried bulbs were cut into quarters and surface sterilisation was performed, using Domestos bleach solution (20%) for 30 min then rinsed (6×) with sterile distilled water. All steps were performed in a laminar hood maintaining strict sterile conditions.

2.2.1.3 Explant preparation

The disinfected bulbs were cut as twin-scales (cut from the basal part of bulbs with 2 or 3 scales of 0.8-1.0 cm in size) in a laminar hood to initiate tissue cultures. A total of 24 twin-scales were obtained from a single bulb. Autoclaving, alcohol and flame sterilisation were used to sterilise the forceps, needles and scalpels to cut the twin-scale explants.

2.2.2 Culture media and culture conditions

Three media combinations were used to initiate tissue culture as shown in Table 2.1, namely MS basal medium (Murashige and Skoog, 1962) and modified MS media MSM1 and MSM2. MS media, ascorbic acid and all the growth regulators were purchased from Duchefa Biochemie (Netherlands), polyvinylpyrrolidone from Sigma-Aldrich (UK) and yeast extract from Thermo Scientific (Germany). Media pH was adjusted to 5.6-5.8 with 1M NaOH and HCL prior to autoclaving (121 ^{o}C, 108 kPa and 30 min). All growth regulators and ascorbic acid were filter-sterilised (0.22 µm) and added to the sterilised media.

Two twin-scale explants were inoculated per Petri plate. Petri plates were sealed with a single layer of parafilm and incubated in a culture room (24±2 ^{o}C, 12 h photoperiod). After one month of inoculation twin-scales became swollen and started to grow small bulblets with green or white shoots, callus and later regenerated bulblets from the callus. Data were recorded after four weeks of incubation. All the tissue culture materials obtained from MS basal, MSM1 and MSM2 were maintained for three to six months by sub-culturing every four to six weeks on the same media until an adequate amount of materials was achieved to set-up experiments for alkaloid extractions, NMR and RNA-seq analyses.

Table 2.1 Media compositions for *N. pseudonarcissus* cv. Carlton tissue culture (Per Litre).

Media	Composition
MS (basal)	MS 4.30 g + sucrose 50 g + agar 8 g
MSM1 (modified MS) High auxin medium	MS (basal) + Yeast Extract (YE) 100 mg + Vitamin C (VC) 50 mg + Polyvinylpyrrolidone (PVP) 30 mg + Kinetin (KN) 0.5 mg + Benzyl Amino Purine (BAP) 1.5 mg + Naphthalene Acetic Acid (NAA) 20 mg
MSM2 (modified MS) Low auxin medium	MS (basal) + Yeast Extract (YE) 100 mg + Vitamin C (VC) 50 mg +Polyvinylpyrrolidone (PVP) 30 mg + Kinetin (KN) 0.5 mg + Benzyl Amino Purine (BAP) 1.5 mg + Naphthalene Acetic Acid (NAA) 4 mg

2.2.3 Calculations and data interpretation

The results of *in vitro* tissue responses (callus, direct bulblets, regenerated bulblets) are represented in percentage. As for example, the percentage of callus (Figure 2.4), regenerated bulblets (RB) (Figure 2.6) and direct bulblets (DB) (Figure 2.7) were calculated based on the number of explants developed to callus or direct bulblets or regenerated bulblets to the total number of explants or callus incubated on the respective media (MS, MSM1 and MSM2).

Representative equations are as follows:

$$\text{Percent (\%) of callus} = \frac{\text{total number of callus obtained from explants grown on MSM1 or MSM2}}{\text{total number of explants incubated in MSM1 and MSM2}} \times 100$$

$$\text{Percent (\%) of RB} = \frac{\text{total number of RB obtained from callus incubated on MSM1 or MSM2}}{\text{total number of callus incubated in MSM1 and MSM2}} \times 100$$

$$\text{Percent (\%) of DB} = \frac{\text{total number of DB obtained from explants grown on MS or MSM1 or MSM2}}{\text{total number of explants incubated in all three media (MS, MSM1, MSM2)}} \times 100$$

2.3 Results

2.3.1 Establishment of culture condition

Initially six Carlton bulbs were used to establish suitable culture conditions; two bulbs for each temperature treatment (45 ^{o}C for 2h, 52 ^{o}C and 54 ^{o}C for 1h) were used. After hot water treatment and surface sterilisation, twenty-four twin-scale explants were cut from each disinfected bulb, incubated up to 10 weeks on MS basal, high auxin (MSM1) and low auxin (MSM2) media. Data were collected and recorded after two weeks of incubation (24 ± 2 ^{o}C, 12 h photoperiod) and continued for another six weeks on a regular basis checking all the cultured plates containing twin-scale explants.

Figure 2.1 shows that 52 ^{o}C, 1h gave the best results as the amount of contamination as well as senescence was lowest. The largest proportion of explants (37.5%) developed as callus or bulblets (successful growth). Results were pooled from the three media for this comparison.

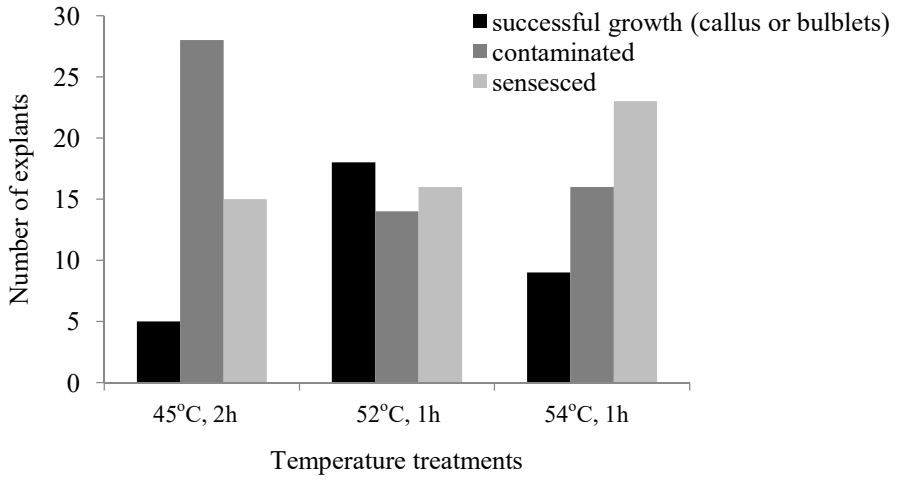

Fig. 2.1 Effect of three temperature pre-treatments (2 h, 45 ^{o}C; 1 h, 52 ^{o}C and 1 h, 54 ^{o}C) after 10 weeks of incubation on MS, MSM1 and MSM2 media.

Figure 2.1 also shows that high levels of contamination as illustrated in Figure 2.2 (A) remained after bulbs were treated at the lower temperature of 45 ^{o}C, while at the higher temperature of 54 ^{o}C, although contamination was reduced, many explants were senesced (Figure 2.2 B). Therefore, 52 ^{o}C, 1h was selected for bulb treatment through this project.

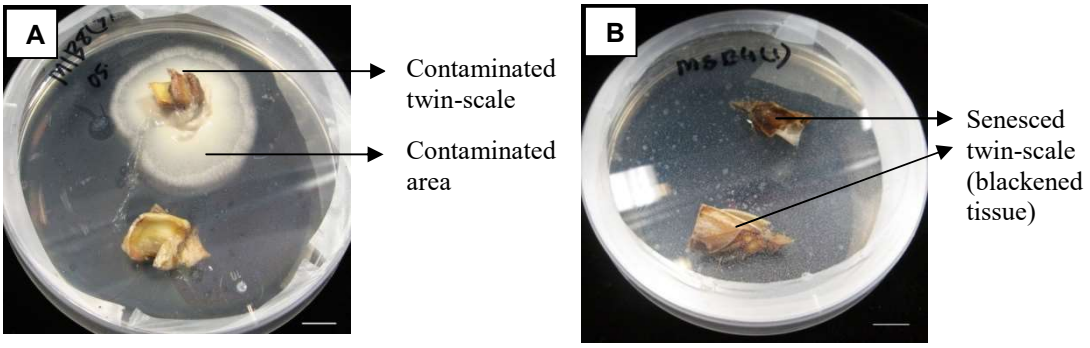

Fig. 2.2 Plate showing contaminated area surrounding twin-scale (A); senesced twin-scale (B) after one week of incubation. Scale bars: 0.5 cm.

2.3.2 *In vitro* tissue differentiation of twin-scale explants

Twenty-five Carlton bulbs were used to initiate *in vitro* culture from twin-scale explants (24 twin-scales per bulb) (Figure 2.3 A).

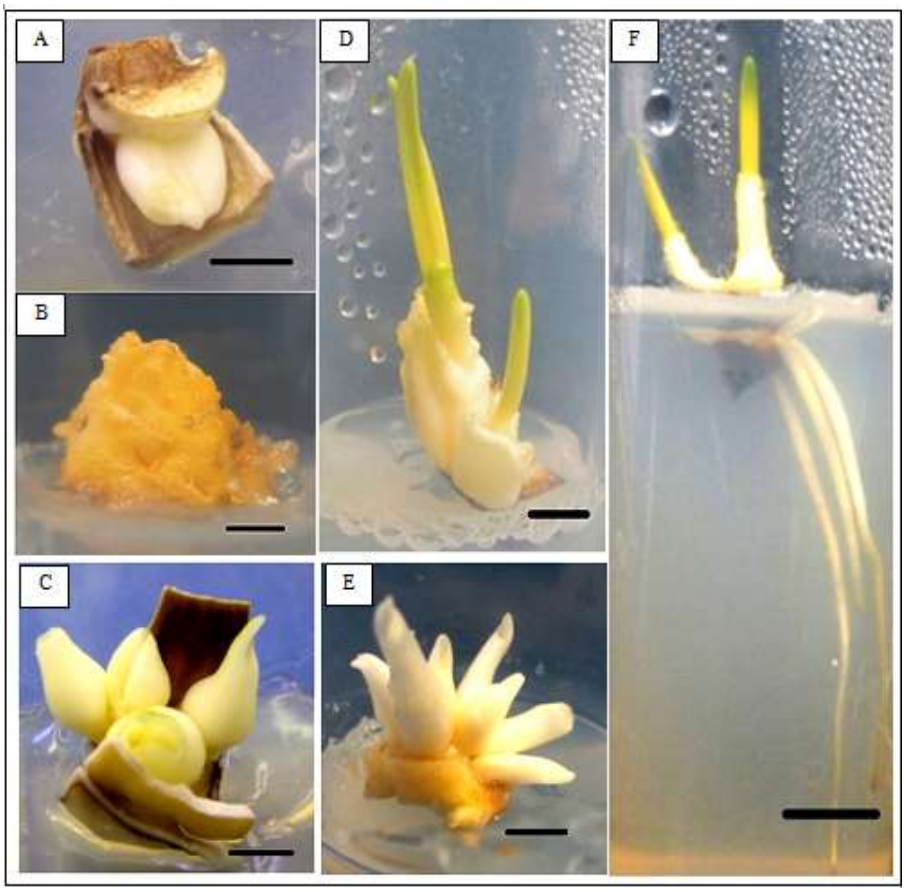

Fig. 2.3 Cultured samples from *N. pseudonarcissus* cv. Carlton: Twin-scale explant with small bulblet (A) from MS basal medium after 4 weeks, callus (B) from MSM1 medium after 8 weeks, white bulblets directly grown from base of twin-scale (C) from MSM2 medium after 8 weeks, small bulblets with green and white shoots (D) from MS basal medium after 12 weeks, bulblets regenerated from callus differentiation (E) on MSM1 medium after14 weeks and bulblets with roots (F) from MS basal medium after 16 weeks of incubation. Scale bars: 0.5 cm.

Among the 24 twin-scale explants obtained from each bulb, 12 twin-scales were cultured on MS basal media, 6 on high auxin (MSM1) and 6 on low auxin (MSM2) media, incubated at 24 ± 2 ^{o}C, 12 h photoperiod.

After about four weeks of culture initiation, some explants had developed into undifferentiated callus (Figure 2.3B). In addition, explants started to give rise to small bulblets with white (Figure 2.3C) and green shoots (Figure 2.3D) directly from the base of a scale, hence named direct bulblets or direct shoots. Calli were sub-cultured on the same media of induction; high (MSM1) and low auxin (MSM2) media, and no callus was grown from MS basal medium. After another six weeks some callus developed into regenerated bulblets (Figure 2.3E). A very few bulblets with roots (16 bulblets with 55 roots) from 11 explants were also obtained from MS basal medium (Figure 2.3F).

2.3.3 Effect of different media and growth regulators on callus induction

MS basal and modified MS media with both high (MSM1) and low (MSM2) auxin were tested for callus induction. No callus found from MS basal medium. Therefore, data presented in Figure 2.4 includes callus from MSM1 and MSM2. In both media, callus initiation from twin-scale occurred about two weeks after culture initiation but better callus formation was recorded from high auxin (MSM1) medium (67% of total explants cultured on MSM1 and MSM2) than the low auxin (MSM2) (13% of total explants cultured on MSM1 and MSM2) (Figure 2.4).

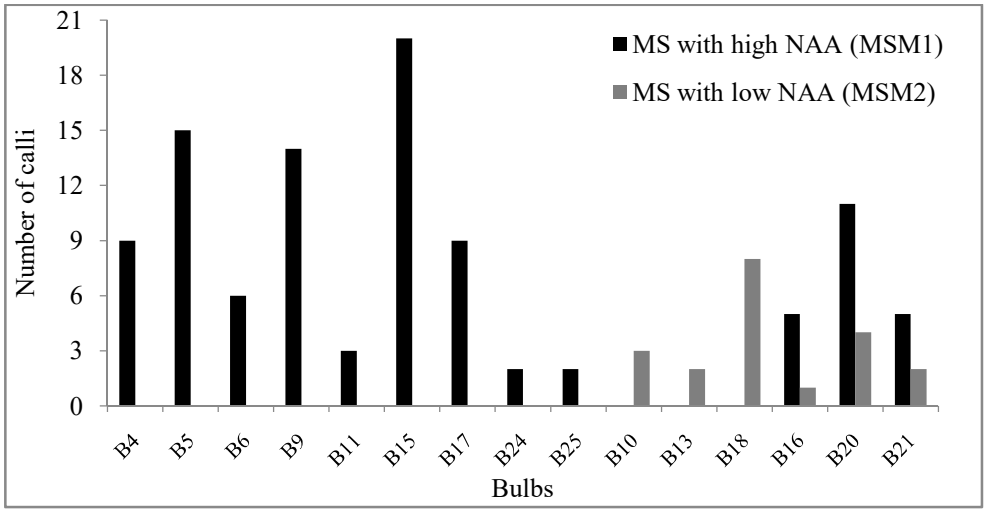

Fig. 2.4 Effect of different media combinations and bulbs on callus induction after twelve weeks of incubation. Multiple bars indicating the response of bulbs in both media with the absence of bars indicating no callus was obtained, due to either contamination or senescence.

2.3.4 Effect of growth regulators on callus growth and maintenance

Calli (1.0 cm×0.8cm) obtained from both high (20 mg/l NAA), MSM1 and low (4 mg/l NAA), MSM2 auxin media were sub-cultured on the same media, six weeks after culture initiation. Each callus was cut into several small pieces (0.5 cm×0.3 cm). They continued to grow and after another eight weeks the size became approximately (1.5 cm×1.0 cm). Calli (1.5 cm×1.0 cm) were again cut into several small pieces (1.0 cm×0.3 cm) before second sub-culture.

After another 4 weeks about 60 pieces of callus obtained from the high auxin (MSM1) medium were tested in two different media combinations for further proliferation and maintenance. Among 60 calli (0.5 cm×0.7 cm), 30 were incubated on high auxin (MSM1) medium (control); 15 were cultured on MS medium supplemented with 10 mg/l 2, 4-D and 15 on MS medium with 8.0 mg/l 2, 4-D (Selles *et al.,* 1999). Figure 2.5 represents the three media combinations, where calli on the control medium showed higher proliferation (73% of total calli inoculated) than on the media with 10 mg/l 2, 4-D (53% of total calli inoculated) and 8.0 mg/l 2, 4-D (33% of total calli inoculated). Therefore, high auxin (20 mg/l NAA) media was used for further sub-culture and calli maintenance for about 1.5 year that was used for alkaloid analysis, RNA extraction and NMR analysis.

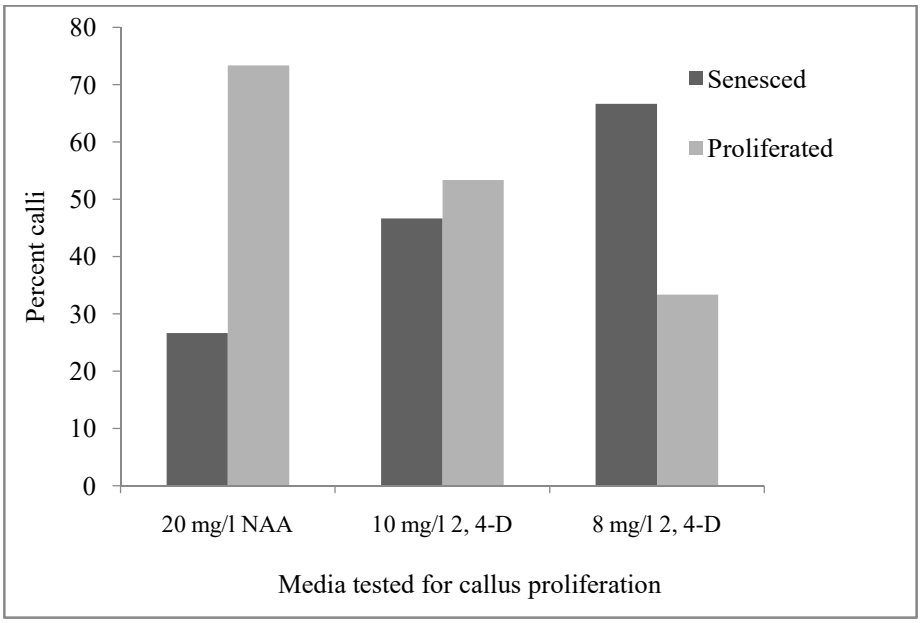

Fig. 2.5 Effect of growth regulators on callus maintenance (20 mg/l NAA, 10 mg/l 2, 4-D or 8.0 mg/l 2, 4-D) after six weeks of sub-culture.

2.3.5 The differentiation of calli (regenerated bulblet and shoot regeneration)

After being on induction media MS with high (MSM1) and low NAA (MSM2) for ten weeks and regular sub-culture (every 3 to 4 weeks) on same media, calli started to form small white bulblets after another six weeks.

Calli regenerated into bulblets with white shoots in both high auxin (45% of total calli inoculated on both media) and low auxin (52% of total calli inoculated on both media) media as illustrated in Figure 2.6. A few calli from MSM1 were sub-cultured on MS basal medium for trial (calli regeneration). However, this resulted in most of the calli senescing growth. In general, a piece of callus could regenerate 2-10 small regenerated bulblets.

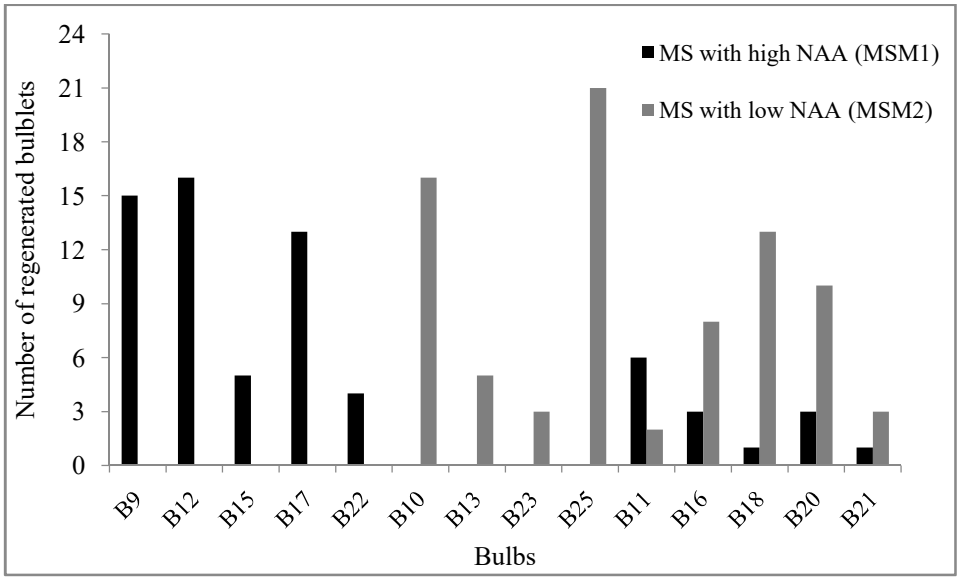

Fig. 2.6 Effect of different media combinations and bulbs on callus differentiation (bulblets/shoots) after sixteen weeks of incubation. Multiple bars indicating the response of bulbs in both media with the absence of bars indicating no callus differentiation was obtained, due to either contamination or senescence.

2.3.6 Effect of different media and growth regulators on direct bulblet and shoot initiation from explants

Figure 2.7 shows that the initiation of small bulblets directly grown from the base of twin-scale explants, was influenced by low concentration of auxin (MSM2) (51% of total explants inoculated in all three media) and MS basal medium (48% of total explants inoculated in all three media). Whereas, a lower percentage (11% of total

explants inoculated in all three media) were obtained from high auxin (MSM1) medium (Figure 2.7).

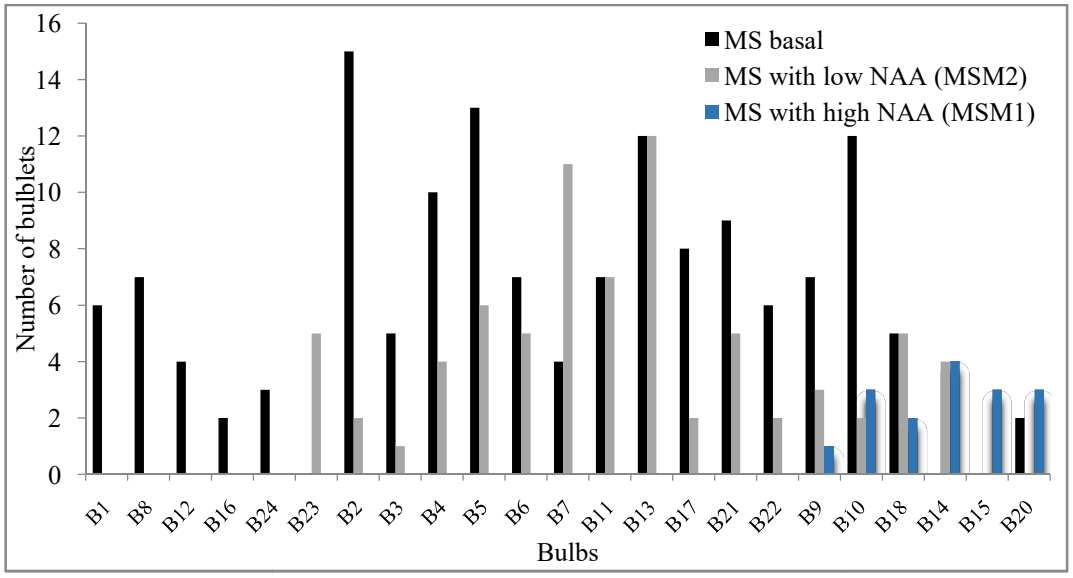

Fig. 2.7 Effect of different media combinations and bulbs on direct bulblet initiation after twelve weeks of incubation. Multiple bars indicating the response of bulbs in multiple media with the absence of bars indicating no bulblets were obtained, due to either contamination or senescence.

In addition green (total 51 shoots) and white (total 160 shoots) shoots were found on bulblets from MS basal and the low auxin (MSM2) medium, whereas the high auxin medium (MSM1) did not give any green shoots, only white shoots (total 76 shoots) were observed.

2.4 Discussion

2.4.1 Twin-scale explants

The results showed that twin-scales could be used as effective explants for *in vitro* callus, shoot and bulblet production in *N. pseudonarcissus* cv. Carlton, as reported for other species of the same genus (Sharma and Kanwar, 2002; Anbari *et al.,* 2007; Taleb *et al.,* 2013).

2.4.2 Pre-culture treatments

The pre-culture treatments of Carlton bulbs with hot water (52 ^{o}C, 1 h) followed by bleach sterilisation showed reduced contamination, senescence and maximum explants differentiation in culture compared to the other two treatments (45 ^{o}C, 2 h and 54 ^{o}C, 1 h). Similar bulb treatments were reported in previous studies; hot water treatment of *Narcissus* bulbs prior to surface sterilisation with bleach solution has been reported as a mandatory step to reduce contamination in culture condition (Sochacki and Orlikowska, 2005; Abu Zahra and Oran, 2007). Sochacki and Orlikowska, 2005, used 54 ^{o}C, 1h for the treatment of *N. tazetta* bulbs in their study, while in the current study this resulted in higher senescence. On the other hand, pre-culture treatment with low temperature for longer duration (44.4 ^{o}C, 3 h) has been used for Carlton bulb sterilisation (Abu Zahra and Oran, 2007). The similar treatment (45 ^{o}C, 2 h) in this study showed higher contamination of explants in culture.

Results indicated that the moderate temperature (52 ^{o}C, 1h) is the most suitable one for bulb treatment, as high temperature might kill the cells (most explants senesced) and a low temperature is not enough to kill contaminants in the tissues (most explants contaminated), even when bulbs were treated for longer period of time (2h).

2.4.3 *In vitro* tissue differentiation

In this study a series of tissues were obtained from the bulb explants, namely callus and bulblets (with white or green shoots) and some roots. Bulblets formed in two ways either directly from the explants or via a callus intermediary stage.

Calli were observed after 4 to 8 weeks and regenerated bulblets from callus were found after 12 to 14 weeks of culture incubation. Formation of small bulblets was observed after 8 weeks of culture on MS media supplemented with auxin and

cytokinin, from the base of scale explants of *N. papyraceus* cv. Shirazi (Anbari *et al.,* 2007). After 9 weeks of culture, tiny bulb-like structures appeared at the base of the twin-scale explants in *N. asturiensis* (Santos *et al.,* 2002). The similar duration (4 to 8 weeks) was also required for small bulblets formation from twin-scale explants of *N. pseudonarcissus* cv. Carlton in this study.

2.4.4 Callus induction

The study showed that a medium containing 20 mg/l NAA + 1.5 mg/l BAP + 0.5 mg/l KN was the most suitable medium for callus induction from twin-scale explants of *N. pseudonarcissus* cv. Carlton. This has a high auxin: cytokinin ratio. Some recent studies showed the similar findings; MS medium containing high auxin was the most suitable medium for callus induction and proliferation derived from bulb explants of *N. tazetta* var. *italicus* (Taleb *et al.,* 2013). Calli derived from bulb and leaf explants of *Narcissus tazetta* var. *italicus* showed the highest growth rate and proliferation when cultured on MS medium supplemented with high NAA (3 mg/l) and low BAP (1.5 mg/l) (Taleb *et al.,* 2013). The best survival, callusing and bulb formation in *N. pseudonarcissus* have been reported when *in vitro* grown shoot tissues were sub-cultured in MS media supplemented with high auxin (2, 4-D), low cytokinin (BAP) and 6% sucrose (El Tahchy *et al.,* 2011a). In our study the medium with high auxin combined with low cytokinin showed better callus induction and proliferation.

Interestingly, Carlton twin-scale explants completely failed to induce callus on MS medium (without growth regulators) which was previously reported in *N. confusus in vitro* cultures from media lacking auxins (Selles *et al.,* 1999). That suggests both auxin and cytokinin are mandatory for cell growth and differentiation (George *et al.,* 2008) and higher concentration of auxins facilitate the callus formation.

2.4.5 Callus differentiation

Adding growth regulators, both auxins and cytokinins to MS basal medium showed the best results for callus differentiation to either shoots or small bulblets. The presence of growth regulators was necessary for callus differentiation as no differentiation or callus induction was observed in MS basal medium deprived of growth regulators, which has previously been reported in *N. pseudonarcissus* (El

Tahchy *et al.,* 2011a). The maximum regeneration of small bulblets from callus was obtained using media supplemented with BAP in Chinese *Narcissus* (Chen *et al.,* 2005). These previous studies strongly support our results of obtaining maximum callus differentiation into small bulblets or shoots from both modified MS media containing different combinations and concentrations of auxins and cytokinins.

2.4.6 Shoot initiation or bulblet formation

Our results showed that the best shoot formation or small bulblets initiation, directly from the base of twin-scale without any callus phase, was obtained from either MS basal medium or MS with low auxin. On the other hand, a few bulblets or shoots were found in the MS medium with high auxin. Similar results were observed in several previous studies of *Narcissus* such as high amounts of cytokinin (5 mg/l BAP) and low amounts of auxin (1 mg/l 2, 4-D) stimulating bulb formation in *N. confusus* (Selles *et al.,* 1997). The best result for shoot proliferation was found in MS medium supplemented with low auxin, especially a NAA to a high cytokinin from *in vitro* bulb scale culture of *N. asturiensis* (Santos *et al.,* 2002). The highest number of adventitious shoots was obtained from MS medium supplemented with high concentration of BAP to low NAA in *N. tazetta* (Abu Zahra and Oran, 2007). That indicates a lower concentration of auxin facilitates the function of cytokinin in inducing differentiation such as shoot formation (George *et al.,* 2008).

Initiation of some white shoots with green parts were also recorded from our study. The green parts (photosynthetic units) of those shoots were separated for use in the metabolomic study i.e. to contrast with white (non-photosynthetic) shoots.

2.4.7 Root development

Only few small plantlets with roots were observed, all from MS medium lacking growth regulators. Similar findings on root development were reported previously in *Narcissus* (Abu Zahra and Oran, 2007; Anbari *et al.,* 2007).

Various tissues including organised, unorganised, photosynthetic and non-photosynthetic have been obtained from *N. pseudonarcissus* cv. Carlton to provide a way to study plant biology related to secondary metabolite production, alkaloid accumulation and associated gene expression.

3 Chapter Three: Alkaloids in bulb, basal plate and tissue culture materials

3.1 Introduction

Most of the Amaryllidaceae alkaloids are obtained from their natural sources and only a few are produced synthetically. The lack of knowledge about their biosynthesis, regulation and transport in natural system, makes their commercial production challenging. The aim of this chapter is to describe a protocol of alkaloid analysis (GC-MS) from a well known natural source of Amaryllidaceae alkaloids, *N. pseudonarcissus* cv. Carlton, comparing two different growth conditions (field versus tissue culture). Field samples of *Narcissus* cv. Andrew's Choice, a variety reported to contain a low level of Gal, also have been included to compare with Carlton field samples. The tissue culture materials (callus, direct bulblets, direct white shoots, green shoots and regenerated bulblets) obtained from Chapter 2 have been used for alkaloid analysis. An authentic galanthamine standard has been used to determine the quantity of galanthamine. Other alkaloids present in these materials has also been assessed based on MS spectra. In addition, to investigate whether different growth regulators and elicitors enhance production of galanthamine by *in vitro* cultures of Carlton is another focus of this chapter.

3.1.1 Biosynthesis of Amaryllidaceae alkaloids

Appendix 3.1 (on disc) contains a list of the major alkaloids that have been identified in the genus *Narcissus*.

3.1.1.1 Initial biosynthetic reactions

The biosynthesis of Amaryllidaceae alkaloids starts with the enzymatic conversation of aromatic amino acids, phenylalanine and tyrosine derived from primary metabolism (Berkov *et al.,* 2014; Singh and Desgagné-Penix, 2014). Both are closely related in their chemical structures but they are not interchangeable (Bastida *et al.,* 2006).

The enzyme PAL catalyses the elimination of ammonia to generate *trans*-cinnamic acid as shown in Figure 3.1 (Singh and Desgagné-Penix, 2014). The conversion of phenylalanine to the C_6C_1 precursor 3, 4-dihydroxybenzyldehyde (3, 4-DHBA) requires the loss of two carbon atoms from the side chain and the introduction of at least two oxygenated subunits into the aromatic ring (Kilgore and Kutchan, 2016),

performed by two hydroxylation reactions, catalysed by cytochrome P450s, cinnamate-4-hydroxylase (C4H) and coumarate-3-hydroxylase (C3H). Tyrosine is decarboxylated to tyramine by tyrosine decarboxylase (TYDC) before incorporation into the Amaryllidaceae alkaloids (Bastida *et al.,* 2006; Takos and Rook, 2013; Singh and Desgagné-Penix, 2014).

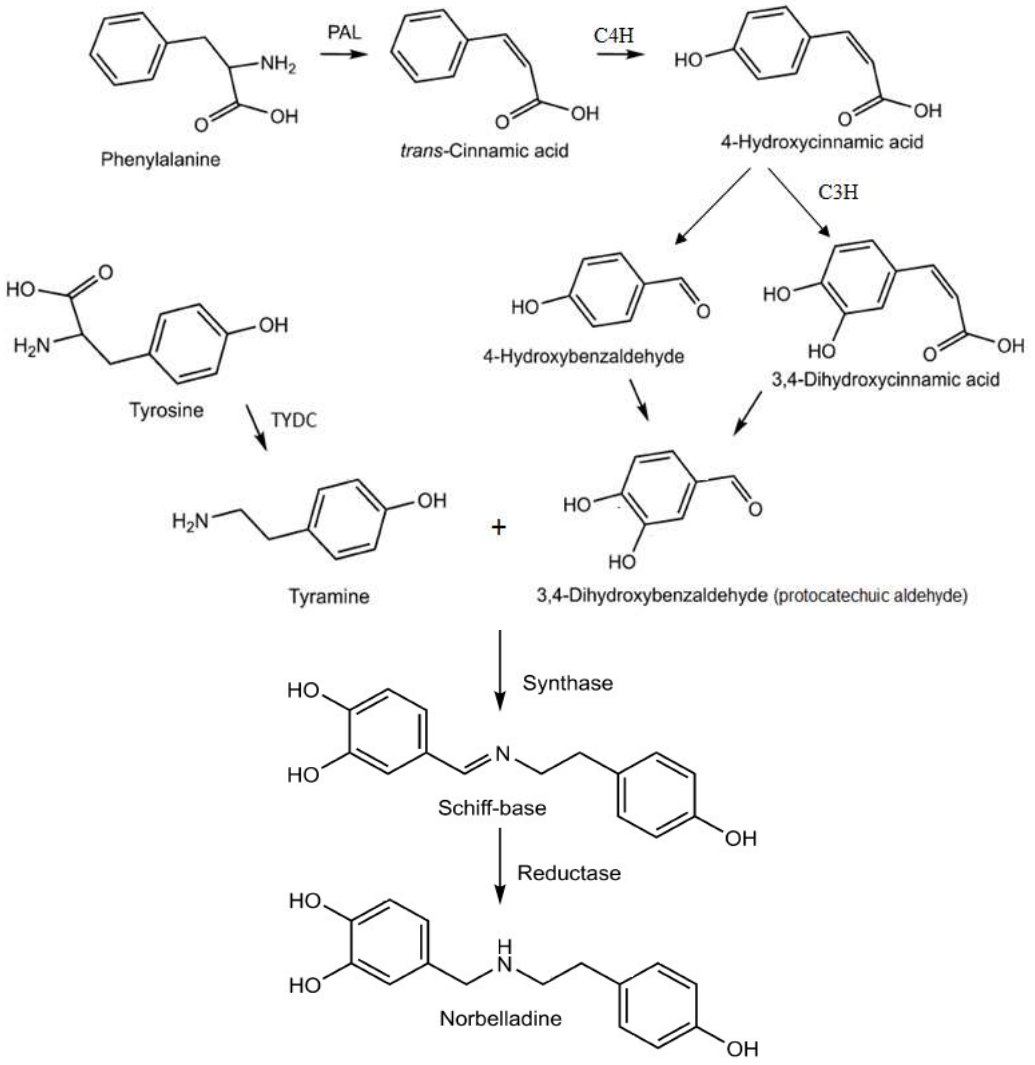

Fig. 3.1 Biosynthetic pathway from the precursors phenylalanine and tyrosine to norbelladine. PAL = phenyl ammonia lyase, TYDC = tyrosine decarboxylase, C4H = cinnamate-4-hydroxylase, C3H = coumarate-3-hydroxylase. (Adapted from Kilgore and Kutchan, 2016; Singh and Desgagné-Penix, 2014; Bastida *et al.,* 2011).

3.1.1.2 Biosynthesis of norbelladine

The enzymatically catalysed condensation of 3, 4-HDBA and tyramine results in the formation of a predicted Schiff-base intermediate through a reaction that can spontaneously occur in solution and which, following reduction, yields norbelladine

as illustrated in Figure 3.1 (Dewick, 2009; Singh and Desgagné-Penix, 2014). Similar condensation reactions have been observed in the biosynthesis of other benzylisoquinoline alkaloids, namely the combination of two tyrosine derivatives, 4-hydroxyphenylacetaldehyde and dopamine by a Pictet-Spengler condensation catalysed by norcoclaurine synthase to produce the alkaloid norcoclaurine (Luk *et al.,* 2007). Another example of Pictet-Spengler condensation has been reported in *Rauvolfia serpentina* to combine tryptamine and secologanin catalysed by strictosidine synthase resulting in strictosidine, the general precursor of monoterpenoid indole alkaloids (Stöckigt *et al.,* 2008).

The Schiff-base exists as three interchanging isomeric structures and the condensation is followed by a reduction of the imine double bond to generate norbelladine. The predicted reductase catalysing this reaction could be an aldo-keto reductase (AKR) or short-chain dehydrogenase/ reductase (SDR) (Kilgore and Kutchan, 2016). Tetrahydroalstonine synthase (a short chain alcohol dehydrogenase) has been reported in *C. roseus* to reduce the imine bond in strictosidine aglycone to form tetrahydroalstonine (Stavrinides *et al.,* 2015).

3.1.1.3 Oxidative phenol coupling

O-methylation reactions are frequent in alkaloid biosynthesis and mostly catalysed by *S*-adenosyl-*L*-methionine (SAM) dependent methyltransferases. *O*-methylation of norbelladine precedes the oxidative phenol coupling which leads to the various structural types of Amaryllidaceae alkaloids (Singh and Desgagné-Penix, 2014). Methylation of norbelladine to 4'-*O*-methylnorbelladine by a 4'-*O*-methyltransferase (*Np*N4OMT) as shown in Figure 3.2 has already been reported in *N. pseudonarcissus* (Kilgore *et al.,* 2014). A crucial step in Amaryllidaceae alkaloid biosynthesis is the cyclisation of this central intermediate, 4'-*O*-methylnorbelladine by oxidative phenol coupling, which can occur in *ortho-para'*, *para-ortho'* and *para-para'* positions (Takos and Rook, 2013; Kilgore *et al.,* 2016a). Several enzymes involved in these reactions have recently been discovered and most were found to be members of the large plant cytochrome P450 family (CYP450s) (Mizutani and Sato, 2011; Chávez *et al.,* 2011; Constantin *et al.,* 2012; Kilgore *et al.,* 2016a).

3.1.1.3.1 Alkaloids proceeding from *ortho-para'*

The alkaloids originating from *ortho-para'* phenol oxidative coupling are mainly lycorine and homolycorine types (Bastida *et al.,* 2006) and their proposed biosynthetic pathway is shown in Figure 3.2.

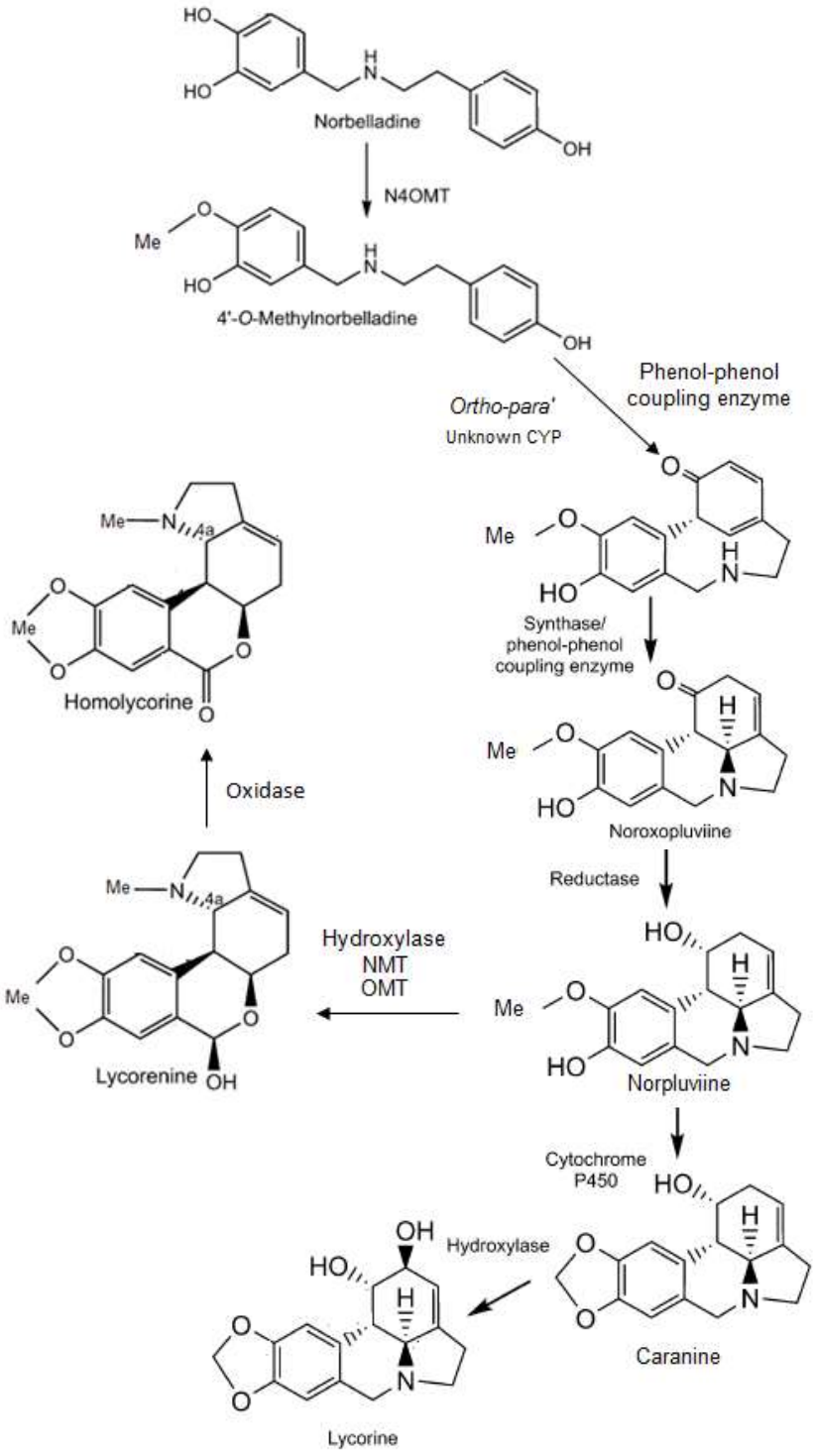

Fig. 3.2 *Ortho-para'* phenol-phenol coupling derived Amaryllidaceae alkaloids. N4OMT = 4'-*O*-methyltransferase, OMT = *O*-methyltransferase, NMT = *N*-methyltransferase (Adapted from Kilgore and Kutchan, 2016; Kilgore *et al.,* 2016a).

Intramolecular coupling of 4'-*O*-methylnorbelladine to form noroxopluviine in lycorine biosynthesis is potentially performed by a cytochrome P450 (Kilgore *et al.,* 2016b). Noroxopluviine is further reduced to form norpluviine. Evidence for this is that, tritiated norpluviine is converted to tritiated lycorine by *Narcissus* cv. Deanna Durbin which indicates that the carbon-2 hydroxyl group of lycorine is derived by allylic oxidation of either norpluviine or caranine (Bastida *et al.,* 2006). Norpluviine was primarily converted to homolycorine type alkaloids in *Narcissus* cv. King Alfred (Bastida *et al.,* 2011; Takos and Rook, 2013).

Figure 3.2 also indicates the predicted types of enzymes involved in the formation of homolycorine from norpluviine namely *N*-methyltranferase (NMT), *O*-methyltransferase (OMT), hydroxylase and an oxidase (Kilgore and Kutchan, 2016). Colclaurine *N*-methyltransferase and caffeine synthase are examples of *N*-methyltransferases that could share homology with the *N*-methyltransferases involved in several Amaryllidaceae alkaloid biosynthesis pathways (Kato *et al.,* 2000; Choi *et al.,* 2002). Homologues of *O*-methyltransferase could also be of potential interest as close homology exists between *N* and *O*-methyltransferases (Raman and Rathinasabapathi, 2003).

3.1.1.3.2 Alkaloids proceeding from *para-para'*

Several Amaryllidaceae alkaloids including the skeleton types of crinine, haemanthamine, tazettine, narciclasine and montanine are derived from *para-para'* C-C phenol coupling and this is illustrated in Figure 3.3 (Kilgore *et al.,* 2016a). An enantiomeric mixture of (10b*R*, 4a*S*) and (10b*S*, 4a*R*)-noroxomaritidine made by a cytochrome P450 *(CYP96T1)* from 4'-*O*-methylnorbelladine has recently been reported in *N. pseudonarcissus* by Kilgore *et al.,* (2016a), where (10b*R*, 4a*S*)-noroxomaritidine serves as the precursor for haemanthamine and others while (10b*S*, 4a*R*)-noroxomaritidine is intermediate in the formation of crinine only.

The predicted step from (10b*R*, 4a*S*)-noroxomaritidine was reduction of the ketone group to synthesize 8-*O*-demethylmaritidine followed by an oxide bridge formation to form vittatine. Vittatine is thought to be converted to haemanthamine through hydroxylation followed by methylation (Figure 3.3). Vittatine is implicated as an intermediate in the biosynthesis of narciclasine and the loss of the ethane bridge

could occur by a retro-Prins reaction (Bastida *et al.,* 2006). The retro-Prins reaction would result in the carbon-2 hydroxyl, cleavage of the 10b-11 bond and migration of the 1-2 carbon double bond to 10b-1 (Kilgore and Kutchan, 2016).

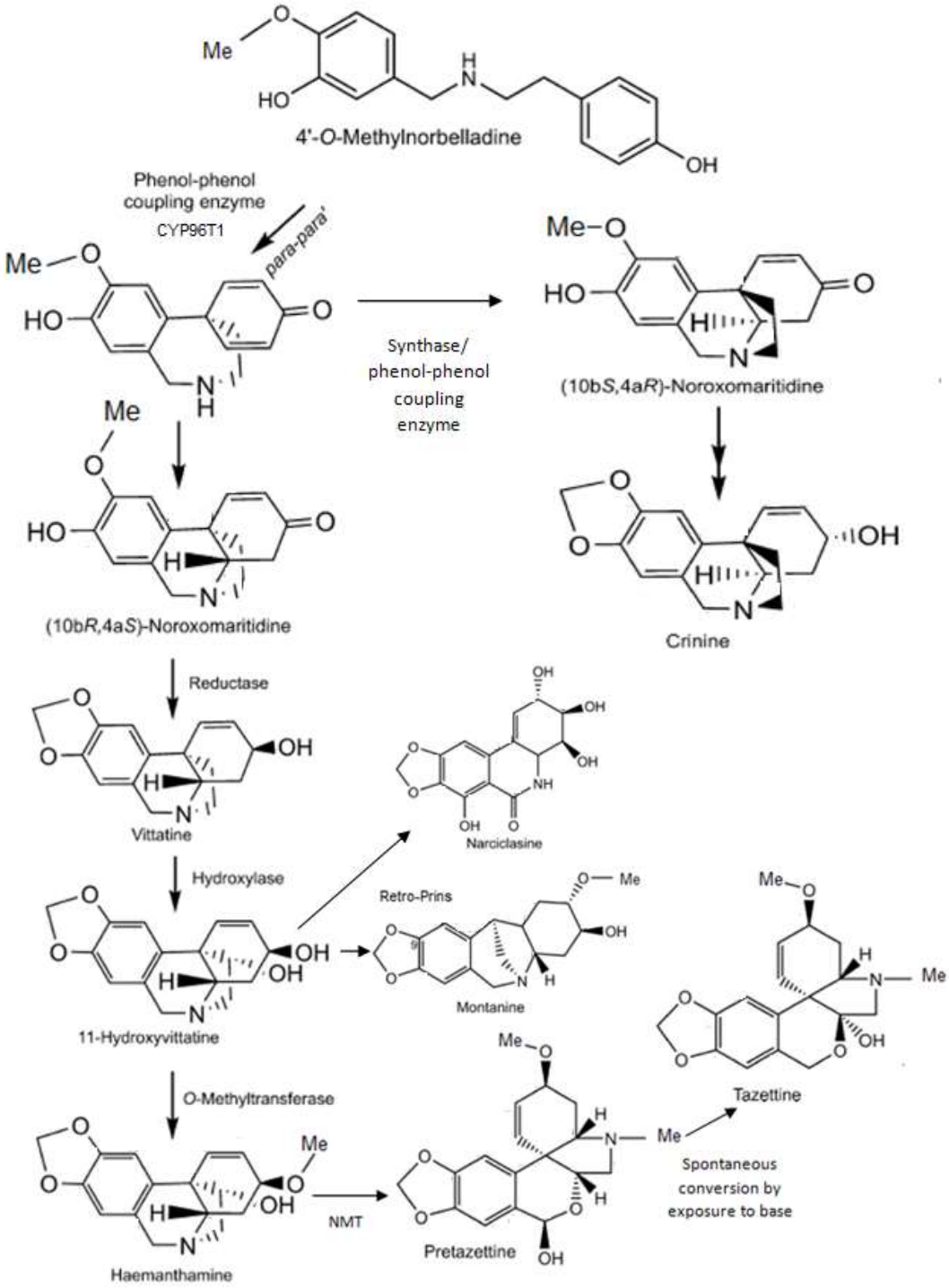

Fig. 3.3 *Para-para'* phenol-phenol coupling derived Amaryllidaceae alkaloids (Adapted from Kilgore and Kutchan, 2016; Kilgore *et al.,* 2016a).

11-hydroxyvittatine has also been proposed as an intermediate in the biosynthesis of haemanthamine via an *O*-methyltransferase and to montanine through the retro-Prins reaction (Figure 3.3) (Kilgore *et al.,* 2016a; Bastida *et al.,* 2011). Hydroxylation of haemanthamine resulted in the formation of haemanthadine and further methylation by an *N*-methyltransferase is supposed to produce pretazettine and then tazettine by spontaneous conversion (Kilgore and Kutchan, 2016).

3.1.1.3.3 Alkaloids proceeding from *para-ortho'*

The galanthamine type alkaloids have a dibenzofuran nucleus, obtained from *para-ortho'* phenol oxidative coupling (Bastida *et al.,* 2011). The biosynthesis of galanthamine as shown in Figure 3.4, involves the phenol oxidative coupling of 4'-*O*-methylnorbelladine to form a postulated dienone. Spontaneous closure of an ether bridge would then yield *N*-demethylnarwedine which is subsequently reduced to *N*-demethylgalanthamine (also called norgalanthamine), which been reported in *L. aestivum* (Eichhorn *et al.,* 1998). In the final step, *N*-demethylgalanthamine is *N*-methylated by an *N*-methyltransferase to galanthamine.

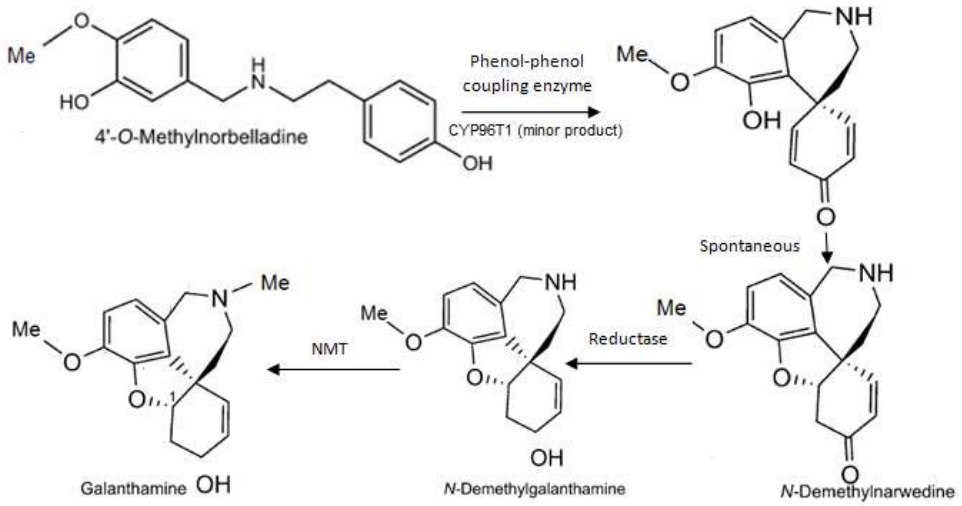

Fig. 3.4 *Para-ortho'* phenol-phenol coupling derived Amaryllidaceae alkaloids.
NMT = *N*-methyltransferase (Adapted from Kilgore and Kutchan, 2016; Kilgore *et al.,* 2016a).

Several recent reviews support this proposed biosynthestic pathway of galanthamine (Eichhorn *et al.,* 1998; Takos and Rook, 2013; Kilgore *et al.,* 2014). It has been reported that the oxygen binding and activativation motif is substantially different between *CYP96T1* and to other hydroxylation catalyzing cytochrome P450 enzymes; the low amount of *N*-demethylnarwedine produced rather than noroxomaritidine by

CYP96T1, which was identified from *Narcissus* transcriptome data, proved that *N*-demethylnarwedine is not the enzyme's primary product (Kilgore *et al.,* 2016a).

3.1.1.4 Chemical synthesis of galanthamine

The current state of the art concerning the chemistry and biology of galanthamine has been reviewed postulating two key reaction protocols; the approach using phenolic oxidative coupling as described in section 3.1.1.3.3 (Figure 3.4) and an alternative via an intramolecular Heck reaction (Marco-Contelles *et al.,* 2006). The formal synthesis of lycoramine, a haemanthamine type alkaloid (Appendix 3.1) produced through *para-para'* phenol coupling reaction by using an intramolecular Heck reaction (Gras *et al.,* 1999), suggested the total synthesis of racemic galanthamine could follow a similar strategy (Guillou *et al.,* 2001) where an intermolecular reaction was used as an alternative for creating the spiro quaternary carbon atom of galanthamine type alkaloids.

The (-)-galanthamine is the form made biologically and isolated from natural sources. A stereo-controlled synthesis of (-)-galanthamine described the involvement of eleven predicted linear steps starting from isovanillin (Satcharoen *et al.,* 2007). Another study described a non-bio mimetic synthetic route for the total synthesis of optically active (+)-galanthamine, starting from D-glucose using Ferrier's carbocyclisation reaction, which can then be converted to (-)-galanthamine (Tanimoto *et al.,* 2007).

3.1.1.5 Galanthamine biosynthesis (field versus tissue culture)

Plants grown in natural conditions generally show a higher accumulation of secondary metabolites than when cultured *in vitro*. A requirement for the enzymes of secondary metabolism to be produced in specific, organised tissues could be a reason for lower secondary metabolites production in tissue culture as it often results into unorganized or undifferentiated tissues (Karuppusamy, 2009; Dias *et al.,* 2016). In the case of galanthamine biosynthesis, higher accumulation in plants from nature than *in vitro* culture has been reported in many studies. For example, leaf and bulb tissues of *Narcissus* field grown ornamental varieties contained 200-4500 µg/g DW Gal (Torras-Claveria *et al.,* 2013) while *N. pseudonarcissus* shoot cultures contained 10-1000 µg/g DW Gal. Another study showed galanthamine production between

2000-4000 µg/g DW in *N. confusus* shoot-clump culture (Colque *et al.,* 2004). The levels in these organised tissues are therefore within the range of field grown tissues, despite being grown in culture.

Nevertheless, the callus cultures often showed very low amount of galanthamine production (0.03-0.11 µg/g DW) in *Narcissus* (Codina, 2002). The interesting question therefore, is how does differentiated tissue differ from callus so that alkaloid biosynthesis can be achieved.

3.1.2 Elicitor treatment during *in vitro* culture

3.1.2.1 Elicitors and elicitation

Elicitation is the process to enhance production or accumulation of metabolites by *in vitro* production systems due to the addition of trace amounts of elicitors (Radman *et al.,* 2003; Dias *et al.,* 2016). Several parameters such as selection of elicitor and concentration, duration of exposure can influence any enhanced production of secondary metabolites (Namdeo, 2007). An appropriate concentration of elicitors and their duration of exposure, play an important role in the elicitation process of plant culture systems. For example, high accumulation of ajmalicine has been reported from *C. roseus* cultures when treated with medium concentrations of fungal elicitors (5.0%) while very low (0.5%) concentrations resulted in lower production and high (10.0%) concentrations had an adverse affect on ajmalicine production (Namdeo, 2002). Several studies showed that high doses of elicitors result in a hypersensitive response leading to cell death, whereas an optimum concentration resulted in successful growth and secondary metabolite accumulation (Roewer *et al.,* 1992; Colque *et al.,* 2004).

3.1.2.2 Elicitors and Amaryllidaceae alkaloid production

Elicitors are complex mixtures of chemicals obtained from either fungal or bacterial extracts, or pure chemical compounds or physical treatments, and have shown effects on alkaloid production in several Amaryllidaceae species including *Narcissus* with enhanced production of alkaloids (Ivanov *et al.,* 2013; Saliba *et al.,* 2015).

The effects of a variety of fungal elicitors on alkaloid production from *N. tazetta* var. *italicus* tissue culture have been investigated. It was observed that culture filtrates

from *F. sporotrichioides* were more effective than other fungal elicitors, showing enhanced growth and total alkaloid production in four weeks old callus cultures (Taleb *et al.,* 2013). Purified fungal components, such as the cell-wall component chitosan, have also been used as more convenient and reproducible fungal elicitors.

Methyl jasmonate, jasmonic acid and salicylic acid have frequently been reported for the production of phenolic compounds by *in vitro* cultures (Mewis *et al.,* 2011) where methyl jasmonate and jasmonic acid induce genes responsible for the synthesis of the enzymes involved in secondary metabolic pathways (Zhao *et al.,* 2005). The phenylpropanoid pathway also requires the action of phenylalanine ammonia lyase to generate cinnamic acid from phenylalanine as an initial step in this branch between primary and secondary metabolism (Dixon and Paiva, 1995), so these elicitors have been tested for effects on alkaloid biosynthesis in several plant species. Methyl jasmonate stimulates enhanced alkaloid production in *Argemone mexicana* cell cultures (Trujillo-Villanueva *et al.,* 2010), *C. roseus in vitro* cultures (Vázquez-Flota *et al.,* 2009), and *Echium rawolfii* hairy root cultures (Abd El-Mawla, 2010). In addition, it was found the best elicitor for galanthamine biosynthesis in seedlings of *L. chinensis* (Mu *et al.,* 2009).

The galanthamine and phenolic acid content of *in vitro* shoot cultures of *L. aestivum* due to the addition of methyl jasmonate and jasmonic acid were investigated (Ivanov *et al.,* 2013). This showed that maximal contents of galanthamine and lycorine were achieved after elicitation with jasmonic acid (5.6 mg/l; day 28 to 35) due to induction of tyrosine decarboxylase activity. Methyl jasmonate (5.6 mg/l; day 28 to 35) also stimulated the yield of *N*-demethylgalanthamine (Figure 3.4), the direct precursor of galanthamine, as well as phenolic acids due to the induction of the activity of phenylalanine ammonia lyase.

Four elicitors (methyl jasmonate, salicylic acid, copper sulphate and silver nitrate) were tested to determine their effects on growth and alkaloid production in *L. aestivum* shoot cultures (Schumann *et al.,* 2013). Seven Amaryllidaceae alkaloids including tyramine, vittatine, narwedine, hordenine, trisphaeridine, *N*-demethylgalanthamine and galanthamine were determined by GC-MS in this study. Methyl jasmonate (22.4 mg/l; 7 weeks of culture) resulted in significantly improved production of galanthamine in both leaves and bulblets whereas the other

elicitors inhibited production of galanthamine. In a separate study, liquid-shake cultured shoot-clumps of *N. confusus* were treated with four elicitors; methyl jasmonate (12 to 112 mg/l), chitosan (50 to 500 mg/l), salicylic acid (14 to 138 mg/l) and arachidonic acid (1 to 10 mg/l). The results showed that high concentrations of these elicitors had an adverse effect on explant growth in general but methyl jasmonate resulted in increased galanthamine and other related alkaloids (Colque *et al.,* 2004).

Several studies have shown the effect of elicitor duration on the growth and accumulation of different Amaryllidaceae alkaloids in culture systems; in most cases the accumulation of high amount of alkaloids were observed after two to six weeks of elicitor exposure (Colque *et al.,* 2004; Ptak *et al.,* 2008; Schumann *et al.,* 2013).

3.1.2.3 Feeding precursors to produce Amaryllidaceae alkaloids

Precursors are chemical compounds, which precede other compounds in a metabolic pathway. The precursors of Amaryllidaceae alkaloids are tyrosine, phenylalanine, tyramine, *trans*-cinnamic acid, norbelladine and 4-*O*-methylnorbelladine (Figure 3.1 and 3.2).

The addition of a precursor often causes higher yields of Amaryllidaceae alkaloids. As for example, concentrations of a well-known Amaryllidaceae alkaloid precursor, *trans*-cinnamic acid (TCIN) (Figure 3.1) ranging from 250 to 1000 mg/l were used to investigate yields of alkaloids in *N. confusus* shoot clump cultures. The highest dose (1000 mg/l TCIN, after two weeks of culture) showed increased production of galanthamine and *N*-formyl-galanthamine although this level inhibited growth of cultures (Bergoñón *et al.,* 1996).

A central precursor of Amaryllidaceae alkaloids is 4'-*O*-methylnorbelladine (Figure 3.2). Studies using deuterium labelling have shown this can be taken up by cultured cells and transformed into Amaryllidaceae alkaloids (Tahchy *et al.,* 2010; El Tahchy *et al.,* 2011b). Incorporation of 4'-*O*-methylnorbelladine into the liquid medium of *L. aestivum* shoot cultures showed highly stimulated biosynthesis of galanthamine (0.5 mg/g DW) and lycorine (0.2 mg/g DW). The maximum production of both alkaloids was observed with precursor concentration of 0.1 g/l after 15 days of culture (Saliba *et al.,* 2015). Another similar study showed increased concentrations of galanthamine

(0.81 mg/g DW) and lycorine (0.54 mg/g DW) over the control (about 0.2 mg/g DW) in *L. aestivum* bulblet cultures derived from leaf explants by providing 4'-*O*-methylnorbelladine (0.3 g/l) for 30 days combined with a physical elicitation (temporary immersion conditions) (Saliba *et al.,* 2016).

Yeast extract is a convenient fungal elicitor and also serves as the major source of nitrogen in culture media (Molnár *et al.,* 2011). It contains the amino acids used as precursors in Amaryllidaceae alkaloid production specifically, 100-118 mg/g nitrogen, 25-26 mg/g phenylalanine and 8-11 mg/g tyrosine (www.bd.com).

These findings of elicitor treatment and precursor feeding were considered for the selection of elicitors/precursors, their concentrations and duration of treatment in this study. Methyl jasmonate (MJ) (5.6 mg/l and 22.4 mg/l); chitosan (100 mg/l); TCIN (1000 mg/l) and an increased amount of yeast extract (800 mg/l) than normally used for culture media (100 mg/l) (Chapter 2, section 2.2.2, Table 2.1) were selected for the treatment of *in vitro* derived callus and regenerated bulblets (Chapter 1). A treatment period of 30 days also has been selected based on previous findings.

3.1.3 Alkaloid analysis methods

Several chromatographic methods have been published for the determination of alkaloids in biological samples including plants (Shawky *et al.,* 2015). The most frequently used are thin-layer chromatography, high performance thin-layer chromatography, gas chromatography (GC), gas chromatography-mass spectrometry (GC-MS) (Shawky *et al.,* 2015), high performance-liquid chromatography (HPLC) (Ivanov *et al.,* 2013) and liquid chromatography-mass spectrometry (LC-MS) (Tahchy *et al.,* 2010). Chromatography methods combined with mass spectrometry is of great value in metabolomic studies as the compounds can be searched against databases for the inference of structure (Kilgore and Kutchan, 2016).

GC-MS is a rapid and robust method and requires very little plant material. It has been used for analysing galanthamine content in both leaves and bulb of 105 ornamental varieties of *Narcissus* (Torras-Claveria *et al.,* 2013; Berkov *et al.,* 2011). Alkaloids extracts of *Narcissus* were analysed by GC-MS, leading to the quantification of galanthamine and to the identification of other alkaloids (El-Tahchy *et al*., 2011a).

The data obtained by GC-MS is highly reproducible as it allows automated sample injection, consistent temperature and a highly standardised gas-flow (Kopka, 2006). GC-MS data analysis is relatively simple because of the standard ionization settings and reproducibility of spectra, while in LC-MS the fragmentation of compounds can greatly vary depending on the instruments and settings (Hopley *et al.,* 2008; Kilgore and Kutchan, 2016). In addition, GC-MS machines are easier to operate and have fewer maintenance issues than LC-MS. Solvents used in liquid chromatography are more expensive in comparison to gases used for gas chromatography and the machine maintenance cost is also high in liquid chromatography (Bhanot, 2014). GC-MS can be used for the detection of small plant metabolites or alkaloids. However, thermolabile metabolites with high molecular weight cannot be detected by GC-MS due to their limited volatility (Tolstikov and Fiehn, 2002). Several HPLC techniques such as hydrophilic interaction liquid chromatography (HILIC) could be an alternative method for the separation of highly polar compounds and reverse-phase liquid chromatography (RP-HPLC) methods for the separation of non polar compounds (Tolstikov and Fiehn, 2002).

Consequently, for all above reasons and due to the availability of GC-MS facilities in the Centre of Proteome Analysis, University of Liverpool, GC-MS was selected for alkaloid analysis in this project.

3.1.4 Basic principle of Gas Chromatography-Mass Spectrometry

GC-MS is divided into two parts; the gas chromatography separates the chemicals based on their volatility and the mass-spectrometry identifies and quantifies the chemicals present in a sample by comparing to an authentic standard (Hussain and Maqbool, 2014).

The samples are injected through an automated injection port, and carried by a non-reactive gas (helium or nitrogen). A high temperature of 250-300 ^{o}C is maintained inside the injection port to convert the chemicals to gases. After that, the gases are separated by partition with the stationary phase coating as they are carried through a capillary column, situated inside an oven. More volatile compounds (small molecules) move faster than less volatile (larger molecules) through the column (Kopka, 2006; Hussain and Maqbool, 2014). After passing through the column, the

chemicals continue to the mass-spectrometry, where they are ionised and fragmented resulting in positively charged ions. After that, they travel through an electromagnetic field containing a specialised filter, which filters the ions based on their mass. The numbers of ions with each selected mass is then counted by a detector and sent to a computer to provide the mass spectrum graphs (Hussain and Maqbool, 2014).

3.2 Materials and methods

3.2.1 Plant materials

Leaves, dormant bulbs (harvested in June/ July), bulbs harvested after flowering (harvested in April/ May) and basal plate tissues of dormant and harvested bulbs from field grown 'Carlton', harvested bulbs and basal plate from Andrew's Choice were cut into small pieces and stored at -80 ^{o}C. All tissue culture derived materials (callus, bulblets directly grown from twin-scale with green or white shoots, regenerated bulblets from callus and roots from small bulblets) were harvested (Chapter 2), weighed and stored at -80^{o}C until final extraction. Carlton bulb tissues were used as control for all sample extractions.

3.2.2 Sample preparation

For alkaloid extraction 100 mg frozen plant material was placed in 2 ml microfuge tube with 500 µl methanol (HPLC grade, Fisher Scientific, UK) and homogenised for 2 × 60 s (Pellet Pestle 1.5 ml Stainless Steel Pk1 with pestle motor, Kimble Chase, USA). Another 500 µl methanol was added and after vortexing for 30 s samples were incubated for 5 h at room temperature assisted by an ultrasonic bath (Grant Instruments Ltd. England) for 15 min of every 30 min. After extraction, the samples were centrifuged at 12,000 *g* for 1 min then 500 µl supernatant was removed slowly without disturbing the pellet to a labelled 2 ml glass vial (Thermo Scientific, Germany).

Extracts could be analysed in two ways depending on the availability of GC-MS machines. For immediate analysis, the freshly prepared sample (65 µl) could be directly transferred to the GC-MS vial (Chromacol Ltd. USA) with 2 µl (5 µg/µl) codeine as an internal standard. For later analysis, the organic solvent could be evaporated at room temperature in a fume hood and the dry extracts stored at -20 ^{o}C until the GC-MS machine was available. Dry extracts needed to be dissolved in methanol (500 µl) and after that the same procedure could be followed as for fresh extract analysis. Freshly prepared sample (65 µl) extracts were used for all GC-MS analysis in this study. Each alkaloid extract was analysed in triplicate.

3.2.3 Standard solutions

Galanthamine (Gal) was used as an external standard (ES) and codeine (Cod) was used as internal standard (IS) (Sigma-Aldrich, UK). Stock powders of galanthamine hydrobromide (2 mg) and codeine (50 mg) were dissolved in methanol (HPLC grade) for final concentrations of galanthamine (0.1 mg/ml) and codeine (5 mg/ml). Standard solutions (Table 3.1) of three different galanthamine concentrations (0.1 mg/ml, 0.05 mg/ml and 0.01 mg/ml) along with codeine (5 mg/ml) were used to obtain the calibration curve for method validation. Each standard concentration was used in three replicates for the Gal peak areas in total ion current (TIC) mode.

Table 3.1 Composition of standard solutions for alkaloid analysis (final volume of each standard = 61 µl).

Standard	Composition	Final Gal conc. (mg/ml)
Standard 1	60 µl Gal (0.1 mg/ml) + 1 µl Cod (5 mg/ml)	0.1
Standard 2	30 µl Gal (0.1 mg/ml) + 1 µl Cod (5 mg/ml) + 30 µl methanol	0.05
Standard 3	6 µl Gal (0.1 mg/ml) + 1 µl Cod (5 mg/ml)+ 54 µl methanol	0.01

3.2.4 Chromatographic conditions

The GC-MS mass spectra were recorded on a Micromass GCT instrument (Waters Ltd. UK) operating in positive ion EI mode with a source temperature of 180 ^{o}C and ionization energy of 70 eV. The chromatography column used was a non-polar general purpose BPX5 (30 m × 0.25 mm × 0.25 µm) (SGE Analytical Science). The temperature program was: 70 ^{o}C for an initial time of 2 min then an increase of 10 ^{o}C min^{-1} to 320 ^{o}C held for a final time of 8 min. The flow rate of the carrier gas (helium) was 0.7 ml min^{-1}. The injection temperature was 250 ^{o}C and the splitless injection mode was used. Aliquots of the extract solutions (1µl) were injected. Retention time for samples from dormant bulb, basal plate and green shoots showed a slight shift to high values due to a change of column to DB5M5 (30 m × 0.25 mm × 0.25 µm) (J & W Scientific, Agilent Science 2000).

3.2.5 GC-MS analysis

After plant sample and standard preparation, all samples were analysed in the Centre of Proteome Research, University of Liverpool using GC-MS. Automated injection of samples to the Micromass GCT instrument and data collection was performed by Mark Prescot, Centre of Proteome Research. All samples were run in a total of 10

batches; 7 batches for plant samples and 3 batches for media extracts, including 24 samples in a single batch. Standards were run with each batch of samples including at the start, in the middle after 12 samples and at the end of 24 samples.

3.2.6 Calculations and data interpretation

The equations generated from the calibration curves (section 3.3.1) of Gal standards (ES) were used to calculate the amount of galanthamine from all sample extracts. As the calibration curve analysis was based on external standards (Gal); calculations were more accurate and reproducible than internal standard (Cod) equations. The amount of galanthamine was calculated based on fresh weight (FW).

To generate calibration curves the Gal concentrations (0.1 mg/ml, 0.05 mg/ml and 0.01 mg/ml) were converted to their corresponding Gal molarities (0.0002671 M, 0.0001336 M and 0.00002671 M), which were calculated based on the molecular weight of galanthamine hydrobromide (368.27 g) used for the preparation of galanthamine solutions. Calibration curves and GC-MS chromatogram graphs were generated using Microsoft Excel 2010 and used to calculate the molarities of the samples. These molarities were then converted to the Gal content (µg) of per g fresh weight plant material. Galanthamine amounts in the Results section have been converted and are shown in µg Gal/g (FW).

3.2.7 Identification of alkaloids

Mass spectrometry of Amaryllidaceae alkaloids by electron impact was reported in the 1960s and 1970s and the fragmentation patterns are consequently well known (Bastida *et al.,* 2006).

Database software, 'the NIST mass spectral library' that was accessed via the Waters Mass Lynx software was used for the identification and quantification of galanthamine in samples under analysis. An authentic galanthamine standard (Table 3.1) was used for quantification purposes.

GC-MS chromatogram data for all samples under analysis for identifying other alkaloids or related compounds were also downloaded from the same database, 'the NIST mass spectral library'. The database contains the mass spectra; compound

structures, and retention index information that allow identifying compound in a single step.

3.2.8 Callus and regenerated bulblets cultured on elicitor treated media

3.2.8.1 Plant materials

Callus and regenerated bulblets from (MSM1) and (MSM2) (Table 2.1; Chapter 2) were used for the elicitor treatments to observe the effect of different elicitors on galanthamine production.

3.2.8.2 Elicitor treatments

All the compounds used as elicitors were purchased from Sigma-ALDRICH (UK). The elicitors and their concentrations (Table 3.2) used were selected according to the recommendation found in the literature (see section 3.1.2.2 and 3.1.2.3). Methyl jasmonate was purchased in liquid form and sterilised through a sterile filter of 0.22 µm, before adding to the autoclaved media. Chitosan was dissolved in 1:3 (HCl: sterile water) by heating and added to the media before autoclaving. *Trans*-cinnamic acid and yeast extract were added directly to the culture media as powder and then the media were autoclaved (121 ^{o}C, 108 kPa and 30 min).

Table 3.2 Media compositions (modified MSM1/ MSM2 with different elicitors) (Details of MSM1 and MSM2 media in Chapter 2, Table 2.1).

Name of media	Compositions
MJ1	MSM1/MSM2 + 5.6 mg/l methyl jasmonate
MJ2	MSM1/MSM2 + 22.43 mg/l methyl jasmonate
CH	MSM1/MSM2 + 100 mg/l chitosan
TCIN	MSM1/MSM2 + 1.0 g/l *trans*-cinnamic acid
YE	MSM1 + 0.8 g/l yeast extract

3.2.8.3 Culture conditions and duration

All elicitor containing sterile media were handled inside a laminar hood; two pieces of calli or regenerated bulblets (each ~200 mg) were placed on each Petri plate, sealed with a single layer of parafilm, incubated in a culture room (24±2 ^{o}C, 12 h photoperiod) for 30 days, including control (media with no elicitors; MSM1/MSM2).

3.2.8.4 Alkaloid extraction and analysis

After 30 days of incubation, the calli and regenerated bulblets were harvested and cut into small pieces, weighed (~100 mg) and transferred into 1.5 ml microfuge tubes. The media surrounding the calli (~5 cm) were also transferred to 5 ml plastic tubes and all samples stored at -80 ^{o}C. The media were vacuum dried and again stored in -80 ^{o}C. Vacuum dried media (~100 mg) were weighed and transferred to a 1.5 ml microfuge tube for extraction. For both tissues and media the same protocol of extraction, standard preparation, calibration curve analysis for calculations and GC-MS chromatographic conditions were used as described in sections 3.2.2 to 3.2.7.

3.3 Results

3.3.1 Calibration curve analysis

The calibration curves were prepared plotting the Gal peak areas (TIC) versus Gal molarities (0.0002671 M, 0.0001336 M and 0.00002671 M). Standards were analysed with the samples and showed high linearity over the range of Gal molarities (0.0002671M to 0.00002671 M) with correlation coefficient (R^2) of 0.902 to 0.954 across the batches of field and *in vitro* plant tissues and slightly lower in the analyses of *in vitro* media (R^2 = 0.785 to 0.889). A typical linear regression equation for standards is shown in Figure 3.5, where *x* represents Gal molarities and *y* is the total Gal peak area.

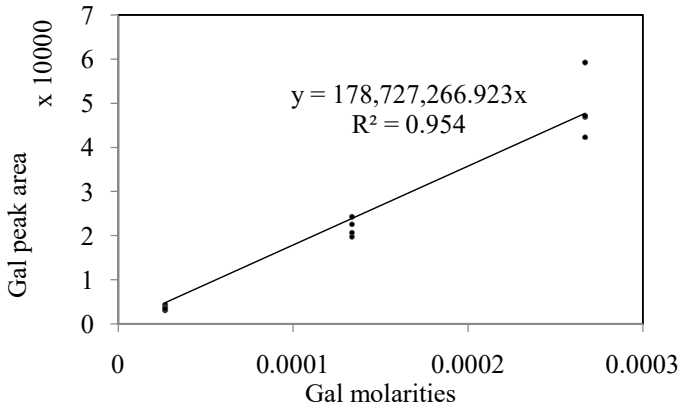

Fig. 3.5 Representative calibration curve of galanthamine standards for first batch of samples showing galanthamine molarities versus Gal peak area.

3.3.2 Quantification of Gal in field samples

The galanthamine content from methanol extracts of field samples were analysed by GC-MS using the method described in section 3.2 and the Gal amounts (Table 3.3) were calculated from the external standard (ES) equations (sections 3.2.6 and 3.3.1).

Table 3.3 Amount of Gal in *N. pseudonarcissus* cv. Carlton and Andrew's Choice field samples (values expressed as µg Gal/g FW ± SD, n = 3 individual bulbs).

Field samples	µg Gal/g FW ± SD
Bulb tissue (Carlton)	1110 ± 0.009
Basal plate tissue (Carlton)	1250 ± 0.005
Leaf tissue (Carlton)	420 ± 0.007
Dormant bulb tissue (Carlton)	960 ± 0.015
Dormant basal plate tissue (Carlton)	830 ± 0.017
Bulb tissue (Andrew's Choice)	630 ± 0.012
Basal plate tissue (Andrew's Choice)	1050 ± 0.006

Field (organised) tissues showed a higher amount of galanthamine than the culture derived samples on a fresh weight (FW) basis (Table 3.3 and 3.4). Among the field samples, Carlton basal plate tissue contained the highest (1250 µg Gal/g FW) amount followed by bulb (1110 µg Gal/g FW) and leaf tissues (420 µg Gal/g FW). Basal plate tissues (830 µg Gal/g FW) and bulb tissues (960 µg Gal/g FW) of Carlton dormant bulb showed slightly less galanthamine content than harvested bulbs (Table 3.3).

Andrew's Choice basal plate tissues (1050 µg Gal/g FW) and bulb tissues (630 µg Gal/g FW) also produced higher amount of galanthamine than tissue culture derived materials but less than Carlton bulb and basal plate (Table 3.3 and 3.4).

3.3.3 Quantification of Gal in tissue culture derived samples

Carlton tissue culture derived samples were also analysed for galanthamine content (Table 3.4) using GC-MS. Carlton bulb tissue was used as control for all tissue culture sample analysis.

Among the tissue culture derived samples, the callus contained low levels or no galanthamine. Table 3.4 shows there was a trace amount (1.0-7.0 µg Gal/g FW) in callus cultured on MS medium supplemented with high auxin but no galanthamine in callus obtained from low auxin medium. However, differentiation of callus to regenerated tissues (small bulblets or white shoots) grown on low auxin medium showed an increased galanthamine level (13-50 µg Gal/g FW).

Direct bulblets grown from the base of twin-scales and white part of their shoots contained less galanthamine (10-215 µg Gal/g FW) than the photosynthetic part (green shoots) (97 and 130 µg Gal/g FW) or the field grown materials. Galanthamine was absent from root extracts (Table 3.4).

Table 3.4 Amount of Gal in *N. pseudonarcissus* cv. Carlton tissue culture derived samples (values expressed as µg Gal/g FW ± SD, n=3).

Tissue culture derived samples	Bulbs No.	Media	µg Gal/g FW ± SD
Callus	Bulb 1	MS with high NAA	1.0 ± 0.0001
	Bulb 2	(20.0 mg/l)	7.0 ± 0.0007
	Bulb 3		7.0 ± 0.0005
Callus	Bulb 4	MS with low NAA	0.0 ± 0.0
	Bulb 5	(4.0 mg/l)	0.0 ± 0.0
	Bulb 6		0.0 ± 0.0
Regenerated bulblets from callus	Bulb 4	MS with low NAA	50.0 ± 0.002
	Bulb 5	(4.0 mg/l)	40.0 ± 0.002
	Bulb 6		0.0 ± 0.0
Regenerated white shoots from callus	Bulb 4	MS with low NAA	13.0 ± 0.001
	Bulb 6	(4.0 mg/l)	0.0 ± 0.0
	Bulb 7		0.0 ± 0.0
Direct bulblets from twin-scale	Bulb 1	MS basal	90.0 ± 0.008
	Bulb 2		10.0 ± 0.007
	Bulb 3		215 ± 0.030
Direct white shoots from twin-scale	Bulb 2	MS with low NAA	10.0 ± 0.0001
	Bulb 3	(4.0 mg/l)	30.0 ± 0.0001
	Bulb 6		10.0 ± 0.0001
Green shoots (Direct shoots)	pooled	MS basal	130.0 ± 0.006
Green shoots (Direct shoots)	pooled	MS with low NAA (4.0 mg/l)	97.0 ± 0.010
Roots	pooled	MS	0.0 ± 0.0

Bulbs are numbered consecutively to show the different materials obtained from the same bulb, e.g. callus and direct bulblets from twin-scale, represented in table for bulb 1 were obtained from the same bulb. The zero values (µg Gal/g FW ± SD) are not due to issues with the culture samples, the cultures were healthy but did not produce any galanthamine.

3.3.4 Identification of other alkaloids and/or related compounds from GC-MS spectra and their mass fragmentation patterns

Other suggested Amaryllidaceae alkaloids, their derivatives and related compounds were identified from the GC-MS chromatograms using GC-MS data system library search using database software 'the NIST mass spectral library' (section 3.2.7). Identification relies on this spectral library since other authentic alkaloids were not

available. In addition, their amounts were not possible to quantify here due to the lack of authentic standards for the specific compounds.

Along with galanthamine (Gal) and the external standard codeine (Cod); lycoramine (galanthamine type), crinamine (haemanthamine type), lycorine and lycorenine (homolycorine type) were found as the top matched compounds from the GC-MS chromatogram of Carlton bulb extracts (Figure 3.6A and 3.6B).

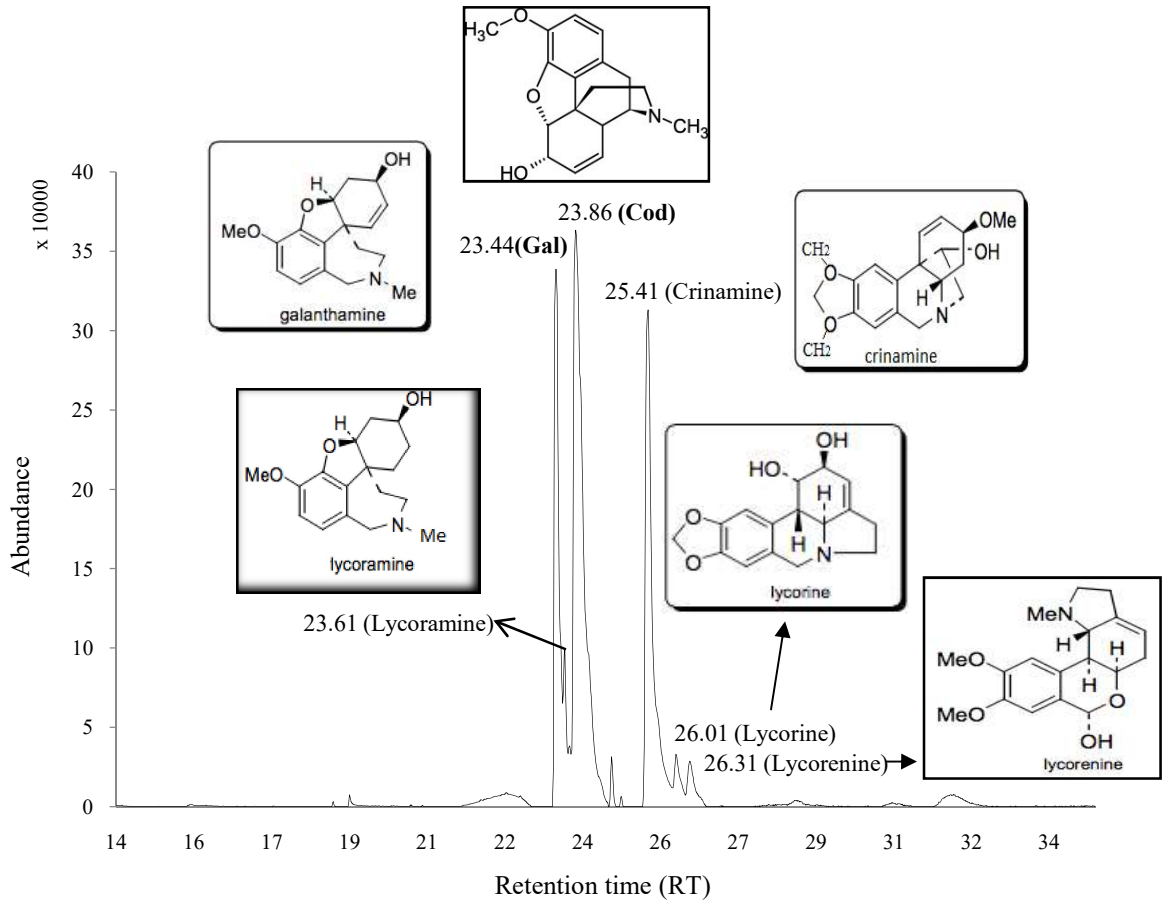

Fig. 3.6A GC-MS chromatogram of alkaloid extract of Carlton bulb tissues (column: BPX5).

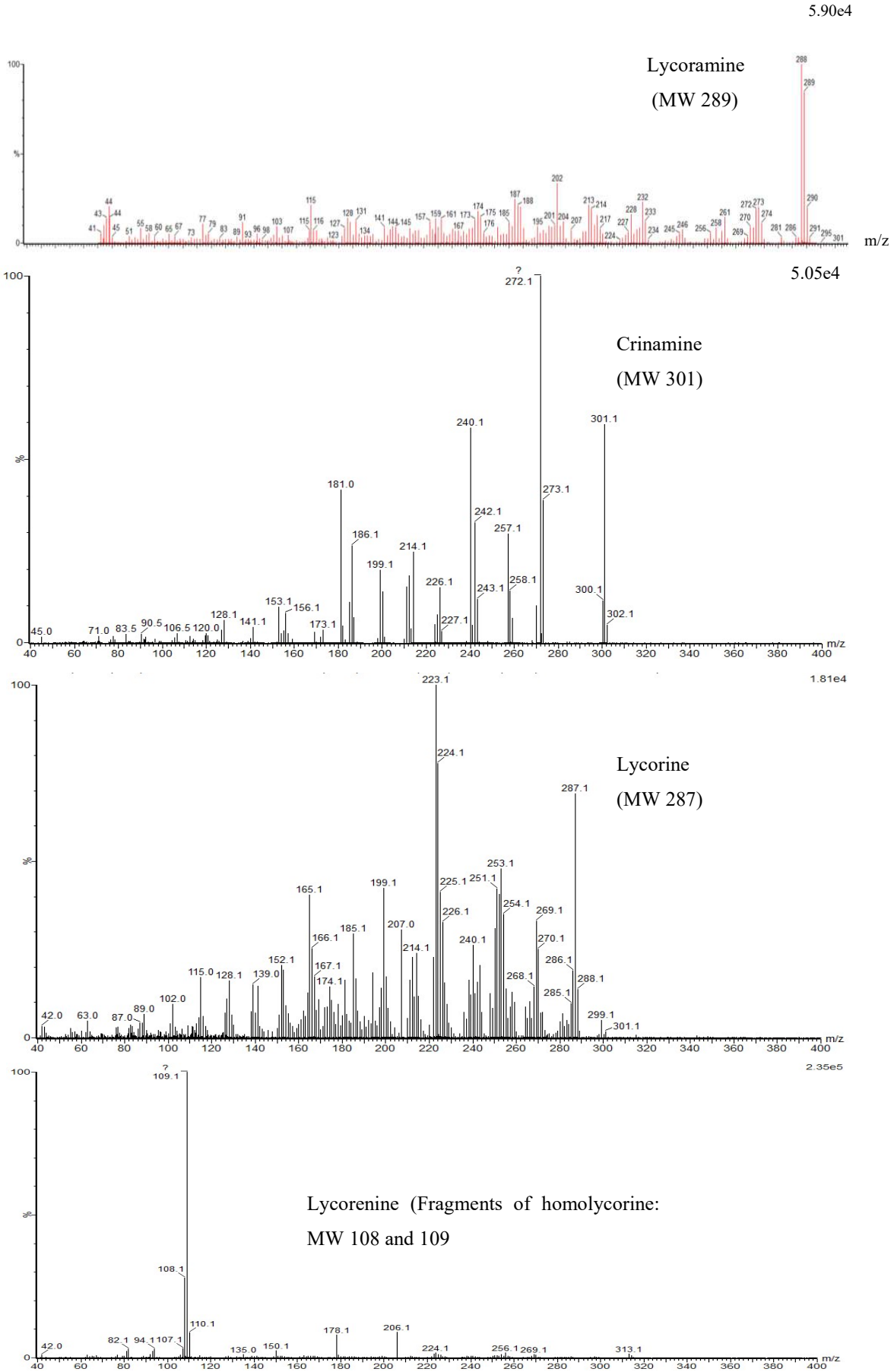

Fig. 3.6B Mass fragmentation pattern of suggested alkaloids found in Carlton bulb extracts (lycoramine, crinamine, lycorine and lycorenine).

The Carlton basal plate chromatogram (Figure 3.7A & 3.7B) also showed top matched mass spectra of related compounds such as lycoramine (Galanthamine type), haemanthamine, pancracine (narciclasine and montamine type) and lycorenine (homolycorine type).

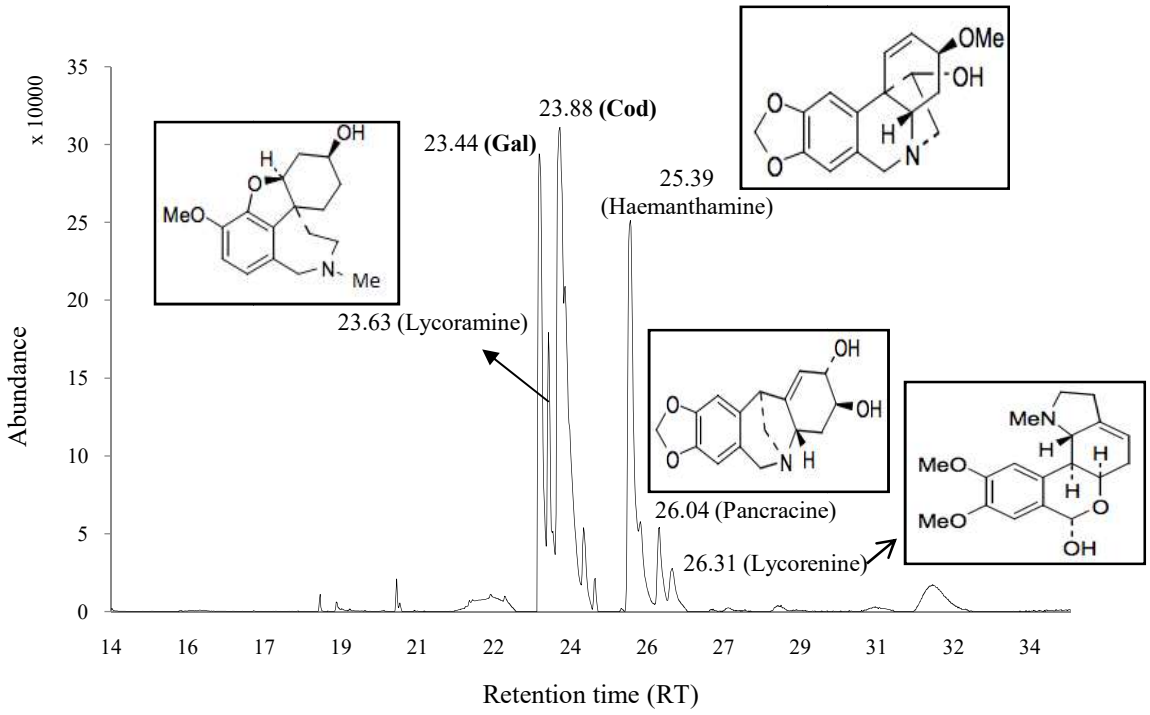

Fig. 3.7A GC-MS chromatogram of alkaloid extract of Carlton basal plate tissues (column: BPX5).

From Figure 3.6A and 3.7A it can be observed that the retention times for crinamine (25.41) and haemanthamine (25.39) are very similar and also for lycorine (26.01) and pancracine (26.08) but the mass fragmentation patterns in Figure 3.6B and 3.7B are very different which indicated they are obviously different compounds.

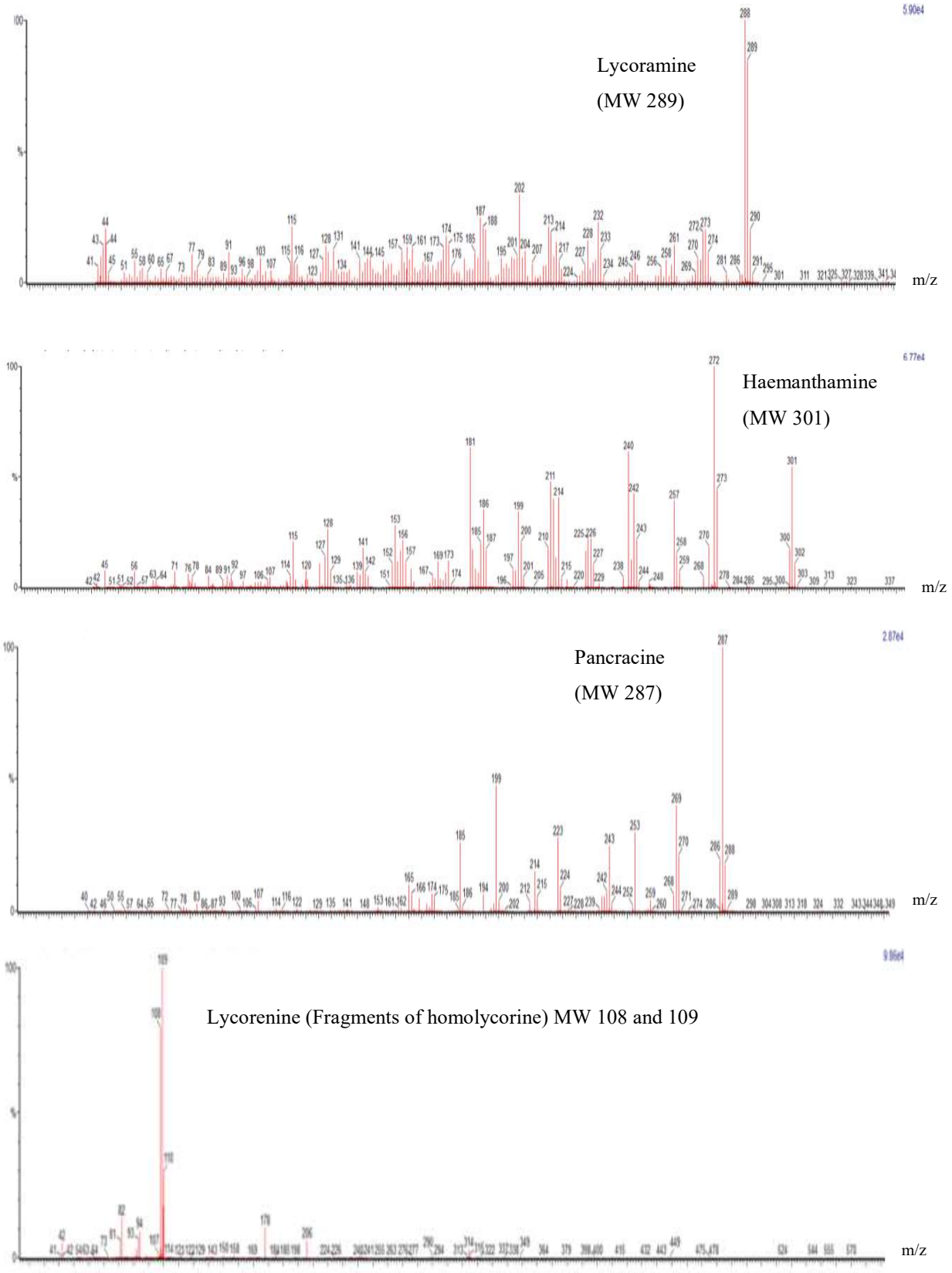

Fig. 3.7B Mass fragmentation pattern of suggested alkaloids found in Carlton basal plate extracts (lycoramine, haemanthamine, pancracine and lycorenine).

The Carlton dormant bulb and basal plate chromatograms showed similar compounds to the harvested bulb and basal plate. The top matched mass spectra from dormant bulb extracts (Figure 3.8A and 3.8B) were lycorenine (homolycorine type) and crinamine (haemanthamine type).

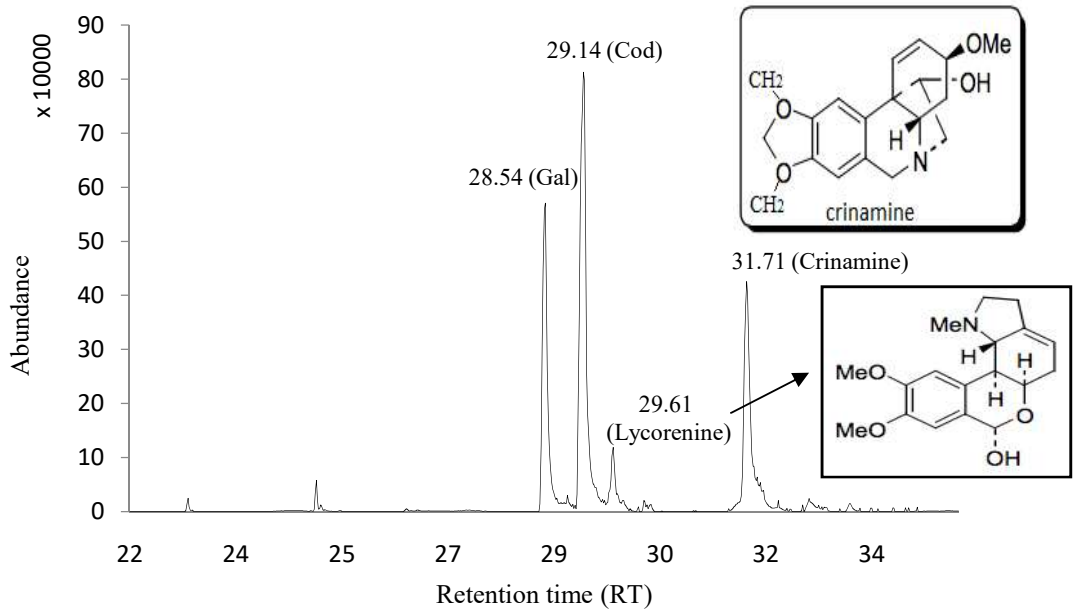

Fig. 3.8A GC-MS chromatogram of alkaloid extract of Carlton dormant bulb (column: DB5M5).

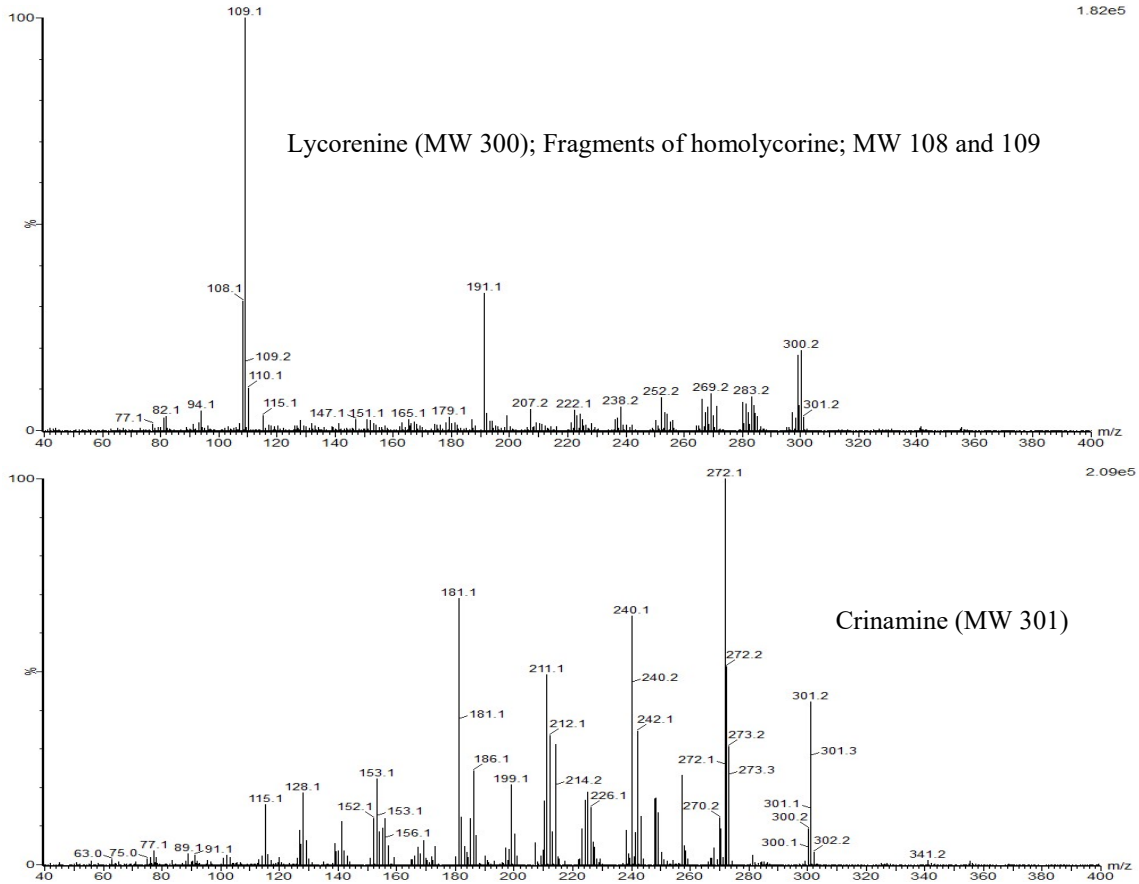

Fig. 3.8B Mass fragmentation pattern of suggested alkaloids found in Carlton dormant bulb extracts (lycorenine and crinamine).

Figures 3.6B, 3.7B, 3.8B and 3.9B showed the fragmentation pattern of suggested alkaloid lycorenine (MW 300), a homolycorine type alkaloid (Appendix 3.1) with the prominent fragments of homolycorine (MW 108 and 109). These mass fragments of homolycorine have been previously reported by Hanks, (2002) and Bastida *et al.,* (2006).

The Carlton dormant basal plate chromatogram (Figure 3.9A) showed one extra matched compound pancracine (narciclasine and montamine type) than bulb that was also present in the harvested basal plate tissue (Figure 3.7A and B).

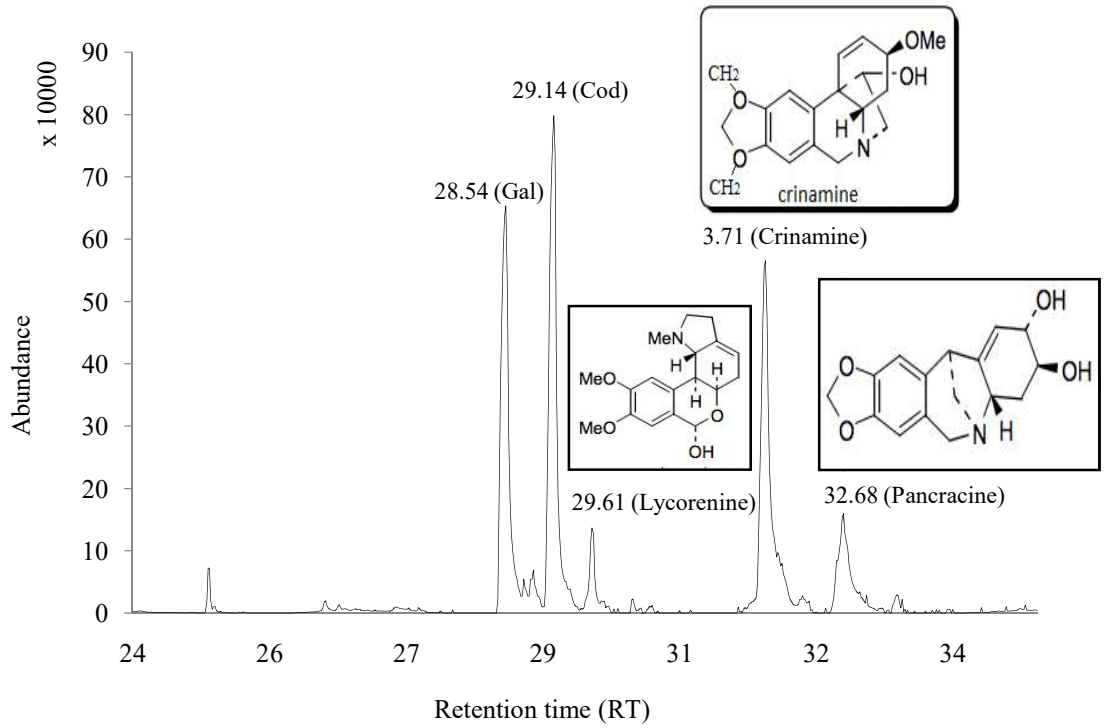

Fig. 3.9A GC-MS chromatogram of alkaloid extract of Carlton dormant basal plate tissues (column: DB5M5).

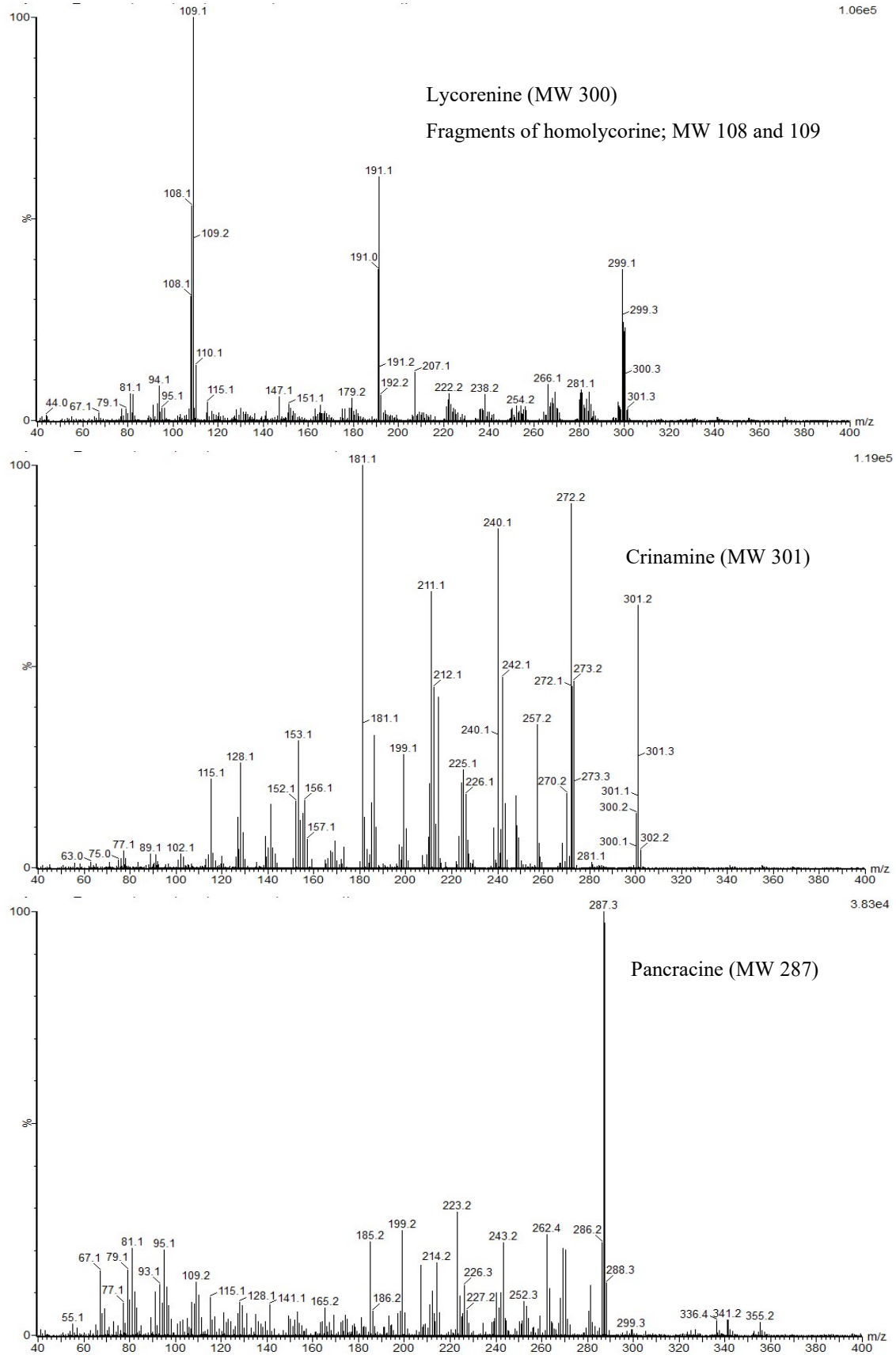

Fig. 3.9B Mass fragmentation pattern of suggested alkaloids found in Carlton dormant basal plate extracts (lycorenine, crinamine and pancracine).

The GC-MS library search and fragmentation pattern showed presence of slightly different compounds in Andrew's Choice chromatograms i.e. tazettine, oxoassoanine and *O*-methyl-macronine.

The Andrew's Choice bulb chromatogram showed top hits of tazettine, oxoassoanine and lycorine (Figure 3.10A and 3.10B). There was no matched compound found from GC-MS database 'the NIST mass spectral library' for 24.29 RT. However, previous studies (Llabrés *et al.,* 1986; Bastida and Viladomat, 2002) reported one lycorine type alkaloid namely oxoassoanine with same molecular weight (281) and similar fragmentation pattern (Figure 3.10B) which was isolated from *N. assoanus*. Therefore, the compound at retention time 24.29 was predicted as oxoassoanine (MW 281).

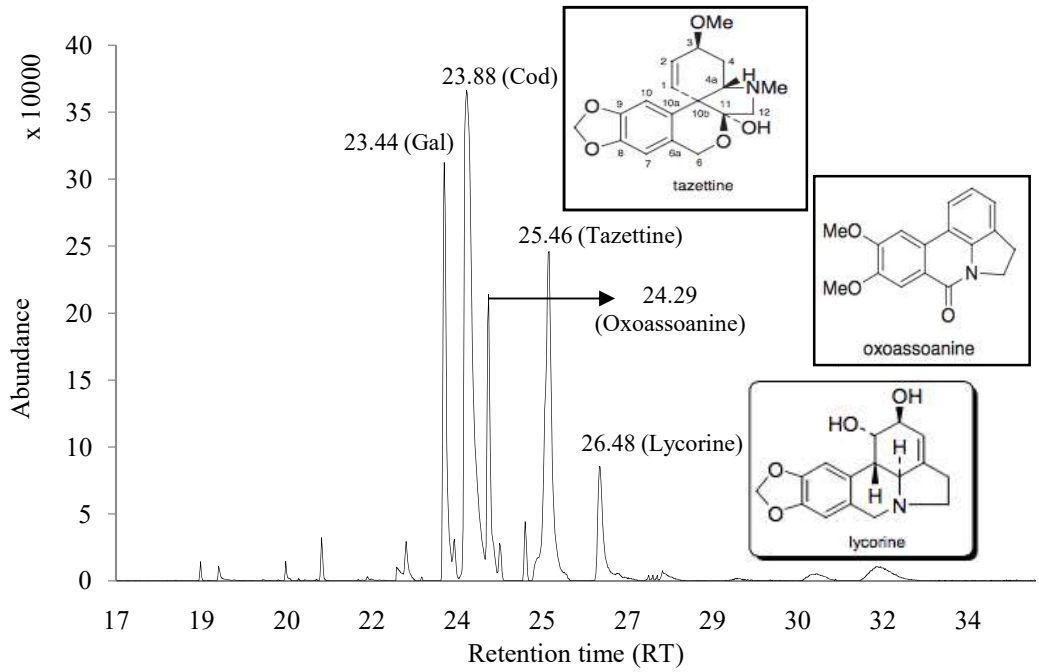

Fig. 3.10A GC-MS chromatogram of alkaloid extract of Andrew's Choice bulb (column: BPX5).

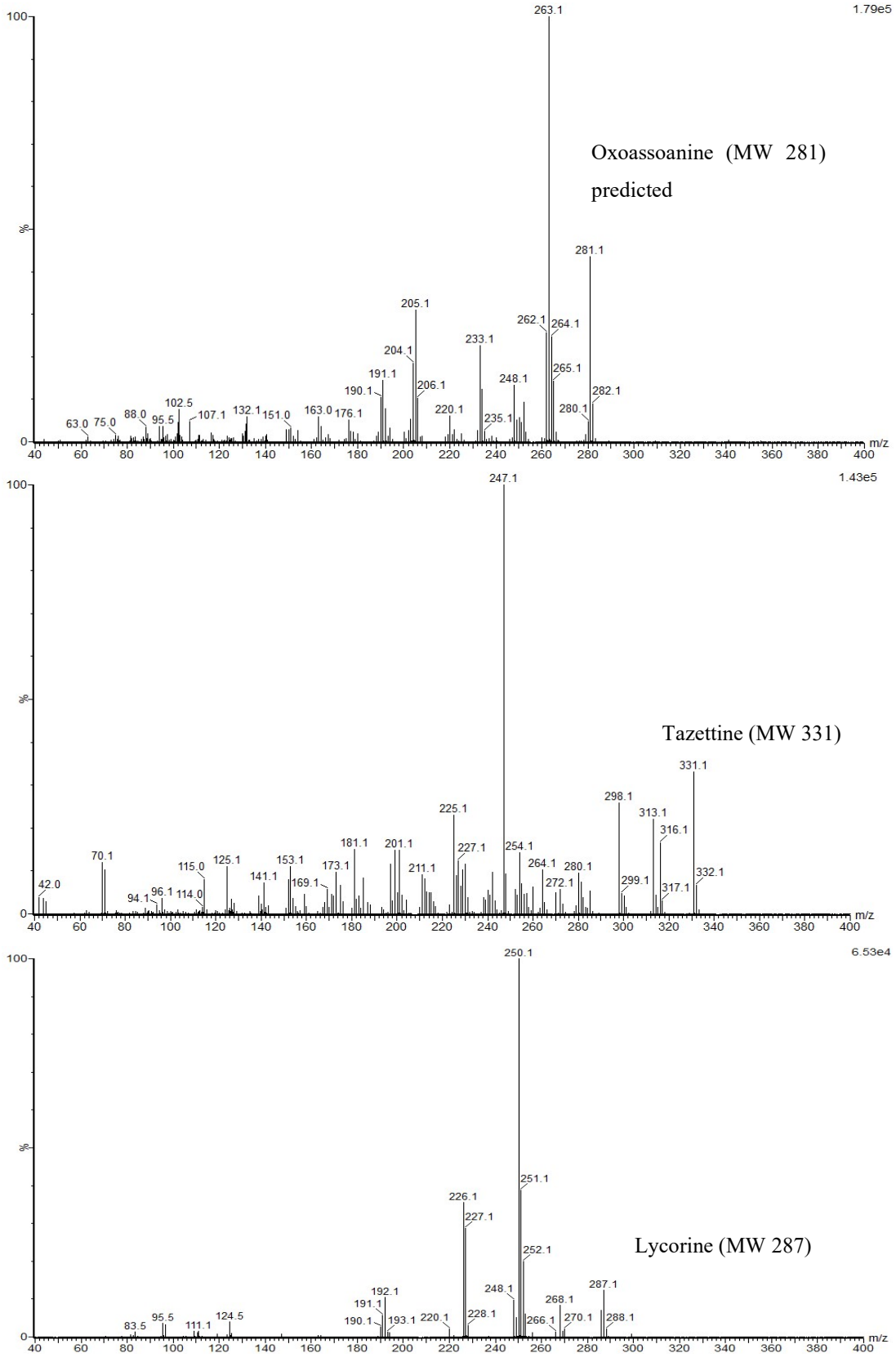

Fig. 3.10B Mass fragmentation pattern of suggested alkaloids found in Andrew's Choice bulb extracts (oxoassoanine, tazettine and lycorine).

The same alkaloids were observed in the Andrew's Choice basal plate chromatogram (oxoassoanine, tazettine and lycorine) as from the bulb extract with an additional suggested compound, O-methyl-macronine (MW 345) (Figure 3.11A and 3.11B). A compound with the same molecular weight has previously been isolated from *N. confusus* bulbs and leaves (Berkov *et al.,* 2011).

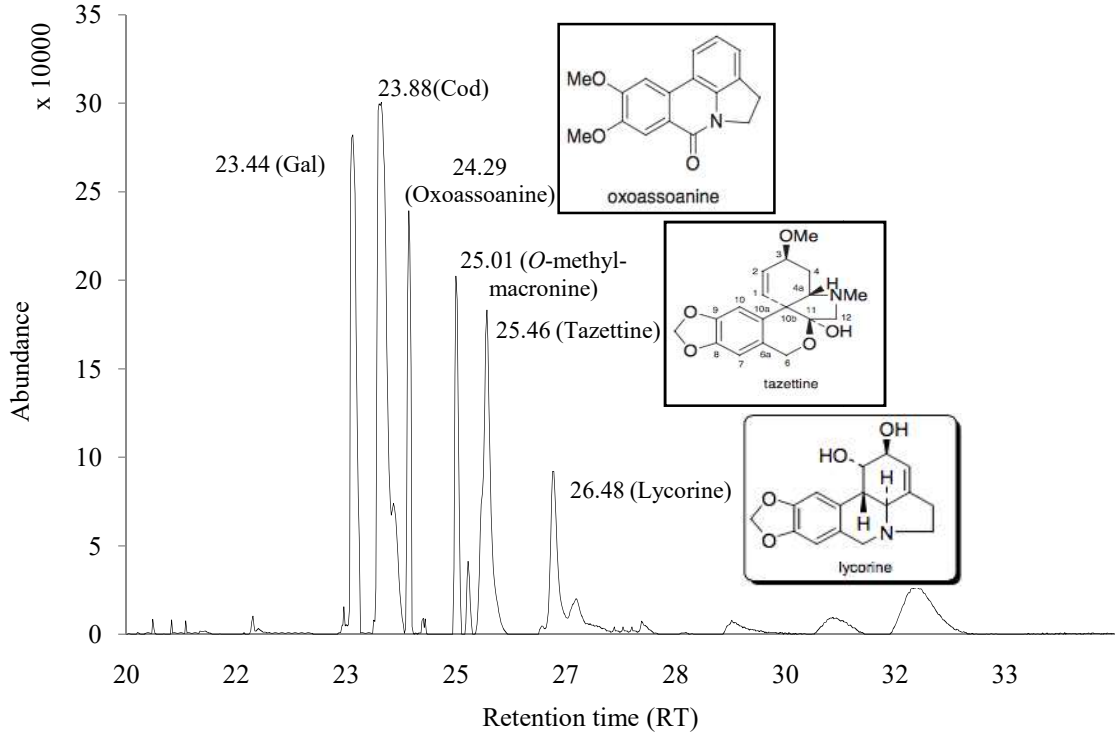

Fig. 3.11A GC-MS chromatogram of alkaloid extract of Andrew's Choice basal plate tissues (column: BPX5).

The slightly increased retention time observed from Figure 3.6A, 3.7A, 3.10A and 3.11A (column: BPX5) in comparison with Figure 3.8A, 3.9A and 3.10A (column: DB5M5) was due to the change of column. That is why, the same compounds showed different retention times. As an example, crinamine was observed at 24.41 RT in Figure 3.6A but at 31.71 RT in Figure 3.8A and 3.9A; but their mass fragmentation patterns are the same (Figure 3.6B and 3.8B, 3.9B) which confirms they are same the compound.

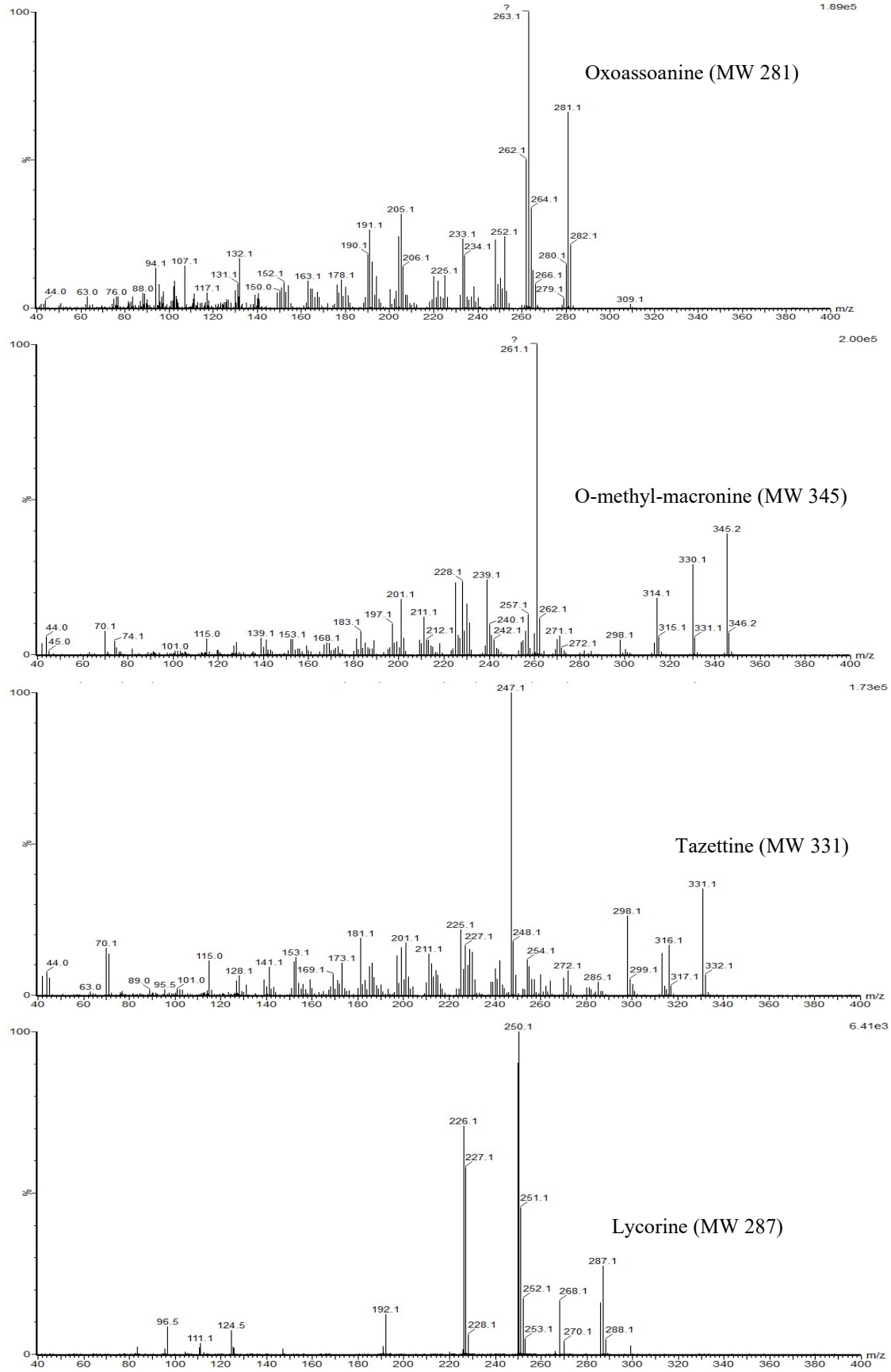

Fig. 3.11B Mass fragmentation pattern of suggested alkaloids found in Andrew's Choice basal plate extracts (oxoassoanine, O-methyl-macronine, tazettine and lycorine).

The GC-MS chromatogram for callus extracts (Figure 3.12A) confirms that callus is the lowest galanthamine producing tissue (section 3.3.3) but it could contain other alkaloids such as tazettine/ pretazettine (MW 331). Tazettine was identified as the top matched from GC-MS database and the next matched compound from same retention time was pretazettine (tazettine type) (Figure 3.12A and 3.12B).

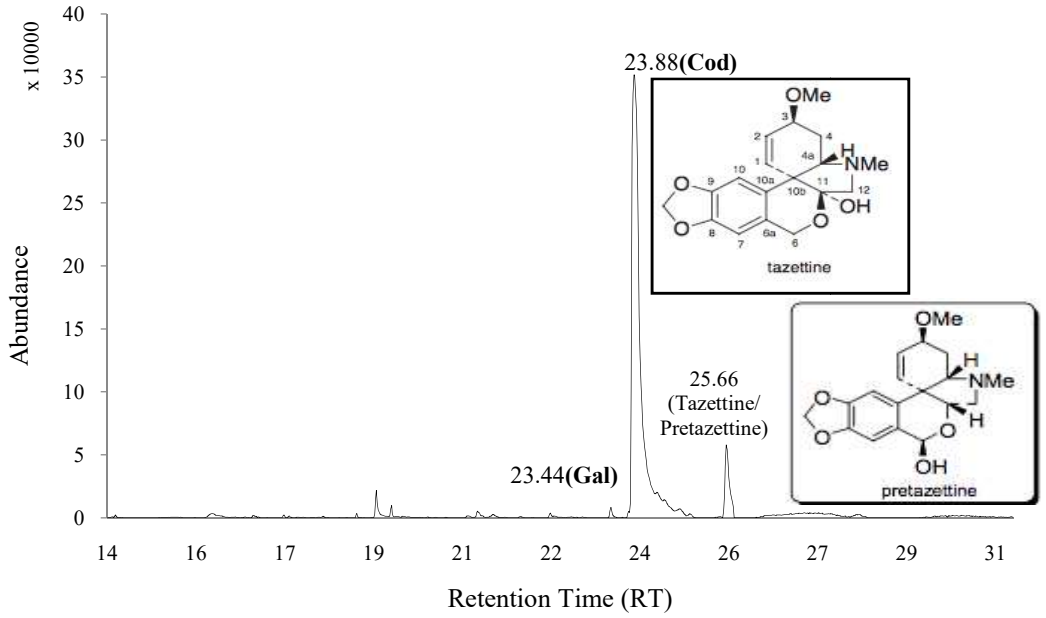

Fig. 3.12A GC-MS chromatogram of alkaloid extract of *N. pseudonarcissus* cv. Carlton culture derived callus (column: BPX5).

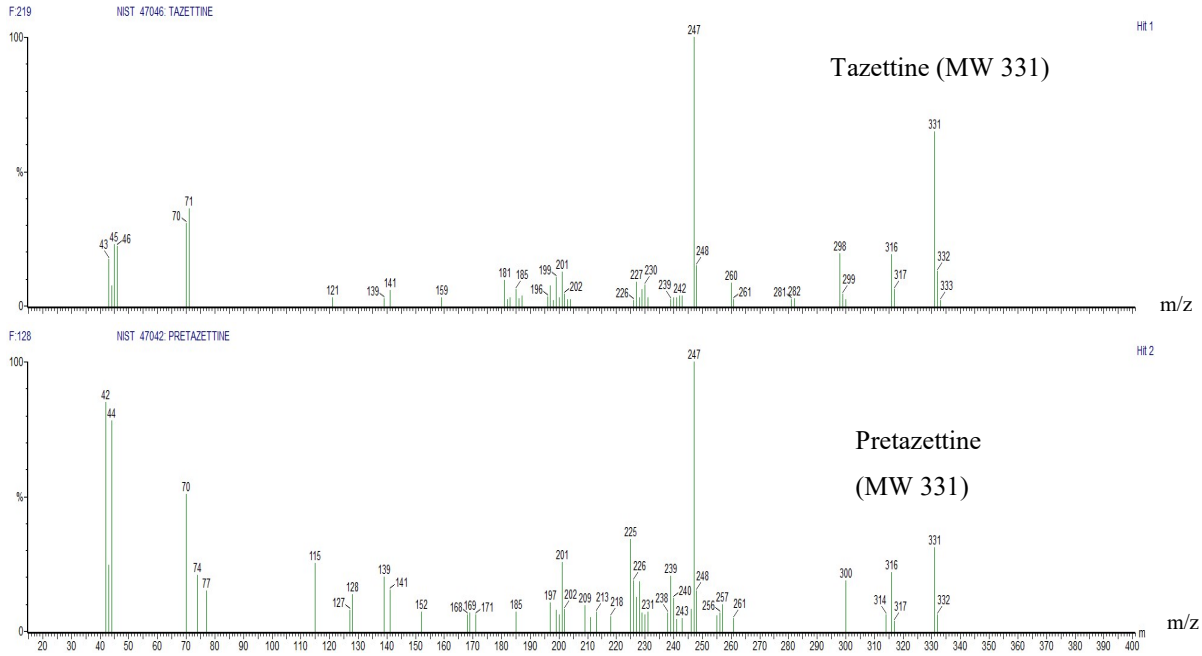

Fig. 3.12B Mass fragmentation pattern of suggested alkaloids found of *N. pseudonarcissus* cv. Carlton culture derived callus extracts (tazettine/ pretazettine).

Organised tissues (small bulblets, white shoots and green shoots) grown *in vitro* often contained galanthamine (Section 3.3.3). The chromatograms of small direct bulblets and green shoot extracts showed the presence of galanthamine. However, the top matched compound for bulblets obtained from chromatogram mass spectra was crinamine (haemanthamine type) as shown in Figure 3.13. The mass fragmentation pattern of crinamine was same as shown in Figure 3.9B.

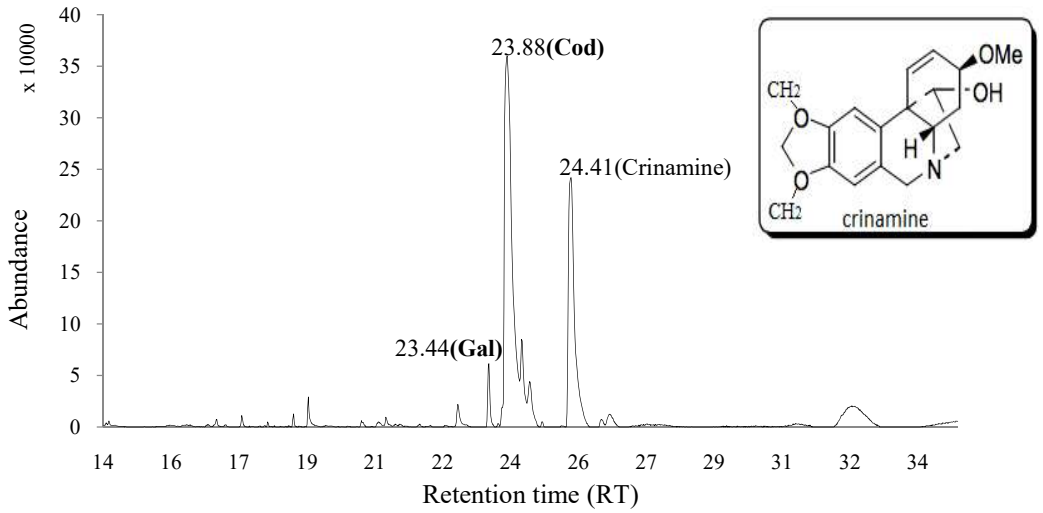

Fig. 3.13 GC-MS chromatogram of alkaloid extract of *N. pseudonarcissus* cv. Carlton culture derived direct bulblets (column: BPX5).

The top matches (Figure 3.14A and 3.14B) found for *in vitro* green shoots are caranine (lycorine type), crinan (*O*-methylnorbelladine type) and lycorenine (homolycorine type). The mass fragmentation pattern of lycorenine was same as shown in Figure 3.8B.

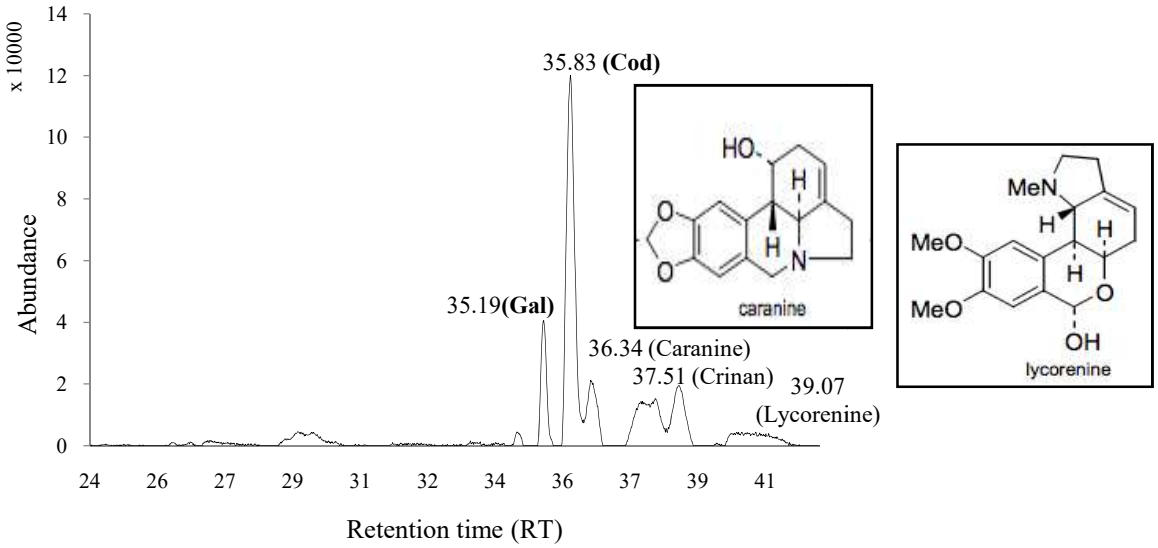

Fig. 3.14A GC-MS chromatogram of alkaloid extract of *N. pseudonarcissus* cv. Carlton culture derived green shoots (column: DB5M5).

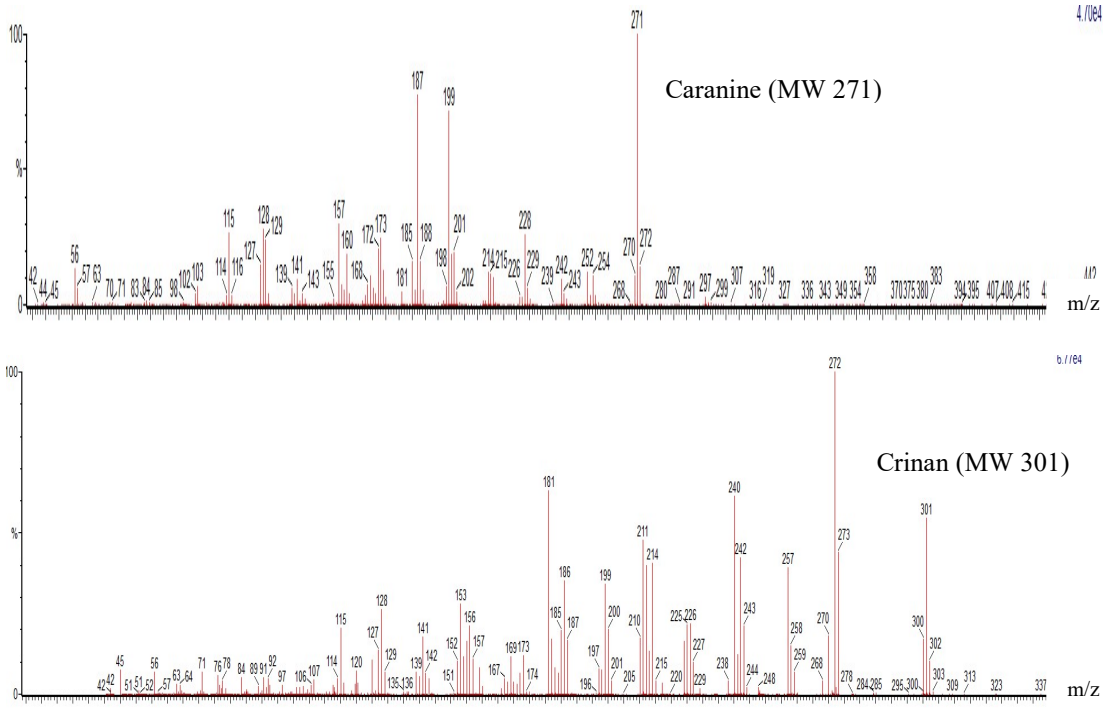

Fig. 3.14B Mass fragmentation pattern of suggested alkaloids found of *N. pseudonarcissus* cv. Carlton culture derived green shoot extracts (caranine and crinan).

3.3.5 Quantification of Gal after elicitor treatment of callus, regenerated bulblets and in their culture media

Methanol extracts of the elicitor (Table 3.2) treated callus, regenerated bulblets and media extracts were analysed by GC-MS to quantify the amount of galanthamine. Carlton bulb (a series of different extracts) was used as control for all analyses and the amount of galanthamine in bulb tissues varied from 600 to 1100 µg/g FW across all sets of calculations (Table 3.5).

Callus grown on media without any elicitors contained a very low level of galanthamine, i.e. 7.8 µg/g FW and 6.8 µg/g FW respectively for MSM1 and MSM2. When medium MSM1/MSM2 was modified with elicitors, the result was an increased level of galanthamine in the callus. An increased level of Gal (44 and 20 µg Gal/g FW) was present in callus cultured on MSM1/MSM2 supplemented with methyl jasmonate (MJ1) but less increase at the higher level of methyl jasmonate (MJ2) (17.0 and 15.0 µg Gal/g FW) (Table 3.5).

There was no remarkable effect on Gal level in callus cultured on MSM1, modified with TCIN (5.0 or 6.0 µg Gal/g FW) along with a complete absence of Gal from MSM1 modified with chitosan or yeast extract. The only exception was callus on

MSM1 treated with chitosan from bulb 17 that contained 23.0 µg Gal/g FW. Galanthamine was also absent in callus from MSM2 modified with either TCIN or chitosan (Table 3.5).

Table 3.5 Amount of galanthamine in *N. pseudonarcissus* cv. Carlton *in vitro* cultured callus and regenerated bulblets (RB) cultured on different elicitor treated media (values expressed as µg Gal/g FW ± SD, n=3; for Carlton bulb samples, n=3 individual bulbs for each batch of elicitor treated samples).

Samples	Bulb	Media modified	Elicitors	µg Gal/g FW ± SD
Carlton Bulb (Field)				600 to 1100 ± 0.005
Callus	B9	MSM1	None	7.8 ± 0.0001
			MJ1	44.0 ± 0.0004
			MJ2	17.0 ± 0.0001
	B12	MSM2	MJ1	20.0 ± 0.0004
			MJ2	15.0 ± 0.0004
	B17	MSM1	TCIN	5.0 ± 0.00006
			CH	23.0 ± 0.0002
	B20	MSM1	TCIN	6.0 ± 0.00001
			CH	0.0 ± 0.0
			YE	0.0 ± 0.0
	B20	MSM2	None	6.8 ± 0.0004
			TCIN	0.0 ± 0.0
			CH	0.0 ± 0.0
Regenerated Bulblets	B9	MSM1	None	7.0 ± 0.0001
			MJ1	9.0 ± 0.00008
			MJ2	6.3 ± 0.00009
	B16	MSM1	TCIN	3.5 ± 0.00005
			CH	5.5 ± 0.00001

Bulbs are the source of callus and regenerated bulblets. MJ1/MJ2 = methyl jasmonate; TCIN = *trans*-cinnamic acid; CH = chitosan and YE = yeast extract (concentrations are in Table 3.2).

Moreover, regenerated bulblets grown on MSM1 modified with elicitors, did not show any difference in Gal content (3.5 to 9.0 µg Gal/g FW) when compared with bulblets grown on MSM1 medium without elicitor treatment (7.0 µg Gal/g FW) (Table 3.5).

All media extracts used for calli and bulblet growth were also analysed for Gal content, but galanthamine was completely absent from all media modified with elicitors. The only exception was a trace amount of galanthamine (3.00 ± 0.0004 µg Gal/g FW) obtained from MSM1 media extract.

3.4 Discussion

3.4.1 Calibration model

The calibration curves showed high linearity at tested Gal standards (0.01-0.1 mg/ml) with correlation coefficient *(R^2)* ranging from 0.78-0.97. Similar calibration methods have previously been used for alkaloid extractions in *Narcissus* (Berkov *et al.,* 2011; Torras-Claveria *et al.,* 2013). Galanthamine has been identified in more than 100 species of *Narcissus* by a calibration curve method using galanthamine standards (1-100 µg/ml) standardised with codeine (Torras-Claveria *et al.,* 2013).

3.4.2 Factors affecting galanthamine biosynthesis

3.4.2.1 Tissue type

The nature of tissue and growth environment often affects secondary metabolite production in plants. Results show the higher amount of galanthamine accumulation in field grown *N. pseudonarcissus* cv. Carlton than tissue culture derived materials. Field derived bulbs (960-1110 µg/g Gal FW), basal plate (830-1250 µg/g Gal FW) and leaves (420 µg/g Gal FW) showed relatively high amounts of Gal, while tissue culture materials showed very low or trace amounts (0 to 130 µg/g Gal FW).

This pattern of galanthamine production supports the previous findings obtained from *Narcissus* species. Previous findings on the amount of galanthamine from field grown bulb (2000-2800 µg/g Gal DW) and leaf tissues (500-800 µg/g Gal DW) of *Narcissus* 'Carlton' showed higher amounts than our findings (Lubbe *et al.,* 2013). This increased amount could be due to the dry weight basis quantification and also they used different method of alkaloid analysis (NMR) that also might affect the results.

Previous analysis of the *in vitro* cultures of several *Narcissus* species showed very low amount of galanthamine accumulation (10-500 µg/g DW) (Selles *et al.,* 1999; El Tahchy *et al.,* 2011a).

3.4.2.2 *In vitro* tissue differentiation

The cell and tissue differentiation strongly influence the galanthamine content. Undifferentiated cell cultures always showed a low level of alkaloids (Pavlov *et al.,* 2007; Berkov *et al.,* 2009). Cellular differentiation or callus dedifferentiation into

organised tissues correlates with increased level of galanthamine (Diop *et al.,* 2006; Ptak *et al.,* 2010).

The pattern of galanthamine accumulation in tissue culture derived samples obtained from our results showed the lowest amount in callus (0-7 µg/g FW) followed by callus differentiated into regenerated white shoot (0-13 µg/g FW) and regenerated bulblets (0-50 µg/g FW). The direct bulblets (10-215 µg/g FW) and white shoot (10-30 µg/g FW) from twin-scale showed slightly increased amount. Finally, the highest amount of Gal was observed in green shoots obtained from twin-scale (97-130 µg/g FW); that indicates the increased production of galanthamine is closely related with cellular differentiation and as photosynthesis occurs in green tissues so that these also have access to greater carbon sources. Gal concentration in *N. pseudonarcissus* shoot cultures has been reported in a previous study showing the amount varied from 10-80 µg/g DW with the highest amount reported 100 µg/g, based on dry weight (El Tahchy *et al.,* 2011a).

Similar findings were observed from *in vitro* cultures of *N. confusus*; where callus had the lowest galanthamine content (0.03 µg/g DW) followed by embryonic callus (0.11 µg/g DW), shoot-clumps (0.14 µg/g DW) and plantlets (1.43 µg/g DW) (Codina, 2002). Another recent study also has reported the similar pattern of lower Gal content in callus (30-60 µg/g DW) than *in vitro* shoots (110-130 µg/g DW) of *N. confusus* (Berkov *et al.,* 2014a).

The high concentration of auxin (20 mg/l NAA) combined with cytokinins showed the accumulation of trace amount of Gal in callus whereas with low auxin (4 mg/l NAA) and cytokinin (1.5 mg/l BAP and 0.5 mg/l KN) none was detected. One previous study showed the similar findings, that MS medium containing low concentration of cytokinin (1.5 mg/l BAP) combined with high auxin (3 mg/l NAA) was the best media for total intracellular alkaloid contents in tissue cultures of *N. tazetta* var. *italicus* (Taleb *et al.,* 2013).

3.4.2.3 Elicitation

Methyl jasmonate has attracted the most attention for its effective elicitation in several plants for the accumulation of different alkaloids (Schumann *et al.,* 2012; Ivanov *et al.,* 2013; Giri *et al.,* 2012). In our study, methyl jasmonate showed the best result for galanthamine accumulation among the four elicitors (MJ, TCIN, CH and YE) used. It increased the amount of galanthamine in callus (15.0-44.0 µg/g FW) and regenerated bulblets (6.3-9.0 µg/g FW) over the control (7.0-7.8 µg/g FW). Methyl jasmonate showed the similar results of enhanced galanthamine production in *N. confusus* shoot-clumps (Colque *et al.,* 2004) and *L. aestivum* shoot cultures (Schumann *et al.,* 2012 and 2013; Ivanov *et al.,* 2013).

3.4.3 Other alkaloids or related compounds in *Narcissus*

The GC-MS chromatograms detected the presence of 12 predicted alkaloids and/ or related compounds along with galanthamine; namely lycoramine, crinan, haemanthamine, crinamine, pancracine, lycorine, oxoassoanine, caranine, lycorenine, tazettine, pretazettine and *O*-methyl macronine (Figure 3.15). Same or similar alkaloids or compounds have been previously identified from *Narcissus* species (Bastida *et al.,* 2006) and recent findings on alkaloids biosynthesis in *Narcissus* (Berkov *et al.,* 2014b) support most of the identified alkaloids obtained from our study.

Lycoramine and homolycorine alkaloids have been previously identified through GC-MS in *N. pseudonarcissus* cv. Carlton bulb extracts (Lubbe *et al.,* 2011; Berkov *et al.,* 2011). Berkov *et al.,* (2011) also reported the presence of tazettine in *N. confusus* bulb. Alternative ways of oxidative phenol coupling (*ortho-para'*, *para-para'* and *para-ortho'*) of *O*-methylnorbelladine produce diverse alkaloids in Amaryllidaceae that differs between species, cultivars and even between the different tissues of same plant (Bastida *et al.,* 2011). Lycorine type (lycorine, caranine) and homolycorine type (homolycorine, lycorenine) alkaloids obtained from the norbelladine precursor by *ortho-para'*; haemanthamine type (crinamine); tazettine type (tazettine, pretazettine) and narciclasine and montanine types (pancracine) by *para-para'*; and galanthamine type (galanthamine, *O*-methyl norbelladine, lycoramine) by *para-ortho'* phenol oxidative coupling, have been reported in various *Narcissus* species (Bastida *et al.,* 2006). Both field and *in vitro* grown samples

derived from *Narcissus* could be a good source of other Amaryllidaceae alkaloids such as lycorine, homolycorine and sanguinine (galanthamine type; Appendix 3.1) (Torras-Claveria *et al.,* 2013 and 2014) and their quantification (sanguinine) through GC-MS is also possible using authentic standards of the related compounds (Singh and Chaturvedi, 2012).

Our results showed that the alkaloids detected in field samples (Carlton and Andrew's Choice) and *in vitro* tissues are well related with the proposed Amaryllidaceae alkaloid biosynthesis pathway; i.e. they are derived from the three phenol coupling pathways (Figure 3.15).

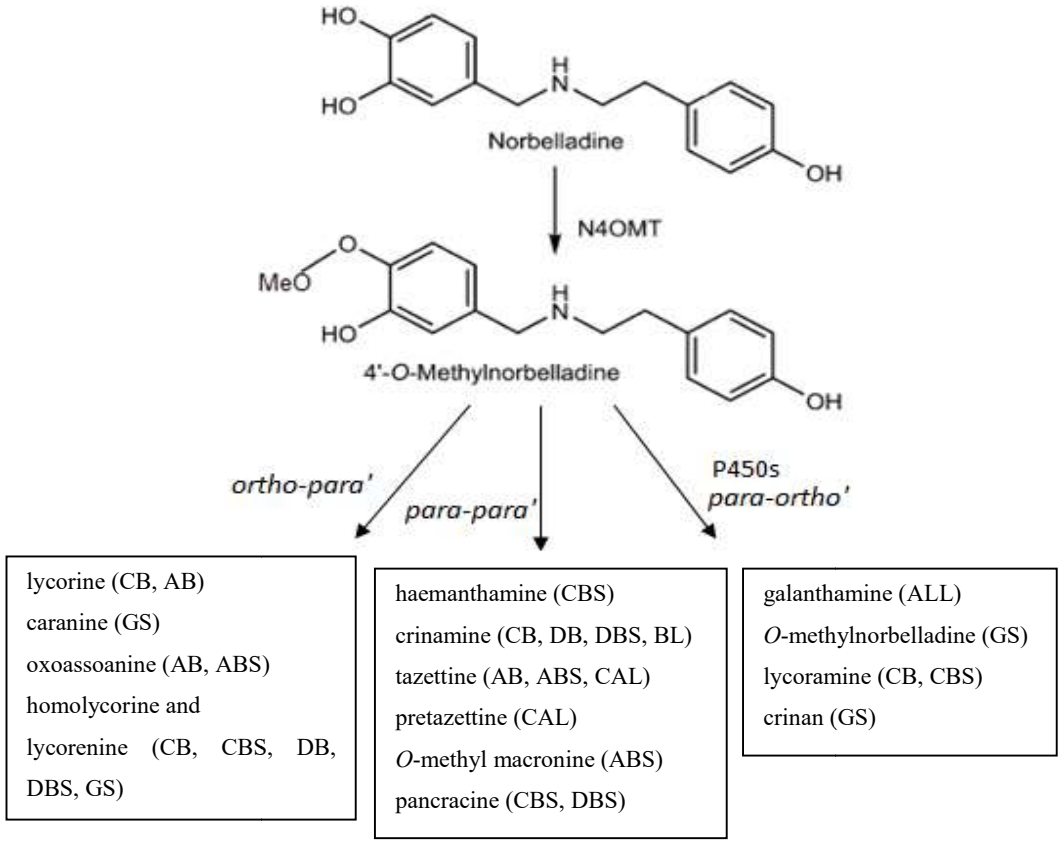

Fig. 3.15 Proposed Amaryllidaceae alkaloid biosynthesis pathway relating the alkaloids detected in *N. pseudonarcissus* cv. 'Carlton' bulb (CB), basal plate (CBS), dormant bulb (DB) and dormant basal plate (DBS); 'Andrew's Choice' bulb (AB) and basal plate (ABS); tissue culture derived callus (CAL), direct bulblets (BL) and green shoot (GS) from Carlton. ALL = galanthamine has been detected (varied from very low to high) in all samples in this study (postulated from Figure 1.5, Chapter 1).

First, Figure 3.15 shows that alkaloid biosynthesis is not directed to a single endpoint. There is evidence that multiple pathways are active (Figure 3.1 to 3.4). Interestingly, alkaloids formed via *ortho-para'* and *para-ortho',* coupling

predominant in the field tissues and in the photosynthetically active green shoots from culture, while *in vitro* callus and bulblets are the only tissues to contain alkaloids formed through *para-para'* route. This may be a consequence of carbon flux, since non-photosynthetic cultured tissues have less carbon resource (Carvalho *et al.,* 2015), and indicate an effect on control of the pathway.

The analysis of Andrew's Choice, a *N. pseudonarcissus* variety included because of its presumed low levels of galanthamine (Pulman, 2015) also shows differential channelling of metabolites through the pathway (Figure 3.15). Although, surprisingly a significant amount of Gal was detected in its bulb and basal plate tissues (630-1050 µg/g Gal FW) in this study, unlike Carlton, it also contained tazettine and oxoassoanine (Figure 3.10A and 3.11A) in these tissues. This could indicate that, for example in a different growth environment the *ortho-para'* and *para-para'* branches would be more active and compete with galanthamine synthesis.

To investigate this further would require confirmation of these alkaloids' identity and further growth trials. In addition, as a different and complementary approach, information on the small metabolites present in the tissues could shed additional light on metabolic differences. Further, alternative approach would be to use RNA-seq to compare field (basal plate) and *in vitro* (callus) tissues, focusing on genes that could be involved in alkaloid biosynthesis.

We therefore, undertook first an NMR-based metabolomic analysis (Chapter 4) and then a transcriptomic analysis (Chapter 5).

4 Chapter Four: Metabolite profiling of *N. pseudonarcissus* cv. Carlton using ^{1}H-NMR

4.1 Introduction

The identification of small molecules is the key factor for understanding the metabolic profile of an organism (Mahmud *et al.,* 2015). This chapter describes the metabolites present in *Narcissus* tissues (field and *in vitro*) using NMR, a highly reproducible and widely used metabolomic analysis method. Another aim is to elucidate the pathways related to the identified metabolites. Comparison of the metabolite patterns and elucidated metabolic pathways between growth conditions, *in vivo* versus *in vitro* is the main objective of this chapter.

4.1.1 Nuclear Magnetic Resonance

NMR spectroscopy is the physical observation of the resonances of magnetic nuclei such as ^{1}H, ^{13}C, ^{15}N or ^{31}P detectable in the presence of an external strong magnetic field (Kim *et al.,* 2010). The NMR signal frequency is determined by the magnitude of the magnetic field at the nucleus, which can act in opposition to the externally applied magnetic field. This can cause the NMR signal frequency to shift, known as chemical shift (Wilson, 2013). It is the resonant frequency of a nucleus compared to an internal standard i.e. TMS (Bertani *et al.,* 2014). The determination of the position and number of chemical shift is important in NMR to determine the molecule's structure (Wilson, 2013).

The key advantage of ^{1}H-NMR spectroscopy is the robustness of the technique. In addition, the integration of signals is quantitative and is truly representative of the relative concentrations of those compounds (Ward *et al.,* 2007). Due to the high sensitivity of the ^{1}H nucleus, the vast majority of NMR-based metabolomic studies have been based on 1D ^{1}H-NMR. Generally, ^{1}H-NMR can detect any metabolite containing hydrogen and provides a rich source of qualitative and quantitative information on compounds of all chemical classes present in biological fluids or tissue extracts (Colquhoun, 2007). The ^{15}N and ^{13}C isotopes have low natural abundance, 0.37% and 1.1% respectively, substantially lower than ^{1}H (99.98%) (Bligny and Douce, 2001). The ^{31}P isotope has very high natural abundance (100%) (Bligny and Douce, 2001) but it can only be used to detect a relatively small number of compounds (Colquhoun, 2007). Therefore, ^{1}H is the preferred nucleus due to its

high magnetic field to provide the best sensitivity and signal dispersion for NMR studies of plant extracts (Colquhoun, 2007).

Nevertheless, the application of NMR-based metabolomic analysis has some limitations. One of the major problems is signal overlap, particularly in crowded regions (Verpoorte *et al.,* 2007), which may hamper signal identification and accurate peak integration. This problem can be solved by using a better resolution method such as two-dimensional (2D) NMR (Kim *et al.,* 2010). The application of 2D NMR spectroscopy for the identification of metabolites from the crowded ^{1}H-NMR signals has been reported for plant *in vitro* culture system (Palama *et al.,* 2010).

Another major disadvantage could be low sensitivity but this can be overcome because of the recent improvements in NMR hardware (Kim *et al.,* 2010). Another risk is that a compound may show changes in its NMR-spectra due to concentration effects or complexity of mixtures (Rhee *et al.,* 2004). Both disadvantages of low-sensitivity and overlapping signals can be overcome by the use of higher field strength magnets (600 MHz or greater), or by using modern cryoprobes (Ward *et al.,* 2007).

4.1.2 NMR of plants

NMR can provide direct molecular structure information in the analysis of complex biological mixtures such as *in vitro* plant extracts (Hendrawati *et al.,* 2006). In addition to identify primary metabolites, NMR can allow simultaneous identification of diverse groups of secondary metabolites without chromatographic separations (Zhi *et al.,* 2012). Signals of metabolites in an NMR spectrum are proportional to their molar concentration, which reflects the real molar levels of metabolites present in plants (Kim *et al.,* 2010). It can also facilitate the identification of unknown products observed in enzyme assays or plant extracts (Kilgore and Kutchan, 2016).

Approximately 50 to 100 compounds can be detected by NMR from plant extracts by using both 1D and 2D NMR spectroscopy (Verpoorte *et al.,* 2007) but studies reporting the identification of 30-150 metabolites with many unknown analytes has also been reviewed (Kim *et al.,* 2011). Because of the presence of large number of compounds, the ^{1}H NMR spectra of plant extracts are inevitably crowded and most

of the compounds also provide multiple signals. In addition, each chemically distinct hydrogen atom, or group of hydrogen atoms give rise to dispersion of signals across the spectrum due to their own chemical shift (Colquhoun, 2007).

NMR has been used as a powerful tool for the classification and characterisation of plant species including quality control of medicinal plants (Choi *et al.,* 2004; Kim *et al.,* 2005; Yang *et al.,* 2006). NMR-based metabolic profiling (1D and/ or 2D) has been used to monitor the response of plants to stress, wounding, and herbivores or infection (Choi *et al.,* 2004 and 2006; Widarto *et al.,* 2006; Simoh *et al.,* 2009).

4.1.3 Practical considerations

Harvesting plant samples is a crucial step for NMR-metabolomic analysis; fresh samples are needed for the best results and rapid cooling is highly recommended to avoid or reduce the degradation of compounds present in samples (Kim *et al.,* 2010). Kim and colleagues (2010) also highlighted biological variation and technical replication as key factors to consider for reproducibility of results. Moreover, the homogeneity of the materials is another important factor, especially when analysing very small amounts (≤ 20 mg).

Choosing an appropriate solvent for sample extraction is a critical step (Kim *et al.,* 2011; Colquhoun, 2007). Aqueous methanol has been reported as an excellent solvent for studying a broad range of primary and secondary metabolites (Verpoorte *et al.,* 2007; Yang *et al.,* 2006). Extraction in acetonitrile has afforded greater reproducibility and equivalent metabolite coverage when samples are subsequently prepared in an aqueous buffer (NMR Centre, University of Liverpool).

The NMR technique requires the presence of deuterated solvent to lock a signal, on which resolution adjustment can be carried out for each sample. Deuterated H_2O (^2H_2O) phosphate buffers are the most frequently used for this purpose as they do not provide any additional signals in the spectrum (Colquhoun, 2007). Special care should be considered during apparatus adjustment, such as shimming which is used as base line correction to optimise the homogeneity of the magnetic field to adjust the resolution of signals, fixed temperature, identical line-width and line-shape for all samples (Defernez and Colquhoun, 2003; Schripsema, 2010). The probe should be tuned to keep the receiver gain at the same level for all samples, but generally tuning

on the first sample is sufficient when all samples have similar ionic strengths (Colquhoun, 2007).

4.1.4 NMR analysis

The absolute metabolite concentration can be calculated by comparison of peak intensity to an internal standard. Tri-methylsilyl propanoic acid (TSP) is a widely used standard for the calibration of NMR shifts for the quantification of metabolites (Kim *et al.,* 2010). The presence of macromolecules with high molecular weight, such as polysaccharides and proteins cause NMR peak broadening. The use of Meiboom-Gill modification of the Carr-Purcell (CPMG) spin-echo sequence method as adopted in the present study can solve this problem by selectively emphasizing the signals from low molecular weight metabolites (Phalaraksh *et al.,* 1999; Kim *et al.,* 2010). Removal of undesired signals caused by residual water is another important factor to consider for better metabolite signals. Presaturation (pre-sat) is a widely used suppression method due to its simplicity and strength (Kim *et al.,* 2010) which shows efficient water signal suppression and generates spectra with good resolution.

4.1.4.1 Metabolite identification

Identification of metabolites is the most important and time-consuming step of NMR-based metabolomic analysis. Comparison of NMR chemical shifts and coupling constants with a database of spectra of authentic substances is the basis for rapid identification of known metabolites (Kim *et al.,* 2010). Several software and database systems are used to assist identification and quantification. Chenomx software (www.chenomx.com) is an integrated set of tools combining sophisticated chemistry with advanced pattern recognition algorithms for identifying and quantifying metabolites. The Chenomx reference libraries contain hundreds of fully searchable pH-sensitive compound models, which are automatically calibrated by adjusting compound line shapes, peak width and chemical shifts to better match with the samples under study.

The identification of common primary metabolites such as sugars, amino acids and organic acids is much easier than the identification of secondary metabolites, where their restricted occurrence, complex structures (Kim *et al.,* 2011) and lack of authentic comparison materials make databases lacking information. As an

alternative strategy, HPLC can be combined with NMR to facilitate identification as it provides NMR data of individual metabolites after separation (Bobzin *et al.,* 2000; Lambert *et al.,* 2007).

Due to the system and software available in NMR Centre, University of Liverpool, a one-dimensional NMR analysis was performed for *N. pseudonarcissus* cv. Carlton samples

4.1.4.2 Data processing

The result of an NMR analysis is a spectrum containing intensities of thousands of data points across the chemical shift scale. It is extremely difficult to make meaningful comparisons of large numbers of spectra or chromatograms visually, especially spectra obtained from plant extracts (Colquhoun, 2007).

The NMR spectrum can be divided into a series of small bins (buckets) after removal of unwanted signals from residual solvents or water. The sum of intensities of signals in each bin is calculated by comparing relative intensities to reference areas or to the sum of total intensities (Kim *et al.,* 2010). The typical bin width (0.04 ppm) is used to minimize the effect of variations (Colquhoun, 2007). Commercial programs such as AMIX TOOLS (Bruker) as used in the present study or ACD NMR Manager (Advanced Chemistry Development, Toronto, Ontario, Canada) can be used to carry out the bucketing of spectra (Kim *et al.,* 2010). After bucketing, the spectra are assembled in a data matrix table, where the columns correspond to the chemical shift values (variables) and rows to the samples (observations) (Colquhoun, 2007).

4.1.4.3 Data analysis and visualisation

Metabolomics experiments often have the aim of classifying samples into different groups. Hence, the differences between datasets are important and both univariate and multivariate statistical methods have a role to discover those differences in the data in an objective way to distinguish between groups (Colquhoun, 2007; Schripsema, 2010). MetaboAnalyst (www.metaboanalyst.ca) as used in the present study is a widely used, freely available, web-based tool that allows comprehensive metabolomic data analysis including univariate and multivariate statistical methods, data visualisation and interpretation.

The raw data obtained from an NMR-spectrum needs to be scaled (normalised) before any statistical analysis (Van Den Berg *et al.,* 2006). The most widely used scaling methods are mean-centering, *Pareto*-scaling and unit variance methods. The unit variance method is particularly suitable to classify the differences between minor compounds (Craig *et al.,* 2006). However, the influence of minor signals may be overlooked with a mean-centering method (Kim *et al.,* 2010). Therefore, the intermediate *Pareto*-scaling method has become more common as it gives some emphasis to weaker or minor signals but still provides interpretable loadings (Colquhoun, 2007), which is used in present study. *Pareto*-scaling is basically mean-centered and each column is divided by the square root of the standard deviation of each variable (Dieterle *et al.,* 2006).

4.1.4.3.1 Univariate data analysis (ANOVA)

Univariate analyses provide a preliminary overview of the data showing potentially significant differences present in the conditions under study. It can be carried out on one variable at a time but many tests would therefore be required for the full dataset (Colquhoun, 2007). One-way Analysis of Variance (ANOVA) by MetaboAnalyst shows whether the overall comparisons are significant followed by *post-hoc* analysis in order to provide the information on which levels are different using Fisher's LSD (Least Significant Difference) and Tukey's HSD (Honestly Significant Difference) (Xia *et al.,* 2015).

4.1.4.3.2 Multivariate data analysis

Multivariate statistical methods can be more useful, since all variables are considered simultaneously in a more managed form. This can assist in visualising how a given sample relates to other samples and leads to the idea of using the distances between samples as a basis for their classification into groups (Colquhoun, 2007).

Principal Component Analysis (PCA) is an unsupervised multivariate method that compares all data variables between samples simultaneously and enables clustering of samples to reduce the dimensionality of the data set (Colquhoun, 2007) and also a useful method to visualise groupings within large datasets (Ward *et al.,* 2007). PCA generates a rotated set of axes using linear combinations of the original axes. The data are summarised into weighted average of the original variables called 'scores',

which are the co-ordinates of the sample. The first principal component (PC1) defines the maximum variation in the data set; the second principal component (PC2) defines the greatest remaining variance in the data set and so on. The first two principal components are considered the most important as they account for the greatest amount of variance, typically over half of the total variance in the data set (Colquhoun, 2007; Xia *et al.,* 2015).

The score plot obtained from PCA is helpful to identify any groupings in the data set and, in addition, to detect any outliers that may be due to errors in sample preparation or instrumentation parameters (Ward *et al.,* 2007). Loadings are coefficients by which the original variables must be multiplied to obtain the scores. The loadings closely resemble the original NMR spectra and thus loading plots can be used to detect and display the spectral areas responsible for the separation in the data. Furthermore, the compounds that are responsible for the differences between groups can be identified (Colquhoun, 2007; Ward *et al.,* 2007).

Hierarchical clustering and clustered heatmaps are very popular data visualisation tools in MetaboAnalyst (www.metaboanalyst.ca). Each sample begins as a separate cluster using two important parameters in hierarchical clustering. The first parameter is a similarity measure and other parameter is clustering algorithms. Clustering algorithms include single linkage (clustering based on the closest pair of observations), average linkage (clustering uses the centroids of the observations), complete linkage (clustering accounts for the farthest pair of observations between two groups) and Ward's linkage (clustering to reduce the sum of squares of any two clusters) (www.metaboanalyst.ca). Clustered heatmaps are often presented in addition to hierarchical clustering. Changing patterns of metabolite concentrations across samples and experimental conditions can be easily visualised by heatmaps as they display the actual data values using colour gradients (Xia *et al.,* 2015).

4.1.4.4 Pathway analysis

The quantitative analysis of a set of metabolites in a selected biochemical pathway, or a specific class of compounds, has attracted increasing interest in recent years in bioinformatics and systems biology (Schuster *et al.,* 2000; Schripsema, 2010). MetPA (Metabolomics Pathway Analysis) (www.metaboanalyst.ca) is a web-based

tool to analyze and visualize the metabolomic data within the biological context of metabolic pathways (Xia *et al.,* 2010).

MetPA requires either metabolite concentration data or a list of metabolites as input and combines several pathway enrichment procedures along with the pathway topological characteristics to identify the most relevant pathways involved in a given metabolomic study (Xia and Wishart, 2010). It contains a library of 1173 metabolic pathways from 15 model organisms including only one plant species, *Arabidopsis thaliana* (Xia *et al.,* 2015). MetaboAnalyst's compound and pathway library was updated in 2015 (Xia *et al.,* 2015) based on the latest version of KEGG (Kanehisa *et al.,* 2014), HMDB (Wishart *et al.,* 2012) and SMPDB (Jewison *et al.,* 2013). Once data have been uploaded to MetPA, the metabolites are mapped to KEGG metabolic pathways for over representation analysis and pathway topology analysis. Over-representation (hypergeomic test) analysis is important to test if a particular group of compounds is represented more than expected (Xia *et al.,* 2015). Topology analysis uses the structure of a known pathway to evaluate the relative importance of the compounds based on their relative locations. Two well-established node centrality measures are used for topology analysis; degree centrality, which is defined as the number of links occurred upon a node and betweenness centrality, which measures the number of shortest paths going through the node (Xia *et al.,* 2015).

4.2 Materials and methods

4.2.1 Plant Materials

Bulbs from the same source (New Generation Daffodil Ltd. UK) were used for NMR. Bulb and basal plate tissues from four different bulbs, *in vitro* grown calli cultured on six different modified MS media (Table 4.1), direct white shoot, green shoot and regenerated white shoot including three replicates for each sample were selected for sample preparation. All field samples and *in vitro* samples were cut into small pieces of equivalent approximate surface area and mass (~20 mg), transferred into 1.5 ml microfuge tubes and stored at -80 ^{o}C until required for extraction of soluble metabolites.

4.2.2 Media and solvents

Six different media (Table 4.1) were used for callus treatment. Callus grown and maintained for 12 weeks on MSM1 (Table 2.1 Chapter 2), was the initial material for all media treatments. MSM1 medium (T1) was modified with five different treatment combinations (T2-T6) as shown in Table 4.1. All chemicals for media preparation are listed and methods of media preparation are described in Chapter 2 (section 2.2.2) and Chapter 3 (section 3.2.8.2).

Sodium phosphate, TSP, NaN_3, HPLC grade acetonitrile and $^{2}H_2O$ were purchased from Sigma Aldrich (UK) and used for sample extraction.

Table 4.1 MSM1 media modified with different elicitors.

Name	Treatment combinations	Duration of treatment
T1	MSM1	0 day, 7 days and 30 days
T2	MSM1 with 2.5% sucrose (25 g/l)	0 day, 7 days and 30 days
T3	MSM1 with increased yeast extract (0.8 g/l)	0 day, 7 days and 30 days
T4	MSM1 with methyl jasmonate (22.50 mg/l)	0 day and 30 days
T5	MSM1 with chitosan (100 mg/l)	0 day and 30 days
T6	MSM1 with TCIN (1000 mg/l)	0 day and 30 days

4.2.3 Soluble metabolites extraction

4.2.3.1 Tissues

Fresh ice-cold solvent solution was prepared with 50% HPLC grade acetonitrile and 50% double distilled water. 500 µl ice-cold solvent solution was added to all 1.5 ml microfuge tubes with ~20 mg frozen samples. The microfuge tubes were placed in an ice bath and sonicated with a probe sonicator (Soniprep 150 plus, MSE, UK) to

disrupt cellular membranes for 3 bursts of 30 seconds at 10 kHz to prevent heating. After vortexing for 1 minute the materials were centrifuged at 12000 *g* for 10 min at 4 ^{o}C and the supernatant was collected, flash frozen in liquid nitrogen and lyophilised overnight in Freeze dryer Lyolab 3000, Thermo Fisher Scientific, UK (Beckonert *et al.,* 2007).

4.2.3.2 Media

The media samples were frozen (20 µl) and stored at -80^{o}C. The same ice-cold solvent as used for tissues was added to all media microfuge tubes. All steps were the same as for tissue extraction except sonication. Sonication was done only once for media and media samples were then processed for lyophilisation in the same way as section 4.2.3.1. All lyophilised samples were stored at -80 ^{o}C until prepared for NMR analysis.

4.2.4 Sample preparation for NMR: Tissues and media

Stock phosphate buffer was prepared with 8 ml 1 mM sodium phosphate pH 7.4, 80 µl of 100 mM tri-methylsilylpropanoate (TSP), 80 µl 1.2 mM NaN_3 and made to a final volume of 80 ml with 100% $^{2}H_2O$. 700 µl of stock phosphate buffer was added to each microfuge tube containing frozen lyophilised samples, vortexed for 1 minute and was centrifuged at 12000 *g* for 2 min at room temperature. 600 µl of supernatant from each sample was removed without disturbing the sample pellet and pipetted carefully into 5 mm NMR tubes (Sigma Aldrich, UK) (Lubbe *et al.,* 2013). The NMR Centre, University of Liverpool, supplied all chemicals (Sigma Aldrich, UK) and instruments to run this experiment.

4.2.5 NMR measurement

NMR spectra were acquired on a Bruker 600 MHz spectrometer (Coventry, UK), equipped with a TCI cryoprobe and Sample Jet auto sampler. Temperatures were calibrated to 25 ^{o}C within a margin of 0.2 ^{o}C for each experiment. The experiment was optimised by lock, tune and shim to achieve the best baseline and water suppression possible for each sample. Two ^{1}H-1D NMR experiments, Carr-Purcell-Meiboom-Gill (CPMG) and Nuclear Overhauser Effect (NOE) were collected for each sample. The NOE displayed ^{1}H from both large and small molecules present in

the extract, The CPMG, however, selected for ^{1}H resonances from small molecules (typically < 500 Da) only. For each experiment 128 scans were recorded using the following parameters: 0.15 Hz/ point, pulse width (PW) 8-9 μs, spectral width of 18.00 ppm and relaxation delay (RD) = 4.0 s. TSP (0.05%, w/v) was used as the internal standard for $^{2}H_2O$.

4.2.6 Data processing and multivariate data analysis

NMR spectra were analysed using AMIX software (Topspin, Bruker) to view and compare spectra by creating pattern files. The region δ 4.7 to 6.2 ppm was excluded from the analysis because of the presence of signal from residual water. After bucketing, the intensities of signals in each bucket were calculated by dividing signal intensities by the relative intensity to a reference area (single peak at 0 ppm from reference compound TSP). The signals of sugars (listed under carbohydrate metabolism in Table 4.2) and crowded weak signals (signals with more than three metabolites at low concentrations) were removed from the bucket tables to reduce the complexity of data interpretation. Initially, one bucket table (matrix table) was created with all identified metabolites for all tissue samples. Then the tissue samples were divided into three groups (see sections 4.3.2.1, 4.3.2.2 and 4.3.2.3) with three modified bucket tables for further analysis. Chenomx (ChenomX Inc. Canada) software used to identify metabolites. Pattern files of identified metabolites with their corresponding ^{1}H chemical shifts (ppm) were generated. The typical format of pattern files generated for tissue and media metabolites are included as Appendix 4.3 (tissue) and Appendix 4.4 (media) on the disc accompanying this thesis.

MetaboAnalyst (Xia *et al.,* 2015) was used for univariate analysis method (one-way ANOVA and *post-hoc* analysis), multivariate analysis method (PCA) and clustering analysis (dendrogram and heatmap) to process the NMR spectra.

4.3 Results

4.3.1 Metabolite identification: tissue and media

NMR-based analysis was performed to detect the metabolites present in *N. pseudonarcissus* cv. Carlton field samples, tissue culture derived samples and media extracts.

Table 4.2 summarizes the important metabolites identified across Carlton basal plate, Carlton bulb, callus, regenerated white shoot grown from callus, direct white shoot grown directly from the base of twin-scale, green shoot and callus treated with six different media combination. Metabolites detected in media extracts of callus treatments (T1-T6) are summarised in Table 4.3. Representative ^{1}H-NMR spectra for field (bulb and basal tissue) and calli showing the major groups of metabolites with their relative positions (ppm) are shown in Figure 4.1. A representative ^{1}H-NMR spectrum for all tissues under analysis with the labelled metabolites is also available in Appendix, Figure 4.2.

A total 129 metabolites including 14 unknown compounds were detected across all tissues under study. In contrast, only 36 metabolites, including 3 unknown compounds, were identified from media extracts. The differential presence of identified metabolites in different tissues and media treatments was generated with the help of one-way ANOVA (heatmap) analysis by MetaboAnalyst 3.0 (Xia *et al.,* 2015) and has been listed in Appendix 4.1 (tissues) and Appendix 4.2 (media) on the disc accompanying this thesis. The metabolites detected are represented with their corresponding identification with the HMDB-ID and KEGG-ID from MetaboAnalyst.

Table 4.2 Metabolite groups of interest identified within at least one *N. pseudonarcissus* field or tissue culture derived sample (ChenomX Inc. Canada databases).

Metabolite Groups	**Metabolites**		
1. Amino acids	3-Chlorotyrosine	β-Alanine	Pantothenate
	4-Aminobutyrate	Creatine	L-Proline
	L-Arginine	Glutamate	Pyroglutamate
	Alanine	Glutamine	Sarcosine
	5-Hydroxylysine	Histidine	Threonine
	4-Aminobutyrate	Homocitrulline	Tyrosine
	Anserine	Isoleucine	Valine
	Asparagine	Leucine	
	Betaine	N6-Acetyl-L-lysine	
2. Amines (derivatives of ammonia)	5-Aminolevulinate	Choline	Serotonin
	4-Pyridoxate	Creatine	Trigonelline
	Cadaverine	Histamine	Trimethylamine
	Carnitine	*O*-Acetylcholine	Tyramine
3. Aromatics	4-Pyridoxate	Indoleacetic acid	Serotonin
	Anserine	Mandelate	Methylhistidine
	Histidine	Melatonin	Trigonelline
	Imidazole		
4. Alcohols and Polyols	4-Pyridoxate	Galactarate	Glucitol
	Chlorogenate	Galactitol	Pantothenate
	Choline	Galactonate	Serotonin
5. Carboxylic acids	3-Methylglutarate	Homovanillate	Thymol
	4-Pyridoxate	Isobutyrate	Trigonelline
	Biotin	*N*-Acetylglutamate	*Trans*-aconitate
	Cis-Aconitic acid	Salicylate	Vanillate
	Formate	Succinate	
	Fumarate	Tartrate	
6. Phenols	3-Chlorotyrosine	Homovanillate	Vanillate
	3-Hydroxymandelate	Salicylate	p-Cresol
	Chlorogenate	Thymol	
7. Carbohydrate metabolism	Sugars		
	D-Fructose	Glucose-6-phosphate	Ribose
	L-Fucose	Lactose	Sucrose
	Glucose	Maltose	
	Sugar derivatives		
	Glucuronate	Galactonate	Galactitol
	Galactarate	Glucitol	
8.Drug components	Acetylsalicylate	Imidazole	Salicylate
9. Plant components	Caffein	Theophylline	Vanillate
	Chlorogenate	Trigonelline	
10. Phenylalanine and tyrosine metabolism	Tyramine	3-Chlorotyrosine	Homovanillate
	Tyrosine	3-Hydroxymandelate	p-Cresol
	γ-Glutamylphenylalanine	3-Phenylpropionate	Succinylacetone

Table 4.3 Metabolite groups from media extracts used for callus treatment (ChenomX Inc. Canada databases).

Metabolite Groups	Metabolites	
1. Amino acids	Anserine Betaine Glycylproline Glutamate Isoleucine	Leucine Pantothenate Threonine Valine
2. Amines	Agmatine Anserine	Histamine *O*-Acetylcholine
3. Carbohydrate metabolism (sugars)	Glucose Galactose	Sucrose Trehalose
4. Carboxylic acids	Acetylsalicylate Formate Isocaproate	Succinate Tartrate 3-Hydroxy-3 methylglutamate
5. Drug components	Acetylsalicylate	Imidazole
6. Fatty acids	Acetate	Butyrate

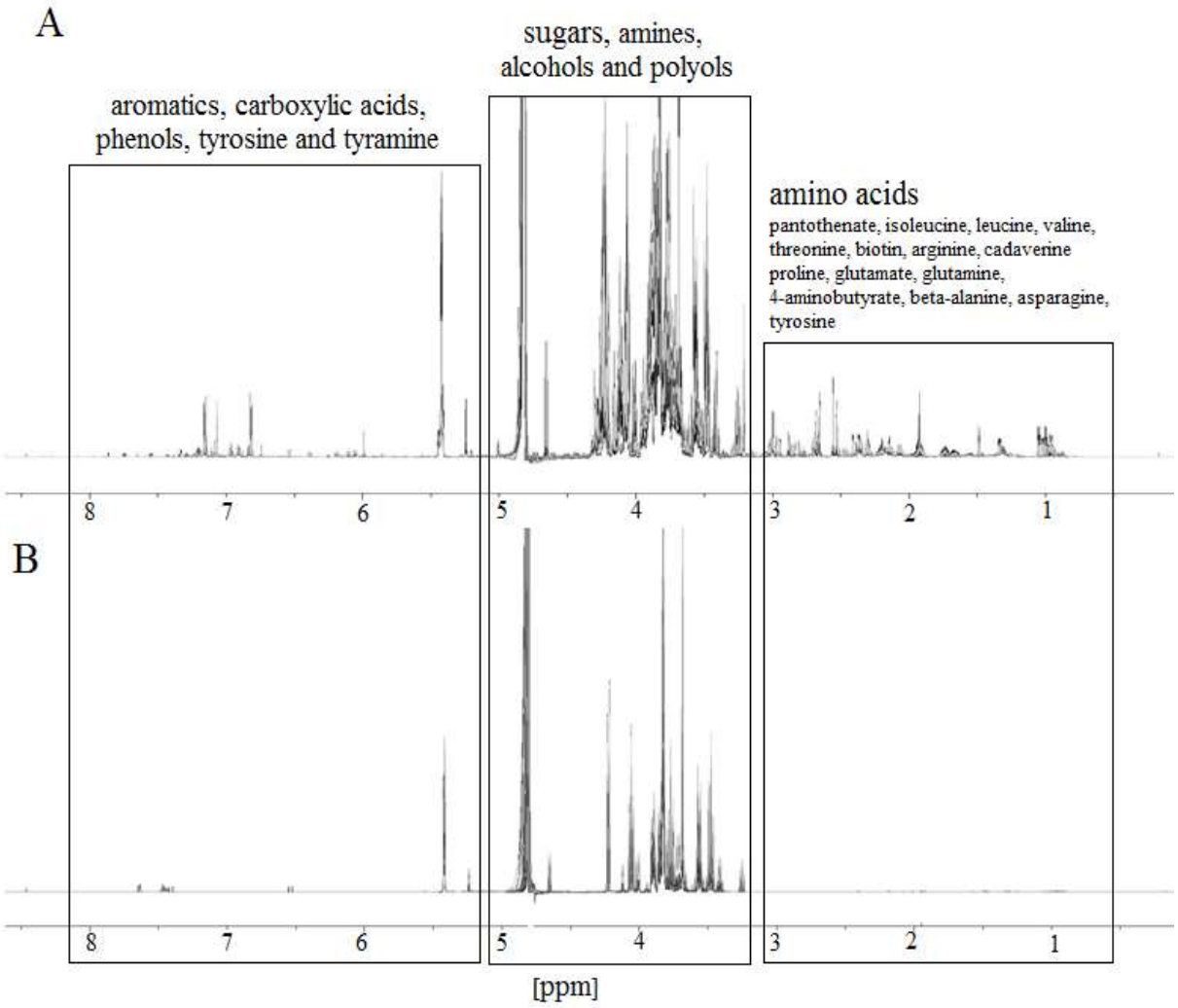

Fig. 4.1 Representative ^{1}H spectra from 0.7 ppm-8.0 ppm of *N. pseudonarcissus* cv. Carlton field (A) and calli (B) samples representing the major groups of metabolites with raw peaks.

4.3.2 Statistical analysis

4.3.2.1 Identification of differences in NMR signal intensity between field and tissue culture derived callus cultured on different media treatments

4.3.2.1.1 Univariate (one-way ANOVA and *post-hoc*) analysis

The ANOVA analysis of the field samples and *in vitro* samples showed which peaks were significantly different between them. MetaboAnalyst parameters were: data normalisation by mean and *pareto*-scaling, p-value threshold set to 0.05 and Fisher's LSD was used for the *post-hoc* analysis. Peaks, which were significantly different between the samples, are shown in black dots with a line representing the p-value threshold (Figure 4.3). The ANOVA showed most (144 out of 156) of the features (metabolites) were significantly different among samples. The *post-hoc* analysis for all metabolites and samples under this section with p-values and Fisher's LSD is in Appendix 4.5.

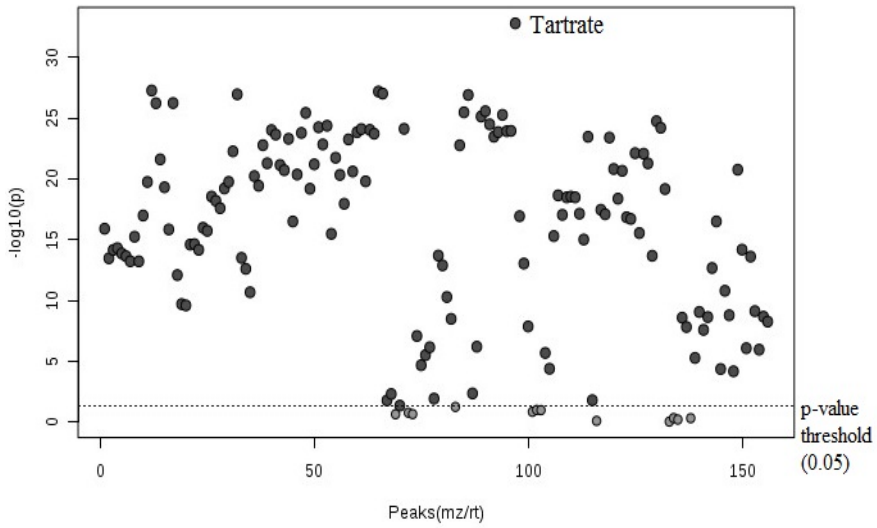

Fig. 4.3 ANOVA in MetaboAnalyst is showing the peak (metabolites) differences among samples under analysis (Carlton basal plate, Carlton bulb and callus) with p-value threshold 0.05. The top compound is labelled. The X-axis is plotting the sizes of NMR peaks, from which the chemical compound can be determined, and whether the ANOVA showed any significant differences among samples.

Metabolites were picked considering the research objectives such as metabolites involved in tyrosine and phenylalanine metabolism, the most significantly different amino acids such as alanine, β-alanine, arginine, asparagine, betaine, proline and threonine, and other compounds such as choline, galactarate, galactitol, glucuronate and glucitol. Figures 4.4, 4.5 and 4.6 represented the ANOVA *post-hoc* analysis showing relative concentrations of selected metabolites in the field tissues and callus.

Most of the metabolites gave more than one signal; weak signals of the same metabolites with very low concentrations are not presented here.

It was observed that the metabolites (Figure 4.4) responsible for tyrosine and phenylalanine metabolism were present in high concentrations in Carlton bulb (CB) and basal plate (CBS), while there were relatively small amounts of tyrosine and trace amounts of other compounds (tyramine, 3-chlorotyrosine, p-cresol, 3-hydroxymandelate and 3-phenylpropionate) in callus (Figure 4.4).

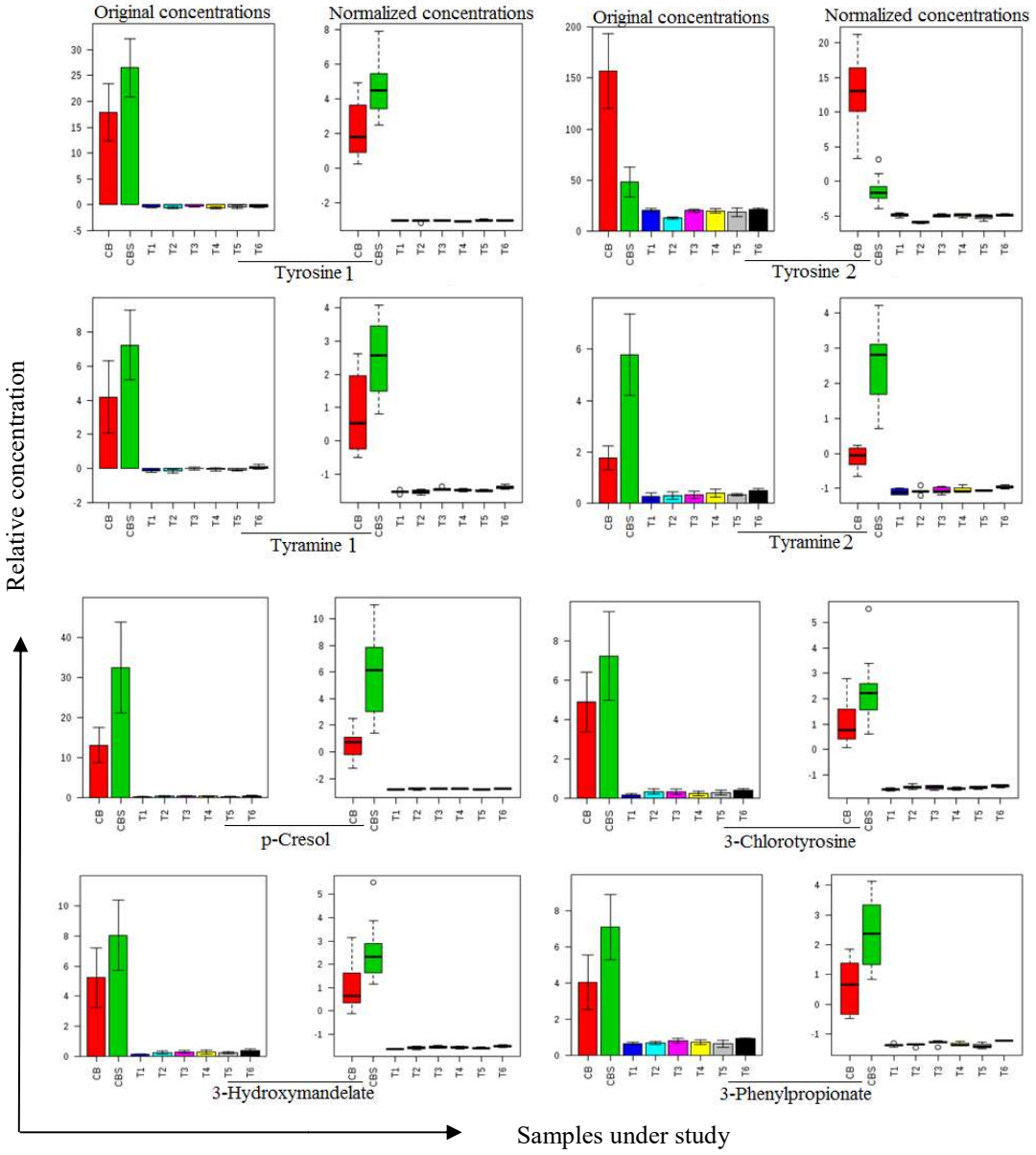

Fig 4.4 ANOVA boxplot representing the relative concentrations of significantly different metabolites (peaks/ signals) involved in tyrosine and phenylalanine metabolism among all samples under study (field samples = CB and CBS; callus = T1-T6). Two sets of tyrosine (1 and 2) and tyramine (1 and 2) are due to the presence of two signals for each compound. Error bars = SE.

The concentration patterns of selected amino acids in all tissues under study showed different distributions. Amino acids such as alanine, asparagine and betaine were present at relatively high concentrations in both field and callus tissues, whereas arginine, beta-alanine and proline were prominent only in field tissues, mainly in bulb. Two signals from threonine represented two different concentration patterns, threonine1 was present across callus, and field samples while threonine 2 mainly in bulb tissues (Figure 4.5).

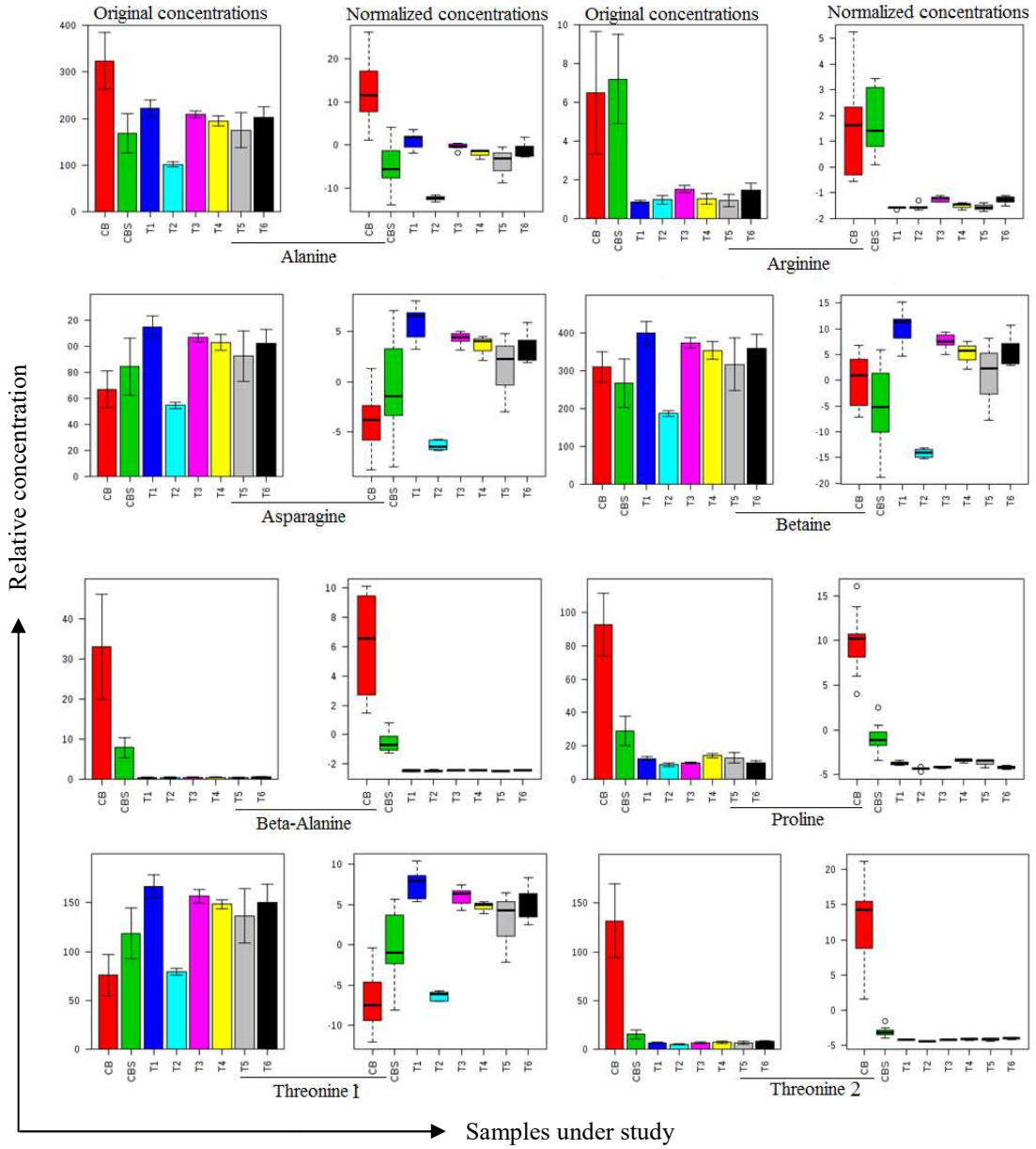

Fig 4.5 ANOVA boxplot representing relative concentrations of significantly different amino acids among samples under study (field samples = CB and CBS; callus = T1-T6). Two sets of threonine (1 and 2) are due to the presence of two signals for threonine. Error bars = SE.

The compounds, which were found to be present in relatively high amounts in both field and callus tissues are represented in Figure 4.6. Choline, galactitol, glucitol and glucuronate showed higher concentrations in all callus tissues except callus from media treatment 2 (T2) and also showed considerably high concentrations in field samples. However, galactarate showed high concentration in bulb tissues only.

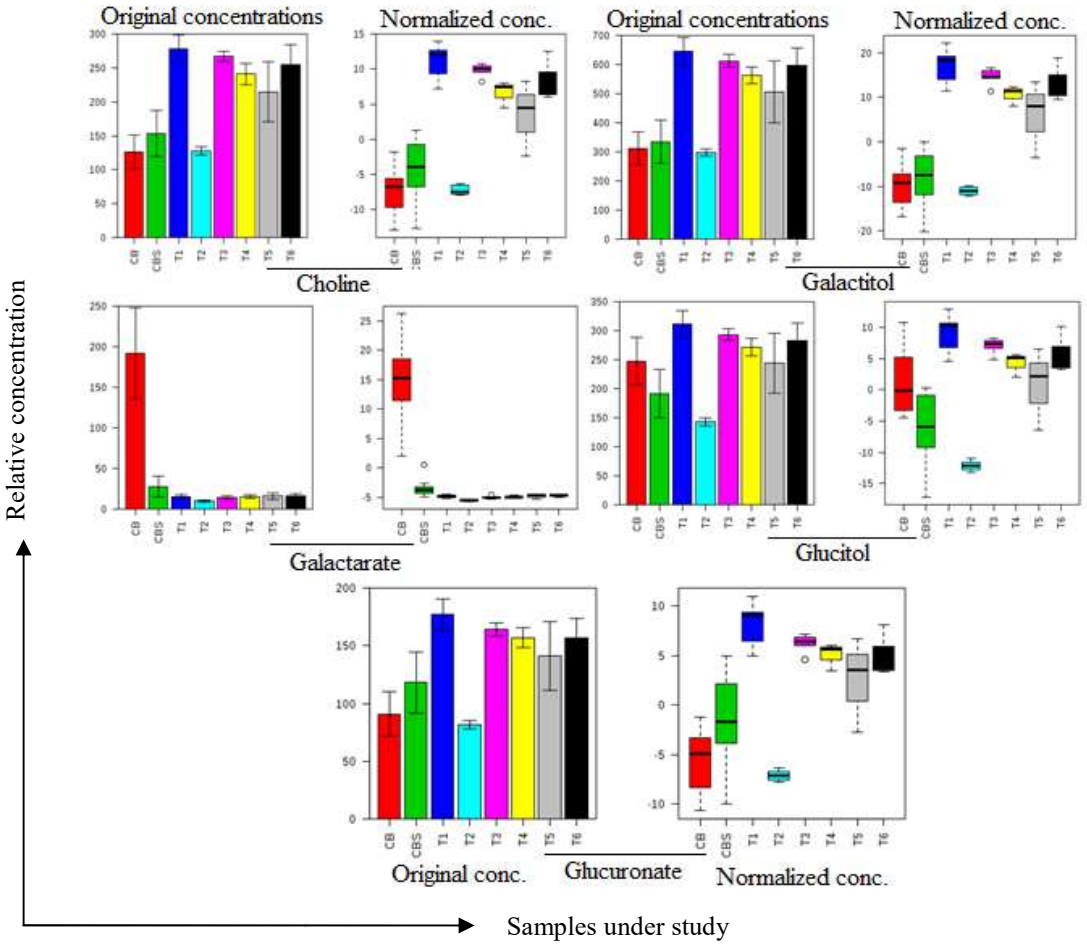

Fig 4.6 ANOVA boxplot representing relative concentrations of choline, galactitol, galactarate, glucitol and glucuronate present in relatively high concentrations in samples under study (field samples = CB and CBS; callus = T1-T6). Error bars = SE.

The compounds or metabolites which were present in relatively high amounts in callus, (treatment T6) compared with field samples (CB and CBS) were melatonin, mandelate, serotonin (5-hydroxytryptamine), *trans*-aconitate, fumarate and unknown compound 9 (Figure 4.7).

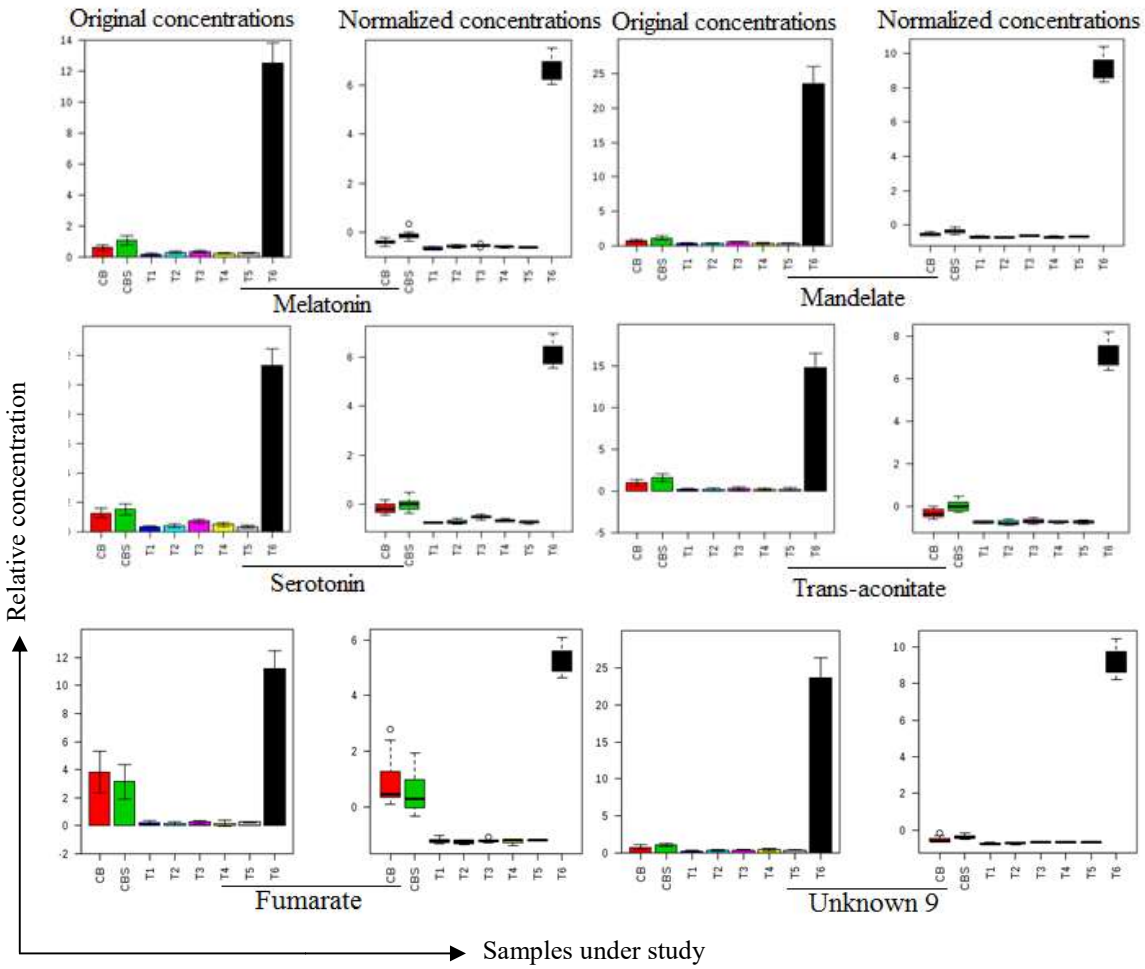

Fig. 4.7 ANOVA boxplot representing relative concentrations of melatonin, mandelate, serotonin, *trans*-aconitate, fumarate and unknown compound 9 highly present in callus samples (field samples = CB and CBS; callus = T1-T6). Error bars = SE.

4.3.2.1.2 ANOVA heatmap for metabolite clustering

A heatmap of the top 30 metabolites selected by the MetaboAnalyst ANOVA *post-hoc* analysis (Figure 4.8) distinguished Carlton basal plate, bulb and callus tissues. High amounts of tyrosine, threonine, proline and 5-aminolevullinate were observed in both Carlton basal plate and bulb tissue. In contrast, tyramine, serotonin, biotin, choline, homocitrulline and beta-alanine were found in Carlton basal plate, and glucuronate, galactitol, lactate, tartrate, pyruvate and *O*-acetylcholine were highly present in Carlton bulb. On the other hand, fumarate, mandelate, UN 9 (Unknown compound 9), serotonin, melatonin and *trans*-aconitate were detected in callus from medium treatment T6 (Figure 4.8).

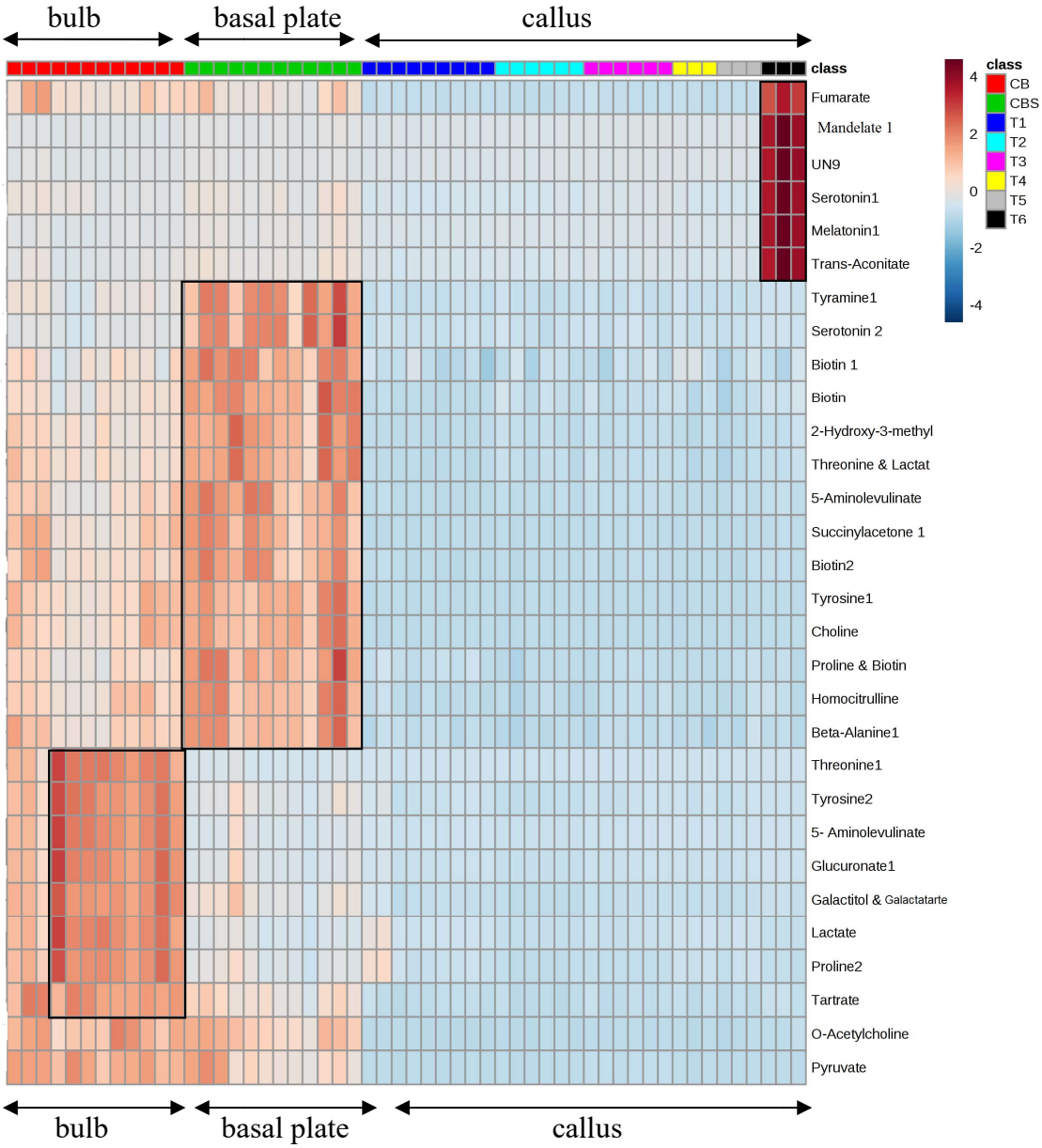

Fig. 4.8 Heatmap representing the top thirty metabolites identified and selected by ANOVA, their distribution, and relative amount in each sample under study. T10D = callus from medium T1 harvested before any treatment, T1-T3 (7D) = callus grown on media T1-T3, harvested after 7days and T1-T6 (30D) = callus from media T1-T6, harvested after 30 days of culture. The boxes on right indicate the relative amount of the metabolites present in samples. The three boxes outlined with black line added to the heatmap represent the metabolites present in relatively high concentrations in bulb, basal plate and callus from T6.

4.3.2.1.3 Multivariate analysis (PCA)

The multivariate data analysis method principal component analysis was also used to compare all the data to obtain an unbiased overview of the differences between field derived tissues and callus. A three-component model explained 93.5% of the total

variance (Figure 4.9) with 4th and 5th components explaining an additional 2.4% and 1% variance respectively.

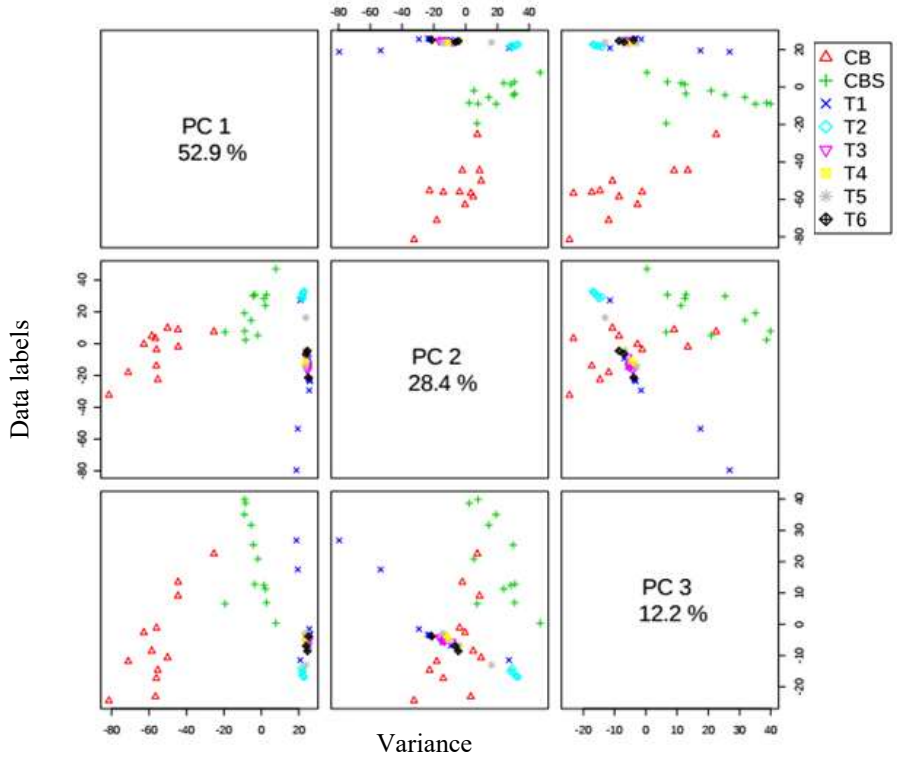

Fig. 4.9 A three-component pair wise score plot between the selected PCs for field and callus tissues, representing variance of each PC in the corresponding diagonal cell. CB= Carlton bulb tissue, CBS= Carlton basal plate and T1-T6= callus derived from different media treatments.

The first two principal components accounted for more than 80% of the variance in the dataset. In the score plot of PC1 and PC2, the field samples (bulb and basal plate) were clearly separated from each other and from the *in vitro* grown callus (T1-T6) along PC1 (Figure 4.10A).

The samples with the lowest score on PC1 (most to the left on score plot) were Carlton bulb samples grown in the field, clearly distinguished from the field grown Carlton basal plate samples with middle scores and from the tissue culture derived callus in the high scoring area of PC1. Thus, the callus cluster was highly correlated with PC1 variance (52.9%), the basal plate showed moderate correlation with both PCs and a highly negative correlation with PC1 variance was observed in the bulb samples. Callus samples showed an overlapping distribution along PC2 and the callus prior to treatments (T1, symbol ×) is clearly distinguishable from others on the PC2 axis and lies in the negative scored area of PC2 (Figure 4.10A).

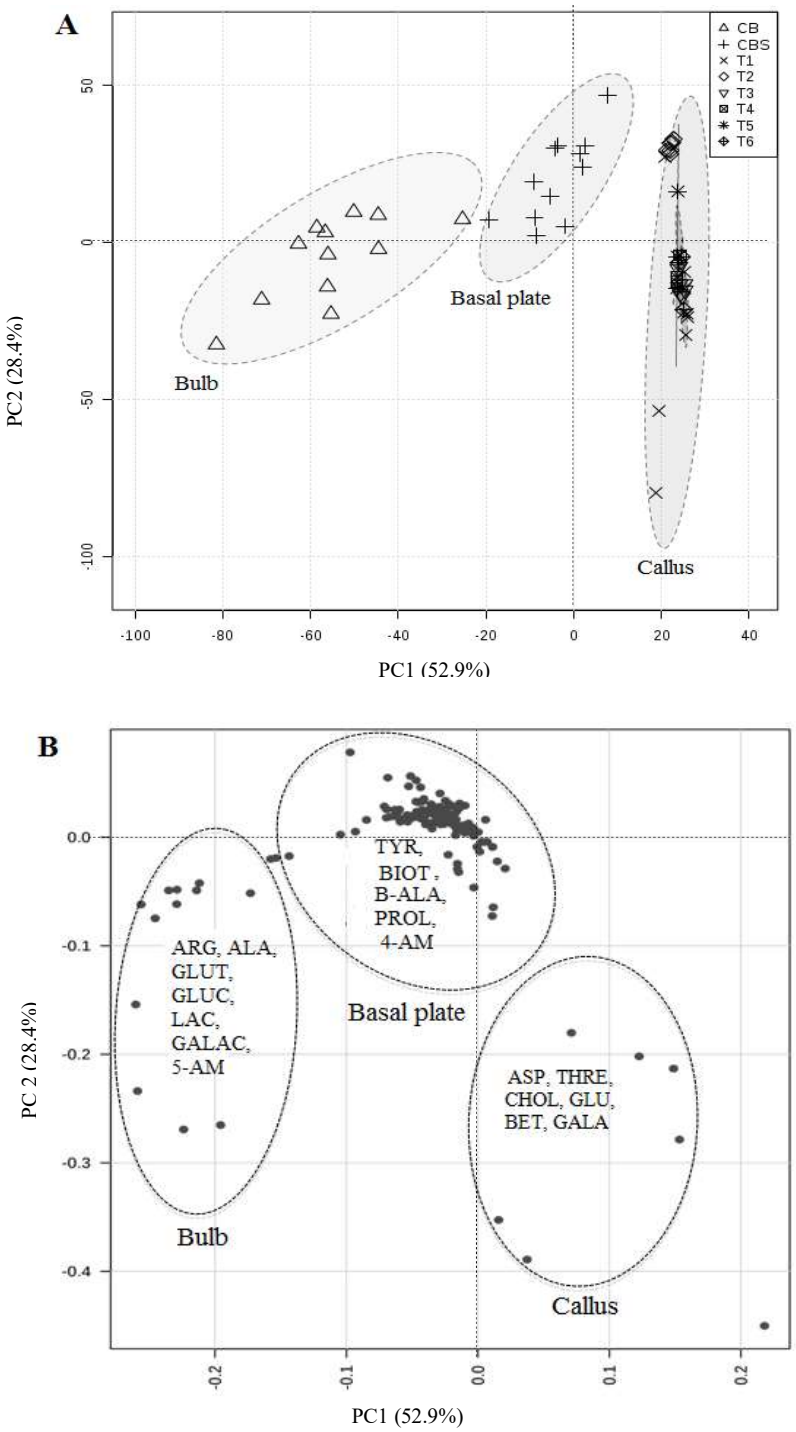

Fig. 4.10 Score scatter plot (A) for principal component analysis (PC1 versus PC2) obtained from ^{1}H NMR spectra of *N. pseudonarcissus* cv. Carlton bulb (CB), basal plate (CBS) and callus (T1-T6); and corresponding loading plot (B). In the loading plot (•) represent ^{1}H NMR signals (metabolites). Those important for discrimination of the assigned classes are labelled, 5-AM: 5-aminolevulinate, 4-AM: 4-aminobutyrate, ARG: arginine, ALA: alanine and ASP: asparagine, BET: betaine, B-ALA: beta-alanine, BIO: biotin, CHOL: choline, GLUT: glutamate, GLUC: glucuronate, GALAC: galactarate, GLU: glucitol, GALA: galactitol, LAC: lactate, PROL: proline, TYR: tyramine and tyrosine, THRE: threonine.

The corresponding loadings of assigned sample clusters in score plot are represented in a loading plot (Figure 4.10B). It showed the distribution of metabolites, which were responsible for the separation of the samples under study. The most important signals assigned to bulb tissues were alanine, arginine, glutamate, glucuronate, lactate, galactarate and 5-aminolevulinate. Metabolites responsible for callus clustering, with no overlapping signals were asparagine, choline, threonine, betaine, glucitol and galactitol. However, most of the metabolite signals were found to be densely situated on the cluster from the basal plate. The metabolites assigned for their separation from the other samples were mainly tyramine, tyrosine, biotin, beta-alanine, proline and 4-aminobutyrate.

The metabolite distribution in field and callus samples found in multivariate statistical analysis (PCA) was found to be in accordance with the results obtained from univariate analysis (AVOVA) comparing the heatmap (Figure 4.8) with the loading plot (Figure 4.10B).

4.3.2.1.4 Clustering analysis (dendrogram)

The dendrogram (Figure 4.11) showed that the bulb and basal plate samples fell into two well discriminated groups (clade 1 and 5).

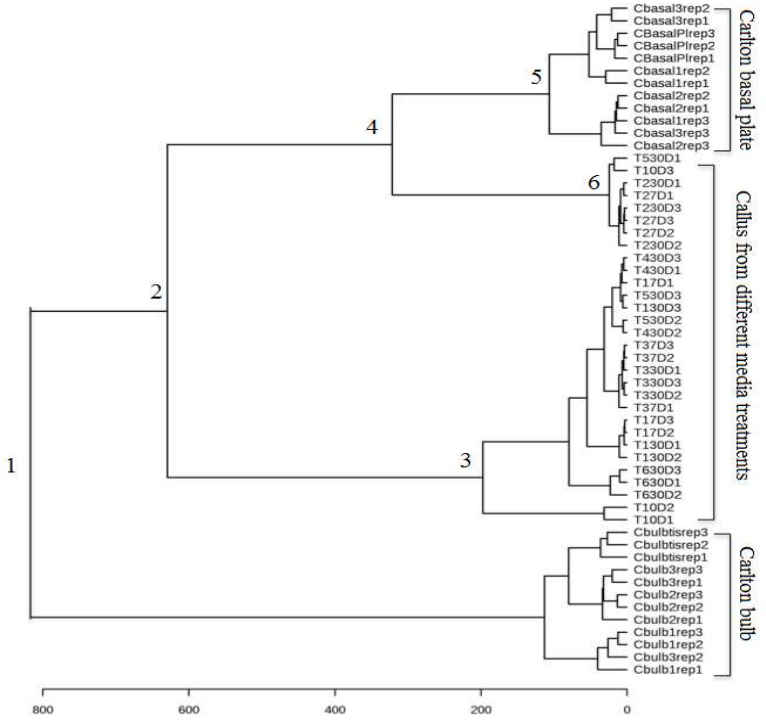

Fig. 4.11 Clustering results shown as dendrogram. Distance measured using Euclidean clustering algorithm using Ward's linkage.

The callus samples formed two clusters (clade 3 and 6), most related to the basal plate samples (clade 2). The majority of callus samples formed a cluster (clade 3) while a small number constituted a separate grouping (clade 6). These samples comprising the entire callus subjected to T2 treatment (media containing 25 g/l sucrose which is half amount of sucrose used for the other treatments) along with one replicate of T1 and T5 (Figure 4.11). The clustering pattern seen in the dendrogram therefore replicates the findings from PCA and ANOVA.

This section of statistical analyses showed the well-defined metabolite distributions or patterns present in both field and callus samples.

The data from the callus samples was grouped for further analysis. Since the calli were subjected to different media treatments and durations, and this sub-grouping could not be observed clearly with the data analysis in this section, a new bucket table was generated taking the data for callus samples separately in the following analyses (section 4.3.2.2).

4.3.2.2 Identification of differences in NMR signal intensity among culture-derived callus subjected to different media (T1-T6) for 7 and 30 days

Calli were cultured on six different media (T1-T6) (Materials and methods, section 4.2.2, Table 4.1), harvested and stored at -80 ^{o}C prior to treatment (0D) from T1, after 7 days (7D) from T1, T2 and T3 and after 30 days of incubation (30D) from T1 to T6. To find the effect of different media and days of treatment on callus metabolite production, data from callus was analysed on its own.

4.3.2.2.1 Univariate (one-way ANOVA and *post-hoc*) analysis

The ANOVA analysis of tissue culture derived callus derived from different media treatments (T1-T6) harvested on day one, after 7 and/ or after 30 days of incubation showed which peaks were significantly different among the samples. MetaboAnalyst parameters were the same as in section 4.3.2.1.1. The ANOVA showed around 52.5% (82 out of 156) of the features (metabolites) were significantly different among calli harvested on different days from the six media treatments (Figure 4.12). The *post-hoc* analysis for all metabolites and calli samples with p-values and Fisher's LSD is in Appendix 4.6.

The top compounds identified by one-way ANOVA for the selected data set of callus were melatonin, mandelate, serotonin, *trans*-aconitate, fumarate and unknown compound 9 (Figure 4.12).

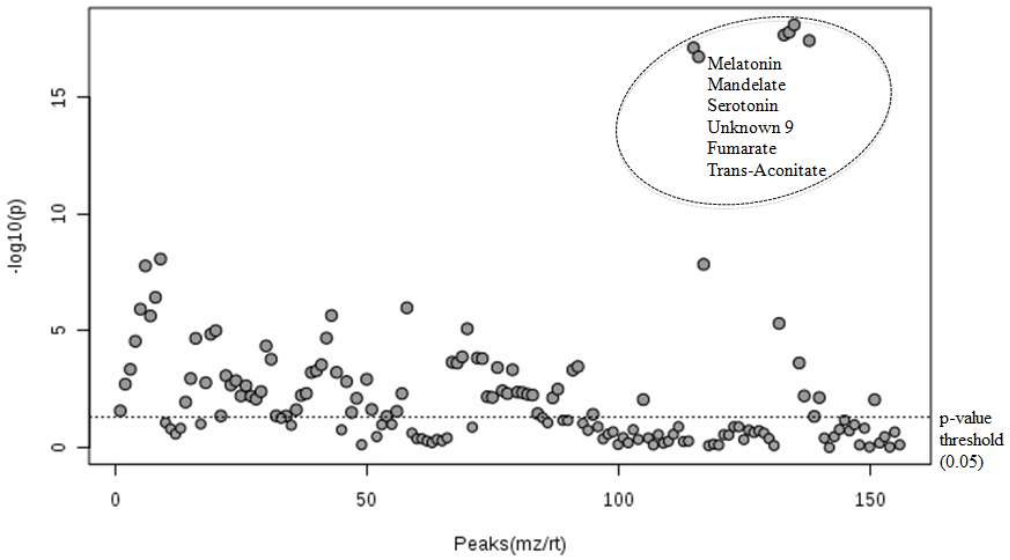

Fig. 4.12 ANOVA in MetaboAnalyst is showing the key peak differences among samples under analysis (callus grown in six different media treatments (T1-T6), harvested on day 1, 7 days and 30 days after incubation) with p-value threshold 0.05. The top compounds are marked with a dotted ring. The X-axis is plotting the sizes of NMR peaks, from which the chemical compound can be determined, and whether the ANOVA showed any significant differences among samples.

4.3.2.2.2 ANOVA heatmap for metabolite clustering

The heatmap generated from ANOVA *post-hoc* analysis (Figure 4.13) represented the top thirty metabolites in callus cultured on different media treatments over three incubation periods. Callus from T1 harvested before treatment (0D) and T6 harvested after 30 days (30D) of incubation showed a clear separation from the other treatments. One replicate of callus (T10D3*) showed only trace amount of metabolites compared with the other two replicates, T10D1 and T10D2 (on left hand side). Metabolites found in callus harvested before treatment (T1) had relative high concentrations of betaine, proline, carnitine, arginine and xylose.

The medium treatment T6 for 30 days (T630D) showed the presence of vanillate, thymol, serotonin, UN9, melatonin, mandelate, fumarate and *trans*-aconitate in relatively high amounts (bottom right hand side). The treatment T3 (callus incubated with increased yeast extract) could also be distinguished from the others based on a spectrum of metabolites (top right hand side). The other samples from treatments T1,

T2, T4 and T5 contained very low levels of metabolites, being grouped by the absence of distinguishing features (Figure 4.13).

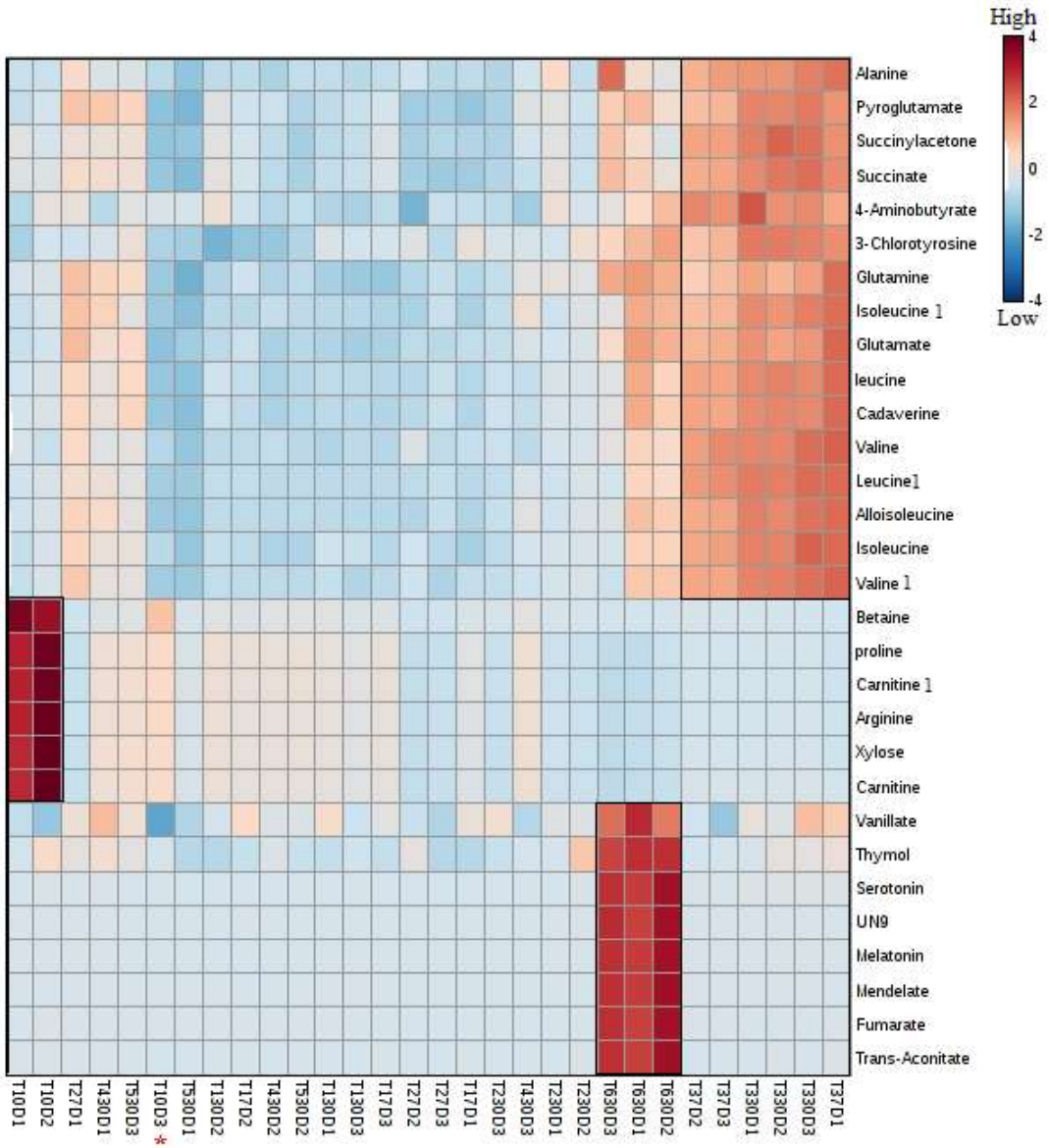

Fig. 4.13 Heatmap representing the top thirty metabolites identified by ANOVA, their distribution and amount in each sample under study; T10D = callus from medium treatment T1 harvested before treatment, T1-T3 (7D) = callus grown on media T1-T3, harvested after 7 days and T1-T6 (30D) = callus from media T1-T6, harvested after 30 days of culture. Numbers 1, 2 and 3 represent three replications for each. The boxes on right indicate the relative amount of the metabolites present in samples. The three boxes outlined with black line added to the heatmap represent the metabolites present in relatively high concentrations in callus incubation with T1 (two replicates), T6 and T3.

4.3.2.2.3 Multivariate analysis (PCA)

A three-component model principal component analysis for callus, grown in different media treatments (T1-T6) explained 96.6% of the variance in total, adding additional

1.6% and 0.6% variance accounted for 4th and 5th components respectively. The first two components accounted for maximum variance of 93.6% in the dataset (Figure 4.14).

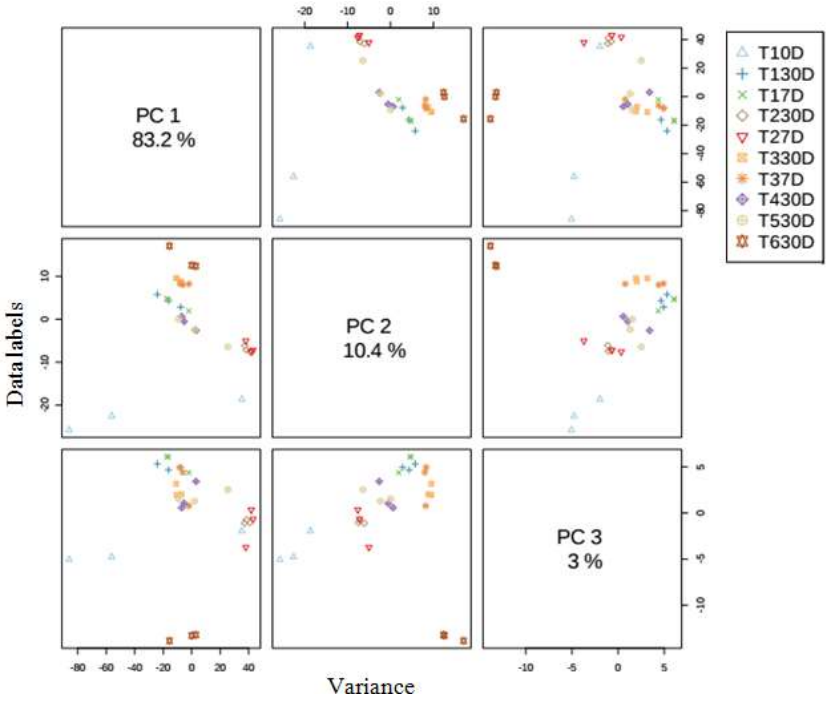

Fig. 4.14 A three-component pair wise score plot between the selected PCs for callus, grown in different media treatments (T1-T6), representing variance of each PC in the corresponding diagonal cell. T10D= callus harvested on day 1 before any treatment, 7D and 30D= callus harvested after 7 days and 30 days of different media treatments.

The score scatter plot of the first two principal components is shown in Figure 4.15A, indicating discrimination of samples according to the media treatments and incubation days of the calli.

The treatment groups showed overlapped clustering except for T10D along PC1 and PC2. However, unlike the heatmap, the other treatment groups could also be resolved. The treatments (T1 to T6, 7 days and 30 days) were separated into closely laying small clusters along PC2 (Figure 4.15). Callus from T1 (MS medium with high auxin, same medium as initial callus induction), harvested prior to treatment showed very clear separation from all other treatments, and it was separately grouped from callus grown on the same medium (T1) harvested after 7 days (T17D) and 30 (T130D) days. The three replications of T10D showed overall distribution along high and low PC1 scored area and on negative scored area along PC2. The T17D and T130 fell into a group, in the same region as the other samples (Figure 4.15A).

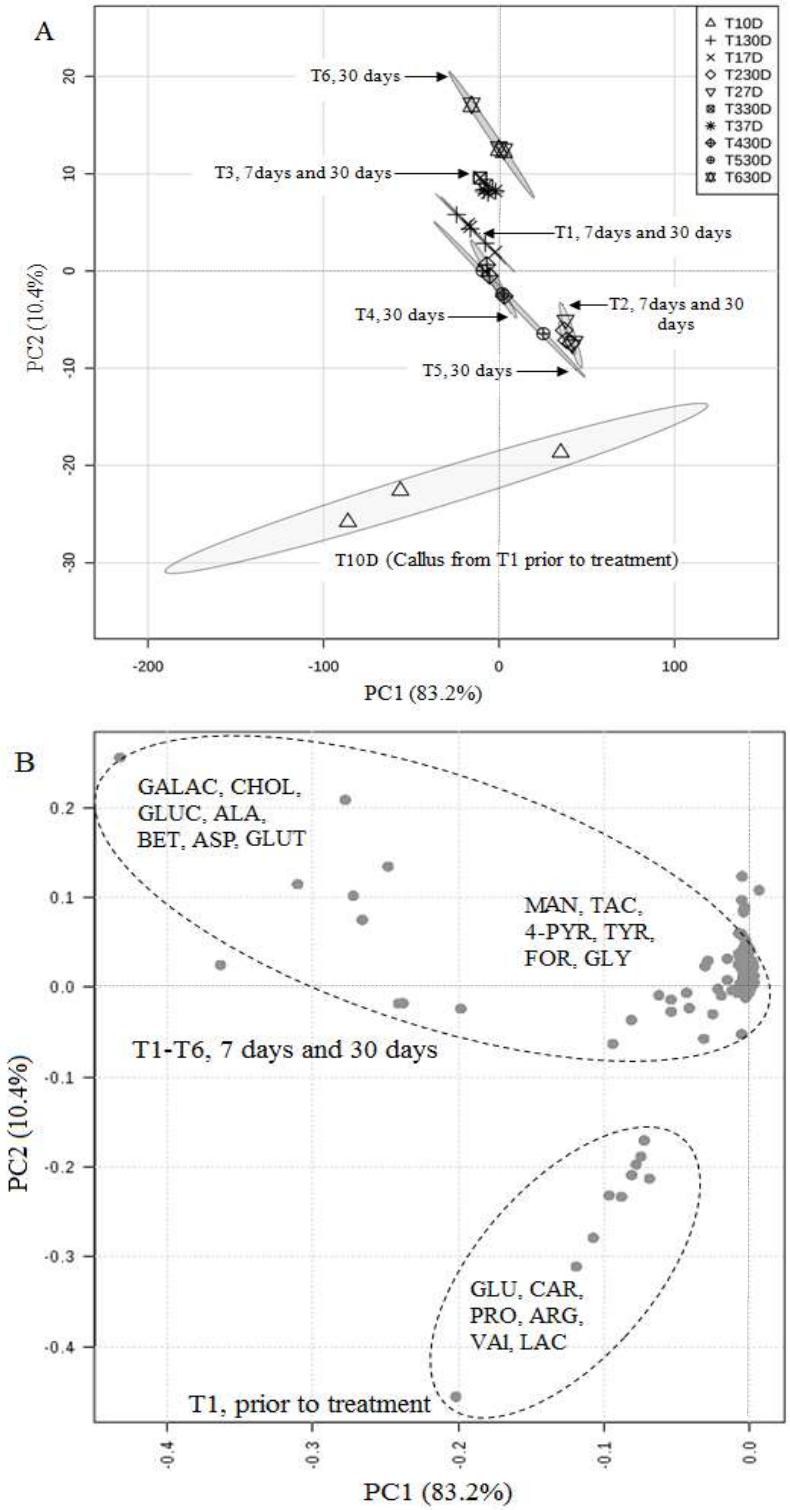

Fig. 4.15 Score scatter plot (A) for principal component analysis (PC1 versus PC2) obtained from ^{1}H NMR spectra of calli grown in different media treatments (T1-T2) harvested prior to treatment (0D), after 7 days (7D) and after 30 days (30D) incubation and corresponding loading plot (B). In the loading plot (•) represent ^{1}H NMR signal buckets (metabolites). Signal buckets (metabolites) important for discrimination of the assigned classes are labelled; ALA: alanine, ASP: asparagine, ARG: arginine, BET: betaine, CHOL: choline, CAR: carnitine, FOR: formate, GALAC: galactitol, GLUC: glucitol, GLUT: glutamate, GLY: glycolate, GLU: glucuronate, LAC: lactate, MAN: mandelate, PRO: proline, TYR: tyrosine, TAC: *trans*-aconitate, VAL: valine, 4-PYR: 4-pyridoxate.

The callus from treatment T2 (medium with 50% less sucrose than T1) and T3 (with increased yeast extract) harvested after both 7 and 30 days, formed two separate groups. Treatments T4 (with methyl jasmonate) and T5 (with chitosan) after 30 days of incubation formed one overlapping distribution in the middle scored region for both PCs. Finally, callus from medium T6 (with TCIN) harvested after 30 days of culture (T630D) lies in one further separate group from the other treatments and on high scored region of PC2 (Figure 4.15A).

The corresponding loading plot (Figure 4.15B) showed that the corresponding ^{1}H NMR signals for the separation of callus cluster on medium T1, harvested prior to treatment were assigned to glucuronate, carnitine, proline, arginine, lactate and valine. All other groups showed overlapped distribution on the loading plot and the distinctive signals belong to galactitol, choline, glucitol, alanine, betaine, asparagine, glutamate, mandelate, *trans*-aconitate, 4-pyridoxate, tyrosine, formate and glycolate.

4.3.2.2.4 Clustering analysis (dendrogram)

The dendrogram (Figure 4.16), generally placing replicates of samples together, shows six separate clusters.

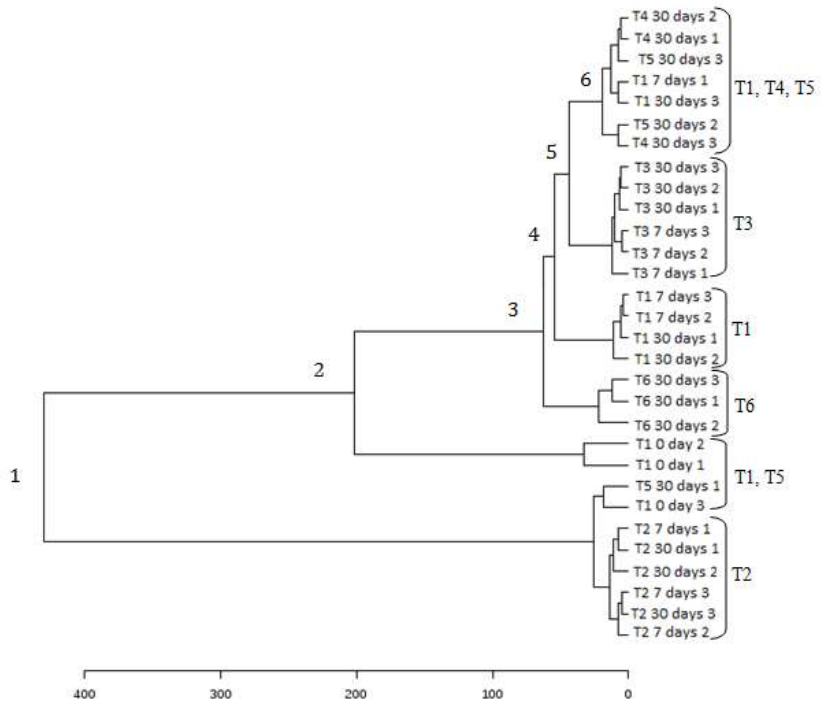

Fig. 4.16 Clustering result shown as dendrogram. Distance measured using Euclidean clustering algorithm using Ward's linkage. T1-T6 = callus incubated in six different media treatments.

All samples from T2 and T3, harvested after both 7 and 30 days, were closely clustered. Samples from T1 were distributed among three clusters. Most of the replicates of T4 and T5 were within clade 6 along with two replicates of T1 (from 7 and 30 days) with the only exception T5, 30 days; replicate1, which clustered with T10D in clade 2. Two replicates of T1 (7 days and 30 days) were also separated from other T1 (clade 6) (Figure 4.16).

4.3.2.3 Identification of differences in NMR signal intensities among field, tissue culture derived direct white shoot, green shoot and regenerated shoots

4.3.2.3.1 Univariate (one-way ANOVA and *post-hoc*) analysis

A one-way ANOVA for data from field derived Carlton bulb, basal plate and tissue culture derived direct white shoot, green shoot and regenerated white shoot was also performed. MetaboAnalyst parameters were the same as stated in section 4.3.2.1.1. The ANOVA showed that most (92.3%) of the features (metabolites) were significantly different (Figure 4.17) among the samples. The most significant metabolites with lowest p-values are circled (4-aminobutyrate, arginine, cadaverine, alanine, leucine and betaine). The *post-hoc* analysis for all metabolites and samples with p-values and Fisher's LSD is shown in Appendix 4.7.

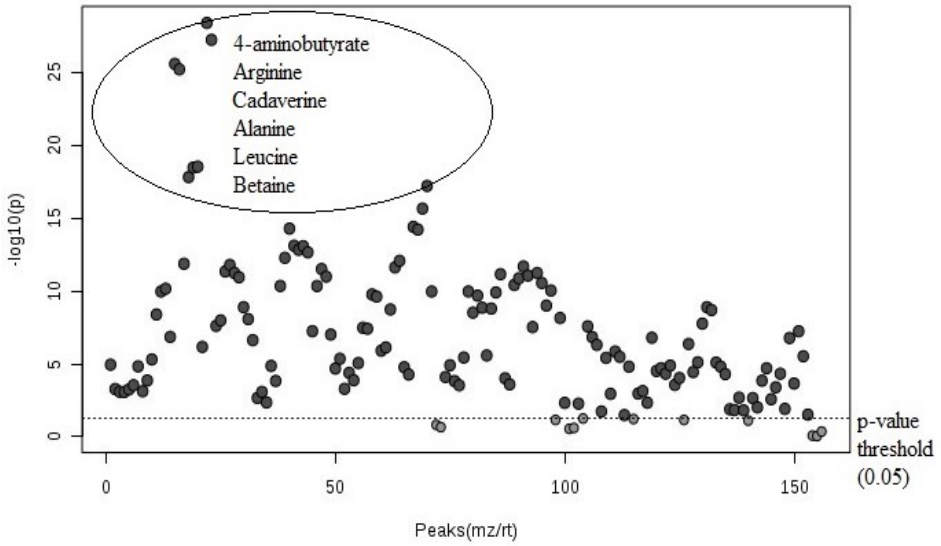

Fig. 4.17 ANOVA in MetaboAnalyst is showing the key peak differences among Carlton basal plate, Carlton bulb and tissue culture derived direct white shoot, green shoot and regenerated white shoot, with p-value threshold 0.05. The X-axis is plotting the sizes of NMR peaks, from which the chemical compound can be determined, and whether the ANOVA showed any significant differences among samples.

4.3.2.3.2 ANOVA heatmap for metabolite clustering

The heatmap for the top fifty metabolites (Figure 4.18) identified from ANOVA *post-hoc* analysis based on Fisher's LSD test are shown here as it showed clear metabolite clustering, while previous heatmaps included top thirty metabolites.

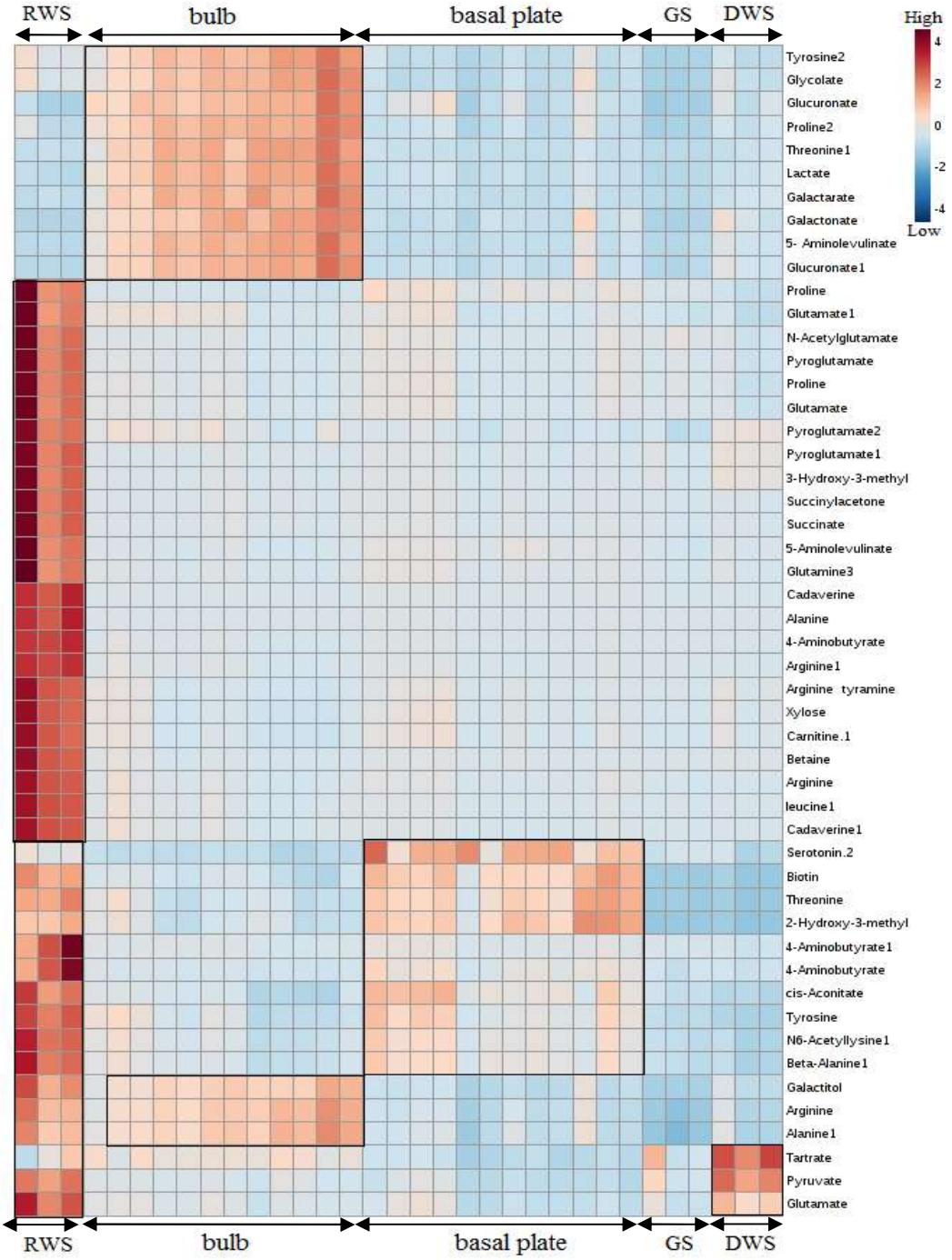

Fig. 4.18 Heatmap representing top fifty metabolites identified by ANOVA, their distribution and amount in each sample under study; CB = Carlton bulb, CBS = Carlton basal plate, DWS = direct white shoot, GS = green shoot and RWS = regenerated white shoot. The boxes on right indicate the relative amount of the metabolites present in samples. The 6 boxes outlined with black line on the heatmap represent the metabolites present in relatively high concentrations in Carlton bulb, basal plate, RWS and DWS.

The heatmap (Figure 4.18) showed that most of the metabolites were present at relatively high amounts in regenerated white shoots. On the other hand, trace amount of tartrate and pyruvate in green shoot and slightly higher amount of tartrate, pyruvate and glutamate in direct white shoot were observed. Moderate to high amount of tyrosine, glycolate, glucuronate, proline, threonine, lactate, galactarate and 5-aminolevulinate were found in Carlton bulb while Carlton basal plate showed trace to moderate levels of serotonin, biotin, threonine, 4-aminobutyrate, tyrosine and beta-alanine (Figure 4.18).

4.3.2.3.3 Multivariate analysis (PCA)

The ^{1}H NMR spectra were submitted to principal component analysis to obtain an overview of the metabolite profiles among all the organised tissue samples under study.

A three-component model (Figure 4.19) principal component analysis for Carlton bulb, basal plate, direct white shoot, green shoot and regenerated shoot explained a total variance of 91.5% in the entire dataset with the 4th and 5th components explaining an additional 2.4% and 1.7% variance respectively. The two first principal components accounted for 79.3% of the total variation.

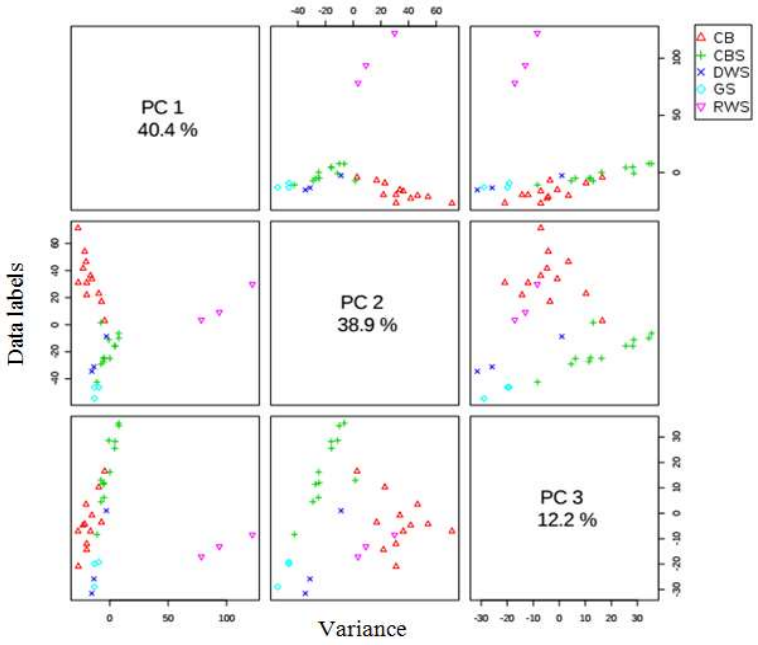

Fig. 4.19 A three-component pair wise score plot between the selected PCs for Carlton bulb, basal plate, direct white shoot, green shoot and regenerated shoot, representing variance of each PC in the corresponding diagonal cell. CB= Carlton bulb tissue, CBS= Carlton basal plate, DWS= direct white shoot from twin-scale, GS= green shoot and RWS= regenerated white shoot from callus.

The score scatter plot of PC1 versus PC2 (Figure 4.20A) showed clear separation of regenerated white shoot (RWS) from the other four tissues. These other tissues were separated along the PC2 axis, with a few overlapping samples

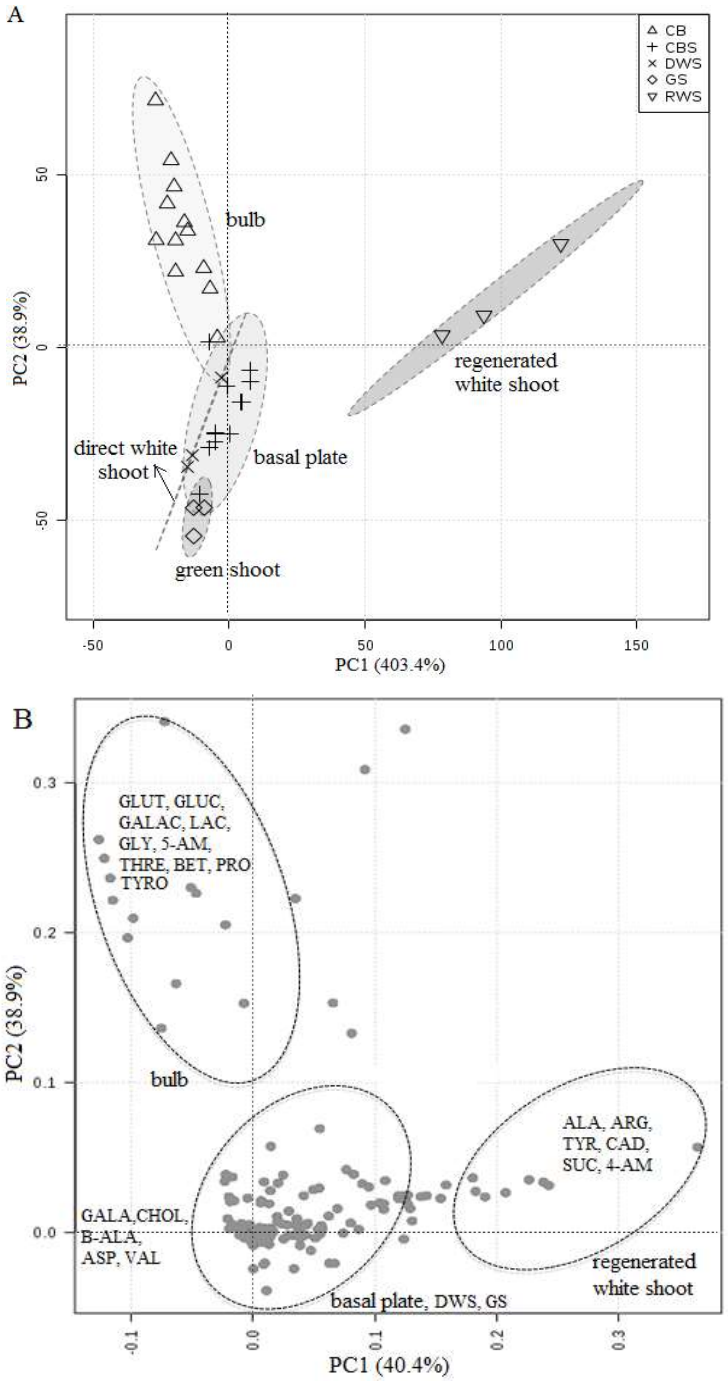

Fig. 4.20 Score scatter plot (A) for principal component analysis (PC1 versus PC2) obtained from ^{1}H NMR spectra of *N. pseudonarcissus* cv. Carlton bulb (CB), basal plate (CBS) tissue culture derived direct white shoot (DWS), green shoot (GS) and regenerated white shoot (RWS). In the loading plot (•) represent ^{1}H NMR signal buckets (metabolites) (B). Signal buckets (metabolites) important for discrimination of the assigned classes are labelled: ALA: alanine, ARG: arginine, ASP: asparagine, BET: betaine, B-AL: beta-alanine, CHOL: choline, CAD: cadaverine, GLUT: glutamate, GLUC: glucuronate, GALAC: galactarate, GLY: glycolate, GALA: galactitol, LAC: lactate, PRO: proline, SUC: succinylacetone, TYR: tyramine, TYRO: tyrosine, THR: threonine, VAL: valine, 5-AM: 5-aminolevulinate and 4-AM: 4-aminobutyrate.

Field samples of bulb tissue were clearly separated from basal plate, green shoot and direct white shoot and positioned on high scored area of PC2 and negative scored area of PC1, hence showing a high correlation with the variance accounted by PC2 (38.9%). The green shoot and direct white shoot had an overlapped grouping along with the basal plate tissue. All were positioned on low scored areas of both PC1 and PC2 and hence had a negative correlation with both PC variances. The regenerated white shoots were positioned on high scored areas of both PCs and had the highest correlation with accounted sample variance as well as a positive effect on sample clustering (Figure 4.20A).

A loading scatter plot of PC1 and PC2 displayed how the different variables contributed to the grouping of samples seen in the PCA score plot (Figure 4.20B). Variables with positive loading scores on both PCs included metabolite signals with high loading scores assigned to alanine, arginine, succinylacetone, 4-aminobutyrate, tyramine and cadaverine for regenerated white shoots. Variables with high positive loading scores on PC2 and negative scores on PC1 represented features for glutamate, glucuronate, galactarate, lactate, glycolate, tyrosine, threonine, glucitol, betaine, and proline and were assigned to Carlton bulb tissues (Figure 4.20B). The variables (metabolites) positioned in the middle region showed densely overlapped distribution of basal plate, green shoot and direct white shoot. Notable variables with positive scores on both PCs were assigned for galactitol, choline, beta-alanine, asparagine and valine (Figure 4.20B).

4.3.2.3.4 Clustering analysis (dendrogram)

The dendrogram (Figure 4.21) showed the separate clustering of Carlton bulb from basal plate, direct white shoot, green shoot and regenerated white shoot. Most of the Carlton bulb replicates showed completely separate clustering from other samples except two replicates that lay closely with CBS (clade 4). Green shoot and direct white shoot were present in the same clade and closely related with CBS replicates (clade 3). Regenerated white shoot samples were in a separate cluster and fell under the same clade with the Carlton bulb replicates (clade 1). The clustering pattern of all samples under study supported all previous analyses (ANOVA and PCA).

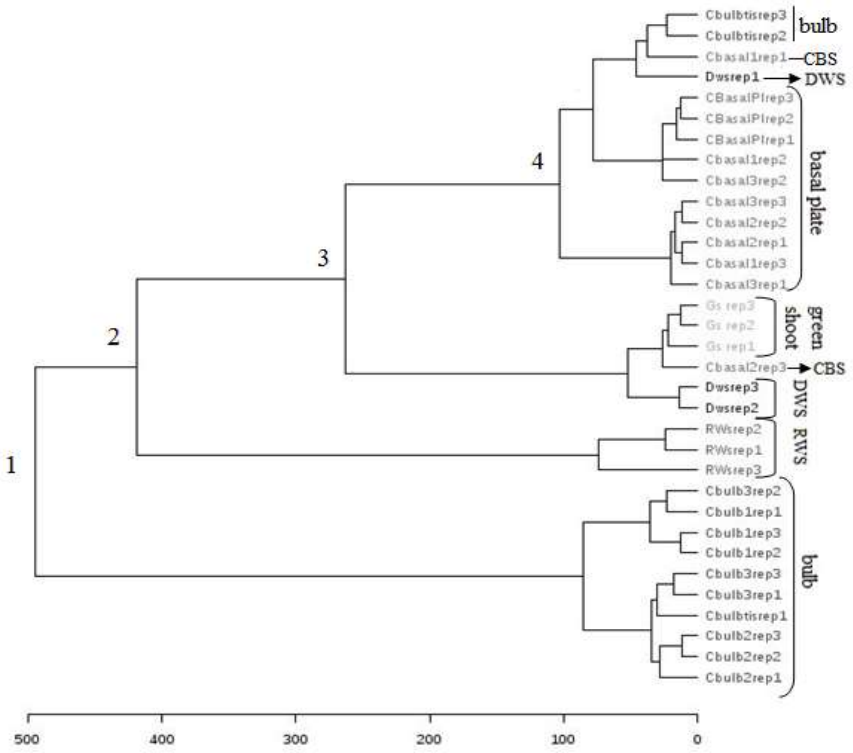

Fig. 4.21 Clustering result shown as dendrogram. Distance measured using Euclidean clustering algorithm using Ward's linkage. CBS = Carlton basal plate, RWS = regenerated white shoot and DWS = direct white shoot.

4.3.2.4 Summary of statistical analyses for tissues (field and tissue culture)

From all the statistical analyses performed, it was observed that more of the identified metabolites were present in the field samples (bulb and basal plate) than the *in vitro* samples (Appendix 4.1). Among tissue culture derived calli, callus from media treatment 6 showed the presence of six metabolites including one unknown compound (UN9) which were present in relatively higher amounts (Table 4.4) and distinguished treatment 6 from the other five treatments. Moreover, a number of the metabolites observed in field samples were also present in relatively high concentrations in regenerated white shoot (Table 4.4).

The signals for identified sugars (Table 4.2) in the tissues under study were removed from the bucket table for all of the above statistical analyses to reduce the complexity of data interpretation. They were analysed separately to observe their relative concentrations in samples under analysis. Most of the sugar signals showed overlapping peaks and were mainly present in callus. The most important sugar signals observed in samples under study with their relative concentrations are shown in Figure 4.22.

Table 4.4 The metabolites found to be differentially present in relatively high amounts in samples under study according to all statistical analyses (ANOVA and PCA).

Samples under study	Metabolites identified in respective samples
Carlton bulb	Alanine, Glycolate, Galactarate, Glucuronate, Galactonate, Glutamate, Lactate, Pyroglutamate, Proline, Tyrosine, 5-Aminolevulinate
Carlton basal plate	Biotin, Cytidine, Caffeine, Chlorogenate, Guanosine, Homocitrulline, Histidine, Histamine, Imidazol, Indole-3-acetate, Lactose, *N*-Acetylserotonin, *N*-Acetyllysine, p-cresol, Succinylacetone, Salicyclate, Tri-methylamine, Tyramine, 3-chlorotyrosine, 3-Hydroxymandelate, 3-Hydroxyvalerate, 2-Hydroxyisobutyrate, 3-Hydroxymethylglutaric acid, 3-Phenylpropionate, 4-Aminobutyrate, 4-Pyridoxate
Callus subjected to media treatments (T1-T5)	Asparagine, Acetoacetate, Betaine, Choline, Glucose, Glucose-6-phosphate, Glucitol, Galactitol, Ribose, Sucrose, Threonine, Valine, Xylose
Callus subjected to media treatment T6	Fumarate, Mandelate, Melatonin, Serotonin, *Trans*-aconitate
Regenerated white shoot	**Alanine**, Arginine, Acetamide, **Betaine**, Carnitine, Cadaverine, **Glutamate**, Glutamine, Hydroxyacetone, Leucine, Pyruvate, **Succinylacetone**, Succinate, **Tyramine, 4-Aminobutyrate**
Direct white shoot	Beta-alanine, Glutamate, Pyruvate, Pyroglutamate, Tartrate, 2-Hydroxyisobutyrate, 3-Hydroxybutyrate
Green shoot	Chlorogenate, Homocitrulline, O-Acetylcholine, Serotonin, Trigonelline

Metabolites in bold were present in higher amounts in regenerated white shoot than field samples (bulb and basal plate). The metabolites listed for direct white shoot and green shoot were present in trace amount when compared with the field samples.

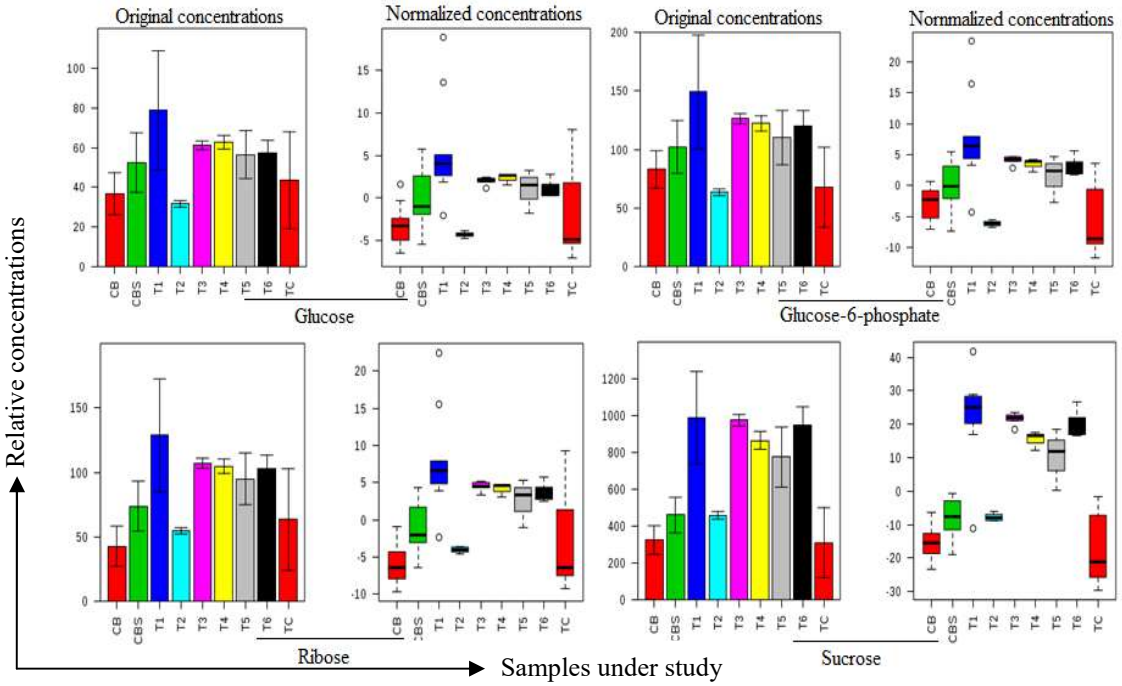

Fig. 4.22 ANOVA boxplot representing relative concentrations of the most abundant sugars (field samples = CB and CBS; callus = T1-T6; TC = regenerated white shoot, direct white shoot and green shoot).

4.3.2.5 Identification of differences in NMR signal intensity among media treatments (T1-T6) after harvesting the calli from media

4.3.2.5.1 Univariate (one-way ANOVA) analysis

The one-way ANOVA of media used for calli treatments indicated the presence of significantly different peaks (metabolites) among media treatments. MetaboAnalyst parameters were the same as stated in section 4.3.2.1.1. The ANOVA showed that most (60 out of 64) of the metabolites were significantly different (Figure 4.23) except aspartate, galactose, trehalose and agmatine.

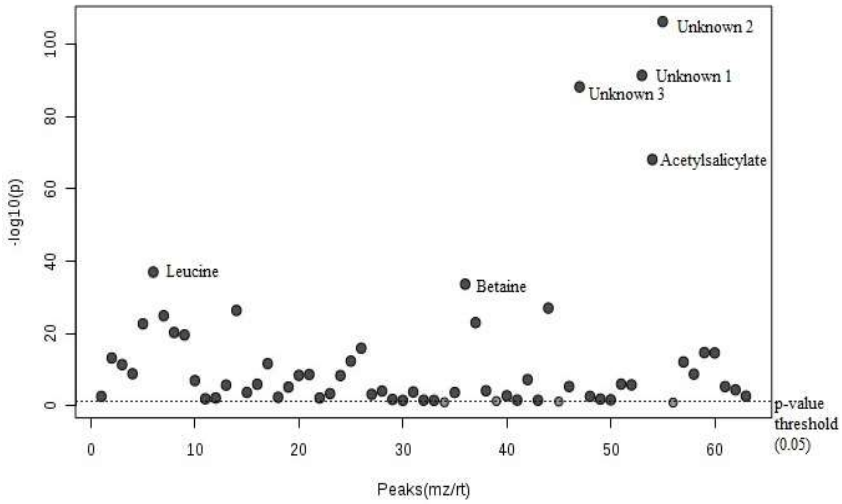

Fig. 4.23 ANOVA in MetaboAnalyst is showing the key peak differences among media with p-value threshold 0.05. The X-axis is plotting the sizes of NMR peaks, from which the chemical compound can be determined, and whether the ANOVA showed any significant differences among samples.

4.3.2.5.2 Multivariate analysis

A two-component model (Figure 4.24) Principal Component Analysis for media treatments explained a total variance of 69.2% in the entire dataset with an additional 11.4%, 6.2% and 5.5% variance accounted for the 3rd, 4th and 5th components, to finally comprise total of 92.3% variance.

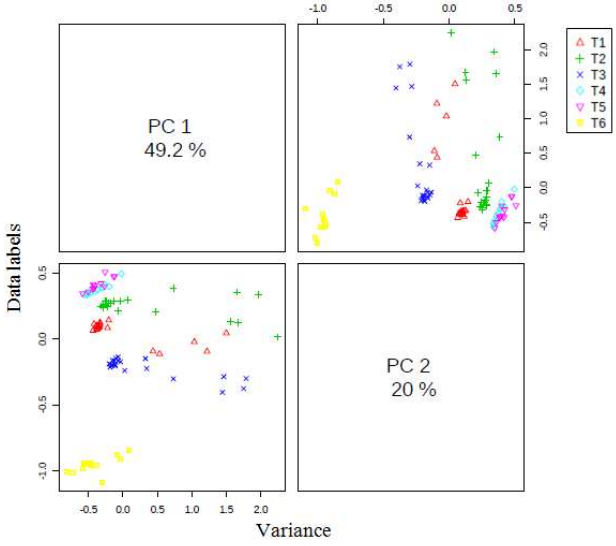

Fig. 4.24 A two-component pair wise score plot between the selected PCs for media, representing variance of each PC in the corresponding diagonal cell. T1-T6 = different media used for callus treatment.

The score scatter plot of PC1 versus PC2 (Figure 4.25A) showed separate clusters of media used for calli treatment. Media T6 and T3 were clearly separated from the others. Media T1, T2, T4 and T5 mainly overlapped while T4 and T5 were in the same cluster.

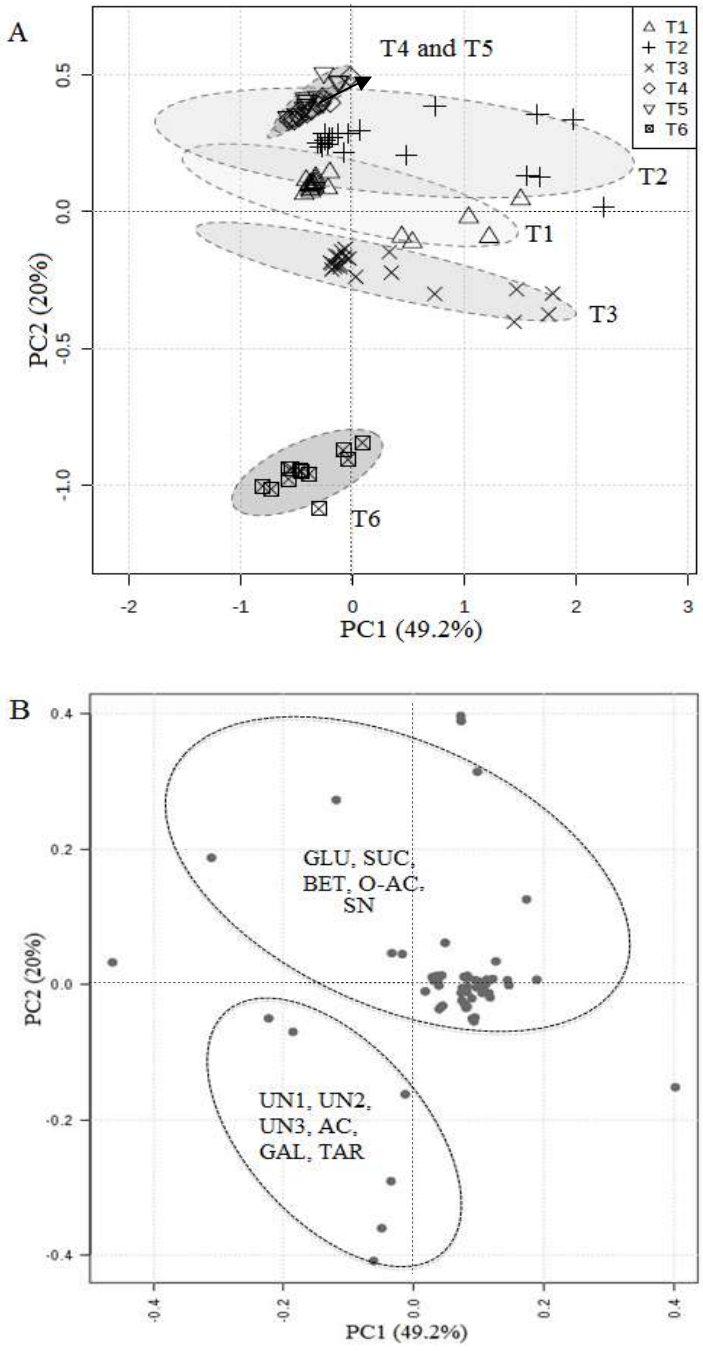

Fig. 4.25 Score scatter plot (A) for principal component analysis (PC1 versus PC2) obtained from ^{1}H NMR spectra of media treatments. In the loading plot (•) represent ^{1}H NMR signal buckets (metabolites) (B). Signal buckets (metabolites) important for discrimination of the assigned classes are labelled; UN: unknown, AC: acetylsalicylate, BET: betaine, GLU: glucose, GAL: galactose, O-AC: O-acetylcholine, SUC: sucrose, SN: sn- glycero-3-phosphocholine.

The corresponding loading plot (Figure 4.25B) was very difficult to determine for the media cluster groups as the media clusters (T1-T5) showed mainly overlapped distribution. They have therefore been divided into two groups; variables (metabolites) for medium T6 were assigned for unknown compound 1, 2 and 3, acetylsalicylate, galactose and tartrate. The other group was for all the other media treatments and the assigned variables with positive loading scores for both PCs were mainly glucose, sucrose, betaine, O-acetylcholine and sn-glycero-3-phosphocholine (Figure 4.25B).

4.3.3 Pathway analysis

Metabolites found in field samples, callus and media treatments were selected separately and submitted to MetPA (www.metaboanalyst.ca), were compared and cross-referenced against hits from the KEGG pathway databases. The pathway library for *Arabidopsis thaliana* was used for all pathway analyses. Over-representation analysis was achieved by the hypergeometric test and the pathway topology analysis was performed using relative-betweenness centrality (section 4.1.4.4) (MetPA, www.metaboanalyst.ca). The individual contribution from each perturbed pathway was visualised by plotting the –log (p) value from the pathway enrichment analysis against the pathway impact values obtained from the pathway topological analysis (Xia *et al.,* 2015; Weng *et al.,* 2015).

4.3.3.1 Pathway analysis for field samples (Carlton bulb and basal plate)

The pathway analysis of metabolites present in bulb and basal plate is summarised in Figure 4.26, suggested the involvement of 52 different metabolic pathways (Table 4.5). Among these, 8 pathways had a high impact on the pathway topology analysis with the maximum number of significant hits (compounds) for the respective pathways (bold in Table 4.5).

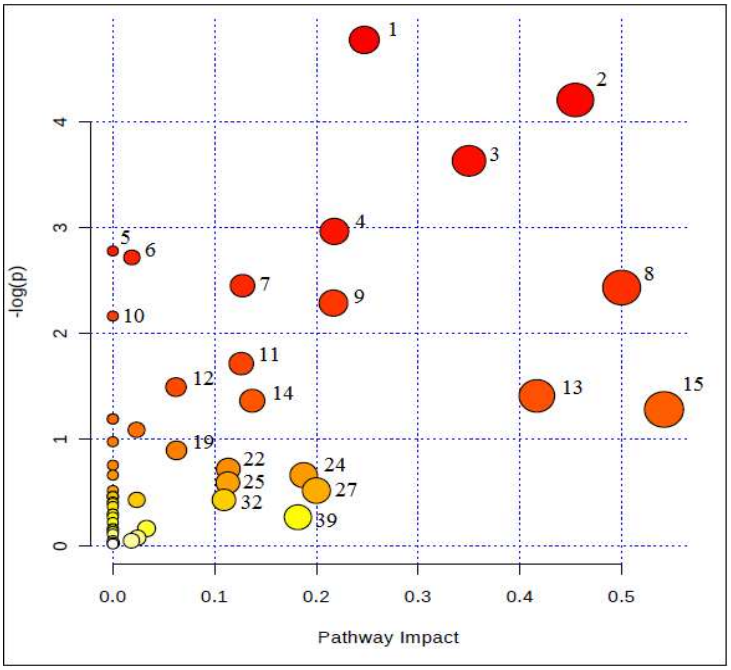

Fig. 4.26 The overview of pathway analysis chart shows the influence of each pathway on tissue metabolism (Pathway impact, X-axis) against how many metabolites in the submitted list appear in that pathway (-log (p), Y-axis). Each circle represents a different pathway, labelled according to Table 4.5. The colour of the circles (yellow to red) represents the presence of metabolites at different levels (low to high) of significance and the size of the circle indicates pathway influence.

Table 4.5 Pathways detected from MetPA for field samples (Total = total number of metabolites involved in each pathway in KEGG, Hits = number of metabolites matched from field sample data. Major pathways are marked in bold).

No.	Suggested pathways	Total	Hits	p-value	Impact
1	**Alanine, aspartate and glutamate metabolism**	**22**	**6**	**0.008**	**0.247**
2	**Tyrosine metabolism**	**18**	**5**	**0.015**	**0.455**
3	**Pantothenate and CoA biosynthesis**	**14**	**4**	**0.027**	**0.35**
4	**Glyoxylate and dicarboxylate metabolism**	**17**	**4**	**0.052**	**0.218**
5	Butanoate metabolism	18	4	0.062	0
6	Valine, leucine and isoleucine biosynthesis	26	5	0.066	0.019
7	Citrate cycle (TCA cycle)	20	4	0.086	0.127
8	**Isoquinoline alkaloid biosynthesis**	**6**	**2**	**0.088**	**0.5**
9	**Arginine and proline metabolism**	**38**	**6**	**0.102**	**0.217**
10	Aminoacyl-tRNA biosynthesis	67	9	0.115	0
11	Galactose metabolism	26	4	0.18	0.126
12	Pyrimidine metabolism	38	5	0.224	0.062
13	**Methane metabolism**	**11**	**2**	**0.244**	**0.417**
14	Glycine, serine and threonine metabolism	30	4	0.255	0.137
15	**beta-Alanine metabolism**	**12**	**2**	**0.277**	**0.542**
16	Synthesis and degradation of ketone bodies	4	1	0.303	0

17	C5-Branched dibasic acid metabolism	4	1	0.303	0
18	Valine, leucine and isoleucine degradation	34	4	0.336	0.023
19	Nitrogen metabolism	15	2	0.376	0
20	Histidine metabolism	16	2	0.407	0.063
21	Indole alkaloid biosynthesis	7	1	0.469	0
22	Starch and sucrose metabolism	30	3	0.486	0.113
23	Tropane, piperidine and pyridine alkaloid biosynthesis	8	1	0.515	0
24	One carbon pool by folate	8	1	0.515	0.188
25	Pyruvate metabolism	21	2	0.553	0.113
26	Stilbenoid, diarylheptanoid and gingerol biosynthesis	10	1	0.595	0
27	Riboflavin metabolism	10	1	0.595	0.2
28	Inositol phosphate metabolism	24	2	0.627	0
29	Cyanoamino acid metabolism	11	1	0.63	0
30	Vitamin B6 metabolism	11	1	0.63	0
31	Glycerophospholipid metabolism	25	2	0.65	0.023
32	Glycolysis or Gluconeogenesis	25	2	0.65	0.109
33	Pentose and glucuronate interconversions	12	1	0.663	0
34	Glutathione metabolism	26	2	0.671	0
35	Tryptophan metabolism	27	2	0.692	0
36	Ascorbate and aldarate metabolism	15	1	0.743	0
37	Propanoate metabolism	15	1	0.743	0
38	Fructose and mannose metabolism	16	1	0.766	0
39	Folate biosynthesis	16	1	0.766	0.182
40	Pentose phosphate pathway	18	1	0.805	0
41	Phenylalanine, tyrosine and tryptophan biosynthesis	21	1	0.852	0
42	Carbon fixation in photosynthetic organisms	21	1	0.852	0.033
43	Glucosinolate biosynthesis	54	3	0.861	0
44	Ubiquinone and other terpenoid-quinone biosynthesis	23	1	0.877	0
45	Terpenoid backbone biosynthesis	25	1	0.897	0
46	Porphyrin and chlorophyll metabolism	29	1	0.929	0.024
47	Cysteine and methionine metabolism	34	1	0.955	0
48	Fatty acid metabolism	34	1	0.955	0.018
49	Purine metabolism	61	2	0.975	0
50	Amino sugar and nucleotide sugar metabolism	41	1	0.977	6E-04
51	Flavonoid biosynthesis	43	1	0.981	0
52	Phenylpropanoid biosynthesis	45	1	0.984	0

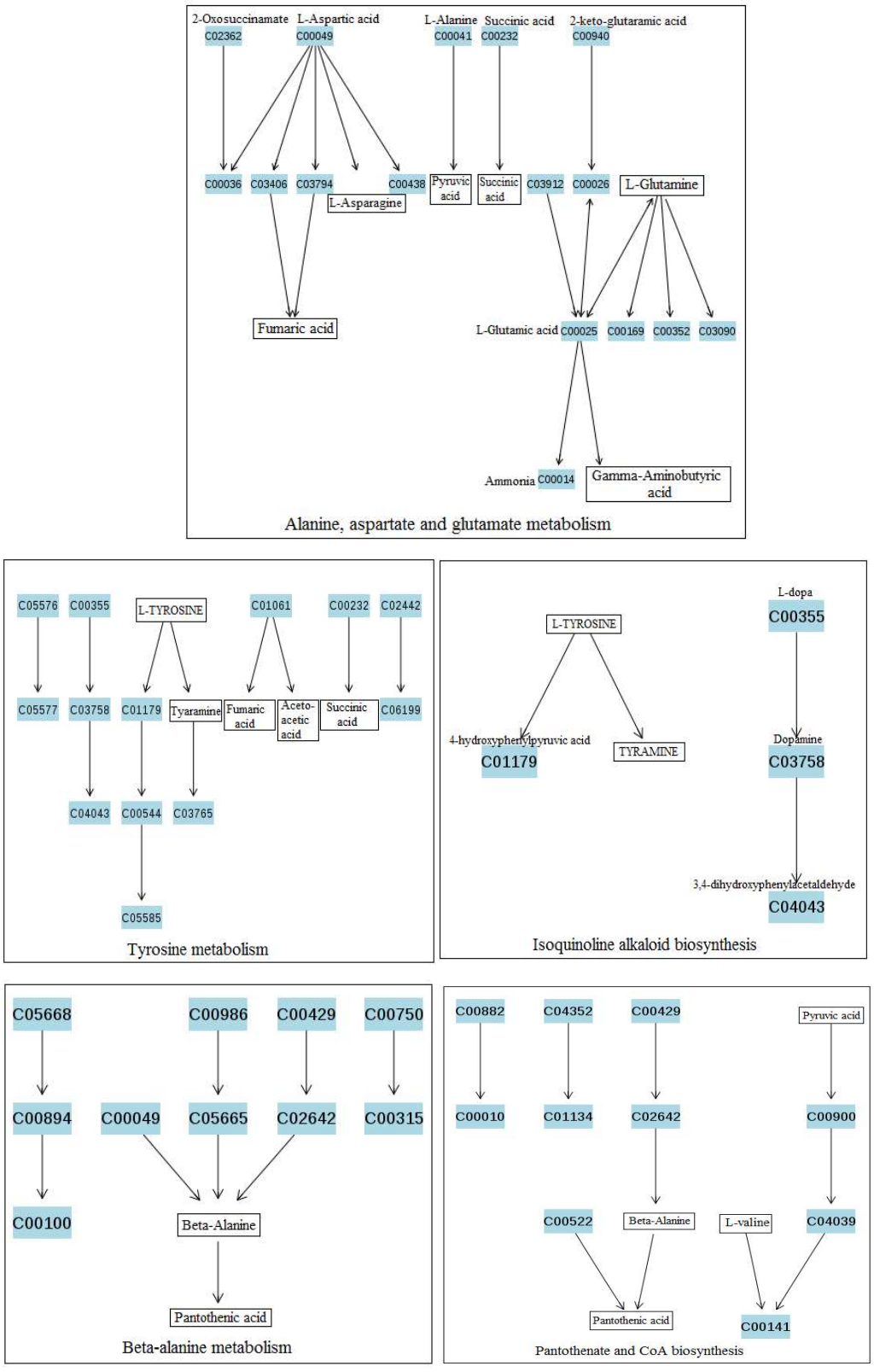

Fig. 4.27 Flowchart generated by MetaboAnalyst showing the significant pathways (alanine, aspartate and glutamate metabolism, tyrosine metabolism, isoquinoline alkaloid biosynthesis, beta-alanine metabolism, pantothenate and CoA biosynthesis) in bulb and basal plate tissues of *N. pseudonarcissus* with pathway contributor metabolites identified by their name box labelled. Other metabolites are labelled with their KEGG IDs.

These major pathways observed in the field tissues were alanine, aspartate and glutamate metabolism, tyrosine metabolism, pantothenate and CoA biosynthesis, isoquinoline alkaloid biosynthesis, beta-alanine metabolism and methane metabolism. Tyrosine metabolism was the pathway that had a contribution of high impact factor (0.45) and the second highest number (5) of identified metabolites as pathway contributors (L-tyrosine, tyramine, fumaric acid, acetoacetic acid and succinic acid) (Figure 4.27).

Alanine, aspartate and glutamate metabolism had the highest number (6) of identified metabolites as pathway contributors (L-asparagine, fumaric acid, succinic acid, pyruvic acid, L-glutamine and gamma-aminobutyric acid) (Figure 4.27). Beta-alanine metabolism (impact factor 0.54) and isoquinoline alkaloid biosynthesis (impact factor 0.50) had the highest impact on pathway analysis but only 2 metabolites for each were identified as pathway contributors. These metabolites were β-alanine and pantothenic acid for β-alanine metabolism and L-tyrosine and tyramine for isoquinoline alkaloid biosynthesis (Figure 4.27).

4.3.3.2 Pathway analysis for *in vitro* callus subjected to different media

Metabolites present in calli subjected to MetaboAnalyst pathway analysis indicated a total of 36 metabolic pathways, summarised in Figure 4.28.

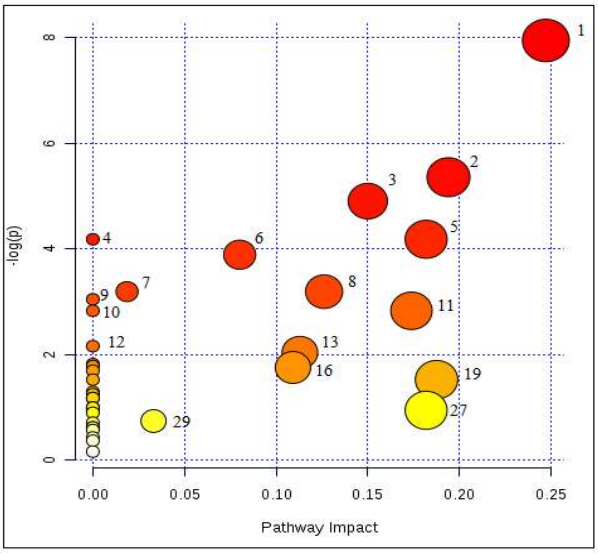

Fig. 4.28 The overview of pathway analysis chart shows the influence of each pathway on tissue metabolism (Pathway impact, X-axis) against how many metabolites in the submitted list appear in that pathway (-log (p), Y-axis). Each circle represents a different pathway, labelled according to Table 4.6. The colour of the circles (yellow to red) represents the presence of metabolites at different levels (low to high) of significance and the size of the circle indicates pathway influence.

Among these, 7 pathways showed a high impact on pathway topology analysis. All the pathways are listed in Table 4.6.

Table 4.6 Pathways detected from MetPA for callus (Total = total number of metabolites involved in each pathway in KEGG, Hits = number of metabolites matched from field sample data. Major pathways are marked in bold).

No.	Suggested pathways	Total	Hits	p-value	Impact
1	**Alanine, aspartate and glutamate metabolism**	**22**	**5**	**0.0003**	**0.25**
2	**Arginine and proline metabolism**	**38**	**5**	**0.005**	**0.20**
3	**Pantothenate and CoA biosynthesis**	**14**	**3**	**0.007**	**0.15**
4	Butanoate metabolism	18	3	0.015	0.0
5	**Tyrosine metabolism**	**18**	**3**	**0.015**	**0.18**
6	**Citrate cycle (TCA cycle)**	**20**	**3**	**0.020**	**0.08**
7	Valine, leucine and isoleucine biosynthesis	26	3	0.041	0.02
8	**Galactose metabolism**	**26**	**3**	**0.041**	**0.13**
9	Aminoacyl-tRNA biosynthesis	67	5	0.047	0.0
10	Glycine, serine and threonine metabolism	30	3	0.059	0.0
11	**Starch and sucrose metabolism**	**30**	**3**	**0.059**	**0.17**
12	C5-Branched dibasic acid metabolism	4	1	0.115	0.0
13	Pyruvate metabolism	21	2	0.130	0.11
14	Inositol phosphate metabolism	24	2	0.162	0.0
15	Isoquinoline alkaloid biosynthesis	6	1	0.168	0.0
16	Glycolysis or Gluconeogenesis	25	2	0.173	0.10
17	Glutathione metabolism	26	2	0.184	0.0
18	Tropane, piperidine and pyridine alkaloid biosynthesis	8	1	0.218	0.0
19	One carbon pool by folate	8	1	0.218	0.19
20	Valine, leucine and isoleucine degradation	34	2	0.275	0.0
21	Cyanoamino acid metabolism	11	1	0.287	0.0
22	Pentose and glucuronate interconversions	12	1	0.310	0.0
23	beta-Alanine metabolism	12	1	0.309	0.0
24	Ascorbate and aldarate metabolism	15	1	0.371	0.0
25	Propanoate metabolism	15	1	0.371	0.0
26	Nitrogen metabolism	15	1	0.371	0.0
27	Folate biosynthesis	16	1	0.390	0.18
28	Glyoxylate and dicarboxylate metabolism	17	1	0.409	0.0
29	Carbon fixation in photosynthetic organisms	21	1	0.478	0.03
30	Glucosinolate biosynthesis	54	2	0.493	0.0
31	Glycerophospholipid metabolism	25	1	0.539	0.0
32	Terpenoid backbone biosynthesis	25	1	0.539	0.0
33	Tryptophan metabolism	27	1	0.567	0.0
34	Cysteine and methionine metabolism	34	1	0.653	0.0
35	Pyrimidine metabolism	38	1	0.694	0.0
36	Purine metabolism	61	1	0.853	0.0

Alanine, aspartate and glutamate metabolism, arginine and proline metabolism, tyrosine metabolism, starch and sucrose metabolism, pantothenate and CoA biosynthesis and galactose metabolism were the notable pathways with relatively high impact and maximum number of identified metabolites as pathway contributors (Figure 4.28, Table 4.6).

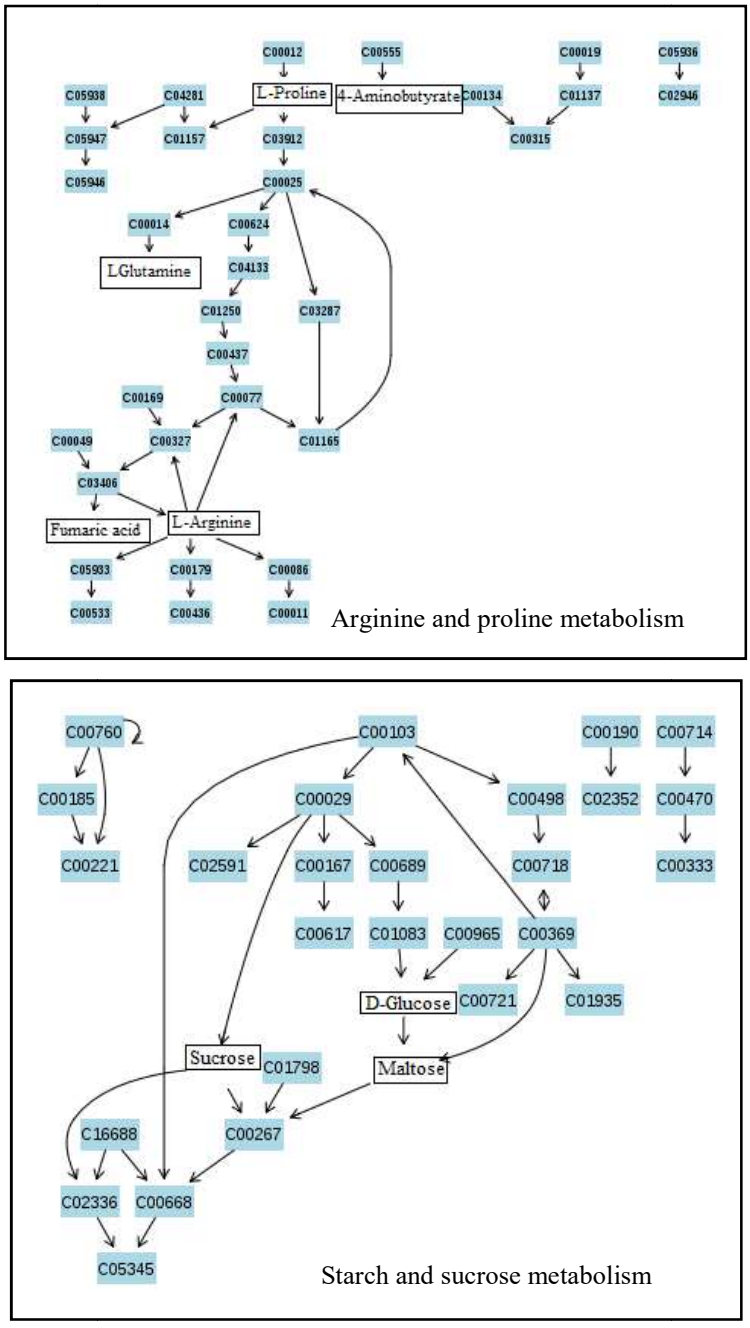

Fig. 4.29 Flowchart generated by MetaboAnalyst showing the significant pathways (arginine and proline metabolism and starch and sucrose metabolism) among tissue culture derived callus of *N. pseudonarcissus* harvested from different media treatments (T1-T6) after different incubation periods with pathway contributor metabolites (box labelled). Other metabolites are labelled with their KEGG IDs. All other pathways have already been listed in Figure 4.27.

Alanine, aspartate and glutamate metabolism showed a combination of highest impact factor (0.25) with highest number (5) of metabolites matched from calli data as pathway contributors. The second pathway which had significant impact (0.20) with highest number of pathway contributor metabolites (5) was arginine and proline metabolism. Tyrosine metabolism also showed high impact (0.18) in callus but less than field samples (0.45) with three matching metabolites as oppose to five.

Alanine, aspartate and glutamate metabolism was defined with the compounds fumaric acid, succinic acid, pyruvic acid, L-glutamine and gamma-aminobutyric acid as pathway identifiers as indicated Figure 4.27. The three metabolites identified as hits for the tyrosine metabolism pathway were tyramine, fumaric acid and succinic acid. The metabolites responsible for pantothenate and CoA biosynthesis pathway were pantothenic acid, pyruvic acid and L-valine (Figure 4.27). The five pathway contributors for arginine and proline metabolism were L-proline, 4-aminobutyrate, L-glutamine, fumaric acid and L-arginine (Figure 4.29). D-glucose, sucrose and maltose were found as pathway contributors in starch and sucrose metabolism pathway (Figure 4.29).

4.3.3.3 Pathway analysis for *in vitro* regenerated white shoot, direct white shoot and green shoot

A total 32 metabolites were present in relatively high amounts in culture derived regenerated white shoots, direct white shoots and green shoots (Figure 4.30). Pathway analysis was performed with all those 32 metabolites, and 21 suggested pathways were obtained from MetPA analysis (Figure 4.30, Table 4.7).

Among them involvement of the three most significant metabolic pathways were identified based on p-values of ≤ 0.05 and an impact factor threshold greater than zero (Table 4.7). They were the arginine and proline metabolism (1), alanine, aspartate and glutamate metabolism (2) and tyrosine metabolism (5) pathways.

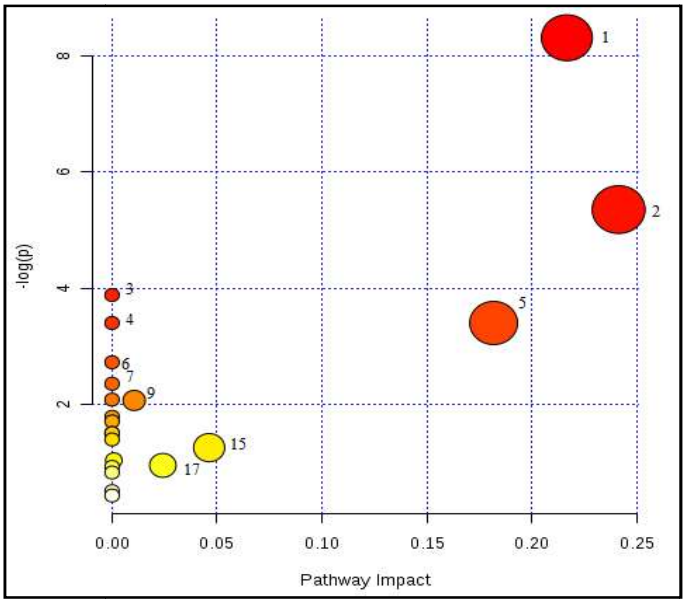

Fig. 4.30 The overview of pathway analysis chart shows the influence of each pathway on tissue metabolism (Pathway impact, X-axis) against how many metabolites in the submitted list appear in that pathway (-log (p), Y-axis). Each circle represents a different pathway, labelled according to Table 4.7. The colour of the circles (yellow to red) represents the presence of metabolites at different levels (low to high) of significance and the size of the circle indicates pathway influence.

The most significant pathway detected was arginine and proline metabolism with the highest number (5) of pathway contributor metabolites and an impact factor of 0.22. The detected metabolites were the same as found in callus (L-proline, 4-aminobutyrate, L-glutamine, L-arginine) except with the addition of L-glutamate (KEGG ID C00624) (Figure 4.29). Alanine, aspartate and glutamate metabolism had highest impact (0.24) on pathway analysis with three matched metabolites; succinic acid, L-glutamine and gamma-aminobutyric acid (Figure 4.27). Tyrosine metabolism was suggested with two pathway contributor metabolites; tyramine and succinic acid (Figure 4.27) with an impact factor of 0.18 in *in vitro* cultured regenerated white shoot, direct white shoot and green shoot.

Table 4.7 Pathways detected from MetPA for tissue culture derived white shoot, green shoot and regenerated shoot (Total = total number of metabolites involved in respective pathways in KEGG, Hits = number of metabolites matched from field sample data. Major pathways are marked in bold).

No.	Suggested pathways	Total	Hits	p-values	Impact
1	**Arginine and proline metabolism**	**38**	**5**	**0.0002**	**0.22**
2	**Alanine, aspartate and glutamate metabolism**	**22**	**3**	**0.005**	**0.24**
3	Aminoacyl-tRNA biosynthesis	67	4	0.02	0.0
4	Butanoate metabolism	18	2	0.03	0.0
5	**Tyrosine metabolism**	**18**	**2**	**0.03**	**0.18**
6	Glutathione metabolism	26	2	0.06	0.0
7	Isoquinoline alkaloid biosynthesis	6	1	0.09	0.0
8	Tropane, piperidine and pyridine alkaloid biosynthesis	8	1	0.12	0.0
9	Pyrimidine metabolism	38	2	0.12	0.01
10	Cyanoamino acid metabolism	11	1	0.16	0.0
11	Pentose and glucuronate interconversions	12	1	0.18	0.0
12	Propanoate metabolism	15	1	0.22	0.0
13	Nitrogen metabolism	15	1	0.21	0.0
14	Glyoxylate and dicarboxylate metabolism	17	1	0.24	0.0
15	Citrate cycle (TCA cycle)	20	1	0.28	0.046
16	Valine, leucine and isoleucine biosynthesis	26	1	0.35	0.0008
17	Porphyrin and chlorophyll metabolism	29	1	0.38	0.024
18	Glycine, serine and threonine metabolism	30	1	0.39	0.0
19	Valine, leucine and isoleucine degradation	34	1	0.43	0.0
20	Glucosinolate biosynthesis	54	1	0.59	0.0
21	Purine metabolism	61	1	0.64	0.0

4.3.3.4 Pathway analysis for media extracts used for the calli treatment

The MetaboAnalyst pathway analysis of metabolites identified in media treatments indicated a total of 29 metabolic pathways (Table 4.8) summarised in Figure 4.31. Among them, involvement of the four most significant metabolic pathways (p-values of ≤0.05; impact factor greater than zero) as shown in Table 4.8 were namely galactose metabolism (1), starch and sucrose metabolism (3) glyoxylate and dicarboxylate metabolism (6) and pantothenate and CoA biosynthesis (5) (Figure 4.31). Alanine, aspartate and glutamate metabolism represented high impact value of 0.2 with p-value 0.07 (2 hits) and methane metabolism had the highest impact (0.4) with the lowest number (1) of hits.

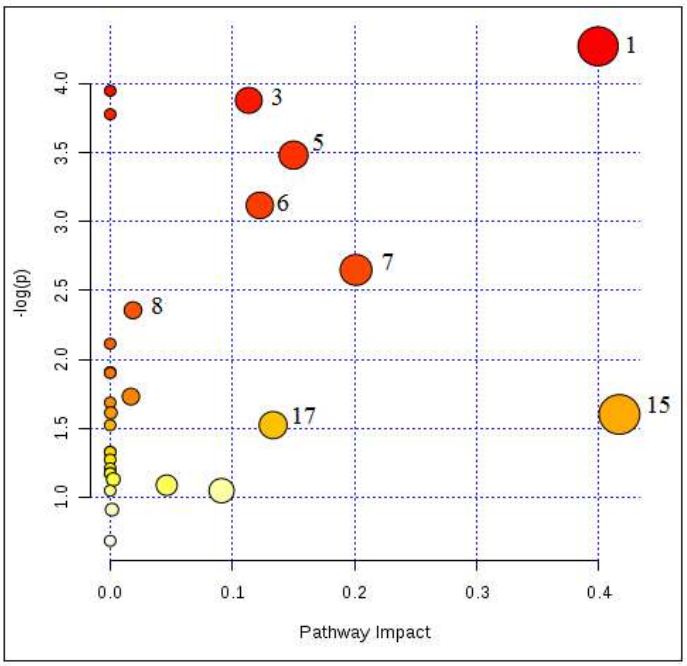

Fig. 4.31 The overview of pathway analysis chart shows the influence of each pathway on tissue metabolism (Pathway impact, X-axis) against how many metabolites in the submitted list appear in that pathway (-log (p), Y-axis). Each circle represents a different pathway, labelled according to Table 4.7. The colour of the circles (yellow to red) represents the presence of metabolites at different levels (low to high) of significance and the size of the circle indicates pathway influence.

Table 4.8 Pathways detected from MetPA for media extracts (T1-T6) (Total = total number of metabolites involved in each pathway in KEGG, Hits = number of metabolites matched from field sample data. Major pathways are marked in bold).

No.	Suggested pathways	Total	Hits	p-values	Impact
1	**Galactose metabolism**	**26**	**3**	**0.014**	**0.399**
2	Cyanoamino acid metabolism	11	2	0.019	0
3	**Starch and sucrose metabolism**	**30**	**3**	**0.021**	**0.113**
4	beta-Alanine metabolism	12	2	0.023	0
5	**Pantothenate and CoA biosynthesis**	**14**	**2**	**0.031**	**0.15**
6	**Glyoxylate and dicarboxylate metabolism**	**17**	**2**	**0.044**	**0.122**
7	**Alanine, aspartate and glutamate metabolism**	**22**	**2**	**0.071**	**0.201**
8	Valine, leucine and isoleucine biosynthesis	26	2	0.095	0.019
9	Glycine, serine and threonine metabolism	30	2	0.121	0
10	Valine, leucine and isoleucine degradation	34	2	0.148	0
11	Aminoacyl-tRNA biosynthesis	67	3	0.149	0
12	Arginine and proline metabolism	38	2	0.177	0.017
13	Lysine biosynthesis	10	1	0.185	0
14	Amino sugar and nucleotide sugar metabolism	41	2	0.199	6E-04
15	**Methane metabolism**	**11**	**1**	**0.202**	**0.417**
16	Nicotinate and nicotinamide metabolism	12	1	0.218	0
17	Sulfur metabolism	12	1	0.218	0.133
18	Propanoate metabolism	15	1	0.265	0
19	Nitrogen metabolism	15	1	0.265	0
20	Histidine metabolism	16	1	0.28	0
21	Glucosinolate biosynthesis	54	2	0.298	0
22	Tyrosine metabolism	18	1	0.309	0
23	Butanoate metabolism	18	1	0.309	0
24	Selenoamino acid metabolism	19	1	0.323	0.003
25	Citrate cycle (TCA cycle)	20	1	0.337	0.046
26	Carbon fixation in photosynthetic organisms	21	1	0.351	0
27	Pyruvate metabolism	21	1	0.351	0.091
28	Glycolysis or Gluconeogenesis	25	1	0.402	0.001
29	Cysteine and methionine metabolism	34	1	0.505	0

The detected metabolites from pathway analysis for galactose metabolism were D-galactose, sucrose and glucose, and for glyoxylate and dicarboxylate metabolism were formate and succinate (Figure 4.32). L-aspartic acid and succinic acid identified alanine, aspartate and glutamate metabolism, while L-valine and pantothenic acid indicated pantothenate and CoA biosynthesis (Figure 4.27). The same metabolites

that were detected in callus were detected for starch and sucrose metabolism in the media pathway analysis (Figure 4.29).

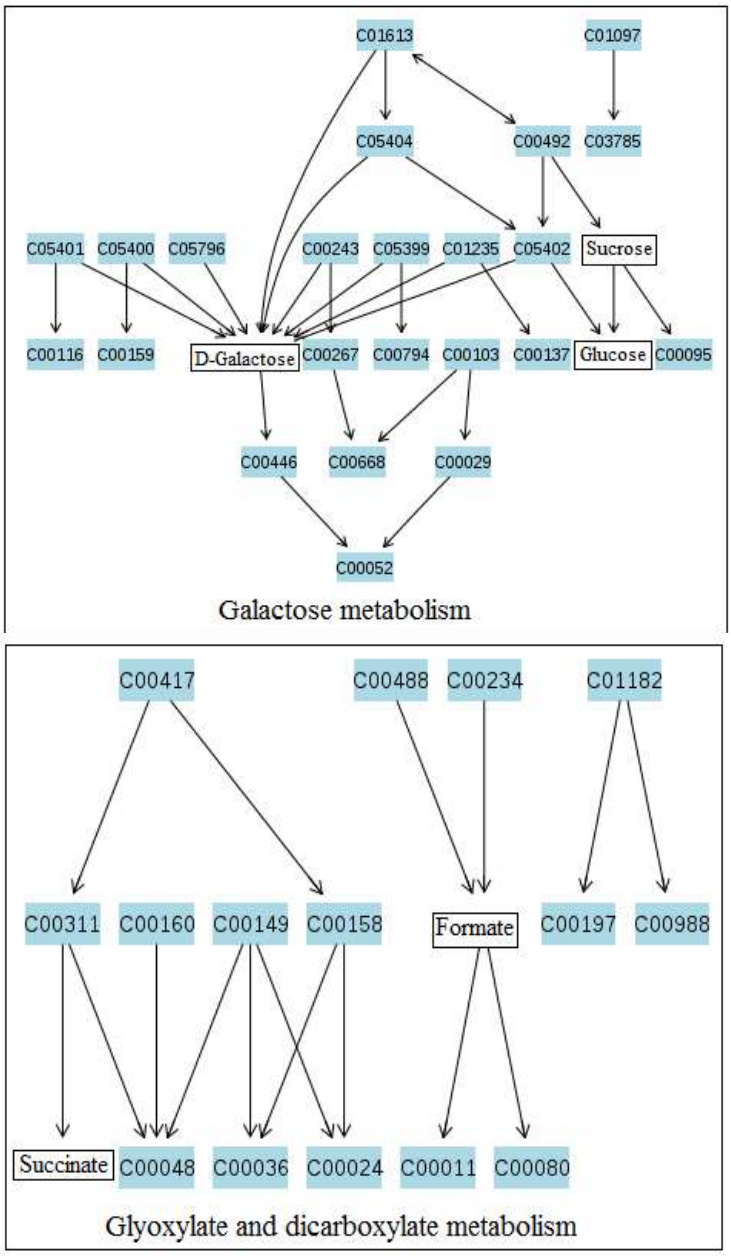

Fig. 4.32 Flowchart generated by MetaboAnalyst showing the significant pathways (galactose metabolism and glyoxylate and dicarboxylate metabolism) in media treatments (T1-T6) (box labelled). Other metabolites are labelled with their KEGG IDs.

4.4 Discussion

NMR-based metabolomics has many applications in plant science including differentiation of plants from different origins, environments or after different treatments (Kim *et al.,* 2010). This study showed NMR-based metabolite identification of *N. pseudonarcissus* cv. Carlton grown in two different

environments; naturally grown field samples and artificially controlled tissue culture samples. Identification of significant pathways involved in the conditions under study is crucial to understand the data obtained from different 'omics' technologies such as metabolomics, genomics and proteomics (Xia and Wishart, 2010). This study revealed the significant metabolic pathways separately for each condition under study i.e. field tissues, *in vitro* calli and shoots; and media extracts for calli treatment.

One of the key challenges of plant secondary metabolite profiling and analysis is to find an optimal balance between accuracy and coverage of metabolite identification (Oksman-Caldentey and Saito, 2005). NMR-based metabolomic analysis often results in mis identification of compounds if an organism specified database is not used (Kim *et al.,* 2011). However, for our study that was not available, although there have been several previous studies of *Narcissus* that have used NMR for metabolite identification (Lubbe *et al.,* 2013; Berkov *et al.,* 2011).

In previous studies, ANOVA has been used to identify significant differences between datasets for the quantification of amino acids, fatty acids, sugars and alkaloids in *Narcissus* bulbs prior to PCA (Lubbe *et al.,* 2012 and 2013). Prior to multivariate analysis, it was important to know if there was any significant difference among signals identified using ^{1}H-NMR, as data with multiple overlapped signals was obtained initially. In the present study, more than 90% signals (metabolites) were found significantly different among tissues (Figure 4.3 and 4.17). On the other hand, calli subjected to different media treatments had only 52.5% significantly different signals (metabolites) (Figure 4.12). The signals (metabolites) detected in media extracts showed around 94% significantly different metabolites among the six media treatments used for calli culture (Figure 4.23). These findings show that there were significant differences in the different datasets that was worthy of further analyses. Therefore, there is a high probability that the majority of assignments were correct. In addition, the pathways were made up with multiple metabolites so false positives would be reduced.

The main findings were the presence of a higher number of metabolites in the field samples than *in vitro* derived calli (Figure 4.1). Naturally, plants contain a huge diversity of primary and secondary metabolites and the most common metabolites observed in plants by NMR were sugars, amino acids, organic acids and other

compounds such as adenine, choline and inositol (Kim *et al.,* 2010). In this study, a total 230 NMR signal peaks of 129 metabolites including 12 unknown compounds (Appendix 4.1, Figure 4.2) were identified in at least one tissue. The majority, 118 metabolites were present in field derived bulb and basal plate tissues (Appendix 4.1).

4.4.1 Field (bulb and basal plate) metabolites

Most of the metabolites detected in the field tissues (bulb and basal plate) in this study were similar to those that have been previously reported in studies using GC-MS (Berkov *et al.,* 2011) and NMR-based extraction from *N. pseudonarcissus* cv. Carlton bulbs (Lubbe *et al.,* 2011, 2012 and 2013). The exception was the presence of sugar alcohols (glucuronate, galactarate, galactonate and galactitol) in our samples. Lubbe et al., used a similar PCA (score and loading) approach as used in this study to identify the metabolic differentiation in *Narcissus* bulbs subjected to different levels of fertilizers (2011); different fungicide treatments (2012) and the effect of growing season on *Narcissus* bulb, leaf and root metabolite production (2013). These studies have also reported the NMR-based identification of Amaryllidaceae alkaloids such as galanthamine, haemanthamine, narciclasine and lycorine. In our study, it was not possible to identify any such alkaloids due to the lack of alkaloids in the reference library (Chenomx). However, precursors for the production of Amaryllidaceae alkaloids such as tyrosine, tyramine, 3-chlorotyrosine, 3-hydroxymandelate, 3-phenylpropionate, homovanillate, p-cresol, succinylacetone and γ-glutamylphenylalanine were detected in field samples.

Our results showed a similar distribution of metabolites in both field derived bulb and basal plate samples but some of the identified metabolites were present in significantly different amounts. The metabolites indicating tyrosine and phenylalanine metabolism were higher in basal plate than bulb tissue (Table 4.4). Another important precursor for Amaryllidaceae alkaloids, tyramine, was also observed in both tissues but was relatively higher in basal plate. The sugar derivatives such as glucitol, galactarate and galactonate were present in high amounts in bulb tissues. This may relate to bulb tissue used for storage therefore, sugar derivatives are high while basal plate is more metabolically active.

The most significant pathway in field samples was tyrosine metabolism with maximum number (5) of hits (-log (p) values from pathway enrichment) and high impact (pathway topology) of 0.455. Moreover, two other significant pathways were detected as isoquinoline alkaloid biosynthesis and beta-alanine metabolism with two hits for each and impact values of 0.5 and 0.54 respectively (Figure 4.26, Table 4.5).

The presence of the most Amaryllidaceae alkaloid precursors in basal plate could relate to the previous findings of higher amount of galanthamine accumulation in basal plate than bulb tissues (Chapter 3). A NMR-based metabolomic study has previously shown the presence of high amount of tyrosine in *Narcissus* bulbs, which was predicted to relate to the up-regulation of the phenylpropanoid pathway for the biosynthesis of phenolic compounds (Lubbe *et al.,* 2013).

MetPA (www.metaboanalyst.ca) have also been reported in field-derived tissues of several plants showing the detection of similar pathways observed in *Narcissus* bulb and basal plate in this study. These include metabolic profiling of tobacco leaves showing proposed metabolic alterations after infection by tobacco mosaic virus (Choi *et al.,* 2006) where the pathways detected as significant were proline, glutamine, alanine and phenylalanine metabolism. Pathway analysis for barley leaves and roots subjected to salt stress has been reported where rice metabolic pathway databases was used as reference (Wu *et al.,* 2013). The results showed that carbohydrate and proline metabolism was enhanced under stressed condition in both leaves and roots, while aspartate metabolism was inhibited.

4.4.2 *In vitro* metabolites

High amount of amino acids, organic acids and carbohydrates have often been reported during callus growth and *in vitro* tissue differentiation (Palama *et al.,* 2010; Yang *et al.,* 2009; Blanc *et al.,* 2002). The metabolites obtained from the tissue culture derived samples in this study showed similar findings to those obtained from NMR based analysis of *C. roseus* calli such as alanine, valine, threonine, asparagine, choline, glucose and sucrose (Yang *et al.,* 2009). NMR-based metabolomics profile of embryogenic and non-embryogenic callus of sugarcane showed the presence of similar metabolites such as valine, leucine, alanine, threonine, 4-aminobutyrate, glutamine, succinate, asparagine, choline, glucose and sucrose (Mahmud *et al.,*

2015). Metabolic pathway analysis (MetPA) was used to elucidate the metabolic networks and the significant pathways detected in sugarcane calli and corresponded to the pathways detected in *Narcissus* calli and other *in vitro* tissues, such as citrate cycle, glycolysis, alanine, aspartate and glutamate metabolism and carbohydrate (glucose, sucrose and fructose) metabolism (Mahmud *et al.,* 2015).

4.4.2.1 Metabolites in callus

The amino acids that were observed in relatively higher amounts in calli than field samples were asparagine, threonine and valine. The relative intensity of notable sugars (e.g. glucose, glucose-6-phosphate, ribose, sucrose and xylose) and sugar alcohols (glucitol and galactitol) signals were higher in callus than all other samples. This could relate to the callus metabolism in media supplemented with sucrose based carbon supply.

Other compounds such as acetoacetate concentrations were higher in callus than field samples. Proline, betaine and choline that have been reported as abiotic stress related metabolites in plants (Rivero *et al.,* 2014) also found to be present in higher amounts in callus than field samples. The high choline content was also reported previously for its relation to plant stress adaptation in *Arabidopsis thaliana* (Tasseva *et al.,* 2004). That could relate to the stressed situation such as high external osmolarity, limited water supply during callus growth as well as tissue disorganisation of callus.

High amounts of asparagine, choline, glucose and sucrose were reported previously in both *C. roseus* and sugarcane calli (Yang *et al.,* 2009; Mahmud *et al.,* 2015). Elevated asparagine, glucose and sucrose has also been reported earlier in an NMR-based metabolomic study on protocorm callus cultures of *Vanilla panifolia* (Palama *et al.,* 2010). In addition, Mahmud *et al.,* (2015) also reported the detection of tyrosine in sugarcane calli, which was also observed in *Narcissus* calli (Figure 4.4). Other metabolites involved in phenylalanine and tyrosine metabolism, such as tyramine and 3-chlorotyrosine were also detected in our *Narcissus* callus but the relative concentrations were less than field samples (Figure 4.4). This may suggest that callus could also be metabolically active for alkaloid biosynthesis since trace amount of Gal was observed as well as other alkaloids (i.e. tazettine/ pretazettine) were also detected in calli (Chapter 3).

The different media treatments (T1-T6) used for calli for 7 and 30 days also showed an important effect on the metabolite distribution pattern in calli. Prior to treatment (Cal0D) callus showed the presence of significantly higher amounts of betaine, arginine, carnitine, proline and xylose (Figure 4.13). The most notable effect was medium T6 (supplemented with *trans*-cinnamic acid, one of the precursors of Amaryllidaceae alkaloids), which showed relatively high amounts of aromatics such as melatonin, serotonin and mandelate as well as carboxylic acids such as fumarate and *trans*-aconitate (Table 4.2 and Figure 4.7). All these compounds showed only trace amounts or were completely absent in callus from the other media treatments as well as field samples (Figure 4.7 and 4.13).

4.4.2.2 Metabolites in differentiated tissues

Clear separation of regenerated white shoot from field (bulb and basal plate), direct white shoot and green shoot was observed through PCA scores and loading plots with significant metabolites responsible for their separate cluster formation. Bulb tissues also formed a separate cluster, but it was difficult to observe signal differentiations between basal plate, direct white shoot and green shoot for their overlapped position on PCA score and loading plots (Figure 4.20A and 4.20B).

It was apparent that after callus differentiation into regenerated white shoot, the tissue contained relatively high amounts of many compounds that were also found in field samples such as tyramine, alanine, glutamate and arginine. Moreover, some compounds such as alanine, betaine, glutamate, succinylacetone, tyramine and 4-aminobutyrate were present in relatively higher concentrations in regenerated white shoots than field samples (section 4.3.2.3.3 and Table 4.4). Shoots, grown directly from twin-scales without any callus formation, showed trace to moderate amounts of tartrate, pyruvate, glutamate, pyroglutamate, chlorogenate, trigonelline, beta-alanine and homocitrulline (Figure 4.18, Table 4.4). Calli differentiation in regenerated white shoots showed increased amount of tyramine and succinylacetone, which are also involved in phenylalanine and tyrosine metabolism (Table 4.4).

Therefore, the detection of these metabolites in differentiated tissues could be in accordance with the higher content of galanthamine in differentiated tissues (direct, regenerated and green tissues) than callus that was observed in Chapter 3, Table 3.2.

The reason behind the higher accumulation of different primary and secondary metabolites in organised tissues (bulblets or shoots) could be related to their nature of accumulation in specific or organised tissues (Dias *et al.,* 2016).

4.4.2.3 Metabolites found in media in which the calli were grown

The data for calli showed a distinct separation of calli from MSM1 medium (T10D) prior to treatment from all the other treatments in PCA score plot (Figure 4.15A). The calli subjected to the other media lay close to each other showing an overlapped distribution of corresponding metabolites (Figure 4.15B).

The media extracts were also analysed and a total of 36 metabolites including three unknown compounds were detected. The major metabolite groups were amino acids, sugars, carboxylic acids, and others (Table 4.3). Therefore, fewer metabolites were identified in media than the calli grown on them. This could relate to the slow release of metabolites from the calli grown on media. The Amaryllidaceae alkaloid precursors observed in callus tissues (Figure 4.4) were completely absent from media extracts. This observation is in agreement with the complete absent of galanthamine in media, which was observed in Chapter 3 (section 3.5.2). A study on biochemical relationship between sugarcane callus tissues and their respective nutrient culture media showed similar results in that there were lower metabolites in respective media than the calli grown on them (Mahmud *et al.,* 2014). Their results showed the presence of proline, glutamine, tyrosine and phenylalanine in calli but not in the respective media. In addition, the major metabolites detected in media extracts were sugars (glucose, fructose, sucrose and maltose), which supports our findings of media metabolites.

The pathway analysis of metabolites obtained from media extracts displayed galactose, starch and sucrose metabolism as the most significant pathways (Figure 4.31 and Table 4.8). That indicates that the sugar metabolites could be released either from callus grown on them or the sugar media containing itself.

5 Chapter Five: Differential gene expression analysis (RNA-seq) of *N. pseudonarcissus* cv. Carlton basal plate versus callus

5.1 Introduction

This chapter describes research aim towards identifying the probable genes or transcripts involved in the biosynthesis of secondary metabolites in *Narcissus*. The first approach towards the aim is the relative expression analysis of the putative genes (*PAL, TYDC,* P450s and *OMT*) involved in the alkaloid biosynthesis in different tissue types (bulb, basal plate, *in vitro* callus and small bulblets) using RT-PCR. However, relative expression analysis alone is not enough for confident gene identification or differential expression.

Therefore, a deep transcriptome analysis is worthwhile to identify the target genes and to detect their expression levels in different tissues in *Narcissus*. A suitable protocol and challenges to extract good quality and quantity of RNA from *Narcissus* basal plate and callus tissue to obtain *Narcissus* transcriptome (RNA-seq) data based on next generation sequencing technology (Illumina) have been described here. Furthermore, the suitable bioinformatic platforms and tools are described to process these RNA-seq data to draw a biologically meaningful conclusion i.e. differential gene/ transcripts expression levels in two conditions (field versus *in vitro*). The difference between basal plate and callus transcripts could give insight into the differences in secondary metabolite profiles.

Consequently, this chapter will provide a rational outline of *Narcissus* transcriptome analysis, which is very complicated and challenging due to the lack of a complete annotated reference genome/ transcriptome or any closely related genome, along with the identification of novel gene/ transcript information related to alkaloid biosynthesis. Information about other genes/ transcripts differentially expressed in field and *in vitro* conditions is another focus of this chapter.

5.1.1 Relative gene expression analysis

5.1.1.1 RT-PCR

RT-PCR is a laboratory technique to produce multiple copies of a particular DNA sequence through an amplification process. The difference between RT-PCR and the traditional PCR is that RNA is first transcribed into its reverse DNA complement

(cDNA) in RT-PCR. The cDNA containing the reversed transcription is then amplified using traditional PCR techniques (Tong *et al.,* 2009).

5.1.1.2 Overview of strategies or methods

The accurate quantification of gene expression depends on many experimental variations such as quality and amount of RNA, appropriate primer and primer design, and selection of appropriate normalisation method (Tong *et al.,* 2009).

5.1.1.2.1 RNA quality

Either total RNA or mRNA can be used for RT-PCR based on the purpose of study. Using mRNA suggests better sensitivity. However, this may lose sensitivity due to the loss of starting material during the purification step (Bustin and Nolan, 2013). Therefore, in most cases using total RNA is preferred as the ultimate sensitivity of RT-PCR is determined by other factors rather than the starting materials (Bustin and Nolan, 2013). It should be free from DNA and any other contaminants (Bustin and Nolan, 2013). Several studies on plants show the use of high quality RNA extraction kits combined with DNase treatment to obtain total RNA for RT-PCR (Delporte *et al.,* 2015; Pant *et al.,* 2015).

5.1.1.2.2 Primer choice and design

There are three priming techniques, which are widely used for cDNA synthesis in RT-PCR i.e. using random primers, oligo-dT, or target-specific primers. Random primers are non-specific but yield the most cDNA and are useful for transcripts with secondary structure. Oligo-dT primers are more specific than random primers but require very high quality mRNA and it can only prime the RNA with poly (A) tail (Bustin and Nolan, 2013). Target-specific primers are the most specific and offer the most sensitive method of quantification (Deprez *et al.,* 2002). However, it requires separate reactions for each target (Bustin and Nolan, 2013).

Designing gene specific primers is the most frequent requirement for the detection of target gene expression using RT-PCR. The structure and sequence of the candidate gene is needed before designing a primer so that it will always amplify the intended product with a specific and appropriate length, i.e. exclude the possibility of amplification of non-functional copies of genes (Kozera and Rapacz, 2013).

The web based primer design is often suggested, such as Primer 3 (www.simgene.com/Primer3), one of the most widely used programmes (Untergasser *et al.,* 2012; Delporte *et al.,* 2015).

mRNA has a cap structure at its 5′ end and a poly (A) sequence at its 3′ end. The sequence information between this cap and the poly (A) is important for the identification of the coding region and the non-coding region. Especially the 5′ regions encode structural information that is relevant to mRNA stability and translational efficiency (Suzuki *et al.,* 1997). A sequence can easily be imposed at the 5' end by priming the reverse transcription with a specific primer or with an oligonucleotide tailored with a poly (dT) stretch (Suzuki *et al.,* 1997). The synthetic oligonucleotides can be used to replace the cap structure (5′-oligo) can serve as the sequence tag for the mRNA start site. Another approach could be to force the specific acquisition of full-length 5' cDNAs consists of ligating an anchor primer to the 5' end of the mRNA before performing the reverse transcription step (Scotto-Lavino *et al.,* 2006). This type of 5′-end-enriched cDNA library may be useful for the isolation of the 5′ ends of mRNAs, which would be difficult to isolate from an oligo-dT primed cDNA library (Suzuki *et al.,* 1997).

5.1.1.2.3 Analysis strategies

RT-PCR data can be expressed relative to an internal standard known as relative quantification or an external standard curve, known as absolute quantification (Bustin and Nolan, 2013). Absolute quantification is required in case of quantifying exact transcript copy number, which employs an external calibration curve to obtain the input template copy number (Yuan *et al.,* 2006). However, relative quantification is the most frequently used technique to determine the expression of target genes. It relies on the comparison of expression of a target gene to a reference gene either quantitatively or semi-quantitatively (Yuan *et al.,* 2006; Che *et al.,* 2006). A housekeeping gene is generally used as the reference. Therefore, relative quantification does not require standards with known concentrations and the gene expressions are expressed as a fold change relative to the reference gene and the number of target gene copies are normalised to the reference gene (Bustin and Nolan, 2013).

Though relative quantification is mainly used for the quantification of gene expression, it has some limitations. It often introduces a significant statistical bias that could results in the wrong biological interpretation (Hocquette and Brandstetter, 2002). This is very common when there is a big difference between the expression levels of target and reference genes or if the expression levels of target genes are very low. In addition, in some cases it is very difficult to find or select a reference gene with constant expression against which the numbers of target gene copy can be normalised (Bustin and Nolan, 2013).

5.1.1.2.4 Selection of reference genes

An accurate and appropriate normalisation method to validate the RT-PCR results is a crucial factor to reach its maximum analytical potential (Kozera and Rapacz, 2013). A reference gene should be expressed at a constant level across experimental conditions in respect of developmental stages or tissue types and its expression should not be affected by any experimental parameters (Schmittgen and Zakrajsek, 2000; Tong *et al.,* 2009). In addition, it should have a similar range of expression to the target gene of interest in the samples under study. If the expression of the reference gene fluctuates then the normalisation will lead to incorrect gene expression profiles of the target gene (Dheda *et al.,* 2005; Tong *et al.,* 2009).

Examples of commonly used housekeeping genes are ribosomal RNA (*rRNA*), β-actin (*ACT*), a cytoskeletal protein and glyceraldehyde-3-phosphate dehydrogenase (*GAPDH*), aglycolytic enzyme (Arya *et al.,* 2013; Tong *et al.,* 2009). However, none of these genes shows constant expressions in some experimental conditions (Bas *et al.,* 2004; Dheda *et al.,* 2005). The reason for these expressional differences may be due to the participation of reference genes not only in the basic cell metabolism but also in other cellular process (Tong *et al.,* 2009). Therefore, it has been suggested to use several reference genes in a single experiment to obtain the mean expression of housekeeping genes for data normalisation (Vandesompele *et al.,* 2002).

On the other hand, *in vitro* produced RNA spike-ins is independent of the biological process and can be used as controls for both the RT-PCR and qPCR (Devonshire *et al.,* 2010). The exogenous RNA standards with known sequences allow determining if an RNA-seq assay accurately represents the composition of known input and also

allow the direct measurement of sequencing error rates, coverage biases, and other variables that could affect the downstream analysis (Devonshire *et al.,* 2010; Jiang *et al.,* 2011). The RNA standards are identical across samples (e.g. constant expression, single isoform, not subject to misannotation, sequence polymorphism between the sample and reference genome, or other biological variation) and hence more reliable and easy to use compared to actin or other housekeeping gene transcripts (Jiang *et al.,* 2011).

The External RNA Control Consortium (ERCC) developed a series of spike-in RNA transcripts with known concentrations and sequence identities (Qing *et al.,* 2013). These RNA standards have been produced by *in vitro* transcription and contain synthetic 3' poly-A sequences. This process enables them to be processed in the same way as mRNA transcripts, using oligo (dT)-based priming strategies which are commonly used in microarray and RT-qPCR protocols (Devonshire *et al.,* 2010).

5.1.1.3 RT-PCR in plant secondary metabolism

Several studies showed the use of RT-PCR for relative expression analysis of secondary metabolism related genes in plants. For example, RT-PCR was used for related gene expression analysis for understanding the phenylpropanoid pathway leading to polyphenols in *Cichorium intybus* (Delporte *et al.,* 2015). Twelve target genes were tested for their relative expressions in cell cultures of *C. intybus* under various experimental conditions. Their results showed the higher expression levels of genes involved in plant secondary metabolite production i.e. *PAL* and *C4H* in cell cultures obtained from media supplemented with methyl jasmonate. Target gene expression was normalised against actin (Delporte *et al.,* 2015). Another study showed the combination of RT-PCR and RNA-seq for the detection of expression level of polyphenol oxidase (POP), which catalyses the oxidation of phenolic compounds into highly reactive quinones, in walnut. 18S ribosomal RNA was used as the internal standard in this study. The results suggested that POP is a central enzyme in the metabolism of tyrosine in walnut (Araji *et al.,* 2014). The expression level of genes involved with biosynthesis of primary and secondary metabolites was detected using RT-PCR in *Arabidopsis*, and the results indicated that the levels of transcripts involved in sugars, amino acids, organic acids, phenylpropanoids and flavonoids biosynthesis increased during phosphorus limitation (Pant *et al.,* 2015).

Therefore, in our study RT-PCR was chosen to detect the relative expression of putative transcripts involved in the Amaryllidaceae alkaloid biosynthesis i.e. two *PALs*, two *TYDCs*, two P450s and *OMT* (Wang *et al.,* 2013; Kilgore *et al.,* 2014, Kilgore *et al.,* 2016a; Kilgore and Kutchan., 2016). A reference housekeeping gene, actin was selected to normalise the expression of target gene specific transcripts.

5.1.2 *Narcissus* transcriptome analysis

5.1.2.1 Next Generation Sequencing (NGS)

The high-throughput DNA sequencing technologies which are collectively known as next generation sequencing (NGS) allow the mass sequencing of genomes and transcriptomes producing a vast array of genetic information (M Perez-de-Castro *et al.,* 2012). The era of first generation sequencing was mainly dominated by 'Sanger sequencing' (Sanger and Coulson, 1975) until the early 21st century and provided the first plant genome sequence with the landmark Arabidopsis genome in 2000 (Arabidopsis Genome Initiative, 2000).

Due to the high cost and time to produce the data, failure to sequence large and complex genomes of important crops, such as wheat and plants with specialised metabolites of medicinal interest (Xiao *et al.,* 2013); Sanger sequencing was superseded with the introduction of second generation sequencing technologies such as the Roche/ 454 platform in 2005, Solexa/ Illumina system in 2006, Applied Biosystems SOLID system in 2007 and Ion Torrent system in 2010 (Ozsolak, 2012). The sequence reads obtained from second generation sequencing technologies are often short (35-500 bp) which allows high quality deep coverage, compared with traditional Sanger sequencing (>700 bp) (Metzker, 2010) and can produce more data at lower cost and less time required for larger genomes or non-model organisms (Xiao *et al.,* 2013).

Among the second-generation technologies Illumina HiSeq is currently the most widely used platform as it is less expensive and provides more reads (up to 5 billion reads per run) (Illumina, Inc. 2016, www.illumina.com) compared to the Roche 454 (~1 million reads per run) (Egan *et al.,* 2012). However, a perceived limitation of the ultra-high-throughput technologies such as Illumina is their short read-lengths that can make assembly more difficult (Paszkiewicz and Studholme, 2010).

5.1.2.2 Illumina sequencing

NGS methods can be grouped into three types: sequencing by synthesis (Sanger, Roche 454 Pyrosequencing, Ion Torrent, Illumina), sequencing by ligation (SOLID, Polonator), and single-molecule sequencing (Helicos, PacBio) (Egan *et al.,* 2012; Hodkinson and Grice, 2015; Maheswari and Ravi, 2016). Illumina sequencing technology works on the principal of sequencing by synthesis initially developed by Solexa; unlike Sanger sequencing it uses nucleotides labelled with both fluorescent dye and a terminating moiety, which allow the use of four nucleotides at a time (Illumina, 2010).

Illumina is based on the application of solid-phase bridge amplification in which 5′ and 3′ adapters are ligated to both ends of genomic DNA. The adapters hybridize to immobilize forward or reverse primers to generate a bridge that facilitates amplification. Thus, several million identical double stranded DNA clusters are produced in each channel of the flow cell (Egan *et al.,* 2012) (Figure 5).

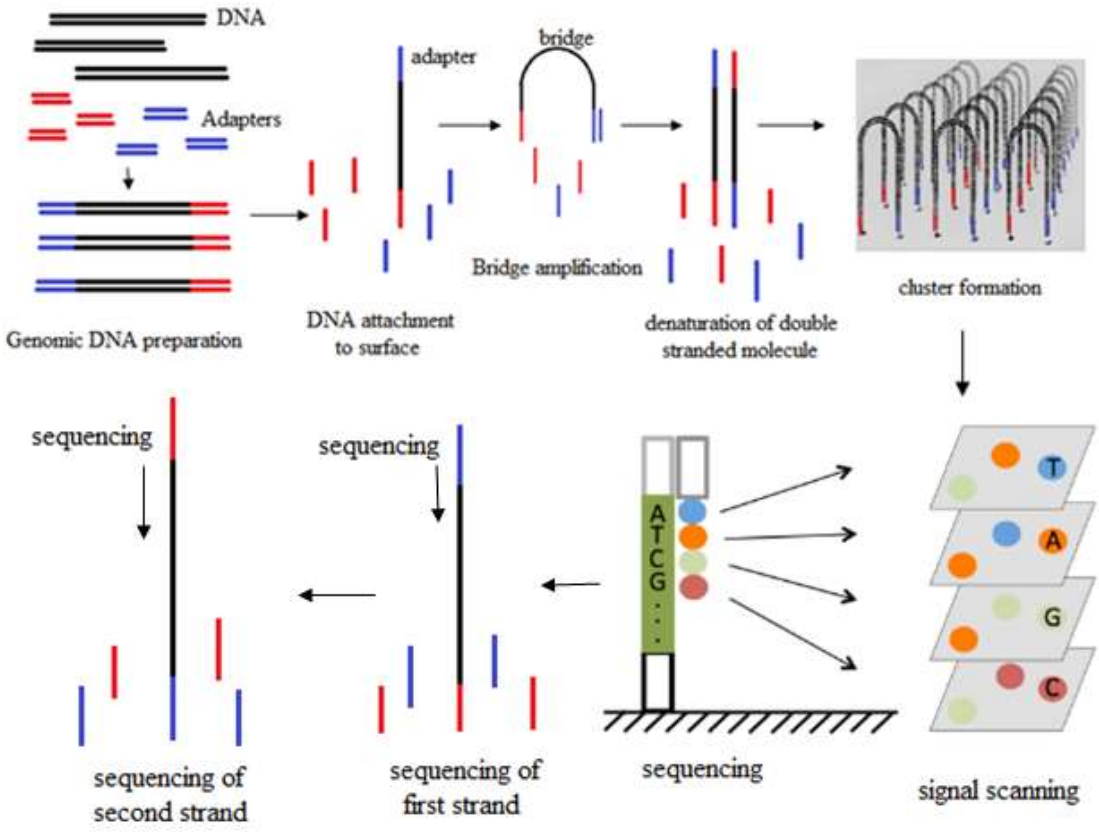

Fig. 5.1 Illumina sequencing workflow (paired-end sequencing) (redrawn from www.illumina.com and www.3402bioinformaticsgroup.com).

After amplification, the DNA amplicons are denatured and primed. The first sequence cycle starts with the addition of a mixture of all four nucleotides, each labelled with a different fluorophore and modified as 3'-O-azidomethyl reversible terminators (Guo *et al.,* 2008) together with primers and DNA polymerases (Illumina, 2010). Elongation is conducted through a series of cyclical washes and continues after the fluorescent dye moiety is cleaved and the 3'-OH group is restored through reaction with tris (2-carboxyethyl) phosphine (Egan *et al.,* 2012). Following laser excitation, the emitted fluorescence from each cluster is captured to identify the first base, second base and so on. This cycle is repeated until the DNA fragment has been synthesized to its target length (Illumina, 2010). The paired end sequencing allows sequencing both ends of a fragment and generate high quality alignable sequence data (Figure 5). The Illumina HiSeq 2500 system on a dual flow cell can produce up to 8 billion paired-end reads in ~5 days run time (www.illumina.com). The large number of reads provides accurate measurement of gene expression levels with very deep coverage to reference genomes (Ekblom and Galindo, 2011).

5.1.2.3 Transcriptome analysis (RNA-seq)

In the beginning of transcriptome analysis, microarray technologies based on nucleic acid hybridization were the only tools available for mass gene expression analysis and for examining transcriptomic features (Wolf, 2013). Prior to microarrays, gene expression researches were limited to small-scale quantitative PCR or northern blot analyses of candidate genes (Naurin *et al.,* 2008; Hirsch *et al.,* 2015). Those technologies were also restricted to previously identified genes and showed inadequacy to determine various kinds of RNA molecules such as previously unknown coding and non-coding RNA, particularly small RNAs (sRNA), including micro RNAs (miRNA), promoter associated RNAs or splice variants that are expressed in specific biological conditions at different time points (Ozsolak and Milos, 2011).

The recent development of next generation sequencing and the maturation of analytical tools have introduced a new era of transcriptome analysis (Trapnell *et al.,* 2012). RNA-seq, often called whole-transcriptome shotgun sequencing, is based on the use of high-throughput sequencing technologies for the characterisation of the RNA content and composition in a biological system. The sequence information

obtained from RNA sequencing is randomly decomposed into short reads of up to several hundred base pairs (Wolf, 2013).

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). This method has been extensively used for non-model organisms 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).

5.1.2.3.1 Advantages and applications

RNA-seq offers a more comprehensive understanding of transcriptome complexity compared to other hybridization-based technologies such as microarrays or Sanger sequencing (Kumar *et al.,* 2012). It allows the investigation of a dynamic range of expression levels (Wang *et al.,* 2009) including the gene expression patterns between populations in the context of speciation (Wolf *et al.,* 2010) or eco-type specific adaptation (Lenz *et al.,* 2013). It can detect the expressed sequences in specific tissues at a specific time and is particularly applicable in non-model organisms, as it does not require a reference genome to gain specific and meaningful transcriptomic information (Strickler *et al.,* 2012; Egan *et al.,* 2012).

In comparison with microarrays, RNA-seq exhibits great depth of sequencing to determine rare transcripts (Strickler *et al.,* 2012) and also it provides a digital measure, which is linearly scaled even at extreme values, whereas microarrays are based on analog-type fluorescent signals (Marioni *et al.,* 2008). It further provides information on gene characterisation (Bräutigam *et al.,* 2011), single-nucleotide polymorphisms (SNPs) (Novaes *et al.,* 2008), and reveals information on novel transcripts (Strickler *et al.,* 2012).

5.1.2.3.2 Limitations

Accurate gene expression analysis using RNA-seq is often challenging for several issues such as the counting of small transcripts become difficult due to the standard size selection during RNA-seq library construction (Hirsch *et al.,* 2015). In some

cases, two different genes have overlapping transcripts. Hence, the transcripts from overlapping genes encoded on different strands are difficult to distinguish (Wolf, 2013). Another issue is the presence of multiple related sequences within the genome, which can affect accurate assigned expression to the correct gene (Hirsch *et al.,* 2015). Measurement of steady-state mRNA levels ignores mRNA stability, or turnover rates, which eventually determine protein abundance (Wolf, 2013).

However, direct information on the originating strand can substantially improve the value of an RNA-Seq experiment. Such information would help to accurately identify antisense transcripts, determine the transcribed strand of other non-coding RNAs, and resolve the correct expression levels of coding or non-coding overlapping transcripts (Levin *et al.,* 2010; Wolf, 2013). The methods developed for strand specific RNA sequencing can be grouped into two classes. The first class relies on the attachment of different adaptors in a known orientation relative to the 5' and 3' ends of the RNA transcripts i.e. RNA ligation, Illumina RNA ligation, SMART (Switching Mechanism at 5' end of RNA Template) and NNSR (Not Not So Random Priming). The other relies on marking one strand by chemical modification such as Bisulfite and dUTP 2nd strand (Levin *et al.,* 2010).

5.1.2.4 Strategies for RNA-sequencing in non-model plants

RNA sequencing is the platform, which makes transcriptomic experiments possible for non-model plants, and other organisms where genomic resources are limited or completely unavailable (Xiao *et al.,* 2013). 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).

5.1.2.4.1 Sample selection

The samples should be selected from appropriate plant tissues and life stages to address the experimental goal. Shock freezing of samples immediately after harvesting has been used as the most reliable method to prevent loss of RNA due to RNase activity (Wolf, 2013). Sampling should include a good number of biological and technical replicates. Biological replicates are important to determine the sample variation and especially if the research goal is to discover differential expression between two conditions (Auer and Doerge, 2010) while technical replicates are used

to reduce the sequencing variations due to instrumental errors, batch or lane effects (Strickler *et al.,* 2012).

In the current study, samples were collected from two different growth conditions (field and *in vitro*) for RNA extraction and biological replicates (3×) (basal plate from three different bulbs, callus from *in vitro* cultures of three different bulbs) were included. Technical replicates for basal plate and callus were not possible to perform because of cost constraints. Therefore, only six samples (three for basal plate and three for callus) were sequenced.

5.1.2.4.2 RNA extraction

Extraction of RNA is the initial and most challenging step to start transcriptome analysis for any organism. Due to the presence and accumulation of secondary metabolites and other compounds, RNA extraction can be difficult in plants (Jaakola *et al.,* 2001). Especially in plants with high levels of phenolic compounds, polysaccharides and lipids, the extraction method has proven to be more problematic (Dang and Chen, 2013).

The CTAB (hexadecyltrimethylammonium bromide) method (Chang *et al.,* 1993) is one of the most widely used RNA extraction methods for plants. The oxidation of phenolic compounds that could bind irreversibly to nucleic acids and cause RNA degradation is a common problem in many plants (Jaakola *et al.,* 2001). The modified CTAB method using polyvinylpyrrolidone (PVP) and β-mercaptoethanol has been used to eliminate the polysaccharides and prevent the oxidation of phenolic compounds (Jaakola *et al.,* 2001) and is a suitable protocol for many plants (Zeng and Yang, 2002; Gambino *et al.,* 2008; Dang and Chen, 2013). DNA contamination from microorganisms is a frequently observed problem. Hence, a high-quality DNase treatment is highly recommended (Wolf, 2013).

A number of kits have also been introduced for the isolation of RNA from plants; i.e. Trizol reagent (Invitrogen) (MacRae, 2007), Qiagen RNeasy Plant Minikit, Qiagen hybrid method (MacKenzie *et al.,* 1997), pBIOZOL Method (Beijing Genomics Institute) and innuPREP Plant RNA Kit (Johnson *et al.,* 2012). Among them, Qiagen RNeasy Plant Minikit (Kurbidaeva and Novokreshchenova, 2011) and innuPREP

Plant RNA Kit (Park *et al.,* 2014; Dobnik *et al.,* 2013) are the most successful kits used for the isolation of RNA from tough plant tissues.

The extraction of total RNA in very short time (~30min) after homogenisation of tissues is possible by using the innuPREP Plant RNA kit. It requires a small amount of starting materials (~100mg) and includes two lysis buffers (PL and RL), which allow the kit to be adapted to a variety of plant species and tissues. The lysis buffer 'PL' is specially optimized for plants with tough tissues, containing high phenols, polysaccharides, lipids and other secondary metabolites. Moreover, the optimized extraction chemistry does not require the use of toxic β-mercaptoethanol, which is widely used for traditional CTAB method (www.analytik-jena.de). Several kits were tested on *Narcissus* in a previous study with the innuPREP Plant RNA kit working the best (Pulman, 2015). Therefore, innuPREP Plant RNA kit was selected to extract RNA from *N. pseudonarcissus* basal plate and calli samples.

5.1.2.4.3 RNA quality

The assessment of RNA integrity or quality is the most critical step in the preparation of a transcriptomic sequencing library to obtain meaningful gene expression measurements (Wolf, 2013). The most widely used techniques for initial RNA quality check are NanoDrop ND-1000 UV-Vis spectrophotometer (NanoDrop Technologies) and an Agilent-2100 Bioanalyser using RNA 6000 NanoChips (Agilent Technologies) (Czechowski *et al.,* 2005). The ratio of sample absorbance at 260 and 280 nm is used to assess the RNA purity in NanoDrop method. Generally, a ratio (260/280) of ~2.0 is accepted as pure RNA. If the ratio is lower than ~2.0 it indicates contamination with protein, phenols or others that absorb at or near 280 (www.nanodrop.com).

In the Agilent Bioanalyser method, the integrity of RNA is detected by the presence of bright 18S and 28S or 26S rRNA bands by analysing the samples on denaturing formaldehyde gels. It also compares a number of RNA integrity metrics; the ratio of the large (28S or 26S) to small (18S) ribosomal RNA (28S or 26S/18S) and provides a RNA integrity number (RIN). A RIN value greater than 8.0 is usually consider high quality RNA, but greater than 7.0 is also accepted as good plant RNA samples for

sequencing (Johnson *et al.,* 2012; Ward *et al.,* 2012). Both methods were implemented to check the initial RNA quality for all samples under this study.

5.1.2.4.4 rRNA depletion

Ribosomal RNA (rRNA) constitutes the major fraction of the transcriptome and hence needs to be removed prior to library preparation (Wolf, 2013). Commercially available kits are used for poly (A) capture or any hybridisation-based method to remove rRNA to obtain a sample highly enriched with mRNA (Ward *et al.,* 2012). Illumina (www.illumina.com) introduced the Ribo-Zero (plant leaf) kit that is compatible with plant leaf samples and the Ribo-Zero (plant seed/root) kit compatible with plant seed or root tissues for use with their sequencing technologies. Ribo-Zero kits utilize a hybridisation capture technique that allows removal of > 99% of cytoplasmic rRNA from 1 µg to 5 µg of intact, partially degraded total RNA samples (Pease and Sooknanan, 2012). It accomplishes the removal of mitochondrial, cytoplasmic and chloroplast rRNA in a single pass to enrich in coding and non-coding RNA species. The plant seed/ root kit contains a higher percentage of cytoplasmic rRNA removal probes compared with the plant leaf kit (www.illumina.com/products/ribo-zero-rrna-removal-plant).

We used underground bulbs of *Narcissus* for *in vitro* callus induction and basal plate samples were also harvested from the base of bulbs. Consequently, the Ribo-Zero (plant seed/ root) kit was selected for the rRNA removal from all samples.

5.1.2.4.5 cDNA synthesis and library preparation

Most of the sequencing platforms require the conversion of RNA to cDNA using an enzymatic reaction for reverse transcription and preparation of a suitable length sequence library (Wolf, 2013). The synthesis is performed using commercially available kits. The most successfully used kits introduced by Illumina are the TruSeq Stranded Total RNA kit (Gutiérrez *et al.,* 2015) and ScriptSeq v2 RNAseq Library Preparation kit. The kit protocol utilizes direct fragmentation of RNA followed by reverse transcription using random hexamer primers for library preparation (www.illumina.com).

The sequencing can be performed by partially sequencing from one end, which is known as single-end strategy or paired-end reads generated by sequencing of short reads from both ends (www.illumina.com). Paired-end sequencing is more helpful for transcriptome assembly, isoform or transcript detection but the insert size should be short (usually < 300 bp) (Wolf, 2013). The ScriptSeq v2 Kit requires about 500 pg to 50 ng of rRNA-depleted sample for the standard kit reaction. It uses a random-primed cDNA synthesis reaction containing a 5' tagging sequence. The 5′ tagged cDNA is tagged at its 3′ end by the terminal tagging reaction, which generates di-tagged, single stranded cDNA. After that the di-tagged cDNA is amplified using PCR. Generally, 10-15 PCR cycles are adequate to produce the sequencing library. The lower number of PCR cycles (10-12) was used to avoid replication of PCR errors which might affect bioinformatics analysis. After purification, the library is ready for cluster generation and sequencing (Pease and Sooknanan, 2012).

The ScriptSeq v2 Kit was used (CGR) for library preparation in this study, as it requires small amount of rRNA depleted sample.

5.1.2.4.6 Sequencing platform

The sequencing technology has an effect on the success of an RNA-seq experiment. Three frequently used sequencing platforms (Roche/ 454 pyrosequencing, Solexa/ Illumina and Ion Torrent) can produce a remarkable amount of data but have particular applications (Ward *et al.,* 2012). The important parameters that should be considered for selecting sequencing platform are price per base pair, error rate and error profiles, total output and read length (Wolf, 2013).

Among the three platforms, Roche/ 454 and Solexa/ Illumina have been reported as successful for transcriptomic studies in non-model plants (Strickler *et al.,* 2012). Both platforms are based on paired-end sequencing, but the orientation of paired reads is different i.e. the orientation of paired-end reads for Roche/ 454 are in the 'forward-forward' orientation, while the Solexa/ Illumina paired-end reads are in the 'forward-reverse' orientation (Strickler *et al.,* 2012). Earlier Illumina sequencing provided shorter reads than 454 pyrosequencing, which caused difficulty for *de novo* assembly (Pop and Salzberg, 2008). Therefore, it was reported in a small number of non-model plant transcriptome studies (Collins *et al.,* 2008; Trick *et al.,* 2009;

Mizrachi *et al.,* 2010), while Roche/ 454 platform was used for several non-model plants (Novaes *et al.,* 2008; Alagna *et al.,* 2009; Barakat *et al.,* 2009; Wang *et al.,* 2009; Guo *et al.,* 2010; Franssen *et al.,* 2011).

However, due to recent improvements in sequencing technology increasing read length, Illumina has become a successful technology for transcriptome assembly as well as detecting novel transcripts in non-model plants due to its deep coverage (Wang *et al.,* 2013; Kilgore *et al.,* 2014). A combined study using both technologies (454 and Illumina) was also successful for gene expression analysis in *Arabidopsis thaliana* and *de novo* assembly in *Eschscholzia californica* and *Persea americana* (Wall *et al.,* 2009).

It has been reported that ~100 million paired-end reads (> 100 bp) are sufficient to capture most of the genes of an RNA sample across a broad range of expression levels and is suitable for many projects (Vijay *et al.,* 2013). The Illumina sequencing platform e.g. HiSeq 2500, generally generates > 500 million reads per lane with a total output of up to 1 terabase (Tb) of data (www.illumina.com). Even though it generates shorter reads than other sequencing platforms (454 pyrosequencing or traditional Sanger or Ion Torrent), due to the large number of reads (pair-end) it allows far greater coverage (Collins *et al.,* 2008). Moreover, the library generated for Illumina does not require any data normalisation like other platforms e.g. 454 pyrosequencing (Collins *et al.,* 2008). Consequently, Illumina HiSeq 2500 sequencing platform was chosen for transcriptome sequencing from *N. pseudonarcissus* cv. Carlton RNA samples.

5.1.2.5 Transcriptome (RNA-seq) data processing and analyses

5.1.2.5.1 Raw data/ sequence processing

The raw data obtained from sequencing should be pre-processed before any downstream analyses such as assembly or mapping (Wolf, 2013), because the raw reads usually come with errors and contain adapters (Strickler *et al.,* 2012). A free command line programme (Cutadapt) for adapter removal from high throughput raw reads has been introduced by Martin, (2011). Other methods of adapter removal are also available such as HTSeq (Anders *et al.,* 2014), Biostrings (Pages *et al.,* 2009), SOAP (Li *et al.,* 2008), MAQ (Li *et al.,* 2008) and Novoalign (Novocraft, 2010;

www.novocraft.com); but they have their specific limitations (Martin, 2011). The Cutadapt programme has been tested and found to work well on small RNA (Illumina) data (Schulte *et al.,* 2010). It supports a variety of file formats generated by next generation sequencing. Besides, multiple adapters are searched in a single run and the best-matched ones can be removed (Martin, 2011). The Cutadapt programme has been recently reported for adapter removal in several plant transcriptomic and genomic studies (Beletsky *et al.,* 2016; Liu *et al.,* 2016; Prabhudas *et al.,* 2015).

The further trimming of sequence reads is also mandatory after the adapter removal prior to any downstream analyses. The deteriorating quality towards both 3' and 5' end of raw reads can negatively affect sequence assembly or mapping. Sickle is a windowed adapted raw read-trimming tool with quality and length thresholds. It dictates when the quality is sufficiently low to trim the 3'-end and determines when the quality is sufficiently high enough to trim the 5'-end of raw reads (Joshi and Fass, 2011). This read trimming tool has been successfully used to trim raw reads before further bioinformatic analyses, reported in plant transcriptomic (Naoumkina *et al.,* 2015) and genomic (Prabhudas *et al.,* 2015) studies.

The raw reads of *Narcissus* in this project have been pre-processed using 'Cutadapt version 1.2.1' and 'Sickle version 1.200' with a minimum window quality score of 20.

5.1.2.5.2 Transcriptome analysis approaches

Next generation sequencing offers two approaches to transcriptome analysis. One is the *de novo* which involves assembly then alignment techniques, in the case of limited or absence of genomic or transcriptomic resources. The other is mapping based on aligning then assembling techniques that are completely dependent upon the existence of a reference genome or transcriptome (Haas and Zody, 2010). Both approaches can be used for the transcriptome analysis in non-model plants (Ward *et al.,* 2012).

5.1.2.5.2.1 *De novo* transcriptome assembly

The *de novo* assembly is often used in non-model systems where reads are assembled into contigs without the guidance of a reference sequence (Kilgore *et al.,* 2014). The

longer reads obtained from 454 pyrosequencing might represent a full-length transcript in a *de novo* assembly. It is often difficult to assemble alternatively spliced variants of the same genes accurately (Strickler *et al.,* 2012). However, the identification and validation of alternative splicing is possible by using sequence alignment to a close relative (if available) to validate the assembly results. That also helps to identify the exon-exon boundaries in transcripts, which can be used for the development of primers for RT-PCR to identify splice variants with specific expression in different tissues or treatments (Ward *et al.,* 2012).

Some studies showed high quality assemblies in non-model plants (Wang *et al.,* 2009; Yang *et al.,* 2010; Der *et al.,* 2011) but the results could not detect gene expression levels due to data normalisation and less sequence depth obtained from 454 sequencing. Recent studies based on Illumina sequencing utilizing non-normalised libraries were successful to identify gene expression (Mizrachi *et al.,* 2010; Ness *et al.,* 2011; Kilgore *et al.,* 2014).

5.1.2.5.2.2 Mapping assembly

The sequence reads can also be mapped using either a reference sequence genome or transcriptome or coding sequences as a template (Strickler *et al.,* 2012; Ghosh and Chan, 2016). Transcriptome alignment to closely related sequences allows direct gene expression comparisons between tissues, treatments and controls and also between replicates. This type of analysis is very useful to identify tissue specific signalling or metabolic pathways (Ward *et al.,* 2012).

The most commonly observed errors in mapping assemblies is detecting an expressed gene, which is not actually expressed, and not detecting one, which is truly expressed. The reason behind this could be either misalignment to divergent sequences or lack of homologous genes in the reference genome or transcriptome (Ward *et al.,* 2012). The first mapping (alignment first) study on a non-model plant was performed by aligning Illumina libraries obtained from the allopolyploid *Pachycladon enysii* to *Arabidopsis thaliana* genome (Collins *et al.,* 2008). This study showed the successful identification of ancestral genes belonging to *P. enysii* in combination with *de novo* assembly. Another study on a non-model plant (*Rubus idaeus* L.) included the detection of gene expression and differential gene expression

levels based on a normalised statistic named FPKM (fragments mapped per kilobase of exon per million reads mapped) (Ward *et al.,* 2012). In this study, the transcriptome data of red raspberry *R. idaeus* L. (red raspberry) was mapped to genome of the closely related *Fragaria vesca* L. (woodland or alpine strawberry).

A good number of free tools are available and frequently used for transcriptome analysis such as Trans-ABySS (Robertson *et al.,* 2010), Trinity (Grabherr *et al.,* 2011) and SOAPdenovo (Xie *et al.,* 2014) for transcriptome assembly. For mapping, the most recurrently used tools are Bowtie (Langmead *et al.,* 2009), BWA (Burrows-Wheeler transform) (Li and Durbin, 2010), TopHat (Trapnell *et al.,* 2009) and Cufflinks (Trapnell *et al.,* 2010).

In this study the Illumina RNA-seq data obtained from *N. pseudonarcissus* cv. Carlton field and callus samples were analysed by mapping to previously constructed transcriptome data of *Narcissus spp*. (Kilgore *et al.,* 2014) available on MedPlant RNA Seq Database (www.medplantrnaseq.org).

5.1.2.5.3 Differential gene expression

Gene expression analysis is one of the most important applications of RNA-seq. It can be applied for the quantification of transcript levels in different tissues and samples across treatment groups (Wolf, 2013). The expression levels can be quantified by mapping to the reference sequence, if one is available (Trapnell *et al.,* 2012), if not the reads need to be assembled first to reconstruct the transcriptome then the assembled reads can be mapped to a reference genome for the identification of transcript levels (Wang *et al.,* 2013). The read counts should be normalised for the detection of differential expression levels, which is accounted for in varying sequence depths between lanes and flow cell (Strickler *et al.,* 2012). A recent approach is to calculate FPKM, which measures the transcript abundances as normalised expected fragments (Trapnell *et al.,* 2010). Transcript abundance and differential expression studies have been reported in several non-model plants (Guo *et al.,* 2010, Swarbreck *et al.,* 2011; Li *et al.,* 2016; Bedre *et al.,* 2016).

5.1.2.5.4 Gene annotation

Information about the genes (annotation/ name assignment) is the most important step to end up with a biologically meaningful RNA-seq experiment and is a special concern in non-model plants (Vijay *et al.,* 2013). In case of *de novo* assemblies, gene annotation is mandatory as the contigs do not provide any information about sequenced genes and their assignment to orthologous genes from distantly related genomes is complicated (Wolf, 2013). Gene information can be obtained only in mapping assembly approaches if an annotated reference genome or transcriptome is available. Hence, gene annotation is also mandatory in case of the presence of a reference genome or transcriptome without any gene information (Wolf, 2013; Vijay *et al.,* 2013).

In most cases, gene annotation is performed by using BLAST to find significant matches to annotated genes (Strickler *et al.,* 2012). BLAST-based gene annotation is appropriate for more distantly related species, while suffix-tree-based methods i.e. *NUCmer* and *PROmer* work well for closely related species (Vijay *et al.,* 2013). In plants, annotation by sequence similarity matches (i.e. BLAST) is sometimes challenging due to the limited number of manually annotated genes as manual annotations are still mainly associated with *Arabidopsis* (Strickler *et al.,* 2012). Combined annotation from different data sets (UniProt, SwissProt, Refseq, TAIR, Interpro and Rfam) (Pulman, 2015) or one approach that integrates different annotation approaches such as Blast2Go (Götz *et al.,* 2008) can be used to overcome this limitation (Strickler *et al.,* 2012). Annotation based on predicted amino acid sequences to allow for divergence is also a useful tool when closely related reference species are absent (Surget-Groba and Montoya-Burgos, 2010).

The BLAST alignment tool was used for the alignment of the *Narcissus* transcripts to sequences from publically available databases (Pulman, 2015) for the transcript annotations.

5.1.2.5.5 Gene function and interaction

Obtaining a set of candidate genes is the ultimate yield of a successful RNA-seq experiment, but not the only interest. The external information on gene functions (information on the potential role they play in the organisms) and interactions (genes

overrepresented in specific metabolic pathways) are also of value to draw a successful biological conclusion to any RNA-seq experiment (Wolf, 2013).

Gene ontology database (www.geneontology.org) is widely used in whole genome or transcriptome annotation projects for comparing gene functions across species using a controlled vocabulary (Gene Ontology Consortium, 2004). The ontology system is split into three functional groups. The 'biological process' represents the biological objectives of the gene accomplished by one or more molecular processes and includes cell growth and maintenance factors. The biochemical activity of genes and gene products are mainly described in the 'molecular process' but their activity in the cell in terms of exactly where and when this occurs, are not included. Finally, the 'cellular component' relates to the location of the gene product activities (Ashburner *et al.,* 2000). Information on gene functions assigned in the GO database is mainly based on inbred strains of model organisms (Wolf, 2013), hence its use is more challenging for non-model plants or organisms as the risk of false positives increases (Rhee *et al.,* 2008; Hill *et al.,* 2010). Therefore, it can be more worthwhile to map genes of interest to candidate metabolic pathways (Wolf, 2013).

The Kyoto Encyclopedia of Genes and Genomes (KEGG) (www.genome.jp/kegg) (Ogata *et al.,*1999) is one of the largest and most comprehensive databases linking metabolism to identified genes and is often used in investigation into biosynthetic pathways (Booth *et al.,* 2013). It includes two main databases: a catalogue of genes and annotations accompany the 'gene database' from complete and partial genomes, while the 'pathway database' includes the graphical representations of cellular processes representing higher order functional information (Ogata *et al.,* 1999). KEGG contains a generalised set of biosynthetic pathways conserved in plants, microbes and animals.

A number of studies on plant secondary metabolism have been reported in non-model plants using the KEGG database to elucidate the gene mapped pathways (Xiao *et al.,* 2013; Sangwan *et al.,* 2013; Guo *et al.,* 2010). Xiao *et al.,* (2013) showed the mapping of putative transcript data obtained from 20 non-model plant species to KEGG metabolic pathways based on EC numbers and the biosynthetic genes represented six specialised metabolic pathways i.e. sesquiterpenes, diterpenes, monoterpenoid indole alkaloids, benzylisoquinoline alkaloids, triterpenes and

polyketides. In another study on a non-model plant, biochemical analysis of *Centella asiatica* leaf transcriptome data was performed using KEGG. All putative transcripts were subjected to a KEGG database query with the BLAST score to retrieve KEGG enzyme codes and pathway maps. Secondary metabolite genes based on KEGG biochemical analysis were assigned to important pathways related to secondary metabolism such as mevalonate pathway, sesquiterpenoid biosynthesis and phenylpropanoid biosynthesis (Sangwan *et al.,* 2013).

5.1.2.6 Statistical and computational platforms and software

Statistical analysis is an indispensable component for RNA-seq data, but there is no standard method available to identify and analyse differentially expressed genes obtained from non-model plants (Egan *et al.,* 2012). Analytical and computational programmes and software for NGS data analyses are developing and most are under evaluation. Several methods for the detection of differentially expressed genes have been reviewed (Kvam *et al.,* 2012). An overview of platforms and software used for the analysis of the *Narcissus* transcriptome data of this study are described below.

5.1.2.6.1 CyVerse (iPlant Collaborative)

CyVerse, previously known as 'iPlant Collaborative', is a cyber infrastructure in support of plant biology research (Goff *et al.,* 2011). This is a United States National Science Foundation funded project that aims to provide an array of computational tools, services and platforms for storing, sharing, and analysing large, complex and diverse biological datasets without the need to master the command-line on a UNIX or Linux system or high-performance computational services (Devisetty *et al.,* 2016). A wide range of learning resources and workshops are also available on CyVerse website to train the scientists with diverse background (www.cyverse.org).

The Discovery Environment (DE) portal in CyVerse offers a web interface for running powerful computing, data exploration, and analysis applications. High-throughput analyses can be performed in DE; the data can be shared within the iPlant community and other scientific data already stored on DE databases are available for the users (Devisetty *et al.,* 2016). DE includes evolution analysis on phylogenetic trees, ultra high-throughput DNA and RNA sequence analysis (pre-processing, assembling, variant detection, transcript abundance), analysis of gene duplication

patterns as compared to species trees (tree reconciliation), and taxonomic name resolution (Goff *et al.,* 2011). It contains many bioinformatic applications for biological data analyses, which are designed to be easy to use (Oliver *et al.,* 2013). Sequence analysis on DE enables users to upload DNA or RNA sequencing data from their desktop, a remote server, or from the NCBI Sequence Read Archive.

Atmosphere is another important CyVerse web portal, which is a virtual machine for the integration of all CyVerse approaches in a cloud-style service (Goff *et al.,* 2011) and is a completely isolated operating system (Smith and Nair, 2005). Atmosphere has a broad array of applications that range from image processing to next generation sequence analysis (Skidmore *et al.,* 2011). The CyVerse website is under continuous monitoring, evaluation and upgrading to make the most use of it by plant scientists as well as scientists with other expertise.

A series of applications are preinstalled in DE. Several of these were used to analyse the *Narcissus* RNA-seq data, in addition to CummeRbund, which is a cloud application on Atmosphere. They are described below.

5.1.2.6.2 TopHat and Bowtie (read alignment)

TopHat is an efficient and fast splice junction mapper designed to align reads from an RNA-seq experiment to a reference genome or transcriptome without relying on known splice sites for RNA-Seq reads (Trapnell *et al.,* 2009). It aligns RNA-seq reads to genome (between 100 bp and several hundred kilobases) using Bowtie (Langmead *et al.,* 2012), and breaks up reads into smaller pieces called 'segments' that Bowtie cannot align on its own (Trapnell *et al.,* 2012). These segments then align to the reference genome and the mapping results can be analysed to identify splice junctions between exons (Oliver *et al.,* 2013).

TopHat can use two steps to find junctions when mapping reads to the reference. Firstly, all reads are mapped to the reference using Bowtie and the unmapped reads are set aside as 'initially unmapped reads' (Trapnell *et al.,* 2012). TopHat allows Bowtie to report more than one alignment for a read, which reports the multi-reads from genes with multiple copies, but excludes alignments to low-complexity sequence, to which failed reads often align. The low complexity reads are not included for further analysis and they are simply leftover. Genes transcribed at low

levels will usually be sequenced at low coverage and the exons of these genes may contain gaps. TopHat has a parameter to control two individual but nearby exons and allow them to be merged into a single exon (Trapnell *et al.,* 2009).

The second step includes the read mapping to spliced junction. Each probable intron is checked against the 'initial unmapped reads' that are detected across the splice junction. TopHat only investigates potential introns in the range of 70 bp to 20 000 bp, as this range describes the vast majority of known eukaryotic introns (Trapnell *et al.,* 2009) and it provides best performance with 75 bp reads or longer (up to 1024 bp) from single or paired end fragments (Ghosh and Chan, 2016). It can build up an index of splice sites without splice site annotations by processing each 'initially unmapped reads' (Trapnell *et al.,* 2012).

One recent study showed the successful read alignment using both Bowtie and TopHat in rice, maize, tomato, potato, soybean and Arabidopsis. It was observed that both mapping techniques could be used for alignment to the full genome, full genome with genome annotation guidance or to the transcriptome sequence only. All these three approaches were tested for *Arabidopsis* and maize and mapping to the full genome or transcriptome without an annotation resulted in the highest number of un-mapped reads (Hirsch *et al.,* 2015).

5.1.2.6.3 Cufflinks (transcript assembly)

The accurate quantification of gene expression levels from RNA-seq reads requires accurate identification of all splice variants (isoforms) of a given gene (Trapnell *et al.,* 2012). A sample may contain reads from multiple splice variants and partial or incorrect transcriptome annotation often results in incorrect quantification of gene and transcript expression levels (Trapnell *et al.,* 2010). Cufflinks is capable to gather all the splicing structures of each gene and assembles individual transcripts from RNA-seq reads that have been aligned to the genome. A few full-length transcript fragments, usually called 'transfrags', are also detected using Cufflinks, which are important to explain all the splicing event outcomes in the input data (Trapnell *et al.,* 2012). RNA-seq alignments are assembled into a parsimonious set of transcripts, then the relative abundance of these transcripts are estimated based on how many reads support each one (Oliver *et al.,* 2013). After that, the expression level of each

transfrag in the sample is quantified and Cufflinks can use its abundance estimates to automatically eliminate them. Cufflinks also offers the quantification of transcript abundances by using a reference annotation rather than assembling the reads (Trapnell *et al.,* 2012). Cufflinks estimates transcript abundances from read counts using the FPKM, which is directly proportional to transcript abundance (Ghosh and Chan, 2016).

5.1.2.6.4 Cuffmerge (merging assembly)

The assembled data obtained from Cufflink analysis should be pooled into a comprehensive set of transcripts before proceeding to differential analysis. It is more appropriate to assemble the samples separately and then merge the resulting assemblies together. Cuffmerge is a tool for merging read assemblies, which is a 'meta-assembler'. It merges the assembled transfrags parsimoniously; the way Cufflinks assembles the reads. Moreover, Cuffmerge can combine reference transcripts into the merged assembly if a reference genome annotation is available. After each sample has been assembled and all samples have been merged, the final assembly is ready for downstream differential analysis (Trapnell *et al.,* 2012).

The main purpose of Cuffmerge is to provide an assembly GTF (Gene Transfer Format) file which can be used by Cuffdiff (differential analysis). A merged, empirical annotation file reflects the expression of rare or novel genes and alternative splicing isoforms; hence, use of such annotation file will be more accurate than using the standard reference annotation for differential analysis (Oliver *et al.,* 2013).

5.1.2.6.5 Cuffdiff (differential analyses)

Cuffdiff is a tool to assemble and evaluate transcripts and their expression levels in more than one condition and test them to identify significant differences between or among them (Oliver *et al.,* 2013). It can deal with multiple technical or biological replicate sequencing libraries for each condition and shows the read count variation for each gene across the replicates. The estimated variance in read counts is used to calculate the significance of observed changes in expression levels (Trapnell *et al.,* 2012).

Cuffdiff output contains numerous output files with information about the differential splicing and expression events across the samples (Ghosh and Chan, 2016). It provides analyses of differential expression and regulation at the gene and transcript levels. The output files include well-known statistics i.e. p-values (both raw and corrected), fold change (in $\log_2$ scale), FPKM-values and also gene and transcript related attributes such as common name and their function (Trapnell *et al.,* 2012).

5.1.2.6.6 CummeRbund (data visualisation)

CummeRbund is user-friendly tool, which deal with the data produced by a Cuffdiff analysis. It helps to integrate all Cuffdiff data and significantly simplifies common data exploration and visualisation tasks such as plotting and cluster analysis of gene and transcript expression data (Trapnell *et al.,* 2012). Additionally, it allows creation of plots, graphs and clustering maps with a single command (Ghosh and Chan, 2016). Finally, Cuffdiff data can be transferred into the R statistical computing environment, making RNA-seq expression analysis with Cuffdiff well suited to many other advanced statistical analysis and plotting packages (Trapnell *et al.,* 2012).

5.2 Materials and methods

5.2.1 Relative expression of gene specific transcripts

5.2.1.1 Plant materials

The samples from Carlton bulb, basal plate, callus and small bulblets directly grown from basal part of twin-scale explants were harvested and stored in -80 ^{o}C. 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.

5.2.1.2 RNA extraction and cDNA preparation

The innuPREP Plant RNA Kit (Analytic Jena, Germany) was used to isolate total RNA from the samples according to the manufacturer's instructions. The quality and quantity of isolated RNA was determined using a NanoDrop Spectrophotometer (Thermo, USA) and samples with a ratio (260/280) from 2.0 to 2.1 were selected for cDNA preparation using the Reverse Transcription Kit (QuantiTect) following the manufacturer's protocol.

5.2.1.3 Primers and PCR reaction

Gene specific primers (Table 5.1) were designed by Pulman (2015). Primers were purchased from Eurofins Genomics (Germany) and used in PCR reactions. Reaction mixtures (10 µl) were prepared with 1µl cDNA template, 1 µl forward (10 µM) and reverse (10 µM) primers, 5 µl BioMix Red (Bioline, UK) and 2 µl PCR water. The following conditions were used for the PCR reactions: pre-denaturation at 94^{o}C (1 min) followed by 35 cycles consisting of 20 s at 94^{o}C (denaturing), 20 s at 57^{o}C (annealing), 30 s at 72^{o}C (extension) and a final extension for 5 min at 72^{o}C. The PCR products were separated on 1% (w/v) agarose gel, stained with Midori Green (Nippon Genetics Europe GmbH, Germany) and recorded using a gel documentation system (U:Genius, Syngene, UK).

The callus and direct bulblets were initiated from three separate bulbs and thus provided independent biological replicates. PCR reactions were carried out in triplicate, as technical replications. Carlton bulb and basal tissues were used as controls for each PCR reaction. For quantification of relative band intensities, PCR

products were analysed using Metamorph (Molecular Devices, USA). All band intensity values were normalised against Actin band intensities for each gel image.

Table 5.1 Primers used for RT-PCR.

Name	Forward Sequences (5'-3')	Reverse Sequences (5'-3')	Length
ACTIN (Housekeeping)	GATAGAACCTCCAATCCAAACACTA	GTGTGATGTGGATATTAGGAAGGAC	25
HDA57 (P450)	ATTTCTCAGCGAGAGCCAAG	CTCCAATTTCTTGGCATGGT	20
Daff 88927 (P450)	CAGTTGGTTTAATTCATCTCTGCTT	ATGACAGAATTCTAGCAGCTTTGTT	25
PAL1	ATGGGAATAAGGAAAAGATGAAAAC	CACAAACCGATACAAAGGATAACAC	25
PAL2	GGGAATAAGGAAAAGATGAAAACAC	GATACAAAGGATAAGACCTGCACTC	25
TYDC1	TGGTTTTAATATTGTGGGTTTCAAT	TTCACTAGCTGTGCCTTGAATTACT	25
TYDC2	GTAATTCAAGGCACAGCTAGTGAAG	ATAAACCACAAGCTTTTCAAGTGAT	25
*Np*N4OMT	AAGACCTGTACGACCATGCA	ATCCACCTCATCTTCCGGAC	20

Primer sequences and designs (Pulman, 2015).

5.2.2 *Narcissus* transcriptome analysis

Figure 5.2 shows an overview of the methods used for cDNA library preparation prior to Illumina sequencing and of the subsequent data analysis.

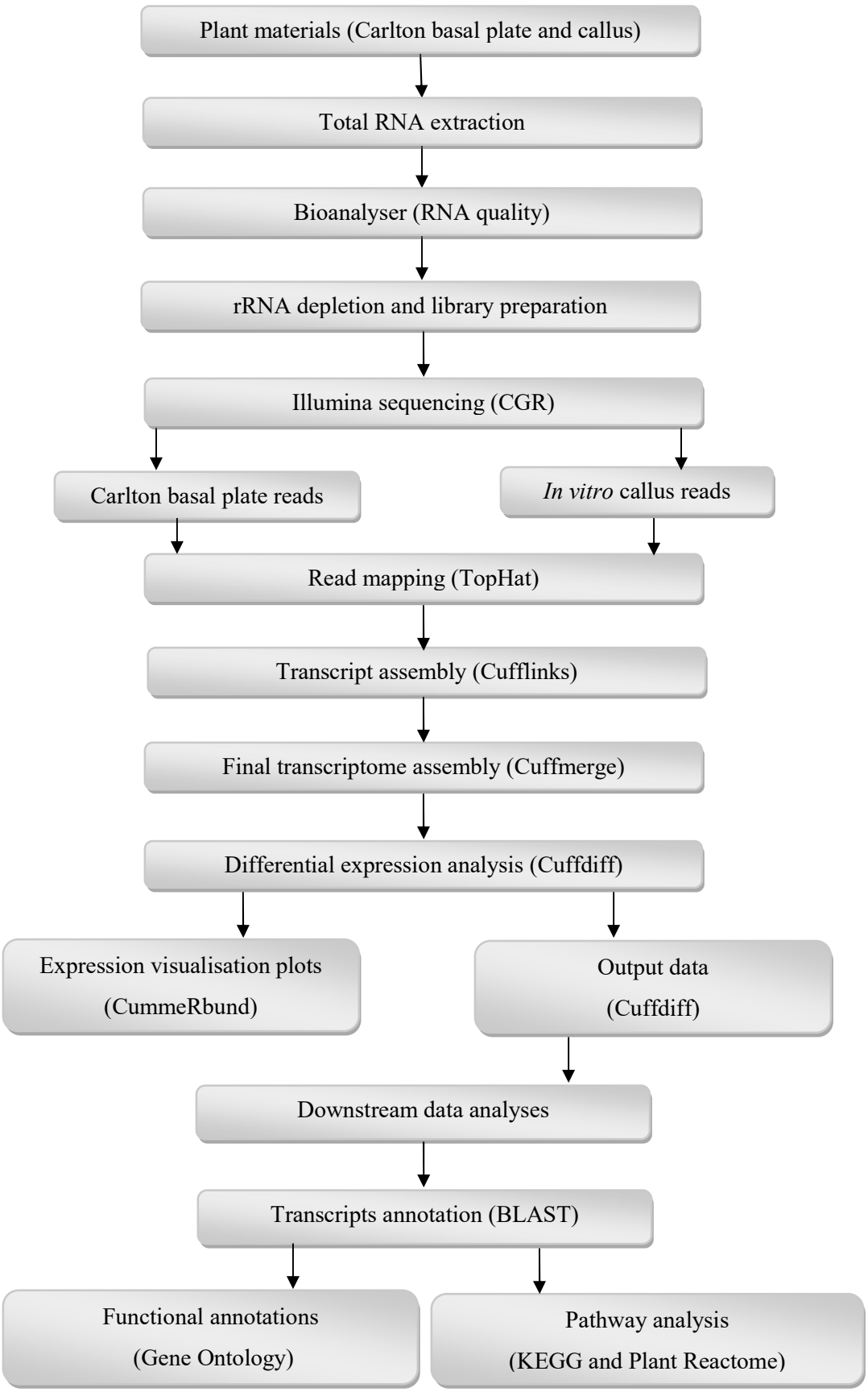

Fig. 5.2 An overview of methods used for the *N. pseudonarcissus* cv. Carlton transcriptome (RNA-seq) data analysis.

5.2.2.1 Library preparation and Illumina sequencing

5.2.2.1.1 Plant materials

The highest and lowest galanthamine containing tissues were selected for RNA extraction. These are basal plate tissues from three bulbs of Carlton and callus grown on MSM1 (Chapter 2, section 2.3, Table 1.2) initiated from three different bulbs. Basal plates and calli were cut into small pieces (1 to 2 mm thick), weighed (~150 mg) and stored in -80^{o}C until required.

5.2.2.1.2 Total RNA extraction and purification

RNA samples were prepared as described in the sections 5.2.1.1 and 5.2.1.2. The manufacturer's protocol of innuPREP Plant RNA Kit (Analytic Jena, Germany) was slightly modified before wash-up stage by adding DNAse treatment (RNase-Free DNase Set, Qiagen) which was a column DNAase treatment, performed following RNeasy Plant mini kit (Qiagen) DNA digestion steps.

5.2.2.1.3 RNA quality and quantity

To determine the quality and quantity of RNA, NanoDrop Spectrometer (Thermo, USA) and Qubit Quant-iT^{TM} RNA Assay Kit (Thermo Fisher Scientific, USA) were used. RNA was extracted in three replicates for each sample and the samples with best quantity and quality of RNA were selected for sequencing. In total six samples, three from each Carlton basal plate and three from each callus were sent to the Centre of Genomic Research (CGR), University of Liverpool, for RNA sequencing.

5.2.2.1.4 Steps performed by CGR

5.2.2.1.4.1 Initial quality and integration of RNA

The quality of selected total RNA samples was initially tested using Agilent RNA 6000 Nano Kit (Agilent Technologies) following the manufacturer's protocol. The first set of Carlton basal plate RNA samples showed significant RNA degradation, so a second set was prepared and submitted to the CGR. 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.

5.2.2.1.4.2 rRNA depletion and library preparation

From total RNA of both basal plate and calli, rRNA was depleted 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.

5.2.2.1.4.3 Illumina sequencing and data processing

The HiSeq 2500 System (Illumina, UK) was used for RNA sequencing, which is a powerful and efficient ultra-high-throughput sequencing system. The raw FASTQ files were trimmed for the presence of Illumina adapter sequences using Cutadapt version 1.2.1 (Martin, 2011). 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 this data processing was completed by CGR. The RNA-seq data in FASTQ format was released by CGR and was available to download for further downstream analyses.

5.2.2.2 RNA-seq analyses

There is no reference genome available for *Narcissus*. Kilgore *et al.,* (2014) reported a *N. pseudonarcissus* transcriptome assembly and raw reads are available at the MedPlant RNA Seq Database (www.medplantrnaseq.org). This *Narcissus* transcriptome data was downloaded from MedPlant RNA Seq Database website and was 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 characterising differential expression. The analyses were performed on Discovery Environment of CyVerse, following the tutorial: DE 002: Characterising 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 visualisation of data.

5.2.2.2.1 Differential expressed genes/ transcripts annotation

The differentially expressed genes/ transcripts (Cuffdiff output) were annotated following a BLAST search by Dr Ryan Joynson (Postdoctoral scientist with Professor Anthony Hall, Institute of Integrative Biology). 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 uncurated and based on top hits from the BLAST output.

5.2.2.2.2 Functional annotations

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).

5.2.2.2.3 Pathway analysis

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.

5.3 Results

5.3.1 RNA samples for relative expression level analysis

Initially the CTAB method was tried to extract RNA from Carlton bulb tissues but the results showed a very poor yield of around 23-55 ng/µl. The innuPREP Plant RNA Kit (Analytic Jena, Germany) was then used and gave adequate quantity and quality RNA from the samples (Table 5.2) for RT-PCR.

Table 5.2 RNA samples selected for RT-PCR.

Sample Name		Nano drop (ng/µl)	Ratio (260/280)	Ratio (260/230)
Carlton bulb		247.31	2.12	2.23
Carlton basal plate		627.67	2.15	2.32
Callus	bulb 4	840.53	2.17	2.31
	bulb 5	340.35	2.17	2.38
	bulb 6	123.62	2.13	2.19
Bulblets	bulb 4	281.93	2.15	2.52
	bulb 5	112.51	2.16	2.30
	bulb 6	416.01	2.15	2.42

5.3.2 Relative expression of gene specific transcripts implicated in alkaloid biosynthesis

All seven gene specific transcripts (Table 5.1) were tested from field (bulb and basal plate tissue) and *in vitro* cultured calli and direct bulblets obtained from three different bulbs (bulbs 4, 5 and 6) where field tissues were used as control. Figure 5.3 and 5.4 show the relative expressions, normalised to actin, for the seven transcripts.

The selected transcripts for P450s (HDA57 and Daff 88927) were expressed almost in all tissue types; both field and *in vitro* grown samples (Figure 5.3A, B, C and D). The only exception was bulb samples (Figure 5.3D) showed no expression at all of Daff 88927. Higher expression of both *PAL* genes was observed in calli (Figure 5.3F and H) than field-grown bulb and basal plate. Low or no expression of PAL1 was identified in field tissues (Figure 5.3E and F) and neither PAL1 nor PAL2 were expressed in bulblets obtained from bulb 5 (Figure 5.3E and G).

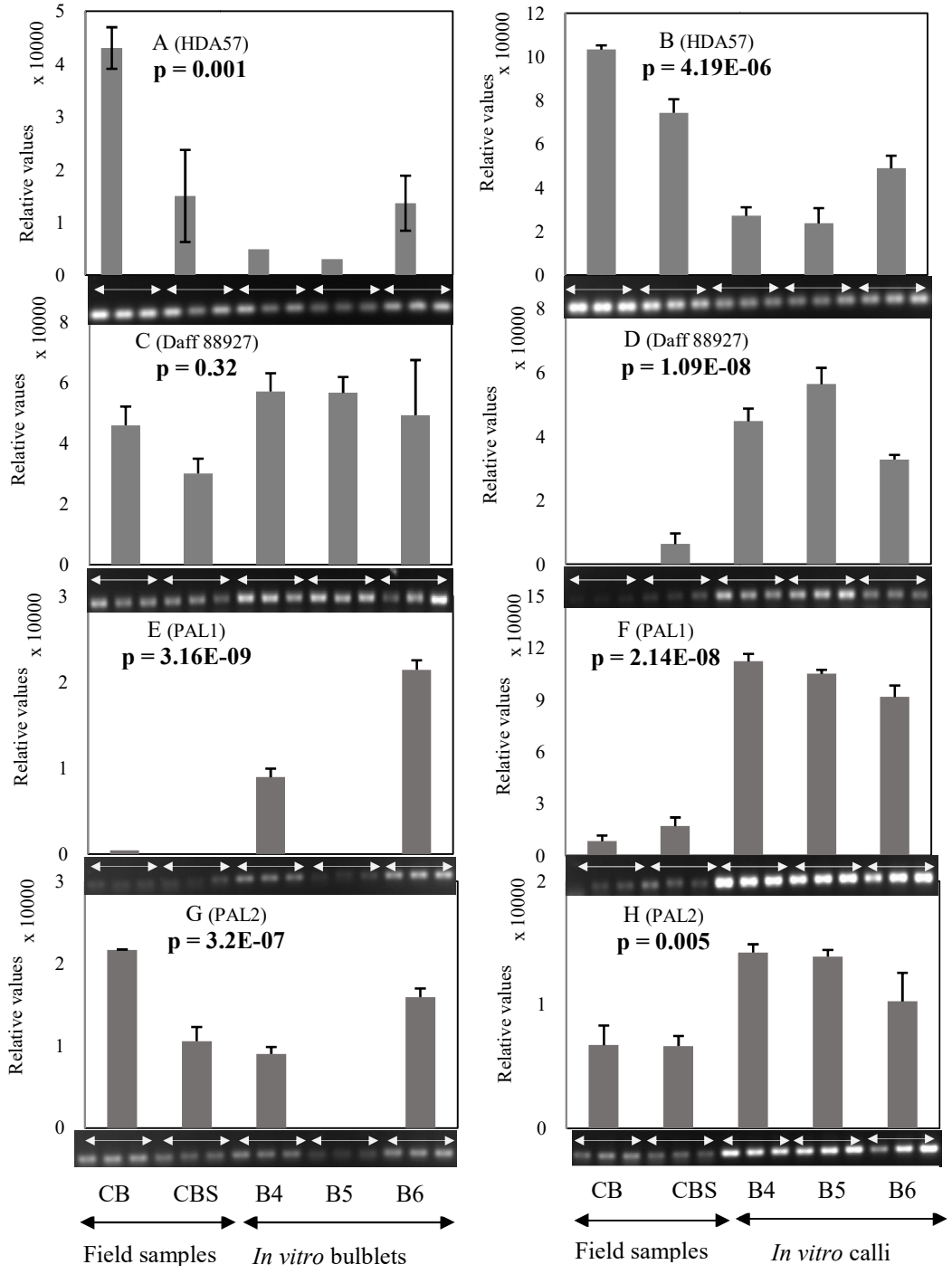

Fig 5.3 Relative expression levels P450s and *PAL* transcripts to ACTIN: HDA57 (A and B); Daff 88927 (C and D); PAL1 (E and F) and PAL2 (G and H) with their corresponding gel images (3 replicates); CB=Carlton bulb, CBS=Carlton basal plate, different bulbs (B4, B5 and B6).
Error bars = SE. The p-values of the ANOVA between the different samples are shown in the figure.

Figure 5.4A, B, C and D showed that the *TYDC* transcripts were present almost in all tissue types, at different levels of expression. Both TYDC1 and TYDC2 showed higher expression level in calli (Figure 5.4B and D) than bulb and basal plate tissue

but lower level in bulblets (Figure 5.4A and C). The expression level of *Np*N4OMT was notably higher in field-grown bulb and basal tissues than *in vitro* grown bulblets and calli (Figure 5.4E and F).

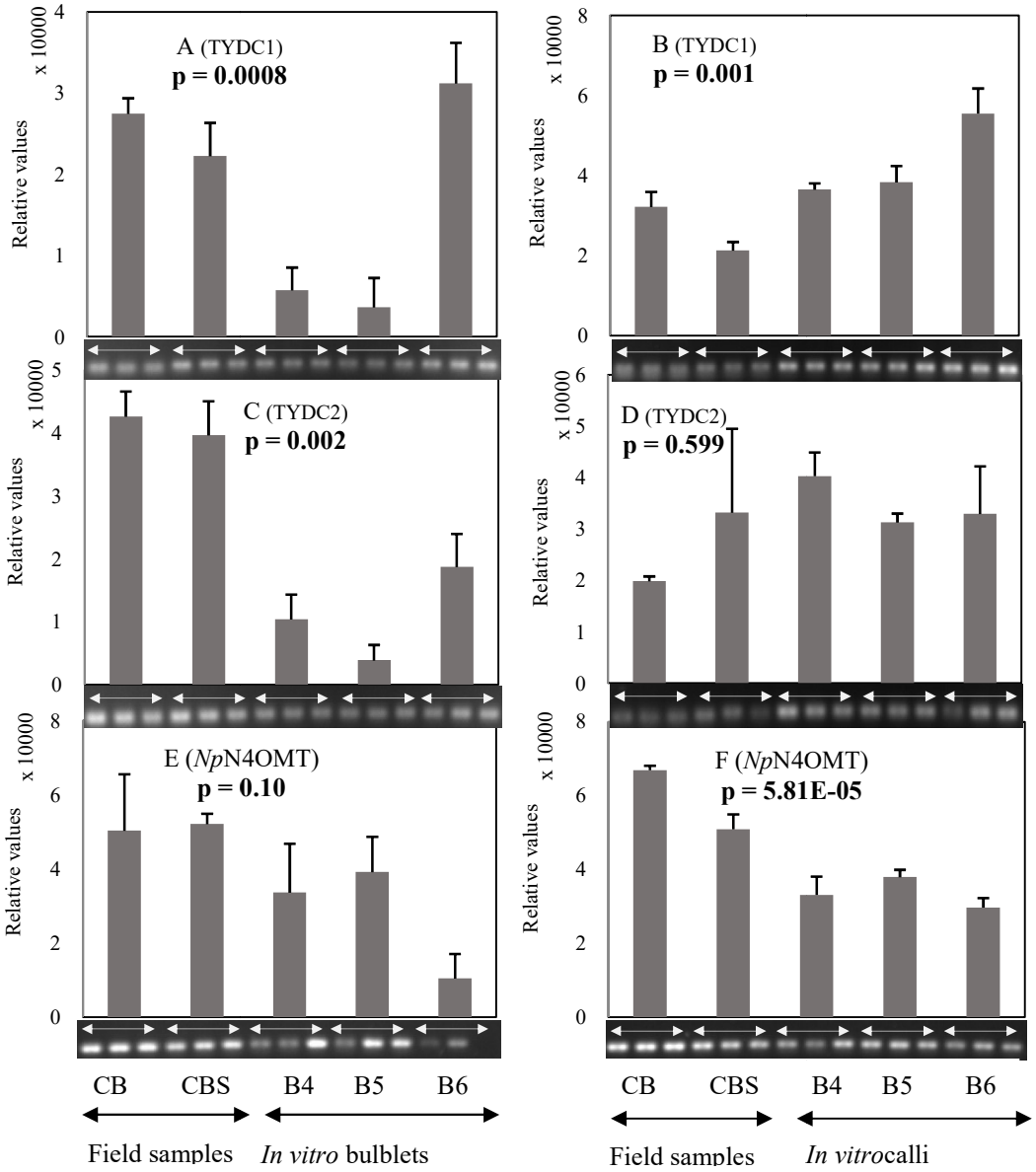

Fig 5.4 Relative expression levels of *TYDC* and *Np*N4OMT transcripts to ACTIN: TYDC1 (A and B); TYDC2 (C and D); *Np*N4OMT (E and F) with their corresponding gel images (3 replicates); CB=Carlton bulb, CBS=Carlton basal plate, different bulbs (B4, B5 and B6). Error bars = SE. The p-values of the ANOVA between the different samples are shown in the figure.

5.3.3 Extraction of RNA for RNA-seq

The Analytic-Jena InnuPREP plant RNA kit with the PL lysis solution gave consistent high yields and 260/280 ratios on NanoDrop. Total RNA of Carlton basal plates from three different bulbs and callus derived from three different bulbs grown

on the same medium (high auxin) with high yield (Qubit and NanoDrop concentrations) and quality (260/230 ratio) were initially selected for RNA sequencing (Table 5.3).

Table 5.3 RNA samples selected for RNA sequencing.

Sample Name	Total Volume (µl)	Qubit (µg/ml)	Nano drop (ng/µl)	Ratio (260/280)	Ratio (260/230)
CBS A	35	461	283.15	2.16	2.06
CBS B	35	458	318.79	2.17	2.31
CBS C	35	1700	954.91	2.17	2.27
CAL1	35	864	537.15	2.18	2.29
CAL2	35	497	312.06	2.19	2.33
CAL3	35	1000	687.42	2.18	2.26

CBS = Carlton basal plate and CAL = Callus.

5.3.3.1 Total RNA quality analysis

The initial Bioanalyser results performed in the institute laboratory before sample submission to CGR showed high degradation with a noisy baseline for basal plate RNA samples. Representative Bioanalyser graphs of basal plate and callus are in Figures 5.5A (ladder), B (basal plate) and C (callus).

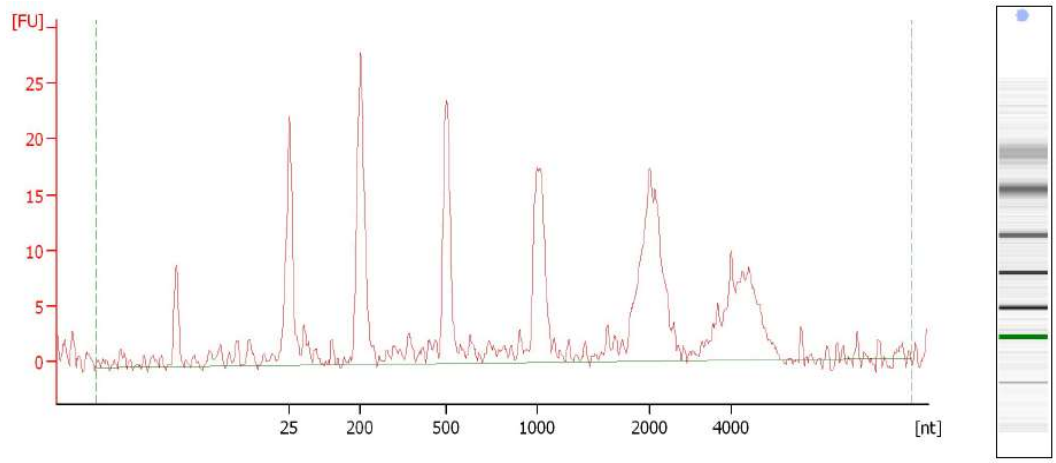

Fig. 5.5A Electropherogram of ladder (RNA area: 261.8 and RNA concentration: 150 ng/µl); FU = fluorescence units; nt = nucleotide.

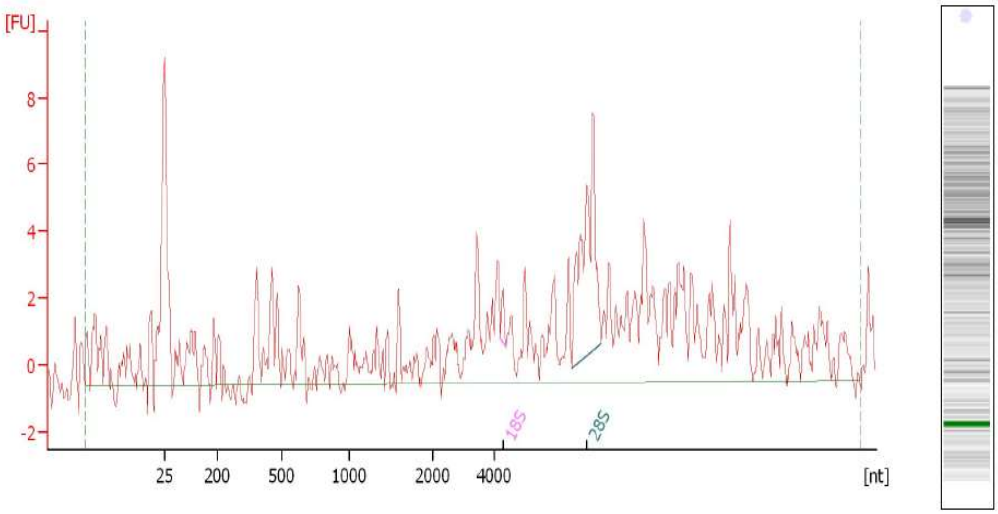

Fig. 5.5B Total RNA electropherogram for Carlton basal plate (CBS A). The rRNA (28S/18S) was 17.1, RIN 6.30 and a final concentration 79.0 ng/µl.

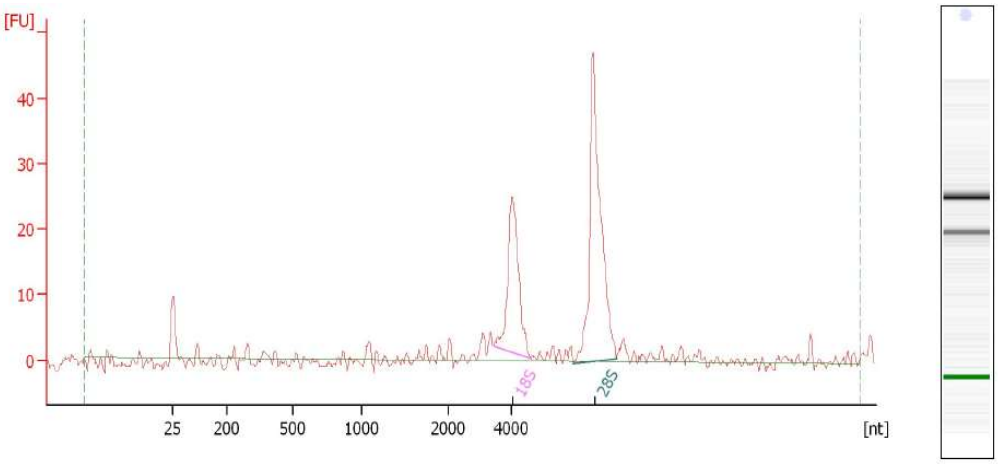

Fig. 5.5C Total RNA electropherogram for callus from CAL1. The rRNA (28S/18S) was 1.8, RIN 8.80 and a final concentration 99.0 ng/µl.

The first RNA samples from all three Carlton basal plate samples (e.g. Figure 5.5B) showed substantial degradation as indicated by the loss of prominent rRNA peaks and the elevated material in the 200-1000 nucleotide range with poor RNA integrity, RIN value 6.3. RNA extractions for basal plates from another three separate bulbs were repeated (Table 5.4).

Table 5.4 Repeated RNA samples of Carlton basal plate selected for the submission to CGR.

Sample name	Total volume (µl)	Nano drop (ng/µl)	Ratio (280/260)	Ratio (280/230)
Carlton basal plate 1 (CBS1)	35	429	2.1	2.2
Carlton basal plate 2 (CBS2)	35	173	2.0	2.2
Carlton basal plate 3 (CBS3)	35	375	2.1	2.3

5.3.3.2 Initial quality check of submitted total RNA sample (performed by CGR)

Figures 5.6A-G represented total RNA quality analysis results obtained from CGR.

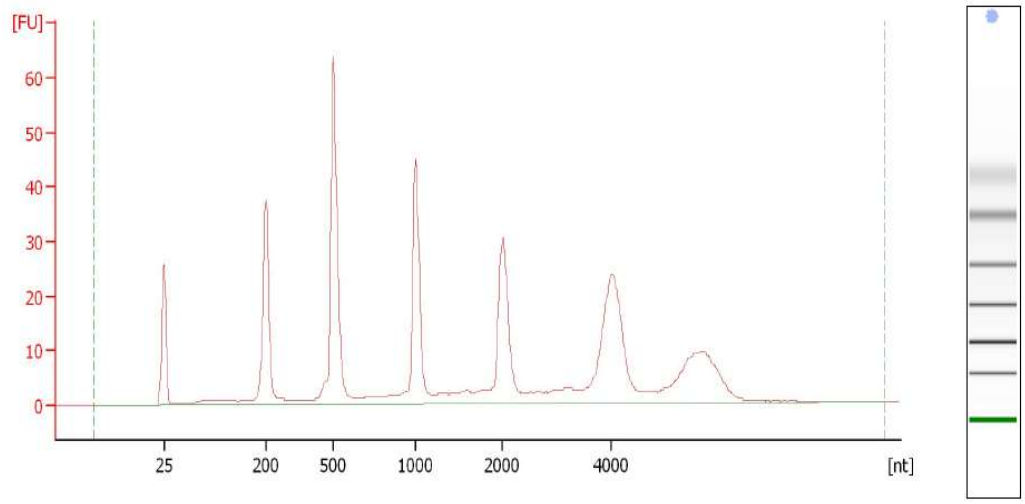

Fig. 5.6A Electropherogram of ladder (RNA area: 459.9 and RNA concentration: 1000 pg/µl); FU = fluorescence units; nt = nucleotide.

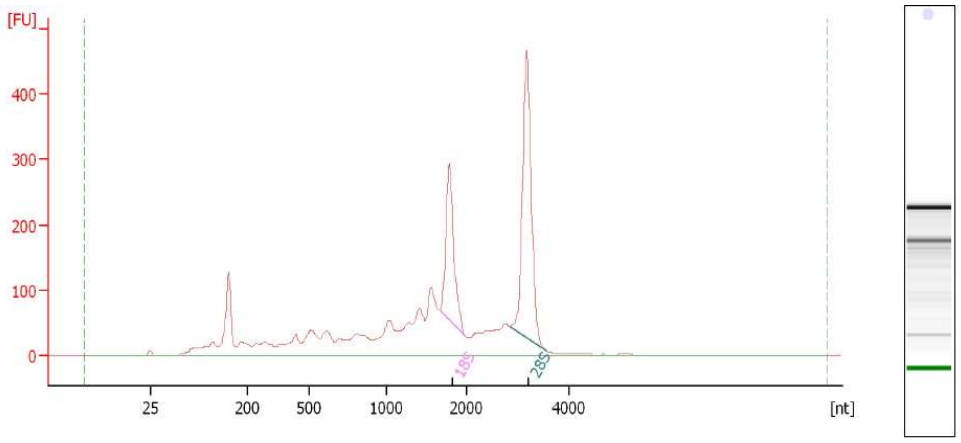

Fig. 5.6B Total RNA electropherogram for Carlton basal plate (CBS1).The rRNA (28S/18S) was 1.8, RIN 7.6 and a final concentration of 6600 pg/µl.

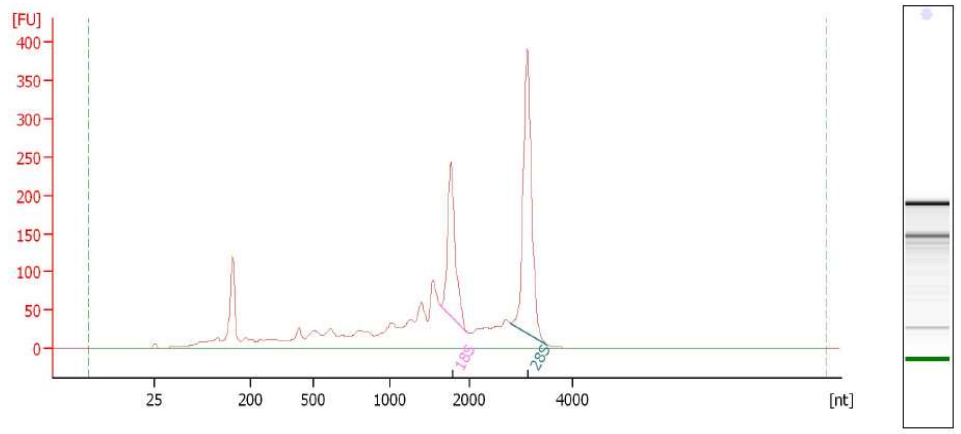

Fig. 5.6C Total RNA electropherogram for Carlton basal plate (CBS2). The rRNA (28S/18S) was 1.8, RIN 7.7 and a final concentration of 4960 pg/µl.

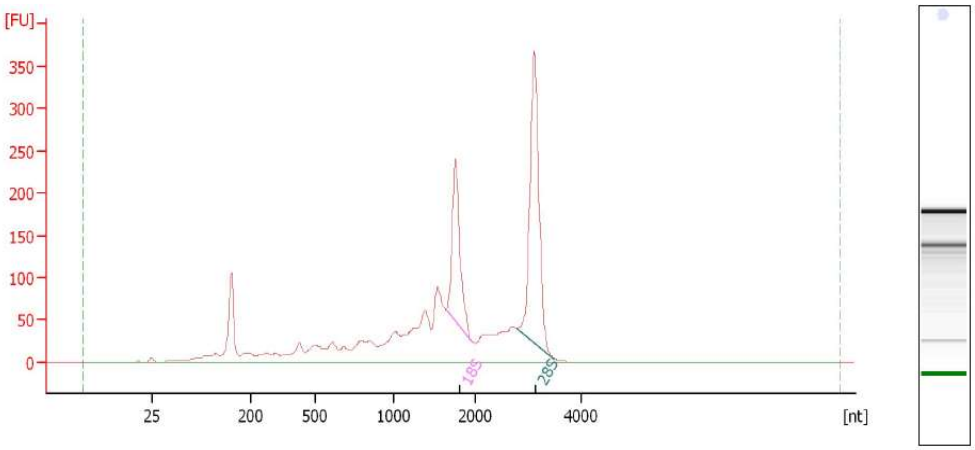

Fig. 5.6D Total RNA electropherogram for Carlton basal plate (CBS3).The rRNA (28S/18S) was 1.8, RIN 7.6 and a final concentration of 5009 pg/µl.

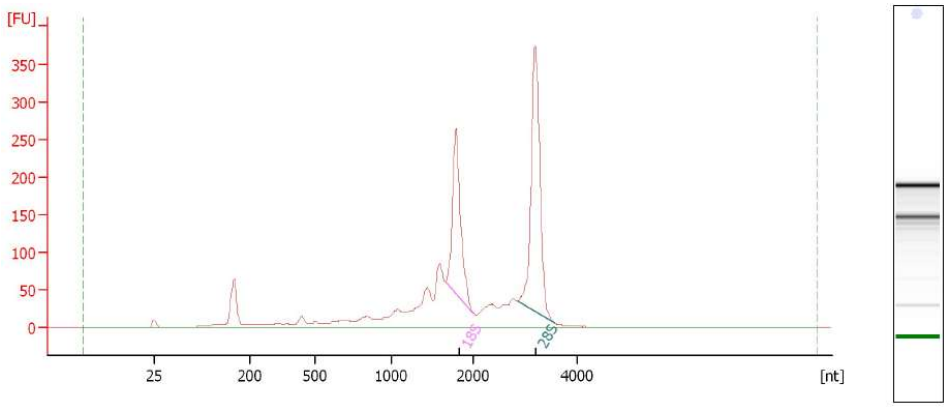

Fig. 5.6E Total RNA electropherogram for callus from bulb 4 (CAL1), same callus sample shown in Figure 5.4C. The rRNA (28S/18S) was 1.5, RIN 8.1 and a final concentration of 4271 pg/µl.

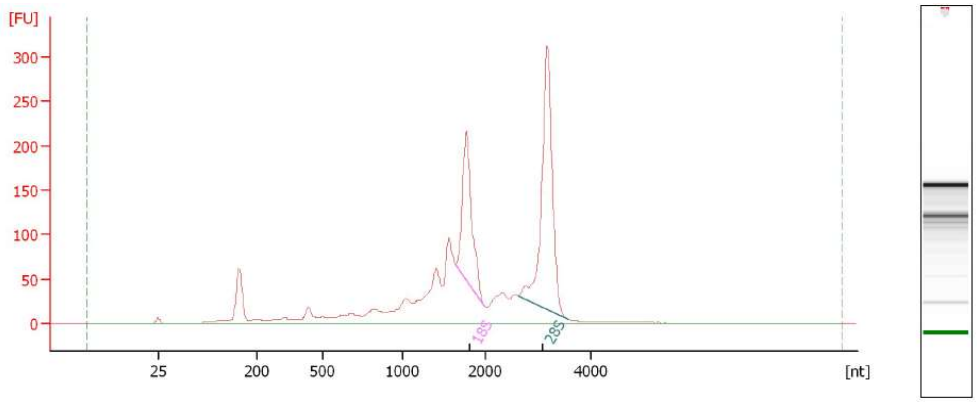

Fig. 5.6F Total RNA electropherogram for callus from bulb 5 (CAL2). The rRNA (28S/18S) was 1.8, RIN 8.0 and a final concentration of 4291 pg/µl.

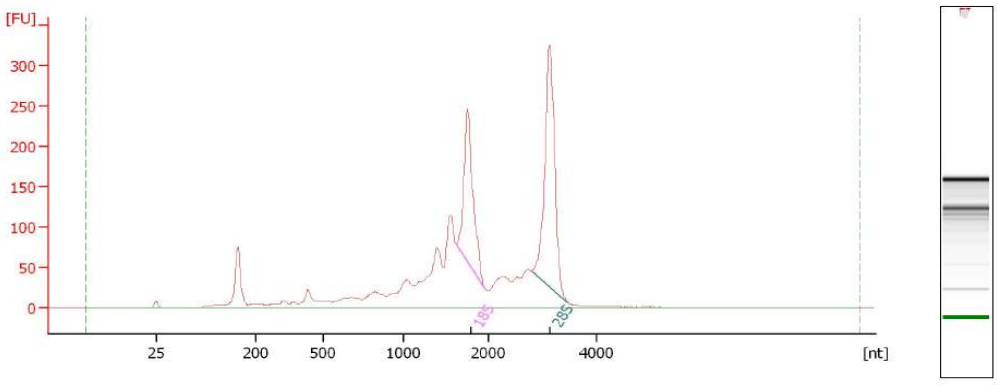

Fig. 5.6G Total RNA electropherogram for callus from bulb 6 (CAL3). The rRNA (28S/18S) was 1.5, RIN 8.1 and a final concentration of 4850 pg/µl.

The RIN numbers for all samples under study were higher than the RNA degradation threshold of 6. The 28S/18S peak ratio also showed values adequately close to 2.0 (ranges from 1.5 to 1.8) and the baseline signals were relatively lower within the

200-1000 nucleotide range (Figure 5.6B-G) than the earlier samples (Figure 5.5B and C). The 28S RNA peaks were clearly more intense than the 18S RNA peaks. Therefore, three replicates of both Carlton basal plate (CBS1, CBS2 and CBS3) and callus (CAL1, CAL2 and CAL3) were deemed to be of high enough yield and quality to continue with the rRNA depletion step prior to cDNA library creation.

5.3.4 rRNA depletion

The Ribo-Zero kits require only of 1-5 μg of total RNA. The yield of depleted RNA depends on the amount of input total RNA, rRNA content of the sample, and the method used to purify the Ribo-Zero-treated RNA. It typically results in less than 8% of the total input RNA (www.illumina.com).

Depletion of rRNA using the Ribo-Zero rRNA removal kit was efficient for all samples under study as the percent recovery of rRNA was less than 1% (Table 5.5) for all samples (callus and Carlton basal plate) under study.

Table 5.5 Post rRNA (Ribosomal) depletion results.

Sample name	Total starting conc. (ng/ul)	Bioanalyser (BA) conc. (ng/ul)	Total recovery ng (BA)	Percent (%) recovery (BA)
Callus 1 (CAL1)	2000	0.723	8.676	0.4338
Callus 2 (CAL2)	2000	0.222	2.664	0.1332
Callus 3 (CAL3)	1000	0.104	1.248	0.1248
Carlton basal plate 1 (CBS1)	1000	0.253	3.036	0.3036
Carlton basal plate 2 (CBS2)	1000	0.615	7.38	0.738
Carlton basal plate 3 (CBS3)	2000	0.501	6.012	0.3006

CAL = callus; CBS = Carlton basal plate.

The rRNA depletion results obtained from the CGR for the callus replicates are shown in Figures 5.7A, B and C and for Carlton basal plate replicates in Figures 5.7D, E and F.

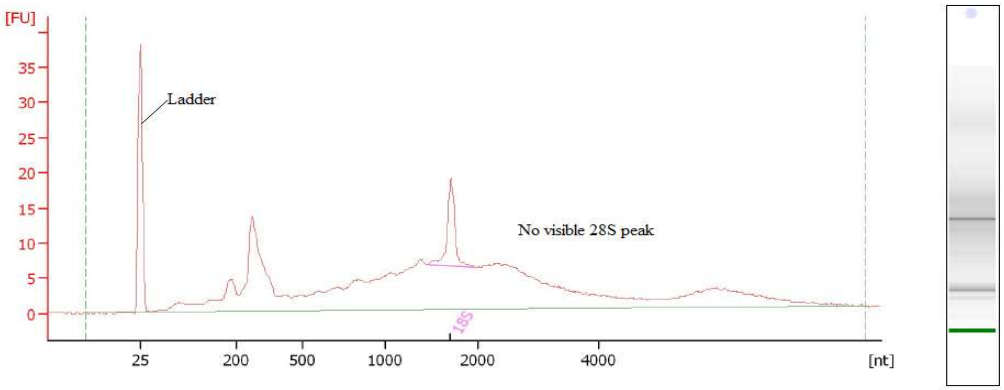

Fig. 5.7A rRNA depleted electropherogram for callus from bulb 4 (CAL1).

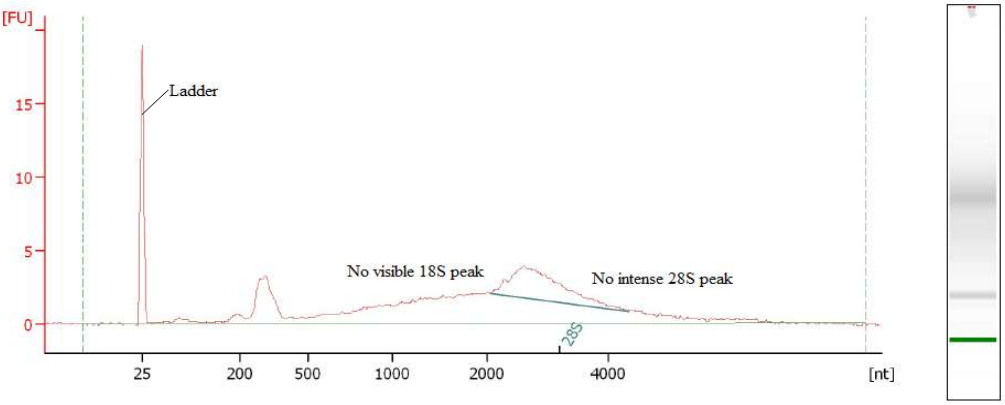

Fig. 5.7 B rRNA depleted electropherogram for callus from bulb 5 (CAL2).

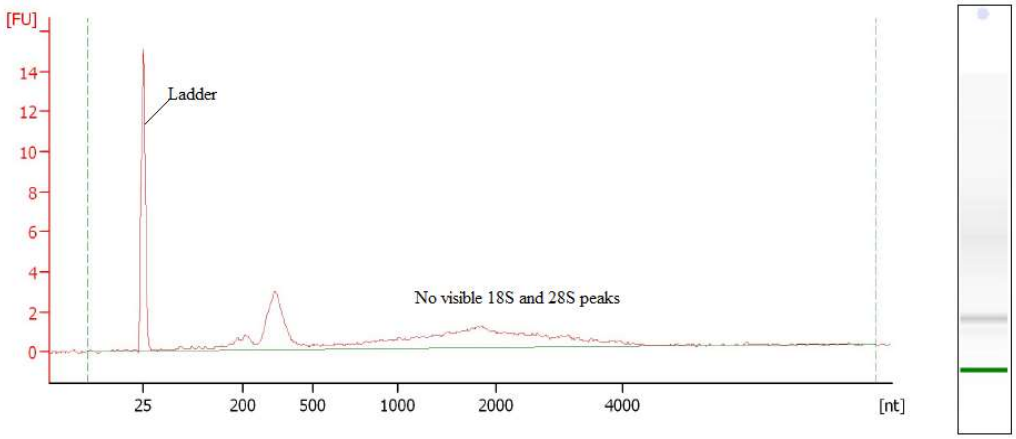

Fig. 5.7 C rRNA depleted electropherogram for callus from bulb 6 (CAL3).

In Figures 5.7A, B and C the lack of obvious rRNA peaks (18S and/ or 28S) with rRNA ratio of 0.0 and the marker peak at 25 nt represents successful ribosomal RNA depletion from callus CAL1, CAL2 and CAL3 samples.

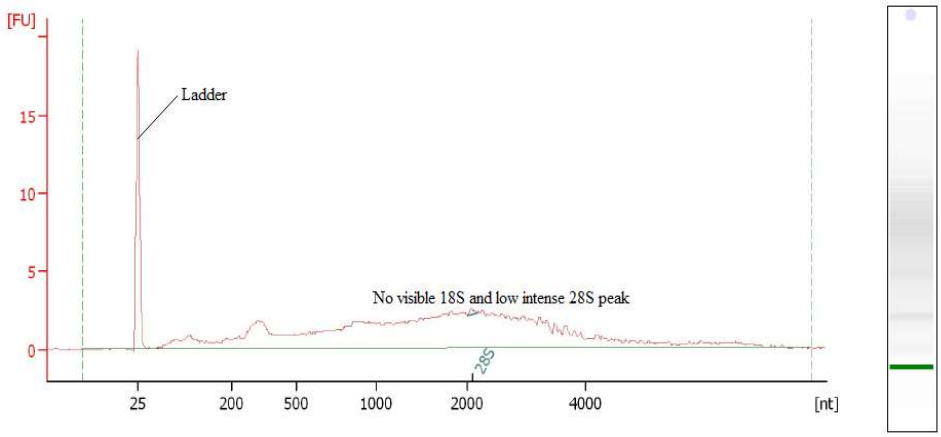

Fig. 5.7D rRNA depleted electropherogram for Carlton basal plate from bulb 1 (CBS1).

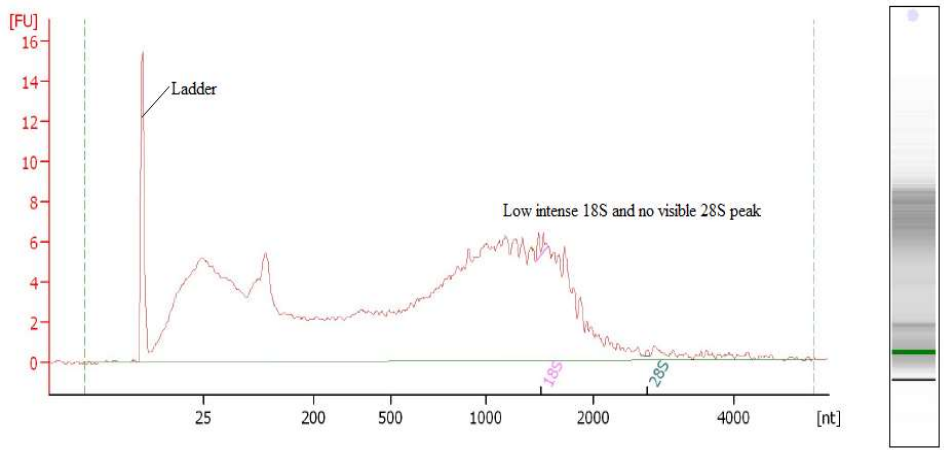

Fig. 5.7E rRNA depleted electropherogram for Carlton basal plate from bulb 2 (CBS2).

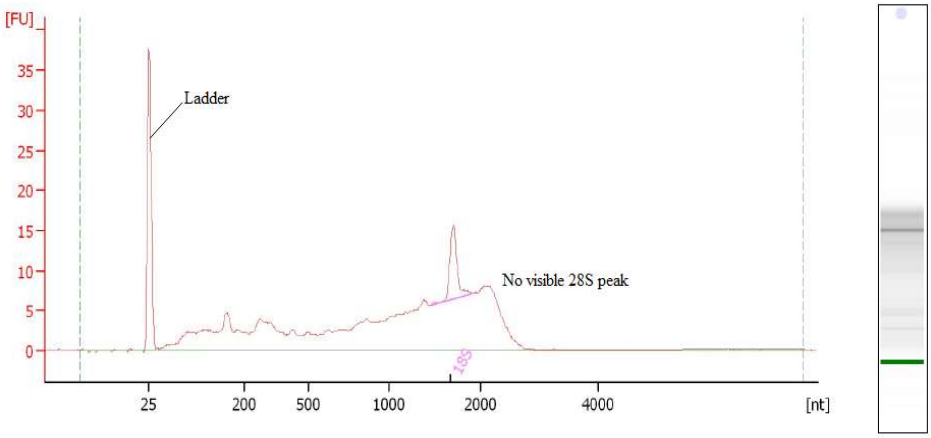

Fig. 5.7F rRNA depleted electropherogram for Carlton basal plate from bulb 3 (CBS3).

Figures 5.7D, E and F show successful rRNA depletion in all three basal plate samples since there are essentially no 28S/18S peaks. For sample CBS2, the rRNA ratio (28S/18S) was 0.2 unlike 0.0 for the other two. However, all were considered acceptable to go forward for library preparation prior to sequencing.

5.3.5 RNA-seq library preparation and sequencing

The Carlton basal plate and callus libraries were generated by Gavin Campbell, CGR, University of Liverpool, using ScriptSeq v2 (Epicentre, USA) (section 5.2.2.4.2). The six libraries were sequenced by CGR (Illumina) and 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 CBS1 and CBS3.

5.3.6 Trimming and filtering of raw reads

The results of the adapter trimming and quality filtering for Carlton basal plate and callus samples are shown in Table 5.6 and Table 5.7 respectively. 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 CAL1, 97.5% in CAL2, 96.05% in CAL3, 96.6% in CBS1 and 97.2% in CBS3, except for CBS2 (92.2%).

Table 5.6 Comparison of Carlton basal plate replicates (CBS1, CBS2 and CBS3) Illumina reads after data trimming and filtering.

Sample	Read		Total number of reads	Maximum read length (bp)	Average read length (bp)	Read length SD (bp)	Minimum base quality score	Maximum base quality score	Average base quality score	Base quality score SD
CBS1	forward	raw	57529368	125	125.0	0.0	2	38	36.5	3.6
		trimmed	55600197	125	107.7	25.5	2	38	36.6	3.3
	reverse	raw	57529368	125	125.0	0.0	2	38	35.8	5.1
		trimmed	55600197	125	104.9	28.4	2	38	36.1	4.2
CBS2	forward	raw	26657577	125	125.0	0.0	2	38	36.6	3.5
		trimmed	24570710	125	103.8	26.7	2	38	36.7	3.2
	reverse	raw	26657577	125	125.0	0.0	2	38	35.9	4.9
		trimmed	24570710	125	101.5	28.9	2	38	36.3	4.0
CBS3	forward	raw	45672176	125	125.0	0.0	2	38	36.5	3.5
		trimmed	44401111	125	102.1	28.2	2	38	36.7	3.2
	reverse	raw	45672176	125	125.0	0.0	2	38	35.6	5.2
		trimmed	44401111	125	99.3	30.5	2	38	36.0	4.4

SD= Standard Deviation; bp= base pair.

Table 5.7 Comparison of callus replicates (CAL1, CAL2 and CAL3) Illumina reads after data trimming and filtering.

Sample	Read		Total number of reads	Maximum read length (bp)	Average read length (bp)	Read length SD (bp)	Minimum base quality score	Maximum base quality score	Average base quality score	Base quality score SD
CAL1	forward	raw	43764750	125	125.0	0.0	2	38	36.5	3.6
		trimmed	42756360	125	109.0	24.2	2	38	36.6	3.4
	reverse	raw	43764750	125	125.0	0.0	2	38	35.8	5.0
		trimmed	42756360	125	106.4	27.3	2	38	36.2	4.1
CAL2	forward	raw	39291034	125	125.0	0.0	2	38	36.6	3.5
		trimmed	38327443	125	107.5	24.3	2	38	36.7	3.2
	reverse	raw	39291034	125	125.0	0.0	2	38	35.9	5.0
		trimmed	38327443	125	105.0	27.1	2	38	36.3	4.0
CAL3	forward	raw	53044767	125	125.0	0.0	2	38	36.5	3.6
		trimmed	50952814	125	112.1	22.1	2	38	36.6	3.3
	reverse	raw	53044767	125	125.0	0.0	2	38	35.8	5.0
		trimmed	50952814	125	109.3	25.7	2	38	36.2	4.1

SD= Standard Deviation; bp= base pair.

Figure 5.8 shows the total number of reads obtained from each replicate of callus and Carlton basal plate showing paired reads, singlet reads and discarded reads due to poor quality or adapter contamination. More than 60 million reads were obtained from all samples under study except one replicate of basal plate (CBS2).

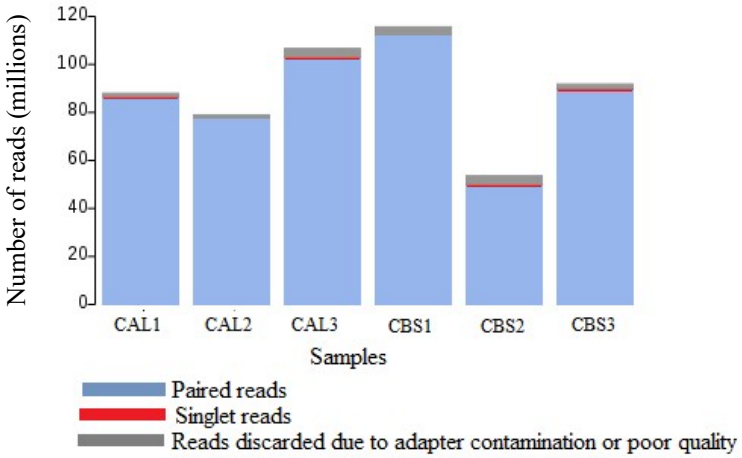

Fig. 5.8 Total number of reads obtained from each replicate of callus (CAL1, CAL2 and CAL3) and Carlton basal plate (CBS1, CBS2 and CBS3).

Figure 5.9 shows the distribution of trimmed read length showing median length (red line), interquartile range (box) while lines extending vertically from the boxes (whiskers) indicate minimum and maximum read lengths.

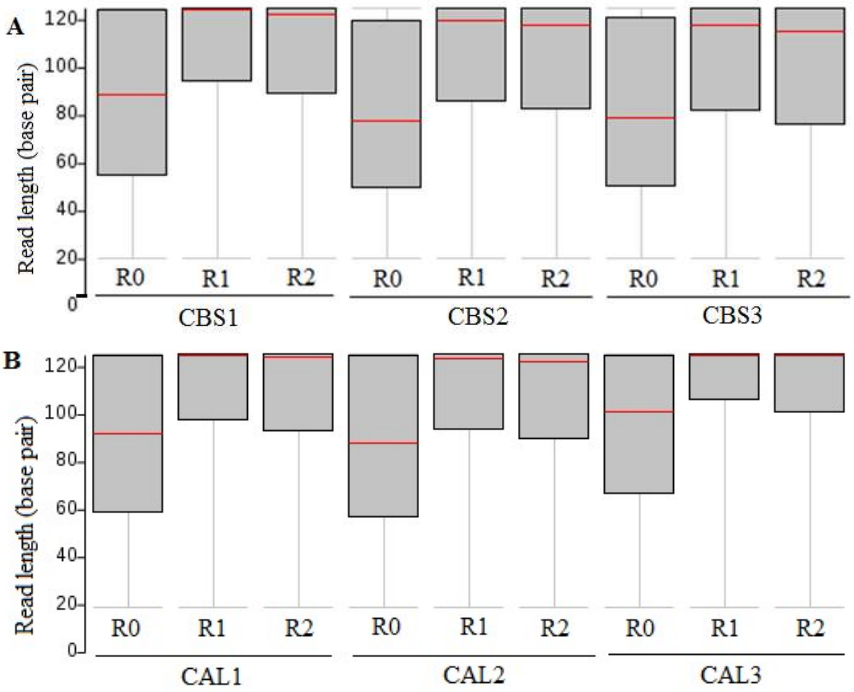

Fig. 5.9 Boxplot showing the distribution of trimmed read lengths: (A) Carlton basal plate replicates (CBS1, CBS2 and CBS3) and (B) callus replicates (CAL1, CAL2 and CAL3) for the forward (R1), reverse (R2) and singlet (R0) reads.

5.3.7 Gene/ transcript expression analysis

5.3.7.1 Sequence alignment

The trimmed RNA-seq reads (FASTQ) obtained from the CGR were initially mapped to the *Arabidopsis* genome data preinstalled in the CyVerse databases and to a *Galanthus spp*. transcriptome assembly downloaded from MedPlant RNA Seq Databases (www.medplantrnaseq.org) using TopHat. The TopHat alignment showed very poor read mapping coverage of only 0.5 to 5.8% (*Arabidopsis*) and 4.1 to 9.8% (*Galanthus*) of total input respectively. A *Narcissus* transcriptome assembly downloaded from MedPlant RNA Seq Databases (www.medplantrnaseq.org) was finally used for RNA-seq read alignment and the TopHat read mapping results are shown in Table 5.8. Moderate mapping coverage ranging from about 35% to 48% of total input was observed in all samples under study except one replicate of basal plate (CBS2) sample (about 19%).

Table 5.8 TopHat output obtained from *Narcissus* RNA-seq data (Discovery Environment, CyVerse) against *Narcissus* transcriptome from MedPlant RNA Seq Databases.

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

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. Minimum intron length, 70 bp and maximum intron length, 50000 bp. Maximum number of alignments to be allowed (alignment score) of 20, number of mismatches allowed in each segment for reads mapped independently was 2, minimum length of read segment 20 bp and meta-pair inner distance of 50bp. These parameters are set for reporting the best possible alignment. This may greatly increase the mapping accuracy at the expense of an increase in running time and rate of alignment.

Some parameters were too strict such the minimum intron length where TopHat will ignore donor or acceptor pairs closer than 70 bases apart and the maximum intron length, TopHat will ignore donor or acceptor pairs farther than 50000 bases apart. This range can be increased for higher alignment rate. The mean expected distance between mate pairs could be set at 200 bp rather than 50 bp. The anchor length, TopHat will report junctions spanned by reads with at least 8 bases (default) on each side of the junction. This must be at least 3, which something might be strict for overall alignment rate. TopHat will allow up to 20 alignments to the reference for a given read, and will choose the alignments based on their alignment scores. TopHat will only report the alignments with the best alignment score. For additional or secondary alignments, this limit can be changed using higher or maximum alignment score.

5.3.7.2 Transcript assembly

The transcript assembly output obtained from Cufflinks detected a total of 2153 differentially expressed genes between basal plate and callus. The CummeRbund figures (Figure 5.10A and 5.10B) of all detected genes illustrated the mapping pattern of all replicates of basal plate and callus samples.

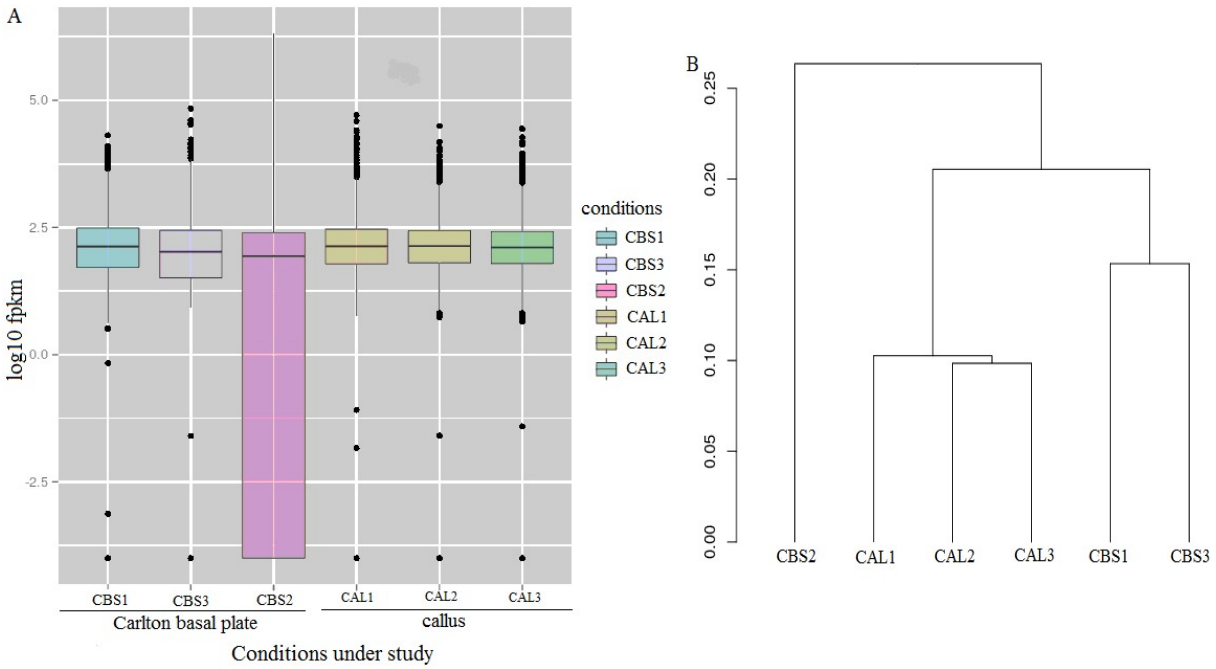

Fig. 5.10 Box plot (CummeRbund) with replicates shows individual replicate FPKM distributions (A); dendrogram (CummeRbund) using JS (Jensen-Shannon) distance with replicates (B) illustrated the relationship among the six RNA-seq samples.

It is obvious (from Figure 5.10) that CBS2 is an outlier from both the dendrogram and CummeRbund box plot. It was already apparent during library preparation and this basal plate replicate showed 74% mapping to rRNA and Figure 5.10A shows it has also high negative mapping (log10 fpkm) values after transcript assembly (Cufflinks). The other two basal plate replicates and all three callus replicates were closely linked with each other (Figure 5.10B).

5.3.7.3 Differential gene expression in basal plate and callus

A total of 2153 genes were found to be differentially expressed and regulated in two conditions under study (field derived basal plate and *in vitro* grown callus). About 64.21% genes were detected as up regulated in callus, whereas 35.21% were up regulated in the basal plate.

However, many of these genes were not significantly different between the two conditions under study, as shown in the volcano plot in Figure 5.11. 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, and 133 genes with q-values of ≤ 0.01 among the total of 2153 differentially expressed genes identified using Cuffdiff analysis. The mapping details of all 206 genes (annotated) with their corresponding $\log_2$ (fold change), total FPKM, regulation pattern and q-values are in Appendix 5.1.

The distribution of all significantly differentially expressed genes (-log10 (p-value) ≥ 2.0) in Figure 5.11 showed that many genes are distinctly differentially expressed between the two conditions. Those genes on the positive $\log_2$ axis are up regulated in callus (down regulated in basal plate), while the negative $\log_2$ axis shows genes up regulated in Carlton basal plate (down regulated in callus).

*q-value is the adjusted p-value considering false discovery rate (FDR), which is necessary for measuring thousands of variables (i.e. gene expression lelels) from a small sample set. Hence, a q-value of 0.05 implies to accept that 5% of the tests found to be statistically significant by p-value will be false positives.

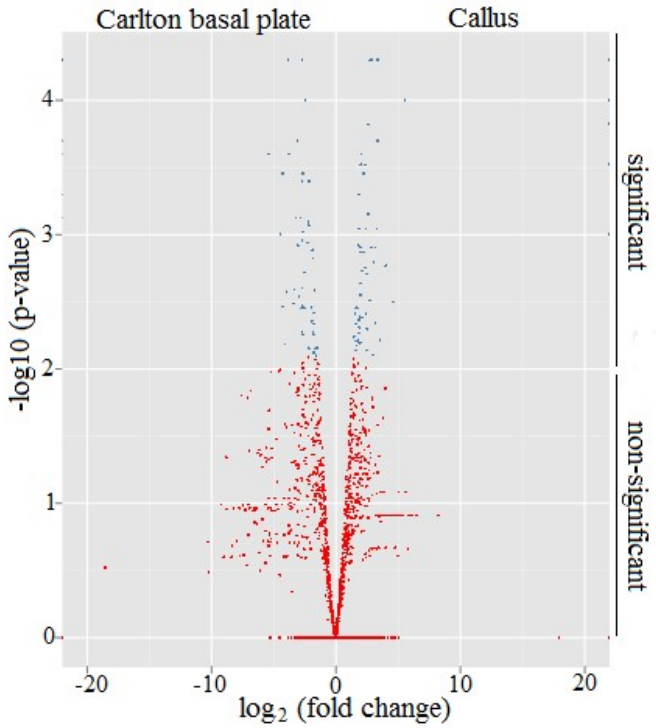

Fig. 5.11 Volcano plot (CummeRbund) indicating the presence of differentially expressed genes between Carlton basal plate and callus. Significantly, expressed genes are indicated by blue dots and red dots are representing the non-significant genes.

Among the 2153 genes, a BLAST search against databases (UniProt, SwissProt, RefSeq and TAIR) provided annotation for about 1700 (78.95%) genes. Eleven most abundant gene groups (Figure 5.12) obtained from the data that contributed 41.7% of the total annotated genes (1700). The largest group was uncharacterised proteins (10.24%) and the second largest gene group was genes responsible for secondary metabolite production (8.88%) which included cytochrome P450s, methyltransferases, CoA-reductases, NADP/ NADH dependant reductases, cinnamate-4-hydroxylases, hydrolases, aldo-keto reductases, oxidoreductases and others (Figure 5.12). The third largest group was tissue differentiation (4.06%) and stress-related proteins (4.06%) such as pathogenesis related, universal stress, zinc finger and chitin or chitinase like proteins (Figure 5.12, Appendix 5.3). Other groups detected as abundant 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/hydolases and sugar and amino acid synthase related proteins.

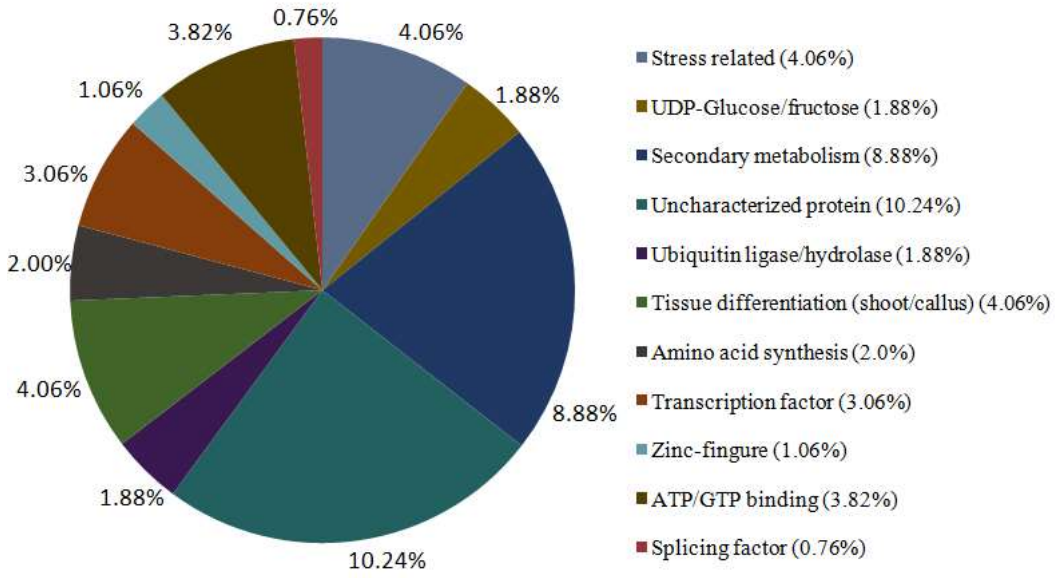

Fig. 5.12 Overall distributions of the eleven abundant gene groups that were differentially expressed between basal plate and callus. The percentages contributed 41.7% to the total annotated genes (78.95%).

A gene list of 106 sorted genes were automatically generated by Cuffdiff software based on total FPKM-values and $\log_2$ (fold change). The representative expression levels (heatmap) of 106 sorted genes are shown in Figure 5.13A, and Figure 5.13B shows the expression of the first fifty genes sorted by Cuffdiff. The heatmaps show genes with q-values of ≤ 0.01.

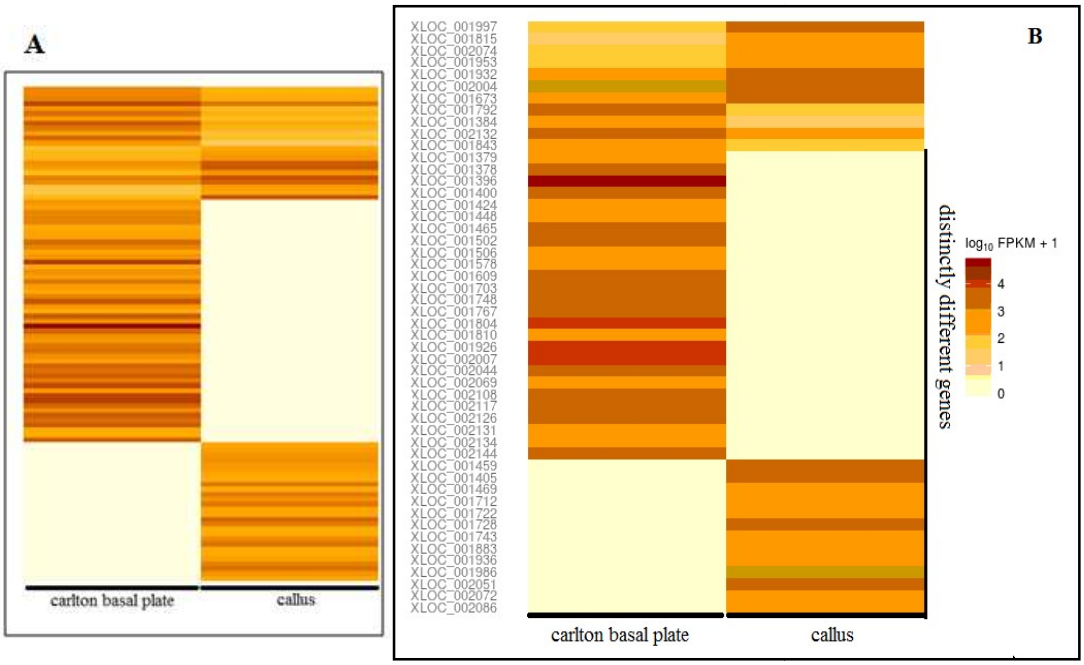

Fig. 5.13 Heatmaps generated using CummeRbund for all 106 sorted genes (A) and fifty genes (B) sorted by Cuffdiff showing differential expression pattern in Carlton basal plate and callus.

5.3.7.4 Genes/ transcripts of interest

The top ten genes (Cuffdiff sorted) according to total FPKM values are listed in Table 5.9 while top ten genes based on fold change are shown in Table 5.10.

Table 5.9 Top ten sorted genes (Cuffdiff) based on total FPKM value.

Gene ID	Gene name	UniProtKB ID	Direction		Total FPKM	q-value	Gene descriptions
			CBS	Callus			
XLOC_001491	PMAA_047660	B6QP93	up	down	15602	0.043	Cytosolic regulator Pianissimo, putative
XLOC_001083	PBANKA_102650	A0A077YGT0	up	down	12469.37	0.028	Metacaspase-like protein
XLOC_001339	LTL	G9M5T0	down	up	9294.78	0.025	Lectin
XLOC_000028	GLYMA_13G0171005	A0A0R0GH81	down	up	9206.69	0.015	Uncharacterized protein
XLOC_001481	GLYMA_13G013900	A0A0R0GR39	down	up	9148.51	0.042	Uncharacterized protein
XLOC_000916	Ccrd_022258	A0A118HAI7	up	down	9136.02	0.008	Non-heme dioxygenase
XLOC_002004	OB10G22000	J3N3V2	down	up	7080.37	0.003	Uncharacterized protein
XLOC_001572	MT2a	I3QP00	up	down	5606.7	0.022	Metallothionein type
XLOC_001932	G7K_1412-t1	A0A0E9NBE8	down	up	5035.22	0.008	Uncharacterized protein
XLOC_001997	PR1	Q05736	down	up	4652.44	0.001	Pathogenesis-related protein

CBS = Carlton basal plate.

Table 5.10 Top ten sorted genes (Cuffdiff) based on fold change.

Gene ID	Gene name	UniProtKB ID	Direction		fold change	q-value	Gene descriptions
			CBS	Callus			
XLOC_001997	PR1	Q05736	down	up	49.09	0.001	Pathogenesis-related protein
XLOC_001792	N/A	D6MK51	up	down	41	0.002	Transcription factor
XLOC_001339	LTL	G9M5T0	down	up	25.58	0.025	Lectin
XLOC_001384	6G-FFT	A0A125SXW6	up	down	21.14	0.009	Fructan:fructan 6G-fructosyltransferase
XLOC_002132	clkA	F4PV79	up	down	18.85	0.003	Putative uncharacterized protein clkA
XLOC_000830	ARP	T2C8V8	up	down	18.33	0.027	Auxin repressed protein
XLOC_001888	LjHb1	Q3C1F4	down	up	16.78	0.014	Nonsymbiotic hemoglobin
XLOC_001491	PMAA_047660	B6QP93	up	down	16.24	0.042	Cytosolic regulator Pianissimo, putative
XLOC_001572	MT2a	I3QP00	up	down	15.09	0.022	Metallothionein type 2a-FL
XLOC_000673	ASR1	Q08655	up	down	13.74	0.0006	Abscisic stress ripening

CBS = Carlton basal plate.

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 (Appendix 5.1).

5.3.7.4.1 Genes/ 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 differentially expressed genes based on their BLAST annotations were grouped into the five major categories phenylpropanoid pathway related enzymes (Cytochrome P450s), methyltransferases, NADP/ NADPH dependent reductases, oxidoreductases and hydrolases (Table 5.11). Other genes related to secondary metabolite accumulation such as phenolic compounds, lectins, polypeptides and others are also listed in Table 5.11.

Table 5.11 Probable genes related to secondary metabolite production detected in *Narcissus* transcriptome data (only significant genes with q-value ≤ 0.05).

Gene ID	UniProt ID	Regulation	log2 (fold change)	q-value	Gene function
1. Phenylpropanoid pathway related					
XLOC_001204	O04892A0A124	CBS-UP	-	0.007	cytochrome P450
XLOC_002108	SCB7	CBS-UP	-	0.007	cytochrome P450
2. Methyltransferases					
XLOC_001033	B9SGP1	CBS-UP	-	0.006	*O*-methyltransferase
XLOC_001843	K4CXY9	CBS-UP	3.05	0.002	*O*-methyltransferase
XLOC_000957	Q5DNB1	CBS-UP	1.78	0.013	*S*-adenosylmethionine synthase
XLOC_000735	F6H5H8	CBS-UP	-	0.006	*O*-methyltransferase
XLOC_002003	A0A0P0YCT2	CAL-UP	2.01	0.011	S-methyltransferase
XLOC_000918	A0A0V0I0R3	CAL-UP	-	0.0006	hydroquinone methyltransferase
3. NADP/ NADPH related					
XLOC_000518	B9T8W8	CBS-UP	2.11	0.004	NADH dehydrogenase
XLOC_001407	Q9SLN8	CBS-UP	2.18	0.006	(NADP(+) reductase
XLOC_001713	E5G6F3	CAL-UP	3.05	0.049	NAD/ NADP binding
4. Oxidoreductase activity related					
XLOC_001731	Q9AV39	CBS-UP	2.09	0.045	Os10g0545200 protein
XLOC_001235	T2DPZ9	CBS-UP	-	0.0006	cytochrome b
XLOC_001407	M0SZ88	CBS-UP	2.18	0.045	uncharacterized
XLOC_001110	M8AU02	CAL-UP	3.21	0.033	aldehyde dehydrogenase
XLOC_000371	M0TCR0	CAL-UP	-	0.0006	uncharacterized
XLOC_002051	F2NYJ0	CAL-UP	-	0.0006	anaerobic reductase
XLOC_002072	O24428	CAL-UP	-	0.0006	stearoyl-ACP desaturase
5. Hydrolase activity related					
XLOC_000089	A0A078EUD2	CBS-UP	2.93	0.023	BnaAnng00280D
XLOC_000792	K4NZ15	CBS-UP	-	0.0006	lipase
XLOC_001578	K4NZ15	CBS-UP	-	0.0006	uncharacterized
XLOC_000006	Q9LLC2	CAL-UP	2.88	0.011	xyloglucan endo-transglucosylase (xe)
XLOC_000580	M0TI33	CAL-UP	-	0.0006	xe
XLOC_001722	A5BND5	CAL-UP	-	0.0006	xe
XLOC_001883	F6GXE7	CAL-UP	-	0.0006	xe

Others					
XLOC_000536	K4P0T2	CBS-UP	-	0.0007	lectin
XLOC_000616	W5VXS2	CBS-UP	-	0.0007	polyphenol oxidase
XLOC_001502	G8XUP0	CBS-UP	-	0.0007	lectin
XLOC_001926	C9W8B3	CBS-UP	-	0.0007	lectin
XLOC_002069	Q40422	CBS-UP	-	0.0007	mannose specific lectin
XLOC_000485	G3GC08	CBS-UP	6	0.0038	lipoxygenase
XLOC_000278	A0A0B0MIR6	CBS-UP	2.64	0.0212	proactivator polypeptide
XLOC_001282	K7P8F2	CBS-UP	2.58	0.0383	aspartic acid protease
XLOC_001804	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	CAL-UP	-	0.0034	3-hydroxy-3-methylglutaryl-coenzyme A reductase
XLOC_000638	J5JKX7	CAL-UP	-	0.0007	PAP2 superfamily
XLOC_001712	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.

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 Appendix 5.2.

5.3.7.4.2 Expression levels of other 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 defence related protein and ATP synthase were detected in *Narcissus* transcriptome data, which are shown in Figure 5.14.

The stress related transcripts were equally expressed in both tissues. Transcripts for ATP binding proteins were more highly expressed in callus than basal plate. Those for heat-shock proteins, which are responsible also for stress factors and callus formation, were detected in callus only. These findings are in accordance with the results obtained from NMR (Chapter 4), since the stress related

metabolites (i.e. proline, choline and betaine) were present in relatively higher concentrations in callus than *Narcissus* field-grown bulb or basal plate.

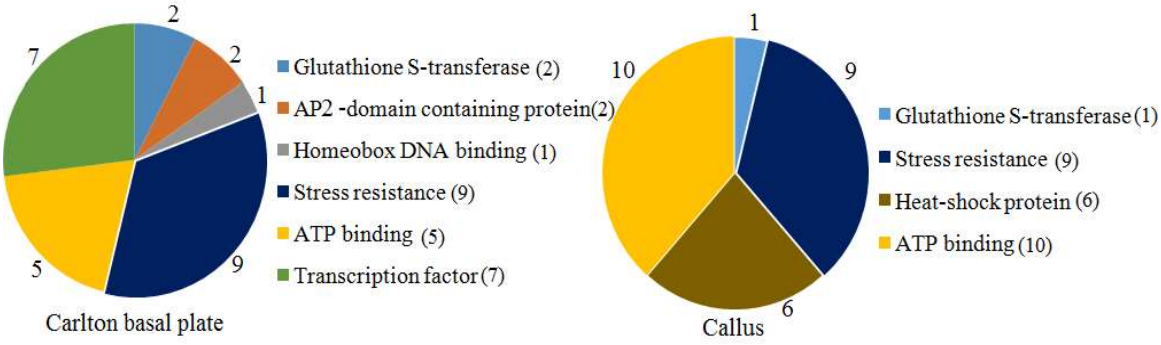

Fig. 5.14 Genes identified as differentially expressed in basal plate and callus related to tissue growth and development (only significant genes with q-values of ≤ 0.05). Numbers on chart are the number of respective protein groups found differentially expressed.

Conversely, transcripts of genes responsible for differentiation such as shoot formation, AP2-domain containing, glutathione *S*-transferases and homeobox DNA binding proteins were only detected in basal plate. All the significant (q-value ≤ 0.05) other transcription factor related genes were detected in basal plate only (Figure 5.15).

Figure 5.15 shows an overview of all other gene groups besides secondary metabolism related genes, which were differentially expressed (both significant and non-significant, q-value ≤ 1.0) in basal plate and callus, with details, listed in Appendix 5.3.

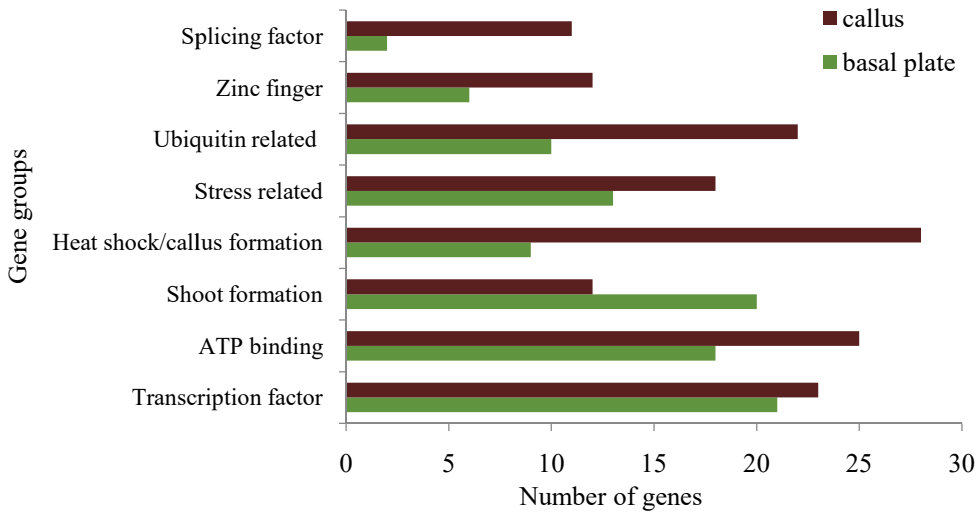

Fig. 5.15 The abundance (total number) of important gene groups detected as differentially expressed in basal plate and callus with q-values ≤ 1.0 (Appendix 5.3).

5.3.8 Functional annotation of *Narcissus* transcripts

5.3.8.1 Functional categorisation of the *Narcissus* basal plate and callus transcripts

The Gene Ontology (GO) via UniProt resulted in the annotation of 100 transcripts (about 16% of total detected transcripts in basal plate) out of 111 transcripts that were detected as significantly (q-value ≤ 0.05) up regulated in basal plate. Some transcripts were assigned to more than one GO category. The results (Figure 5.16) showed 55 transcripts being assigned to molecular function with the top two categories being binding and catalytic activity. A total of 35 transcripts were assigned to cellular component with the top three categories being cell part, organelle and membrane/membrane part. The biological process category was assigned to a total of 49 transcripts showing metabolic process, cellular process and single-organism process as the top three categories.

The functional annotation of 90 (about 8% of total transcripts detected in callus) transcripts out of the 95 significantly (q-value ≤ 0.05) up regulated transcripts detected in callus, were annotated via UniProt GO analysis (Figure 5.17). Figure 5.16 shows that in callus, 62 transcripts were assigned to molecular function with catalytic activity and binding as the top two categories. Forty 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 top three categories being the cellular process, metabolic process and single-organism process.

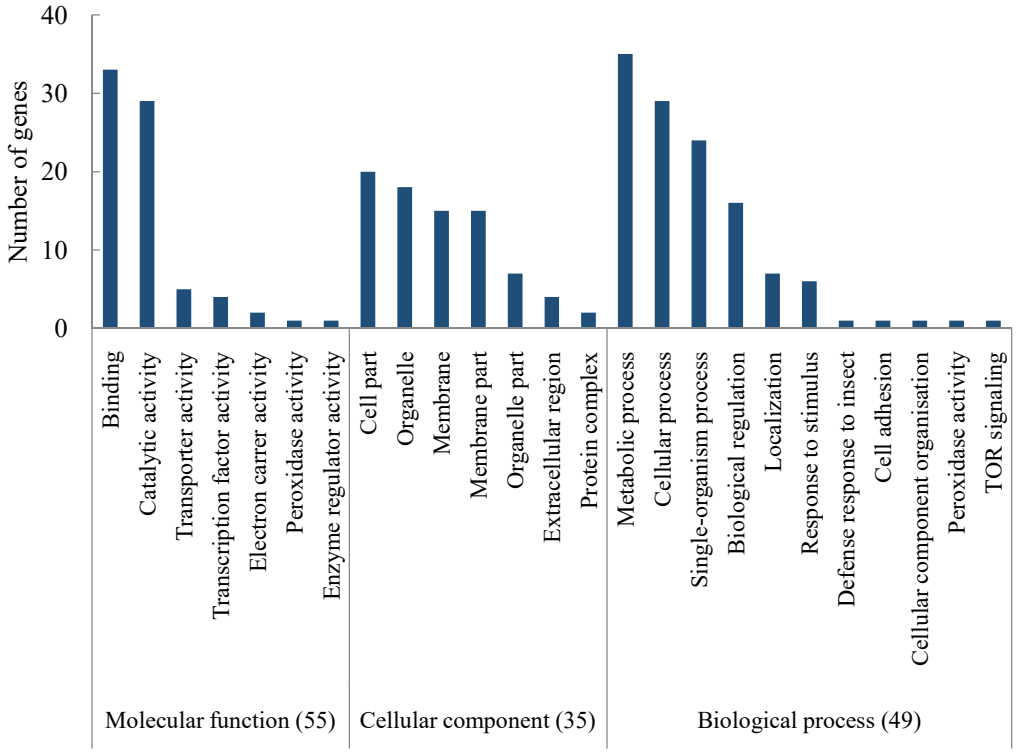

Fig. 5.16 The UniProt Gene Ontology (GO) analysis of significantly up-regulated transcripts detected in basal plate. The numbers in bracket represent the total transcripts assigned for the respective GO categories. Several transcripts were assigned to more than one category.

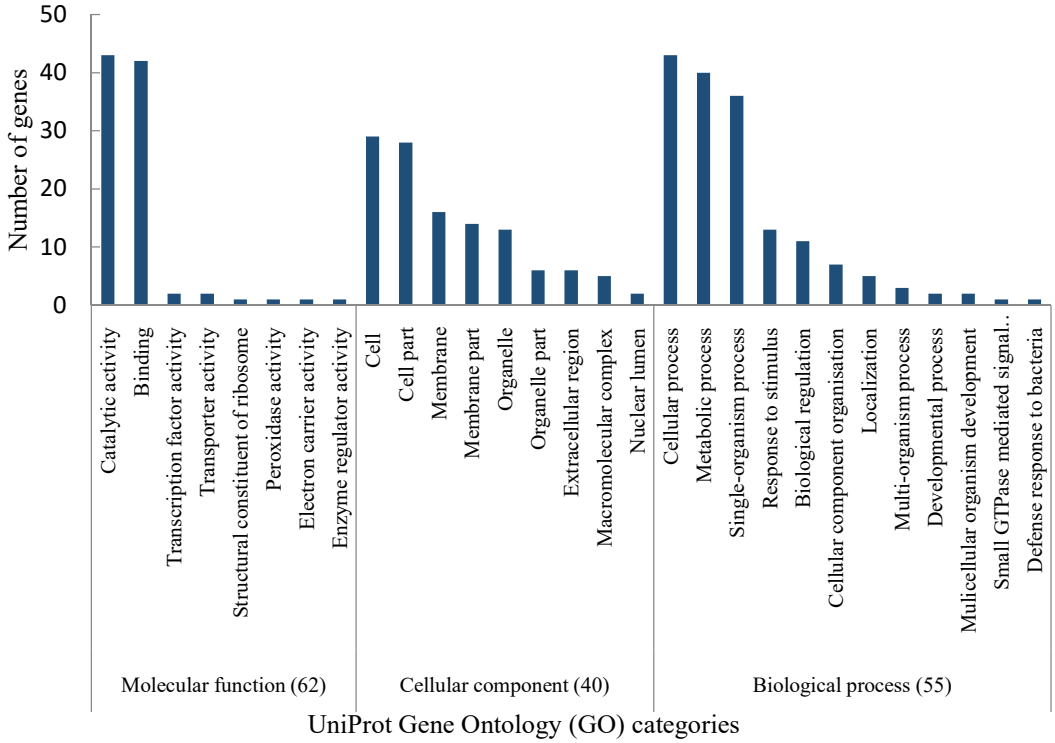

Fig. 5.17 The UniProt Gene Ontology (GO) analysis of significantly up-regulated transcripts detected in callus. The numbers in bracket represent the total transcripts assigned for the respective GO categories. Several transcripts were assigned to more than one category.

5.3.8.2 The assignment of GO terms to *Narcissus* transcripts (basal plate and callus) using Quick GO-Beta

The EMBL-EBI Quick GO-Beta annotation (www.ebi.ac.uk/QuickGO-Beta/annotations) was used for further annotation to assign the detected transcripts to complement GO functionality predictions with GO terms. This resulted in 365 GO terms for 111 basal plate transcripts and 618 GO terms for 95 callus transcripts, which were significantly differentially expressed in both tissues, substantially more than from UniProt. As shown in Figure 5.18, for 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. For the 618 GO terms annotated to transcripts from callus, molecular function category was assigned to 275 (44.5%), 196 (31.72%) for biological process and 147 (23.79%) were assigned for cellular component. The proportions assigned to each category were therefore essentially the same for both tissues but with several terms to each transcript.

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 (Figure 5.18, Appendix 5.4). 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), transferase activity (GO:0016740) were the top assigned molecular functions related GO terms detected in callus (Figure 5.18, Appendix 5.4).

A full list of all annotated GO terms (Quick GO-Beta) detected in basal plate and callus with their corresponding gene descriptions are in Appendix 5.4.

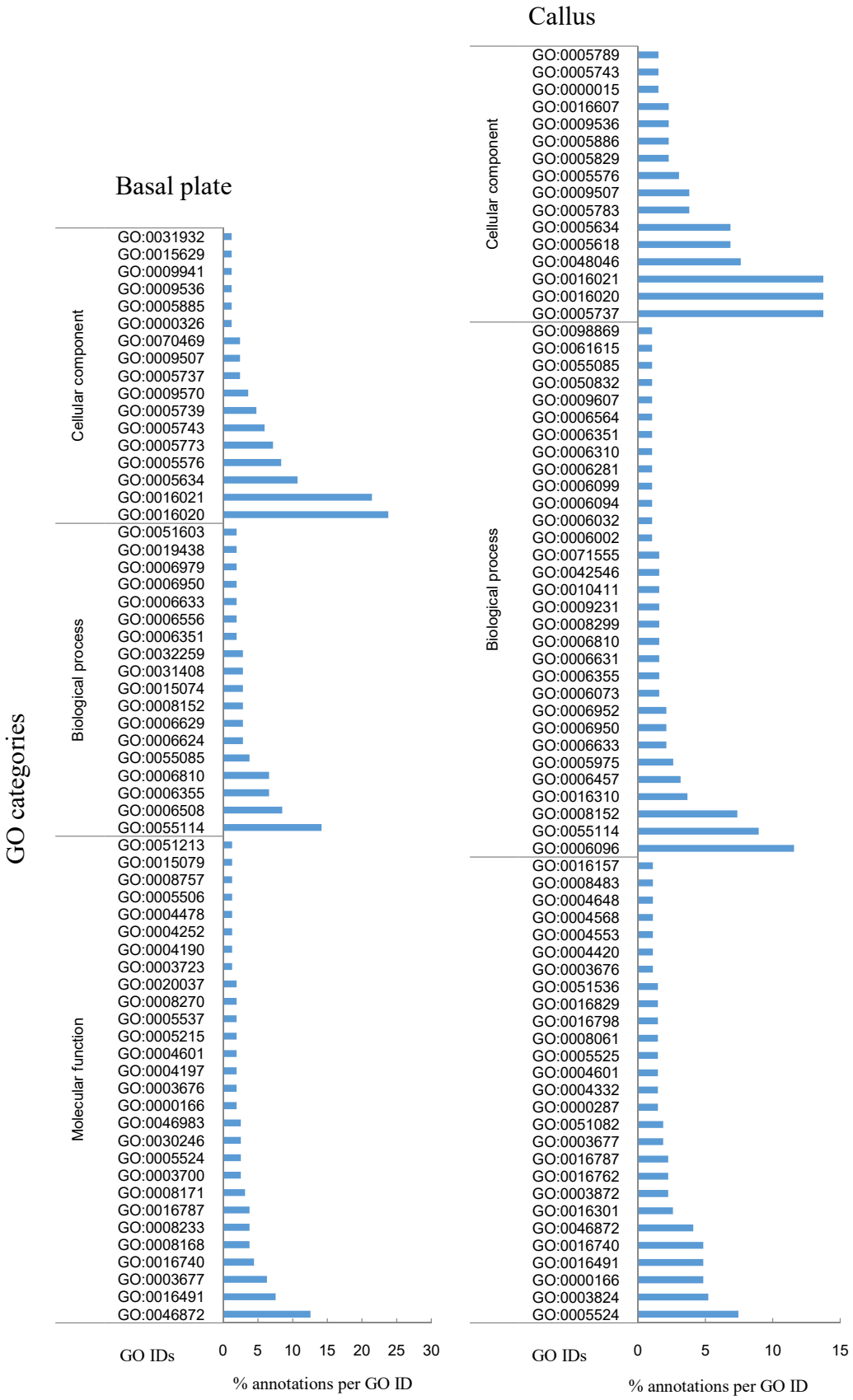

Fig. 5.18 An overview of GO assignment for basal plate and callus GO IDs annotated against significant up regulated transcripts at q-value ≤ 0.05; using EMBL-EBI Quick GO-Beta annotations. GO terms contributing less than 1% are not included.

5.3.9 Pathway analyses of *Narcissus* transcripts

5.3.9.1 Pathway annotation of *Narcissus* transcripts (KEGG)

The UniProt IDs of up regulated 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 (Appendix 5.5) for basal plate and 78 KEGG pathways for callus (Appendix 5.6).

Important pathways based on maximum number of mapped hits (in brackets) are shown in Table 5.12 (basal plate) and 5.13 (callus). From Table 5.12 it was observed 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.

Table 5.12 Summary of KEGG pathway mapping results of *Narcissus* basal plate transcripts.

KO IDS	Descriptions	Enzyme class (EC)
1. Metabolic pathways (11)		
ko:K00001	alcohol dehydrogenase	[EC:1.1.1.1]
ko:K00517	Oxidoreductases	[EC:E1.14.-.-]
ko:K00789	metK; S-adenosylmethionine synthetase	[EC:2.5.1.6]
ko:K00818	acetylornithine aminotransferase	[EC:2.6.1.11]
ko:K01611	S-adenosylmethionine decarboxylase	[EC:4.1.1.50]
ko:K03878	ND1; NADH-ubiquinone oxidoreductase chain 1	[EC:1.6.5.3]
ko:K07513	ACAA1; acetyl-CoA acyltransferase 1	[EC:2.3.1.16]
ko:K09753	CCR; cinnamoyl-CoA reductase	[EC:1.2.1.44]
ko:K09754	CYP98A; coumaroylquinate(coumaroylshikimate) 3' monooxygenase	EC:1.14.13.36]
ko:K13066	caffeic acid 3-O-methyltransferase	[EC:2.1.1.68]
ko:K14423	SMO1; 4,4-dimethyl-9beta,19-cyclopropylsterol-4alpha-methyl oxidase	[EC:1.14.13.72]
2.Biosynthesis of secondary metabolites (10)		
ko:K00001	alcohol dehydrogenase	[EC:1.1.1.1]
ko:K00517	Oxidoreductases	[EC:E1.14.-.-]
ko:K00789	metK; S-adenosylmethionine synthetase	[EC:2.5.1.6]
ko:K00818	acetylornithine aminotransferase	[EC:2.6.1.11]
ko:K07513	ACAA1; acetyl-CoA acyltransferase 1	[EC:2.3.1.16]
ko:K09753	CCR; cinnamoyl-CoA reductase	[EC:1.2.1.44]
ko:K09754	CYP98A; coumaroylquinate(coumaroylshikimate) 3'monooxygenase	[EC:1.14.13.36]

ko:K11818	CYP98A; coumaroylquinate(coumaroylshikimate) 3'monooxygenase	[EC:1.14.-.-]
ko:K13066	CYP98A; coumaroylquinate(coumaroylshikimate) 3'monooxygenase	[EC:2.1.1.68]
ko:K14423	CYP98A; coumaroylquinate(coumaroylshikimate) 3'monooxygenase	[EC:1.14.13.72]
3. Phenylpropanoid biosynthesis (3)		
ko:K09753	CCR; cinnamoyl-CoA reductase	[EC:1.2.1.44]
ko:K09754	CYP98A; coumaroylquinate(coumaroylshikimate) 3' monooxygen	[EC:1.14.13.36]
ko:K13066	caffeic acid 3-O-methyltransferase	[EC:2.1.1.68]
4. Biosynthesis of amino acids (2)-central metabolism		
ko:K00789	S-adenosylmethionine synthetase	[EC:2.5.1.6]
ko:K00818	acetylornithine aminotransferase	[EC:2.6.1.11]
5. Tryptophan metabolism (2)-central metabolism		
ko:K11818 ko:K14338	CYP83B1; cytochrome P450, family 83, subfamily B, polypeptide 1 cytochrome P450 / NADPH-cytochrome P450 reductase	[EC:1.14.-.-] [EC:1.14.14.1 1.6.2.4]
6. Tyrosine metabolism (1)-central metabolism		
ko:K00001	alcohol dehydrogenase	[EC:1.1.1.1]

Numbers in brackets represent the number of hits correspond to the pathways.

Like the basal plate issues, the KEGG pathways mapped in callus were (Table 5.13) also related to secondary metabolism. This showed a higher number of hits (13) (Table 5.13) for biosynthesis of secondary metabolites than basal plate (10) (Table 5.12). Moreover, carbon metabolism and glycolysis related pathways were detected in callus that were not found in basal plate. This may be related to the more active heterotrophic metabolism of the callus tissues. Pathways related to aromatic amino acid metabolism were present in both basal plate and callus (Table 5.12 and 5.13).

Table 5.13 Summary of KEGG pathway mapping results of *Narcissus* callus transcripts.

KO IDs	Descriptions	Enzyme Class (EC)
1. Metabolic pathways (14)		
ko:K00021 ko:K00264	HMGCR; hydroxymethylglutaryl-CoA reductase (NADPH) GLT1; glutamate synthase (NADPH/NADH)	[EC:1.1.1.34] [EC:1.4.1.13 1.4.1.14]
ko:K00517 ko:K00549	Oxidoreductases metE; 5-methyltetrahydropteroyltriglutamate--homocysteine methyltransferase	[EC:E1.14.-.-] [EC:2.1.1.14]
ko:K00831	serC; phosphoserine aminotransferase	[EC:2.6.1.52]
ko:K00889	PIP5K; 1-phosphatidylinositol-4-phosphate 5-kinase	[EC:2.7.1.68]
ko:K01623	ALDO; fructose-bisphosphate aldolase, class I	[EC:4.1.2.13]
ko:K01803	TPI; triosephosphate isomerase (TIM)	[EC:5.3.1.1]
ko:K01904	4CL; 4-coumarate--CoA ligase	[EC:6.2.1.12]
ko:K07513	ACAA1; acetyl-CoA acyltransferase 1	[EC:2.3.1.16]

ko:K09753	CCR; cinnamoyl-CoA reductase	[EC:1.2.1.44]
ko:K12502	VTE3; MPBQ/MSBQ methyltransferase	[EC:2.1.1.295]
ko:K14454	GOT1; aspartate aminotransferase, cytoplasmic	[EC:2.6.1.1]
ko:K15633	gpmI; 2,3-bisphosphoglycerate-independent phosphoglycerate mutase	[EC:5.4.2.12]
2. Biosynthesis of secondary metabolites (13)		
ko:K00021 ko:K00264	HMGCR; hydroxymethylglutaryl-CoA reductase (NADPH) GLT1; glutamate synthase (NADPH/NADH)	[EC:1.1.1.34] [EC:1.4.1.13 1.4.1.14]
ko:K00517 ko:K00549	Oxidoreductases metE; 5-methyltetrahydropteroyltriglutamate--homocysteine methyltransferase	[EC:E1.14.-.-] [EC:2.1.1.14]
ko:K01623	ALDO; fructose-bisphosphate aldolase, class I	[EC:4.1.2.13]
ko:K01803	TPI; triosephosphate isomerase (TIM)	[EC:5.3.1.1]
ko:K01904	4CL; 4-coumarate--CoA ligase	[EC:6.2.1.12]
ko:K07513	ACAA1; acetyl-CoA acyltransferase 1	[EC:2.3.1.16]
ko:K09753	CCR; cinnamoyl-CoA reductase	[EC:1.2.1.44]
ko:K12502	VTE3; MPBQ/MSBQ methyltransferase	[EC:2.1.1.295]
ko:K14454	GOT1; aspartate aminotransferase, cytoplasmic	[EC:2.6.1.1]
ko:K15397 ko:K15633	KCS; 3-ketoacyl-CoA synthase gpmI; 2,3-bisphosphoglycerate-independent phosphoglycerate mutase	[EC:2.3.1.199] [EC:5.4.2.12]
3. Biosynthesis of amino acids (7) -central metabolism		
ko:K00264	GLT1; glutamate synthase (NADPH/NADH)	[EC:1.4.1.13 1.4.1.14]
ko:K00549	metE; 5-methyltetrahydropteroyltriglutamate--homocysteine methyltransferase	[EC:2.1.1.14]
ko:K00831	serC; phosphoserine aminotransferase	[EC:2.6.1.52]
ko:K01623	ALDO; fructose-bisphosphate aldolase, class I	[EC:4.1.2.13]
ko:K01803	TPI; triosephosphate isomerase (TIM)	[EC:5.3.1.1]
ko:K14454 ko:K15633	GOT1; aspartate aminotransferase, cytoplasmic gpmI; 2,3-bisphosphoglycerate-independent phosphoglycerate mutase	[EC:2.6.1.1] [EC:5.4.2.12]
4. Carbon metabolism (5)		
ko:K00831	serC; phosphoserine aminotransferase	[EC:2.6.1.52]
ko:K01623	ALDO; fructose-bisphosphate aldolase, class I	[EC:4.1.2.13]
ko:K01803	TPI; triosephosphate isomerase (TIM)	[EC:5.3.1.1]
ko:K14454 ko:K15633	GOT1; aspartate aminotransferase, cytoplasmic gpmI; 2,3-bisphosphoglycerate-independent phosphoglycerate mutase	[EC:2.6.1.1] [EC:5.4.2.12]
5. Glycolysis / Gluconeogenesis (3) -central metabolism		
ko:K01623	ALDO; fructose-bisphosphate aldolase, class I	[EC:4.1.2.13]
ko:K01803 ko:K15633	TPI; triosephosphate isomerase (TIM) gpmI; 2,3-bisphosphoglycerate-independent phosphoglycerate mutase	[EC:5.3.1.1] [EC:5.4.2.12]
6. Phenylpropanoid biosynthesis (2)		
ko:K01904	4CL; 4-coumarate--CoA ligase	[EC:6.2.1.12]
ko:K09753	CCR; cinnamoyl-CoA reductase	[EC:1.2.1.44]
7. Phenylalanine metabolism (2) -central metabolism		
ko:K01904	4CL; 4-coumarate--CoA ligase	[EC:6.2.1.12]

ko:K14454	GOT1; aspartate aminotransferase, cytoplasmic	[EC:2.6.1.1]
8. Tryptophan metabolism (1) -central metabolism		
ko:K14338	cypD_E; cytochrome P450 / NADPH-cytochrome P450 reductase	[EC:1.14.14.1 1.6.2.4]
9. Phenylalanine, tyrosine and tryptophan biosynthesis (1) -central metabolism		
ko:K14454	GOT1; aspartate aminotransferase, cytoplasmic	[EC:2.6.1.1]
10. Tyrosine metabolism (1)-central metabolism		
ko:K14454	GOT1; aspartate aminotransferase, cytoplasmic	[EC:2.6.1.1]

Numbers in brackets represent the number of hits correspond to the pathways.

5.3.9.2 Pathway analyses of *Narcissus* transcripts (Plant Reactome)

The Plant Reactome database (plantreactome.gramene.org) for plant metabolomic and regulatory pathways is a freely accessible, curated and peer reviewed pathway database. The UniProt IDs for genes of interest, gene identifier lists, microarray data and metabolomics data can be used as input to map to a curated *Oryza sativa* pathway database. This provides an alternative way to map metabolic pathways in our data.

The KO IDs obtained from UniProtKB Retrieve/ID mapping tool as mentioned in section 5.3.9.1, were mapped back to UniProtKB IDs for both basal plate and callus. The result yielded 45 UniProt genes corresponding to *Oryza sativa* for basal plate and 120 UniProt genes for callus. These UniProt genes for basal plate and callus were separately mapped to the Plant Reactome *Oryza sativa* pathway database.

The Plant Reactome pathway analyses resulted in a total of 14 pathways for basal plate and 26 pathways for callus. The pathways related to secondary metabolism (according to database) and central metabolism are represented in Table 5.14. The details of all detected pathways are in Appendix 5.7.

The postulated pathway (based on KEGG and Plant Reactome results) (Figure 5.19) of secondary metabolite biosynthesis in *Narcissus* 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 and pathways related to amino acid (phenylalanine, tyrosine and tryptophan) metabolism.

Table 5.14 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).

Carlton basal plate						
Pathway name	**Entities found**	**Entities total**	**Entities p-value**	**Entities FDR**	**Mapped entities**	***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.

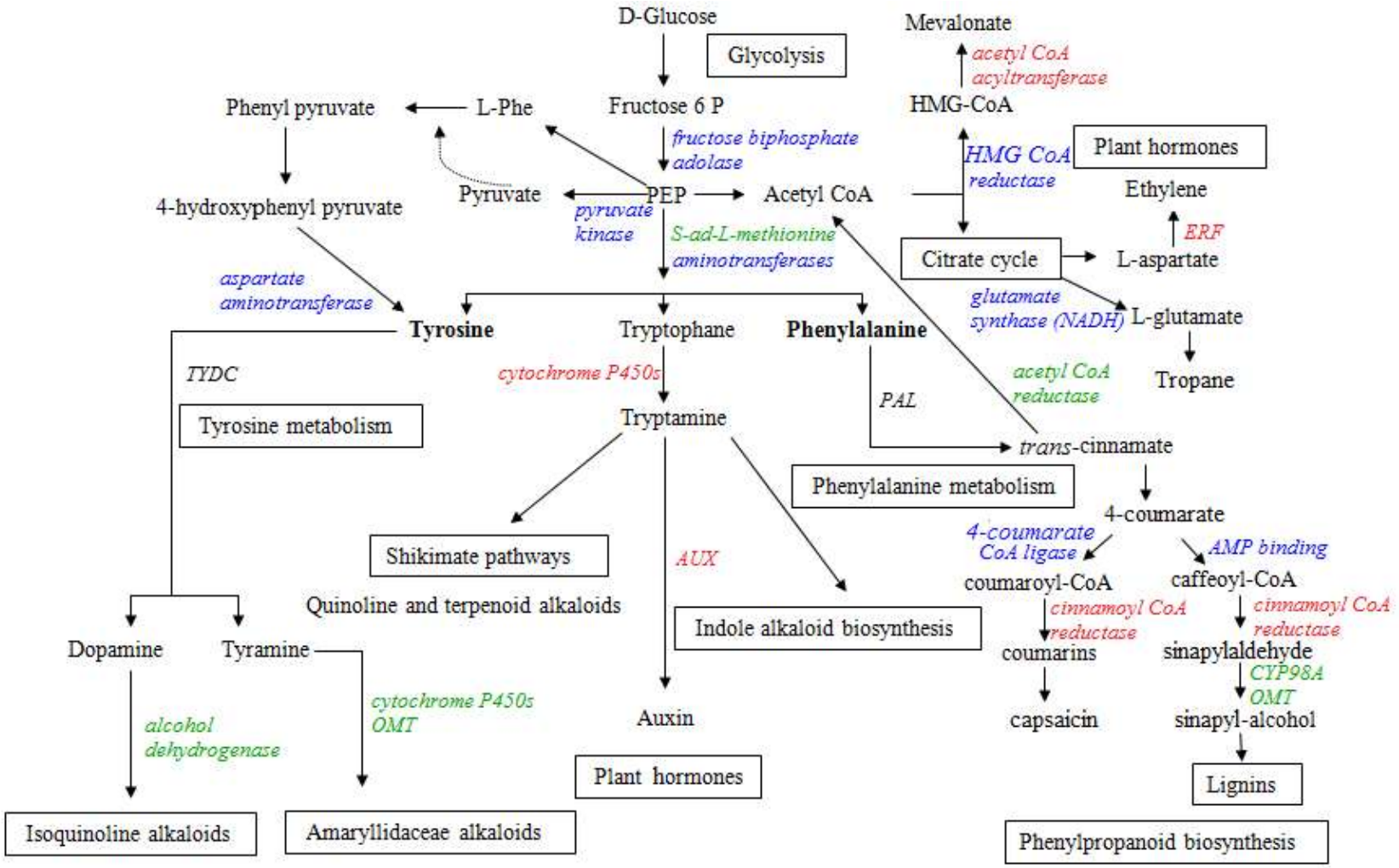

Fig. 5.19 Simplified schematic overview of the biosynthesis of main plant secondary metabolites in *N. pseudonarcissus* cv. Carlton. The postulated transcripts (italics), are labelled in different colours; callus (blue), basal plate (green) and both callus and basal plate (red). HMG= hydroxymethylglutarate; ERF= Ethylene response factor; AUX=Auxin related enzyme. The schematic was developed from KEGG and Plant Reactome pathway analyses.

5.4 Discussion

5.4.1 RNA extraction

The isolation of RNA from plants with high level of sugars, phenols, and secondary metabolites such as *N. pseudonarcissus* cv. Carlton is a challenging process. The use of the traditional CTAB method yielded very low quantity (20- 55 ng/µl) and quality (A_{260}/A_{280}: 1.80 and A_{260}/A_{230}: 0.22) RNA, indicating a high level of contamination due to DNA, protein, phenols and other contaminants in the extracted RNA samples (www.nanodrop.com).

Using the innuPREP Plant RNA Kit with the PL buffer resulted in satisfactory amounts of RNA (~100-800 ng/µl) of acceptable quality (A_{260}/A_{230}: 2.00-2.42) from all samples under study. These were considered suitable for further analyses using RT-PCR and also for library preparation for RNA sequencing.

5.4.2 Relative expression analysis (RT-PCR)

The relative expression levels of putative Amaryllidaceae alkaloid biosynthesis related transcripts such as cytochrome P450s, *PAL*, *TYDC* and *OMT* were tested in several *Narcissus* tissues using RT-PCR. The results (section 5.3.2) showed that more or less all transcripts were expressed in field and *in vitro* tissues although to different levels.

It was already known (Chapter 3) that the amount of galanthamine was either very low or absent in *in vitro* grown callus and direct bulblets. So, the relative expression levels of putative transcripts related to galanthamine accumulation observed in section 5.3.2 suggest the expression of putative genes in *in vitro* tissues along with field tissues, which is really intriguing. Hence, a deep transcriptome analysis of *Narcissus* basal plate (highest galanthamine producing tissue) and callus (lowest galanthamine producing tissue) was undertaken to obtain further knowledge about the differential genes/ transcripts expression levels in field (i.e. basal plate) versus *in vitro* tissues (i.e. callus).

The RT-PCR results might affect due to the technical errors such as using of 1µl cDNA template, which could result in pipetting error. The cDNA sample dilution and use of 5µl template could be reconsidered to minimise the pipetting error.

5.4.3 Transcriptome analysis

5.4.3.1 RNA quality and integrity

RNA samples with good quality and the required concentration were achieved using innuPREP Plant RNA Kit, with the addition of DNAse treatment (section 5.2.2.1.2) and sent for sequencing to the CGR. All samples fulfilled the initial quality requirements (RIN > 7; clear 18S and 28S peaks and 28S/ 18S ratios 1.5-2.0) on Bioanalyser analyses by the CGR. The Agilent bioanalyser is an extensively used method for RNA quality check, which measures a number of RNA integrity metrics to generate an RNA integrity number (RIN) (Ward *et al.,* 2012). It has been reported that a RIN > 8 indicates high quality but RIN > 7 is also considered as sufficient (Ward *et al.,* 2012; Johnson *et al.,* 2012).

The RIN value alone is not enough to assess RNA quality (Ward *et al.,* 2012). The ratio of 28S and 18S ribosomal RNA molecules is also an important factor. The degradation of the larger ribosomal rRNA (26S or 28S) occurs faster than the smaller (18S), therefore a 28S/ 18S or 26S/ 18S ratio below the accepted ratio (≥ 1) is considered as a measure of RNA degradation (Johnson *et al.,* 2012).

5.4.3.2 rRNA depletion and library preparation

The total RNA within the cell consists more than 80% of ribosomal RNA, which is necessary to remove before library preparation for deep sequencing to avoid wasting sequence reads (Benes *et al.,* 2011). Our results showed successful rRNA depletion in all samples under study, which was indicated by the absence of 28S and 18S peaks and 18S/18S ratio of 0.0 (Benes *et al.,* 2011) after using Ribo-Zero rRNA removal kit (www.illumina.com).

In total six libraries were generated and sequencing was performed for all six samples by CGR. One replicate of basal plate (CBS2) samples 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. This indicates that the measurements of RNA quality unfortunately cannot always predict successful sequencing. Repeating the sequencing of one sample was not possible due to the expense and limitation of CGR sequencing protocol set-up. Therefore, further

analysis was continued including CBS2 as a replicate of Carlton basal plate, which could be a weakness for further differential expression analysis.

5.4.3.3 RNA-seq data analysis

5.4.3.3.1 Sequence mapping

The *Narcissus* basal plate and callus transcriptome sequences were mapped to *Narcissus spp*. transcriptome data published on MedPlant RNA Seq Database (www.medplantrnaseq.org). A mapping strategy was chosen instead of *de novo* assembly as it has been reported more applicable for differential expression analysis (Vijay *et al.,* 2013), which was the main aim of this project. However, a *de novo* assembly strategy has also been reported as successful in *N. pseudonarcissus* transcriptome analysis to identify transcripts involved in the alkaloid biosynthesis (Kilgore *et al.,* 2014, Kilgore *et al.,* 2016a). Differential gene expression analysis for the *Narcissus* transcriptome data was performed on Discovery Environment (Oliver *et al.,* 2013) and Atmosphere (Skidmore *et al.,* 2011) of a well documented cyber-infrastructure CyVerse which was previously known as iPlant Collaborative (Goff *et al.,* 2011).

Initially, the *Arabidopsis* genome (from CyVerse database) and *Galanthus sp*. transcriptome (from MedPlant RNA Seq database) data were used to map *Narcissus* transcriptome data but statistical analyses for differential transcript expression were discarded due to very poor mapping or read alignment (less than 10% of total input, section 5.3.7.1).

The sequence alignment of basal plate and callus transcripts to MedPlant *Narcissus* transcriptome using TopHat showed overall mapping coverage ranging from around 35% to 48% of total input sequence. The only exception was the CBS2, basal plate replicate with an overall alignment rate of about 19%. The mapping coverage can be considered as moderate. Typically, read alignment rate should be 70% to obtain an accurate 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).

Therefore, the mapping rate obtained from our results might be due to the lack of complete annotated transcriptome or poor read quality of CBS2. 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 rates ranging from 17% to 94% using TopHat has been reported previously (Trapnell *et al.,* 2009).

As more than 15% mapping coverage was obtained from CBS2 (19%) and moderate coverage in the other samples (35% to 48%), it was assumed to be an acceptable and reasonable differential transcript expression result from our *Narcissus* transcriptome data and further analysis was worthwhile.

5.4.3.3.2 Differential expression analysis

The aligned reads after assembly resulted in a total of 2153 Cuff genes, although CBS2 showed divergent assembly results from the other samples and was obviously an outlier. Cufflinks 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 to study differential expressed and regulated genes and sorted genes (Trapnell *et al.,* 2012).

The Cuffdiff detected 2153 differentially expressed genes/ transcripts among which 64.21% genes/ transcripts were up regulated in callus and 35.21% were up regulated in basal plate. Moreover, only 206 genes/ transcripts were detected as significantly differentially expressed (q-values ≤ 0.05) in the two conditions under study; 95 genes were up regulated in callus and 111 genes were up regulated in basal plate.

5.4.3.4 Genes of interest

About 79% (1700) of Cuffdiff listed differentially expressed genes, were annotated using a BLAST search against known databases. Eleven abundant gene groups were

detected within these annotated differentially expressed genes, comprising about 42% of the annotated genes. Most significant were uncharacterised 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%). 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 defence/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 5.11 and 5.14; Table 5.11).

5.4.3.4.1 Basal plate transcripts

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/ alkaloid biosynthesis; such as cytochrome *P450s*, *OMTs*, *NADP/NADPH* dehydrogenases/ reductases and *S*-adenosylmethionine synthetases/ decarboxylases and lectins. Other important proteins were also found up regulated in basal plate with q-values higher than 0.05 such as 3-ketoacyl-CoA, acyl-CoA, cinnamoyl-CoA, cinnamate 4-hydroxylase, alcohol dehydrogenase and caffeic acid *O*-methyltransferase. All these are involved in plant secondary metabolite production (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*) catalysing 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 catalysed by *SAM*-dependant methyltransferases (Liscombe *et al.,* 2012). Other probable enzymes related to alkaloid biosynthesis in Amaryllidaceae were also detected in this study 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).

5.4.3.4.2 Callus transcripts

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, hydroxymethylglutarate 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.

In our study, the methyltransferases detected in callus were either probable *N*-methyltransferases or *(SAM)* dependant *N*-methyltransferases. Similar *N*-methyltransferases 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 alkaloid related putative transcripts *OMT*, *NMT*, *SAM* and others 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*-transferase (Gong *et al.,* 2005), which was identified, in both basal plate and callus. A homeobox protein was recognised 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).

5.4.3.5 Functional categorisation of *Narcissus* transcripts using GO

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*. Therefore, the GO annotation results may contain both expected and unexpected annotations.

One hundred out of 111 basal plate transcripts (significant at q values ≤ 0.05) and 90 out of 95 callus transcripts (significant at q values ≤ 0.05) were annotated using the UniProt Gene Ontology tool. In both cases, most of the transcripts were assigned to molecular function, followed by biological process and cellular component. 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).

The most important part of GO analysis is the assignment of transcripts to their corresponding GO terms. A total 365 GO terms for basal plate transcripts and 618 GO terms for callus were obtained from Quick GO-Beta annotations (www.ebi.ac.uk/QuickGO-Beta/annotations). 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 that were not detected via UniProt GO. All these were detected in basal plate only.

5.4.3.6 Pathway mapping

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).

The basal plate and callus transcripts were mapped to KEGG mapper (search pathway) and provide 55 pathways for basal plate and 78 pathways for callus. In both tissues central metabolic pathways and biosynthesis of secondary metabolites were the top pathways with the highest number of assigned enzymes. A total of 13 enzymes were detected assigned to biosynthesis of secondary metabolites pathway in callus whereas 10 enzymes were found in basal plate.

Some human disease (e.g. Parkinson's and Huntington's disease) related pathways were also detected in both tissues. This 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. Identification of such unwanted pathways is therefore likely.

Therefore, it was worthwhile to compare this result with a plant-specific database, the Plant Reactome, which was also used to map the *Narcissus* transcripts for further comparison and confirmation of the pathways detected by KEGG. Plant Reactome is a recently developed plant pathway database based on curated *Oryza sativa* reference pathways (Tello-Ruiz *et al.,* 2015). The mapping results yielded 14 mapped pathways for basal plate and 26 pathways for callus. The most notable pathways were phenylpropanoid biosynthesis, amino acid metabolism, flavonoid biosynthesis and *S*-adenosyl-*L*-methionine cycle. It was observed that in most cases a higher number of enzymes corresponding to the pathways was detected in callus than basal plate, which is in accordance with the KEGG results.

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 5.18) generated in this study of the biosynthesis of secondary metabolites in *N. pseudonarcissus* cv. Carlton, based on KEGG and Plant Reactome

results, suggests that transcripts (cytochrome P450s and *OMTs*) possibly involved in alkaloid biosynthesis were up regulated in basal plate. Enzymes involved in the initial primary pathways leading to tyrosine, phenylalanine and tryptophan were up regulated in callus and transcripts leading to other alkaloids (isoquinoline, indole and quinoline) were basically detected in both tissues.

5.4.4 Field (basal plate) versus *in vitro* (callus) transcripts

The relative expression (RT-PCR) as well as differential transcripts expression (transcriptome) analyses showed the expression of putative genes involved in the alkaloid production in both field and *in vitro* tissues of *Narcissus*. However, it was observed that most of the probable transcripts involved in secondary metabolism were expressed in basal plate rather than callus (section 5.4.3.4.1 and 5.4.3.4.2).

The link between expression of pathways 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.

The mapping approaches taken in this project has allowed the identification of a series of genes that are differentially regulated between basal plate and callus. Several of these are members of gene families likely to be involved in alkaloid biosynthesis. However, to identify these more clearly and confidently will require further work, as described in the next chapter.

6 Chapter Six: Conclusions and future work

6.1 *N. pseudonarcissus* cv. Carlton tissue culture

The success of tissue culture in *Narcissus* is often difficult and challenging due to the high rate of contamination in culture condition and slow growth rate (Hanks, 2000; Sochacki and Orlikowska, 2005; Anbari *et al.,* 2007). However, this study showed a successful and reproducible method to obtain different tissue types with comparatively low contamination rate in *N. pseudonarcissus* cv. Carlton.

One of the crucial benefits of plant tissue culture is it provides different types of plant materials with different morphological, physiological and biological properties (Bhatia and Dahiya, 2015). In addition, it offers a completely aseptic and controlled growth environment without any effect of natural conditions; i.e. climate, natural contaminants such as insects and pathogens (Loyola-Vargas *et al.,* 2008). Tissues under *in vitro* culture usually exhibit a high degree of plasticity, which offers the initiation of different tissue types (Bhatia and Dahiya, 2015). This study represented the achievement of different organised, unorganised, photosynthetic and non-photosynthetic tissues from *in vitro* cultures of *Narcissus*. Plant cell growth and development processes are parallel to the environmental conditions (Bhatia and Dahiya, 2015). Comparing tissues originated from two growth environments (field versus *in vitro*) and different tissue types would provide a well-adapted knowledge for insight into the plant biology related to metabolite production in *Narcissus*, which was considered as the main aim of this study.

A suitable medium combined with appropriate concentrations and combinations of growth regulators is the basic requirement for a successful plant tissue culture (Kaya and Aki, 2013; Rezaei *et al.,* 2016). Current study showed the effect of different media combinations on callus induction, shoot and bulblet formation in *Narcissus*. It was found that the auxin had the crucial effect on callus induction (high auxin medium) (Taleb *et al.,* 2013) and shoot and bulblets initiation (no or low auxin media) (Santos *et al.,* 2002) with the supplement of low cytokinin (George *et al.,* 2008).

6.2 Alkaloid analysis (GC-MS)

The field-derived materials from *N. pseudonarcissus* cv. Carlton showed the presence of higher amount of galanthamine than tissue culture derived materials. That might be due to the longer transitional growth phase of naturally grown plants, which allows the plant to accumulate higher secondary metabolites than *in vitro* grown tissues (Dias *et al.,* 2016). Moreover, the organisation of tissues is often related with the accumulation of secondary metabolites (Karuppusamy, 2009).

This study showed organised tissues such as bulblets, white shoots and green shoots contained higher amounts of galanthamine than callus and regenerated shoots from callus. In addition, it was observed regenerated white shoots, which were produced through the modification of callus showed slightly higher galanthamine content than callus. Other Amaryllidaceae alkaloids have also been detected from field and *in vitro* cultured tissues, which would provide the knowledge about the alkaloid accumulation in different tissues derived from three phenol-coupling pathways involved in the biosynthetic pathway of Amaryllidaceae alkaloid production (Bastida *et al.,* 2006; Berkov *et al.,* 2014).

6.3 Metabolomic analysis

Metabolomic analysis using GC-MS requires the presence of authentic standards for specific metabolite identification (Singh and Chaturvedi, 2012). On the other hand, NMR-based metabolomic study provides information simultaneously about small molecules or metabolites present in tissues or organs by comparing with a reference database (Kim *et al.,* 2010), which can provide an overview of a biological system.

The NMR-based identification of metabolites in this study revealed the presence of relatively higher levels of molecules involve in the biosynthesis of phenylalanine and tyrosine metabolism in field samples (Carlton bulb and basal plate) than *in vitro* samples. Biosynthesis of phenylalanine and tyrosine are the initial biosynthetic pathways involved in the galanthamine and other Amaryllidaceae alkaloid production.

6.4 Transcriptome analysis

Transcriptome analysis (RNA-seq) is an effective platform for the identification and functional characterisation of novel candidate genes as well as to identify genes encoding uncharacterised enzymes (Desgagné-Penix *et al.,* 2010). In this study, differential gene expression analysis of *Narcissus* was obtained from two conditions i.e. field and *in vitro* callus. It showed the expression of genes involved in the biosynthesis of alkaloids were present in both conditions i.e. cytochrome *P450s*, *OMTs*, *NADP/ NADPH* dehydrogenases/ reductases, *SAM*-synthetases/ decarboxylases, 3-ketoacyl-CoA, acyl-CoA, cinnamoyl-CoA, cinnamate 4-hydroxylase, alcohol dehydrogenase, caffeic acid, *NMTs* and *NADPH*-cytochrome *P450s* (Góngora-Castillo *et al.,* 2012; Desgagné-Penix *et al.,* 2010).

However, transcripts (i.e. cytochrome P450s and *OMTs*) involved in the Amaryllidaceae alkaloid biosynthesis (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, hydroxymethylglutarate and glutamate synthase) leading to the biosynthesis of precursors (tyrosine, phenylalanine and tryptophan) for secondary metabolites were up-regulated in callus.

The alkaloid analysis (GC-MS) confirmed high amounts of galanthamine in basal plate while it was very low or absent from callus. This could be due to the lower expression levels of cytochrome P450s and *OMTs* transcripts in callus tissue; which have been reported as the main transcripts involved in galanthamine biosynthesis (Kilgore *et al.,* 2014; Kilgore *et al.,* 2016a). The metabolomic (NMR) study also showed the presence of higher concentrations of precursors in basal plate than callus. Therefore, the *Narcissus* transcriptome analysis results (Chapter 5) are in agreement with the GC-MS and NMR findings (Chapter 3 and Chapter 4)

6.5 Future work

The data illustrated in this study are characterised with the prediction of metabolites and transcripts that could be involved in alkaloid biosynthesis in *Narcissus* but had several limitations and hence implies a great scope for further work. There is often a compromise of cost and time to acquire appropriate coverage of project aim and objectives. The following sections are designed to illustrate the limitations and future prospects.

6.5.1 Tissue culture

The tissue culture method can be used for large-scale propagation of *Narcissus* as it suggests minimum contamination and senescence rate in culture along with the use of twin-scale explants, which allows obtaining a large number of explants (24-30) from a single bulb that would be useful screening of methods to induce alkaloid production in culture.

In addition, this protocol allows growing whole plant with green/ white shoot or bulblets with roots. Only three media combinations were tried in this study based on previous works. Nevertheless, there is a great opportunity to test other auxin or cytokinins combining in different concentrations, which might implicate positive effect on *in vitro* plantlets/ tissues growth and galanthamine or other alkaloids or secondary metabolite accumulation (Selles *et al.,* 1999; Colque *et al.,* 2004; El Tahchy *et al.,* 2011a; Taleb *et al.,* 2013). Application of a library of small bioactive molecules could be used to find more efficient induction of alkaloid biosynthesis.

6.5.2 Alkaloid analysis

GC-MS has been used for the quantification of the amount of galanthamine in field and *in vitro* tissues using an authentic galanthamine standard. The results showed the clear difference of galanthamine content in different growth conditions and different tissue types. However, several other Amaryllidaceae alkaloids were detected based on fragmentation patterns in field as well as *in vitro* tissues. Confirmation of their identity and quantification was not possible due to the absence of authentic standards. If this was confirmed, it would provide a new material as a source of these alkaloids. Several Amaryllidaceae alkaloids other than Gal have pharmacological potential and this could make *in vitro* tissues i.e. callus as a great source of other

Amaryllidaceae alkaloids i.e. tazettine/ pretazettine (Masi *et al.,* 2015). Moreover, due to the availability of GC-MS only, high polar alkaloids were not possible to detect. HPLC methods could be an alternative for the detection of high polar and non-polar alkaloids (Tolstikov and Fiehn, 2002) in future.

In this study, only four elicitors were tested for effects on galanthamine accumulation in *in vitro* tissues due to the time limitation and GC-MS cost. However, our results with NMR indicate that it would be a cost effective method for initial screening of the effect of small molecules on alkaloid production, with GC-MS used as a second stage in analysis.

6.5.3 NMR metabolomics

NMR analysis of *Narcissus* field and *in vitro* tissues was successful in detecting a large number of small molecules in the tissue extracts that indicate the activity of metabolic pathways. Knowledge about the small molecules detected in different tissue types provided an overall chemical assay of *Narcissus* and their accumulation pattern in different tissue types. In addition, it has given an idea of the metabolic pathways involving these compounds, based on the metabolism of model systems.

There were a number of limitations in this NMR-based metabolomic analysis, as this was the first plant metabolomic study performed in the NMR Centre, University of Liverpool. The software used for metabolite identification (Chenomx) was based on human metabolite data, including only one plant species, *Arabidopsis*. Hence, the database did not include any plant specific metabolites and the detection of Amaryllidaceae alkaloids was not possible. However, several previous studies reported the identification of plant alkaloids based on plant-specific databases (Berkov *et al.,* 2011; Lubbe *et al.,* 2011, 2012 and 2013). Identification of metabolites using a plant-specific database would be worthwhile to test whether it provided better secondary metabolites production pattern in *Narcissus*.

Plants usually produce a large number of complex structured secondary metabolites (Colquhoun, 2007), resulting in an inevitably crowded signal for *Narcissus* samples on NMR spectra. Resolving these signals can be impossible in one-dimensional NMR (Colquhoun, 2007) as in our study. In the most crowded regions, any more than three overlapped signals with very low intensities were removed from the

bucket tables for statistical analyses. Future analysis using two-dimensional NMR (Palama *et al.,* 2010) would be helpful to overcome the problem of identifying and analysing the crowded signals to reveal for their metabolites.

6.5.4 Transcriptome analysis

In our study, the main aim of the *Narcissus* transcriptome analysis was to detect the differential expression of genes/ transcripts possibly involved in secondary metabolism in two different tissues i.e. basal plate versus callus. The results showed a total of 2153 differentially expressed transcripts and 78.95% of those were annotated.

Only 206 (9.5% of total) transcripts were found to be significantly differentially expressed (q-value $\leq$0.05). Differential expression analysis excluded many transcripts present in the samples. For example, transcripts for *PAL* and *TYDC*, involved in phenylalanine and tyrosine metabolism pathways were not detected among the differentially expressed genes, which were previously detected in *Narcissus* (Pulman, 2015). Our transcriptome data showed low mapping (19 to 48%) to reference *N. pseudonarcissus* transcriptome data (www.medplantrnaseq.org), which could also exclude many transcripts from our data. 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). A complete annotation of this reference transcriptome would be better for the confident identification of novel transcripts involved in secondary metabolite production in *Narcissus*. However, due to the time limitation performing these analyses was not possible in this project.

The time and cost of this project did not allow a repeat of the one replicate of Carlton basal plate, which showed higher mapping to rRNA in Illumina post run and showed divergent results than other samples under analysis (Chapter 5). Repeating the sequencing could be another issue need to reconsider.

The candidate genes predicted in this project, involved in Amaryllidaceae alkaloid biosynthesis, such as *PAL*, *TYDC*, cytochrome P450s, *OMT*s, have been previously identified using experimental methods in *N. pseudonarcissus* (Kilgore *et al.,* 2014, Kilgore *et al.,* 2016a) and *L. aurea* (Wang *et al.,* 2013).

6.5.6 Future prospects

6.5.6.1 Plant response to defence or stress linked to alkaloid biosynthesis

Most of the plant secondary metabolites derived from the isoprenoid, phenylpropanoid, alkaloid or fatty acid/polyketide pathways. The rich diversity of plant secondary metabolites results in improved defence against microbial attack or insect or animal predation. However, such diversity has made the application of conventional molecular and genetic techniques difficult to address the functions of natural products in plant defence, or to improve plant disease resistance using metabolic pathway engineering (Dixon, 2001). The present study showed a link between Amaryllidaceae alkaloid pathways to the defence or stress response in plants. The pathways (Plant Reactome and KEGG) observed in *Narcissus* field and callus samples were linked to phenylpropanoid, hormone, flavonoid biosynthesis and mevalonate pathway, which are the pathways leading to a diversity of secondary metabolites related to plant defence or stress response (Dixon, 2001). The putative transcripts involved in these pathways (Chapter 5, Figure 5.15 and 5.19) are also discovered in current study, which could be the further resources for metabolic engineering in *Narcissus* linked to plant stress or defence response.

As for example, targeted transgenic approaches allow for evaluation of effects of directly altered phytoalexin profiles in *Medicago sativa* (alfalfa) and such approaches have been undertaken in the case of stilbenoids and isoflavonoids. The study showed that the introduction of a novel phytoalexin (resveratrol) into alfalfa by constitutive expression of a grapevine stilbene-synthase gene resulted in reduced symptoms following infection by the leaf spot pathogen *Phoma medicaginis* (Dixon, 2001). Transferred DNA (T-DNA) activation tagging has been applied to the characterisation of transcriptional regulators of natural product pathways such as phenylpropanoid biosynthesis (Borewitz, 2001) and this approach allows the identification of regulatory genes without the need to understand the individual enzymatic steps of the pathway. Hence, it can offer an exciting opportunity to develop new molecular tools for pathway engineering for improved plant defence (van der Fits and Memelink, 2000).

Environmental factors such as freezing temperature, humidity, light intensity, drought, salinity, minerals and CO2 influence the growth of a plant and secondary

metabolite production. Such metabolites often accumulate in plants subjected to different biotic and abiotic stresses (Akula and Ravishankar, 2011). Therefore, future experiments on cultures challenged with different environmental and pathogen stresses would be worthwhile to perform in terms of metabolomic and transcriptomic (RNA-seq) analyses. Stresses such as pathogen attack, UV-irradiation, high light, wounding, nutrient deficiencies, temperature and herbicide treatment often increase the accumulation of phenylpropanoids (Dixon, 1995). As for example, salinity often creates both ionic and osmotic stress in plants, resulting in accumulation or decrease of specific secondary metabolites in plants (Parida and Das, 2005). Study showed the salinity treated *C. roseus* plants accumulated an increased amount of total indole alkaloid in shoots and roots when compared to untreated plants (Jaleel *et al.,* 2007). Drought causes oxidative stress and was reported to show an increase in the amounts of flavonoids and phenolic acids in willow leaves (Akula and Ravishankar, 2011). Another study showed the combined effect of pathogen and drought stress on the ajmalicine content in *C. roseus*. The results showed the ajmalicine content was increased due to *Pseudomonas fluorescens* treatment to the drought stressed plants (Jaleel *et al.,* 2007).

6.5.6.2 Partitioning of alkaloids linked to the availability of photosynthate

The alkaloid (galanthamine) partitioning knowledge in different organs of *Narcissus* throughout the plant growth cycle could provide an overall idea of alkaloid biosynthesis pattern in *Narcissus*. Therefore, determining the source-sink relationship linked to the availability of photosynthate and the repartitioning or recycling of alkaloids in tissues as part of carbon flux would be sensible to take into account for future work.

Carbon isotopic labelling techniques have been used for the determination of the carbon balance of a whole plant and to understand the relation between source and the sink. Techniques using ^{14}C or ^{13}C labelling have been used for many years to study the allocation of carbohydrates produced by photosynthesis between different parts of the plant (Barzegar and Nekounam, 2016). The regulation of carbon partitioning at the whole plant level is directly linked to the cellular pathways of assimilate transport and the metabolism in source leaves, and sink organs such as roots/bulb and fruit (Osorio *et al.,* 2014).

The ^{14}C is a long-lived radioactive (half-life of 5,700 years), relatively cheap and easy to use for labelling experiments. The labelled plant material allows carbon allocation into different tissues and accurate quantification of carbon partitioning into major metabolite classes and biosynthetic end-products (Kölling *et al,* 2013).

The partitioning of major primary metabolite fractions and distribution of sugars, amino acids, and organic acids and the simultaneous utilization of those assimilates for total alkaloid production has been studied at whole plant level in *C. roseus* by using ^{14}C-saccharose labelling. The results showed roots as the important metabolic sink with highest ^{14}C content followed by leaves and stems. Hence, roots were the major accumulators of metabolites accompanied by higher biosynthetic utilization for alkaloid accumulation (Srivastava and Srivastava, 2006).

6.5.6.3 Genetic mapping for *Narcissus* breeding linked to alkaloid production

The functional genomics approaches are powerful tools to accelerate complete investigations of cellular metabolism in specialised tissues. A vast majority of medicinal plants or plants containing diversified secondary compounds are lacking extensive genomic data (Goossens *et al.,* 2003). PCR-based DNA fingerprinting techniques such as random amplified polymorphic DNA analysis (RAPD), inter simple sequence repeat (ISSR) and amplified fragment length polymorphism (AFLP) are widely used for assessing genetic diversity for a wide range of plants without requiring any prior knowledge of extensive genomic data (Badfar-Chaleshtori *et al.,* 2012). The PCR-based marker systems have markedly increased the efficiency and reduced the cost of transferring genetic information across species. In this approach, oligonucleotide primers are designed from sequences of conserved regions such as gene exons that span polymorphic introns or microsatellites to produce ITAPs (intron targeted amplified polymorphic sequence) (Phan *et al.,* 2007).

The construction of functional maps, composed of genes of known function, is an important component of the candidate gene approach. Therefore, identification of SNPs and ESTs (Expressed Sequence Tags) could be the important resources for gene discovery, molecular marker development, QTL (Quantitative Trait Loci) analysis and genetic linkage (Pulman, 2015). Marker development linked to the alkaloid production would be highly applicable for the commercial alkaloid

extraction and improve breeding programs of *Narcissus*. Genetic mapping in *Lupinus albus* L. allowed the determination of quantitative trait loci (QTLs) governing the alkaloid content along with other important traits such as anthracnose resistance and flowering time (Phan *et al.,* 2007).

There are few changes would be worthwhile to take into account if this work would start afresh. Those could be listed as follows:

1. *In vitro* sample would be treated with several elicitors under a series of time course to enhance the production of galanthamine and other important alkaloids.
2. Determining the amount of other alkaloids such as tazettine/ pretazettine detected in callus with the use of authentic standards.
3. RNA-seq work should be started earlier (end of second year, as in this study it was started by the end of third year) allowing more time to focus on cultures grown under different stress factors and look into the transcript level differences between field, *in vitro* cultures and stress factors treated cultured samples.
4. More emphasis would be attempted on the pathway analysis and that knowledge could be used for metabolomic engineering.
5. Would have been revisiting the experimental errors including some method modification (e.g. section 5.4.2), replicates for RNA-sequencing, method of RNA-seq analysis (e.g. section 5.3.7.1), and selection of genetic markers to identify the genes underpinning the high galanthamine levels.

Finally, this was the first study combining tissue culture, metabolomics and transcriptomics in *Narcissus* that would be a rational to understand the plant biology related to alkaloid production.

8 References

Abd El-Mawla, A. 2010. Effect of certain elicitors on production of pyrrolizidine alkaloids in hairy root cultures of Echium rauwolfii. *Die Pharmazie-An International Journal of Pharmaceutical Sciences,* 65**,** 224-226.

Abu Zahra, H. & Oran, S. Micropropagation of the Wild Endangered Daffodil Narcissus tazetta L. I International Medicinal and Aromatic Plants Conference on Culinary Herbs 826, 2007. 135-140.

Ahloowalia, B., Prakash, J., Savangikar, V. & Savangikar, C. Low cost options for tissue culture technology in developing countries. Proceedings of a Technical Meeting organized by the Joint FAO/IAEA Division of Nuclear Techniques in Food and Agriculture and held in Vienna, 26, 2002.

Akula, R. and Ravishankar, G.A., 2011. Influence of abiotic stress signals on secondary metabolites in plants. *Plant signaling & behavior*, *6*(11), pp.1720-1731.

Alagna, F., D'agostino, N., Torchia, L., Servili, M., Rao, R., Pietrella, M., Giuliano, G., Chiusano, M. L., Baldoni, L. & Perrotta, G. 2009. Comparative 454 pyrosequencing of transcripts from two olive genotypes during fruit development. *BMC Genomics,* 10**,** 1.

Alves, M. N., Sartoratto, A. & Trigo, J. R. 2007. Scopolamine in Brugmansia suaveolens (Solanaceae): defence, allocation, costs, and induced response. *Journal of chemical ecology,* 33**,** 297-309.

Anand, S. 2010. Various approaches for secondary metabolite production through plant tissue culture. *Pharmacia,* 1**,** 1-7.

Anbari, S., Tohidfar, M., Hosseini, R. & Haddad, R. 2007. Somatic embryogenesis induction in Narcissus papyraceus cv. Shirazi. *Plant Tissue Culture and Biotechnology,* 17**,** 37-46.

Anders, S., Pyl, P. T. & Huber, W. 2014. HTSeq–a Python framework to work with high-throughput sequencing data. *Bioinformatics***,** btu638.

Anjusha, S. & Gangaprasad, A. 2016. In vitro propagation and anthraquinone quantification in Gynochthodes umbellata (L.) Razafim. & B. Bremer (Rubiaceae)—A dye yielding plant. *Industrial Crops and Products,* 81**,** 83-90.

Araji, S., Grammer, T. A., Gertzen, R., Anderson, S. D., Mikulic-Petkovsek, M., Veberic, R., Phu, M. L., Solar, A., Leslie, C. A. & Dandekar, A. M. 2014. Novel roles for the polyphenol oxidase enzyme in secondary metabolism and the regulation of cell death in walnut. *Plant physiology,* 164**,** 1191-1203.

Arya, M., Shergill, I. S., Williamson, M., Gommersall, L., Arya, N. & Patel, H. R. 2014. Basic principles of real-time quantitative PCR. *Expert review of molecular diagnostics*.

Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., Davis, A. P., Dolinski, K., Dwight, S. S. & Eppig, J. T. 2000. Gene Ontology: tool for the unification of biology. *Nature genetics,* 25**,** 25-29.

Auer, P. L. & Doerge, R. 2010. Statistical design and analysis of RNA sequencing data. *Genetics,* 185**,** 405-416.

Babu, C. P. 2000. *Elicitation of tissue cultures for production of Biomedicines.* Ph. D. Thesis Kakatiya University, Warangal, AP, India.

Badfar-Chaleshtori, S., Shiran, B., Kohgard, M., Mommeni, H., Hafizi, A., Khodambashi, M., Mirakhorli, N. and Sorkheh, K., 2012. Assessment of genetic diversity and structure of Imperial Crown (Fritillaria imperialis L.) populations in the Zagros region of Iran using AFLP, ISSR and RAPD markers and implications for its conservation. *Biochemical Systematics and Ecology*, *42*, pp.35-48.

Barzegar, T. and Nekounam, F., 2016. Plant sink-source relationships and carbon isotopic labeling techniques. *Iranian Journal of Plant Physiology*, *6*(2).

Bas, A., Forsberg, G., Hammarström, S. & Hammarström, M. L. 2004. Utility of the Housekeeping Genes 18S rRNA, β-Actin and Glyceraldehyde-3-Phosphate-Dehydrogenase for Normalisation in Real-Time Quantitative Reverse Transcriptase-Polymerase Chain Reaction Analysis of Gene Expression in Human T Lymphocytes. *Scandinavian journal of immunology,* 59**,** 566-573.

Baskaran, P., Moyo, M. & Van Staden, J. 2014. In vitro plant regeneration, phenolic compound production and pharmacological activities of Coleonema pulchellum. *South African Journal of Botany,* 90**,** 74-79.

Bastida, J., Berkov, S., Torras, L., Pigni, N. B., De Andrade, J. P., Martínez, V., Codina, C. & Viladomat, F. 2011. Chemical and biological aspects of Amaryllidaceae alkaloids. *Recent Advances in Pharmaceutical Sciences,* 3**,** 65-100.

Bastida, J., Lavilla, R. & Viladomat, F. 2006. Chemical and biological aspects of Narcissus alkaloids. *The alkaloids: chemistry and biology*, *63*, pp.87-179.

Bastida, J. & Viladomat, F. 2002. 6 Alkaloids of Narcissus. *Narcissus and daffodil: the genus Narcissus***,** 141.

Beckonert, O., Keun, H. C., Ebbels, T. M., Bundy, J., Holmes, E., Lindon, J. C. & Nicholson, J. K. 2007. Metabolic profiling, metabolomic and metabonomic procedures for NMR spectroscopy of urine, plasma, serum and tissue extracts. *Nature protocols,* 2**,** 2692-2703.

Bedre, R., Mangu, V. R., Srivastava, S., Sanchez, L. E. & Baisakh, N. 2016. Transcriptome analysis of smooth cordgrass (Spartina alterniflora Loisel), a monocot halophyte, reveals candidate genes involved in its adaptation to salinity. *BMC Genomics,* 17**,** 657.

Beemster, G. T., Fiorani, F. & Inze, D. 2003. Cell cycle: the key to plant growth control? *Trends in plant science,* 8**,** 154-158.

Benes, V., Blake, J. & Doyle, K. 2011. Ribo-Zero Gold Kit: improved RNA-seq results after removal of cytoplasmic and mitochondrial ribosomal RNA. *Nat Meth,* 8.

Beletsky, A. V., Filyushin, M. A., Gruzdev, E. V., Mazur, A. M., Prokhortchouk, E. B., Kochieva, E. Z., Mardanov, A. V., Ravin, N. V. & Skryabin, K. G. 2016. De novo transcriptome assembly of the mycoheterotrophic plant Monotropa hypopitys. *Genomics Data*.

Bergoñón, S., Codina, C., Bastida, J., Viladomat, F. & Melé, E. 1992. The shake liquid culture as an alternative way to the multiplication of Narcissus plants. VI International Symposium on Flower Bulbs 325, 447-452.

Bergoñón, S., Codina, C., Bastida, J., Viladomat, F. & Melé, E. 1996. Galanthamine production in "shoot-clump" cultures of Narcissus confusus in liquid-shake medium. *Plant Cell, Tissue and Organ Culture,* 45**,** 191-199.

Berkov, S., Bastida, J., Viladomat, F. & Codina, C. 2011. Development and validation of a GC–MS method for rapid determination of galanthamine in Leucojum aestivum and Narcissus ssp.: A metabolomic approach. *Talanta,* 83**,** 1455-1465.

Berkov, S., Georgieva, L., Kondakova, V., Atanassov, A., Viladomat, F., Bastida, J. & Codina, C. 2009. Plant sources of galanthamine: phytochemical and biotechnological aspects. *Biotechnology & Biotechnological Equipment,* 23**,** 1170-1176.

Berkov, S., Georgieva, L., Kondakova, V., Viladomat, F., Bastida, J., Atanassov, A. & Codina, C. 2013. The geographic isolation of Leucojum aestivum populations leads to divergation of alkaloid biosynthesis. *Biochemical Systematics and Ecology,* 46**,** 152-161.

Berkov, S., Ivanov, I., Georgiev, V., Codina, C. & Pavlov, A. 2014a. Galanthamine biosynthesis in plant in vitro systems. *Engineering in Life Sciences,* 14**,** 643-650.

Berkov, S., Martínez-Francés, V., Bastida, J., Codina, C. & Ríos, S. 2014b. Evolution of alkaloid biosynthesis in the genus Narcissus. *Phytochemistry,* 99**,** 95-106.

Berli, F. J., Moreno, D., Piccoli, P., Hespanhol-Viana, L., Silva, M. F., Bressan-Smith, R., Cavagnaro, J. B. & Bottini, R. 2010. Abscisic acid is involved in the response of grape (Vitis vinifera L.) cv. Malbec leaf tissues to ultraviolet-B radiation by enhancing ultraviolet-absorbing compounds, antioxidant enzymes and membrane sterols. *Plant, cell & environment,* 33**,** 1-10.

Bertani, P., Raya, J. & Bechinger, B. 2014. 15 N chemical shift referencing in solid state NMR. *Solid state nuclear magnetic resonance,* 61**,** 15-18.

Bhatia, S. & Dahiya, R. 2015. Concepts and Techniques of Plant Tissue Culture Science. *Modern Applications of Plant Biotechnology in Pharmaceutical Sciences*.

Bhojwani, S. S. & Razdan, M. K. 1986. *Plant tissue culture: theory and practice*, Elsevier.

Blanc, G., Lardet, L., Martin, A., Jacob, J. & Carron, M. 2002. Differential carbohydrate metabolism conducts morphogenesis in embryogenic callus of Hevea brasiliensis (Müll. Arg.). *Journal of Experimental Botany,* 53**,** 1453-1462.

Bligny, R. & Douce, R. 2001. NMR and plant metabolism. *Current Opinion in Plant Biology,* 4**,** 191-196.

Bobzin, S., Yang, S. & Kasten, T. 2000. LC-NMR: a new tool to expedite the dereplication and identification of natural products. *Journal of Industrial Microbiology and Biotechnology,* 25**,** 342-345.

Booth, S. C., Weljie, A. M. & Turner, R. J. 2013. Computational tools for the secondary analysis of metabolomics experiments. *Computational and structural biotechnology journal,* 4**,** 1-13.

Borewitz, J., Xia, Y., Blount, J. W., Dixon, R. A. & Lamb, C (2001). Activation tagging identifies a conserved MYB regulator of phenylpropanoid biosynthesis. *Plant Cell*, 12, 2383–2393.

Bowes, S., Edmondson, R., Linfield, C. & Langton, F. 1992. Screening immature bulbs of daffodil (Narcissus L.) crosses for resistance to basal rot disease caused by Fusarium oxysporum f. sp. narcissi. *Euphytica,* 63**,** 199-206.

Böttcher, C., Von Roepenack-Lahaye, E., Schmidt, J., Schmotz, C., Neumann, S., Scheel, D. & Clemens, S. 2008. Metabolome analysis of biosynthetic mutants reveals a diversity of metabolic changes and allows identification of a large number of new compounds in Arabidopsis. *Plant physiology,* 147**,** 2107-2120.

Branch, I. K. 2013. Growth and physiological characteristics of Narcissus pseudonarcissus at different nitrogen levels. *Intl J Farm & Alli Sci*. Vol., 2 (S2): 1325-1329, 2013.

Bräutigam, A., Mullick, T., Schliesky, S. & Weber, A. P. 2011. Critical assessment of assembly strategies for non-model species mRNA-Seq data and application of next-generation sequencing to the comparison of C3 and C4 species. *Journal of Experimental Botany,* 62**,** 3093-3102.

Broeckling, C. D., Huhman, D. V., Farag, M. A., Smith, J. T., May, G. D., Mendes, P., Dixon, R. A. & Sumner, L. W. 2005. Metabolic profiling of Medicago truncatula cell cultures reveals the effects of biotic and abiotic elicitors on metabolism. *Journal of Experimental Botany,* 56**,** 323-336.

Brooker, N., Windorski, J. & Bluml, E. 2007. Halogenated coumarin derivatives as novel seed protectants. *Communications in agricultural and applied biological sciences,* 73**,** 81-89.

Bustin, S. A. & Nolan, T. 2013. Analysis of mRNA expression by real-time PCR. *Real-Time PCR: Adv Technol Appl,* 51**,** 111-135

Carvalho, L., Santos, P. & Amâncio, S. 2015. Effect of light intensity and CO 2 concentration on growth and the acquisition of in vivo characteristics during acclimatization of grapevine regenerated in vitro. *VITIS-Journal of Grapevine Research,* 41(1), P.1.

Chang, S., Puryear, J. & Cairney, J. 1993. A simple and efficient method for isolating RNA from pine trees. *Plant molecular biology reporter,* 11**,** 113-116.

Chaturvedi, H., Jain, M. & Kidwai, N. 2007. Cloning of medicinal plants through tissue culture-A review. *Indian journal of experimental biology,* 45**,** 937.

Chávez, M. L. D., Rolf, M., Gesell, A. & Kutchan, T. M. 2011. Characterization of two methylenedioxy bridge-forming cytochrome P450-dependent enzymes of alkaloid formation in the Mexican prickly poppy Argemone mexicana. *Archives of Biochemistry and Biophysics,* 507**,** 186-193.

Che, P., Lall, S., Nettleton, D. & Howell, S. H. 2006. Gene expression programs during shoot, root, and callus development in Arabidopsis tissue culture. *Plant physiology,* 141**,** 620-637.

Chen, L., Zhu, X., Gu, L. & Wu, J. 2005. Efficient callus induction and plant regeneration from anther of Chinese narcissus (Narcissus tazetta L. var. chinensis Roem). *Plant Cell Reports,* 24**,** 401-407.

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 research,* 14**,** 1147-1159.

Cherkasov, O. & Tolkachev, O. 2002. Narcissus and other Amaryllidaceae as sources of galanthamine. *Medicinal and Aromatic Plants±Industrial Profiles,* 21**,** 242-55.

Choi, K.-B., Morishige, T., Shitan, N., Yazaki, K. & Sato, F. 2002. Molecular cloning and characterization of coclaurinen-methyltransferase from cultured cells of Coptis japonica. *Journal of Biological Chemistry,* 277**,** 830-835.

Choi, Y. H., Tapias, E. C., Kim, H. K., Lefeber, A. W., Erkelens, C., Verhoeven, J. T. J., Brzin, J., Zel, J. & Verpoorte, R. 2004. Metabolic discrimination of Catharanthus roseus leaves infected by phytoplasma using 1H-NMR spectroscopy and multivariate data analysis. *Plant physiology,* 135**,** 2398-2410.

Choi, Y. H., Kim, H. K., Linthorst, H. J., Hollander, J. G., Lefeber, A. W., Erkelens, C., Nuzillard, J.-M. & Verpoorte, R. 2006. NMR Metabolomics to Revisit the Tobacco Mosaic Virus Infection in Nicotiana t abacum Leaves. *Journal of natural products,* 69**,** 742-748.

Chow, Y., Selby, C. & Harvey, B. 1992. A simple method for maintaining high multiplication of Narcissus shoot cultures in vitro. *Plant Cell, Tissue and Organ Culture,* 30**,** 227-230.

Christophersen, C. 1995. Theory of the origin, function, and evolution secondary metabolites. *Studies in natural products chemistry,* 18**,** 677-737.

Codina, C. 2002. 7 Production of galanthamine by Narcissus tissues in vitro. *Narcissus and Daffodil (The Genus Narcissus).* Taylor & Francis London.

Colque, R., Viladomat, F., Bastida, J. & Codina, C. 2004. Improved production of galanthamine and related alkaloids by methyl jasmonate in Narcissus confusus shoot-clumps. *Planta medica,* 70**,** 1180-1188.

Colquhoun, I. J. 2007. Use of NMR for metabolic profiling in plant systems. *Journal of Pesticide Science,* 32**,** 200-212.

Collins, L. J., Biggs, P. J., Voelckel, C. & Joly, S. 2008. An approach to transcriptome analysis of non-model organisms using short-read sequences. *Genome informatics,* 21**,** 3-14.

Constantin, M.-A., Conrad, J. & Beifuss, U. 2012. Laccase-catalyzed oxidative phenolic coupling of vanillidene derivatives. *Green Chemistry,* 14**,** 2375-2379.

Craig, A., Cloarec, O., Holmes, E., Nicholson, J. K. & Lindon, J. C. 2006. Scaling and normalisation effects in NMR spectroscopic metabonomic data sets. *Analytical chemistry,* 78**,** 2262-2267.

Creelman, R. A. & Mullet, J. E. 1997. Biosynthesis and action of jasmonates in plants. *Annual review of plant biology,* 48**,** 355-381.

Czechowski, T., Stitt, M., Altmann, T., Udvardi, M. K. & Scheible, W.-R. 2005. Genome-wide identification and testing of superior reference genes for transcript normalisation in Arabidopsis. *Plant physiology,* 139**,** 5-17.

Dal Toso, R. & Melandri, F. 2011. Sustainable sourcing of natural food ingredients by plant cell cultures. *AGRO FOOD INDUSTRY HI-TECH,* 22**,** 30-32.

Dang, P. M. & Chen, C. Y. 2013. Modified method for combined DNA and RNA isolation from peanut and other oil seeds. *Molecular biology reports,* 40**,** 1563-1568.

De Bruyn, M., Ferreira, D., Slabbert, M. & Pretorius, J. 1992. In vitro propagation of Amaryllis belladonna. *Plant Cell, Tissue and Organ Culture,* 31**,** 179-184.

Defernez, M. & Colquhoun, I. J. 2003. Factors affecting the robustness of metabolite fingerprinting using 1 H NMR spectra. *Phytochemistry,* 62**,** 1009-1017

Delporte, M., Legrand, G., Hilbert, J.-L. & Gagneul, D. 2015. Selection and validation of reference genes for quantitative real-time PCR analysis of gene expression in Cichorium intybus. *Frontiers in plant science,* 6:651.

Deprez, R. H. L., Fijnvandraat, A. C., Ruijter, J. M. & Moorman, A. F. 2002. Sensitivity and accuracy of quantitative real-time polymerase chain reaction using SYBR green I depends on cDNA synthesis conditions. *Analytical biochemistry,* 307**,** 63-69.

Der, J. P., Barker, M. S., Wickett, N. J. & Wolf, P. G. 2011. De novo characterization of the gametophyte transcriptome in bracken fern, Pteridium aquilinum. *BMC Genomics,* 12**,** 1.

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 biology,* 10**,** 1.

Devisetty, U. K., Kennedy, K., Sarando, P., Merchant, N. & Lyons, E. 2016. Bringing your tools to CyVerse Discovery Environment using Docker. *F1000Research,* 5.

Devonshire, A. S., R. Elaswarapu, et al. (2010). Evaluation of external RNA controls for the standardisation of gene expression biomarker measurements. *BMC Genomics*, 11(1): 662.

Dewick, P. M. 2009. *Medicinal natural products: a biosynthetic approach (3rd ed.)* John Wiley & Sons.

Dheda, K., Huggett, J., Chang, J.-S., Kim, L., Bustin, S., Johnson, M., Rook, G. & Zumla, A. 2005. The implications of using an inappropriate reference gene for real-time reverse transcription PCR data normalization. *Analytical biochemistry,* 344**,** 141-143.

Dias, M. I., Sousa, M. J., Alves, R. C. & Ferreira, I. C. 2016. Exploring plant tissue culture to improve the production of phenolic compounds: A review. *Industrial Crops and Products,* 82**,** 9-22.

Dieterle, F., Ross, A., Schlotterbeck, G. & Senn, H. 2006. Probabilistic quotient normalisation as robust method to account for dilution of complex biological mixtures. Application in 1H NMR metabonomics. *Analytical chemistry,* 78**,** 4281-4290.

Diop, M. F., Ptak, A., Chretien, F., Henry, M., Chapleur, Y. & Laurain-Mattar, D. 2006. Galanthamine content of bulbs and in vitro cultures of Leucojum aestivum L. *Natural Product Communications,* 1**,** 475-479.

Dixon, R. A. & Paiva, N. L. 1995. Stress-induced phenylpropanoid metabolism. *The plant cell,* 7**,** 1085.

Dixon, R.A., 2001. Natural products and plant disease resistance. *Nature*, *411*(6839), pp.843-847.

Dobnik, D., Baebler, Š., Kogovšek, P., Pompe-Novak, M., Štebih, D., Panter, G., Janež, N., Morisset, D., Žel, J. & Gruden, K. 2013. β-1, 3-glucanase class III promotes spread of PVYNTN and improves in planta protein production. *Plant biotechnology reports,* 7**,** 547-555.

Dudareva, N. & Pichersky, E. 2008. Metabolic engineering of plant volatiles. *Current opinion in biotechnology,* 19**,** 181-189.

Egan, A. N., Schlueter, J. & Spooner, D. M. 2012. Applications of next-generation sequencing in plant biology. *American journal of botany,* 99**,** 175-185.

Eichhorn, J., Takada, T., Kita, Y. & Zenk, M. H. 1998. Biosynthesis of the Amaryllidaceae alkaloid galanthamine. *Phytochemistry,* 49**,** 1037-1047.

Eilert, U., Ehmke, A. & Wolters, B. 1984. Elicitor-induced accumulation of acridone alkaloid epoxides in Ruta graveolens suspension cultures. *Planta medica,* 50**,** 508-512.

Ekblom, R. & Galindo, J. 2011. Applications of next generation sequencing in molecular ecology of non-model organisms. *Heredity,* 107**,** 1-15.

El Tahchy, A., Bordage, S., Ptak, A., Dupire, F., Barre, E., Guillou, C., Henry, M., Chapleur, Y. & Laurain-Mattar, D. 2011a. Effects of sucrose and plant growth regulators on acetylcholinesterase inhibitory activity of alkaloids accumulated in shoot cultures of Amaryllidaceae. *Plant Cell, Tissue and Organ Culture (PCTOC),* 106**,** 381-390.

El Tahchy, A., Ptak, A., Boisbrun, M., Barre, E., Guillou, C., Dupire, F. O., Chrétien, F. O., Henry, M., Chapleur, Y. & Laurain-Mattar, D. 2011b. Kinetic study of the rearrangement of deuterium-labeled 4′-O-methylnorbelladine in Leucojum aestivum shoot cultures by mass spectrometry. Influence of precursor feeding on Amaryllidaceae alkaloid accumulation. *Journal of natural products,* 74**,** 2356-2361.

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 physiology,* 151**,** 1096-1113.

Farnsworth, N. R. 1988. Screening plants for new medicines. *Biodiversity,* 1**,** 83-97.

Fiehn, O., Kopka, J., Dörmann, P., Altmann, T., Trethewey, R. N. & Willmitzer, L. 2000. Metabolite profiling for plant functional genomics. *Nature biotechnology,* 18**,** 1157-1161.

Franssen, S. U., Shrestha, R. P., Bräutigam, A., Bornberg-Bauer, E. & Weber, A. P. 2011. Comprehensive transcriptome analysis of the highly complex Pisum sativum genome using next generation sequencing. *BMC Genomics,* 12**,** 1.

Gambino, G., Perrone, I. & Gribaudo, I. 2008. A rapid and effective method for RNA extraction from different tissues of grapevine and other woody plants. *Phytochemical Analysis,* 19**,** 520-525.

Gamborg, O., Murashige, T., Thorpe, T. & Vasil, I. 1976. Plant tissue culture media. *In Vitro Cellular & Developmental Biology-Plant,* 12**,** 473-478.

Gao, S., Zhu, D., Cai, Z., Jiang, Y. & Xu, D. 1999. Organ culture of a precious Chinese medicinal plant–Fritillaria unibracteata. *Plant Cell, Tissue and Organ Culture,* 59**,** 197-201.

George, E. F., Hall, M. A. & De Klerk, G.-J. 2008. Plant growth regulators I: introduction; auxins, their analogues and inhibitors. *Plant propagation by tissue culture.* Springer.

George, E. F., Hall, M. A. & De Klerk, G.-J. 2008. Plant growth regulators II: cytokinins, their analogues and antagonists. *Plant propagation by tissue culture.* Springer.

Georgiev, V., Ivanov, I., Berkov, S., Ilieva, M., Georgiev, M., Gocheva, T. & Pavlov, A. 2012. Galanthamine production by Leucojum aestivum L. shoot culture in a modified bubble column bioreactor with internal sections. *Engineering in Life Sciences,* 12**,** 534-543.

Ghosh, S. & Chan, C. K. K. 2016. Analysis of RNA-Seq Data Using TopHat and Cufflinks. *Plant Bioinformatics: Methods and Protocols***,** 339-361.

Giri, L., Dhyani, P., Rawat, S., Bhatt, I. D., Nandi, S. K., Rawal, R. S. & Pande, V. 2012. In vitro production of phenolic compounds and antioxidant activity in callus suspension cultures of Habenaria edgeworthii: a rare Himalayan medicinal orchid. *Industrial Crops and Products,* 39**,** 1-6.

Goff, S. A., Vaughn, M., Mckay, S., Lyons, E., Stapleton, A. E., Gessler, D., Matasci, N., Wang, L., Hanlon, M. & Lenards, A. 2011b. The iPlant collaborative: cyberinfrastructure for plant biology. *Frontiers in plant science,* 2**,** 34.

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.

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 molecular biology,* 57**,** 53-66.

Goossens, A. & Rischer, H. 2007. Implementation of functional genomics for gene discovery in alkaloid producing plants. *Phytochemistry Reviews,* 6**,** 35-49.

Goossens, A., Häkkinen, S.T., Laakso, I., Seppänen-Laakso, T., Biondi, S., De Sutter, V., Lammertyn, F., Nuutila, A.M., Söderlund, H., Zabeau, M. and Inzé, D., 2003. A functional genomics approach toward the understanding of secondary metabolism in plant cells. *Proceedings of the National Academy of Sciences*, *100*(14), pp.8595-8600.

Gotti, R., Fiori, J., Bartolini, M. & Cavrini, V. 2006. Analysis of Amaryllidaceae alkaloids from Narcissus by GC–MS and capillary electrophoresis. *Journal of pharmaceutical and biomedical analysis,* 42**,** 17-24.

Götz, S., García-Gómez, J. M., Terol, J., Williams, T. D., Nagaraj, S. H., Nueda, M. J., Robles, M., Talón, M., Dopazo, J. & Conesa, A. 2008. High-throughput functional annotation and data mining with the Blast2GO suite. *Nucleic acids research,* 36**,** 3420-3435.

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. *Nature biotechnology,* 29**,** 644-652.

Graham, S. W. & Barrett, S. C. 2004. Phylogenetic reconstruction of the evolution of stylar polymorphisms in Narcissus (Amaryllidaceae). *American journal of botany,* 91**,** 1007-1021.

Gras, E., Guillou, C. & Thal, C. 1999. A formal synthesis of (±) lycoramine via an intramolecular Heck reaction. *Tetrahedron letters,* 40**,** 9243-9244.

Guillou, C., Beunard, J. L., Gras, E. & Thal, C. 2001. An Efficient Total Synthesis of (±)-Galanthamine. *Angewandte Chemie International Edition,* 40**,** 4745-4746.

Guo, J., Xu, N., Li, Z., Zhang, S., Wu, J., Kim, D. H., Marma, M. S., Meng, Q., CAO, H. & LI, X. 2008. Four-color DNA sequencing with 3′-O-modified nucleotide reversible terminators and chemically cleavable fluorescent dideoxynucleotides. *Proceedings of the National Academy of Sciences,* 105**,** 9145-9150.

Guo, S., Zheng, Y., Joung, J. G., Liu, S., Zhang, Z., Crasta, O. R., Sobral, B. W., Xu, Y., Huang, S. & Fei, Z. 2010. Transcriptome sequencing and comparative analysis of cucumber flowers with different sex types. *BMC Genomics,* 11**,** 1.

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.

Gupta, V. & Datta, S. 2001. Influence of gibberellic acid (GA3) on growth and flowering in chrysanthemum (Chrysanthemummorifolium, Ramat) cv. Jayanti. *Indian journal of plant physiology,* 6**,** 420-422.

Gutiérrez, P. A., Alzate, J. F. & Montoya, M. M. 2015. Complete genome sequence of an isolate of Potato virus X (PVX) infecting Cape gooseberry (Physalis peruviana) in Colombia. *Virus genes,* 50**,** 518-522.

Haas, B. J. & Zody, M. C. 2010. Advancing RNA-seq analysis. *Nature biotechnology,* 28**,** 421.

Hadacek, F. 2002. Secondary metabolites as plant traits: current assessment and future perspectives. *Critical Reviews in Plant Sciences,* 21**,** 273-322.

Hale, B. K., Herms, D. A., Hansen, R. C., Clausen, T. P. & Arnold, D. 2005. Effects of drought stress and nutrient availability on dry matter allocation, phenolic glycosides, and rapid induced resistance of poplar to two lymantriid defoliators. *Journal of chemical ecology,* 31**,** 2601-2620.

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 physiology,* 161**,** 600-616.

Hanks, G. R. Narcissus bulb morphology and twin-scale propagation. IV International Symposium on Flower Bulbs 177, 1985. 309-314.

Hanks, G. R. 2002. Commercial production of Narcissus bulbs. *Narcissus and daffodil: the genus Narcissus***,** 53-130.

Hartmann, T. 1996. Diversity and variability of plant secondary metabolism: a mechanistic view. *Entomologia Experimentalis et Applicata,* 80**,** 177-188.

Hasunuma, T., Miyazawa, S. I., Yoshimura, S., Shinzaki, Y., Tomizawa, K. I., Shindo, K., Choi, S. K., Misawa, N. & Miyake, C. 2008. Biosynthesis of astaxanthin in tobacco leaves by transplastomic engineering. *The Plant Journal,* 55**,** 857-868.

Hayashi, H., Huang, P. & Inoue, K. 2003. Up-regulation of soyasaponin biosynthesis by methyl jasmonate in cultured cells of Glycyrrhiza glabra. *Plant and cell physiology,* 44**,** 404-411.

He, M., Qu, C., Gao, O., Hu, X. & Hong, X. 2015. Biological and pharmacological activities of amaryllidaceae alkaloids. *RSC Advances,* 5**,** 16562-16574.

Hendrawati, O., Yao, Q., Kim, H. K., Linthorst, H. J., Erkelens, C., Lefeber, A. W., Choi, Y. H. & Verpoorte, R. 2006. Metabolic differentiation of Arabidopsis treated with methyl jasmonate using nuclear magnetic resonance spectroscopy. *Plant science,* 170**,** 1118-1124.

Herbert, R. B. 2001. The biosynthesis of plant alkaloids and nitrogenous microbial metabolites. *Natural product reports,* 18**,** 50-65.

Herranz, J., Copete, M. & Ferrandis, P. 2013. Environmental regulation of embryo growth, dormancy breaking and germination in Narcissus alcaracensis (Amaryllidaceae), a threatened endemic Iberian daffodil. *The American Midland Naturalist,* 169**,** 147-167.

Hill, D. P., Berardini, T. Z., Howe, D. G. & Van Auken, K. M. 2010. Representing ontogeny through ontology: a developmental biologist's guide to the gene ontology. *Molecular reproduction and development,* 77**,** 314-329.

Hirsch, C. D., Springer, N. M. & Hirsch, C. N. 2015. Genomic limitations to RNA sequencing expression profiling. *The Plant Journal,* 84**,** 491-503.

Hocquette, J. F. & Brandstetter, A. M. 2002. Common practice in molecular biology may introduce statistical bias and misleading biological interpretation. *The Journal of nutritional biochemistry,* 13**,** 370-377.

Hodkinson, B. P. & Grice, E. A. 2015. Next-generation sequencing: a review of technologies and tools for wound microbiome research. *Advances in wound care,* 4**,** 50-58.

Hol, G. & Van Der Linde, P. 1992. Reduction of contamination in bulb-explant cultures ofNarcissus by a hot-water treatment of parent bulbs. *Plant Cell, Tissue and Organ Culture,* 31**,** 75-79.

Hopley, C., Bristow, T., Lubben, A., Simpson, A., Bull, E., Klagkou, K., Herniman, J. & Langley, J. 2008. Towards a universal product ion mass spectral library–reproducibility of product ion spectra across eleven different mass spectrometers. *Rapid Communications in Mass Spectrometry,* 22**,** 1779-1786.

Horgan, R. P. & Kenny, L. C. 2011. 'Omic'technologies: genomics, transcriptomics, proteomics and metabolomics. *The Obstetrician & Gynaecologist,* 13**,** 189-195.

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 research,* 43**,** D1057-D1063.

Hussain, S. Z. & Maqbool, K. GC-MS: Principle, Technique and its application in Food Science. *International Journal of Current Science*. 2014, 13: E 116-126

Hussey, G. 1982. In vitro propagation of Narcissus. *Annals of Botany,* 49**,** 707-719

Ikeuchi, M., Sugimoto, K. & Iwase, A. 2013. Plant callus: mechanisms of induction and repression. *The plant cell,* 25**,** 3159-3173.

Ikezawa, N., Iwasa, K. & Sato, F. 2008. Molecular cloning and characterization of CYP80G2, a cytochrome P450 that catalyzes an intramolecular C–C phenol coupling of (S)-reticuline in magnoflorine biosynthesis, from cultured Coptis japonica cells. *Journal of Biological Chemistry,* 283**,** 8810-8821.

Ingkaninan, K., Hazekamp, A., De Best, C., Irth, H., Tjaden, U., Van Der Heijden, R., Van Der Greef, J. & Verpoorte, R. 2000. The application of HPLC with on-line coupled UV/MS-biochemical detection for isolation of an acetylcholinesterase inhibitor from narcissus' Sir Winston Churchill'. *Journal of natural products,* 63**,** 803-806.

Islam, M. M., Ahmed, M. & Mahaldar, D. 2005. In vitro callus induction and plant regeneration in seed explants of rice (Oryza Sativa L.). *Research Journal of Agriculture and Biological Sciences,* 1**,** 72-75.

Ivanov, I., Georgiev, V. & Pavlov, A. 2013. Elicitation of galanthamine biosynthesis by Leucojum aestivum liquid shoot cultures. *Journal of plant physiology,* 170**,** 1122-1129.

Jaakola, L., Pirttilä, A. M., Halonen, M. & Hohtola, A. 2001. Isolation of high quality RNA from bilberry (Vaccinium myrtillus L.) fruit. *Molecular biotechnology,* 19**,** 201-203.

Jacobs, D. I., Snoeijer, W., Hallard, D. & Verpoorte, R. 2004. The Catharanthus alkaloids: pharmacognosy and biotechnology. *Current medicinal chemistry,* 11**,** 607-628.

Jaleel, C.A., Manivannan, P., Sankar, B., Kishorekumar, A. and Panneerselvam, R., 2007. Calcium chloride effects on salinity-induced oxidative stress, proline metabolism and indole alkaloid accumulation in Catharanthus roseus. *Comptes Rendus Biologies*, *330*(9), pp.674-683.

Jaleel, C.A., Manivannan, P., Sankar, B., Kishorekumar, A., Gopi, R., Somasundaram, R. and Panneerselvam, R., 2007. Pseudomonas fluorescens enhances biomass yield and ajmalicine production in Catharanthus roseus under water deficit stress. *Colloids and Surfaces B: Biointerfaces*, *60*(1), pp.7-11.

Jewison, T., Su, Y., Disfany, F. M., Liang, Y., Knox, C., Maciejewski, A., Poelzer, J., Huynh, J., Zhou, Y. & Arndt, D. 2013. SMPDB 2.0: big improvements to the small molecule pathway database. *Nucleic acids research***,** gkt1067.

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.

Jones, M., Pulman, J. & Walker, T. 2011. Drugs from daffodils. *CHEMISTRY & INDUSTRY***,** 18-20.

Joshi, N. & Fass, J. 2011. Sickle: A sliding-window, adaptive, quality-based trimming tool for FASTQ files (Version 1.33)[Software]. *Available at httpslégitbubcoménajoshiéchkle QOll*.

Jiang, L., Schlesinger, F., Davis, C. A., Zhang, Y., Li, R., Salit, M., Oliver, B. (2011). Synthetic spike-in standards for RNA-seq experiments. *Genome Research*, *21*(9), 1543–1551. http://doi.org/10.1101/gr.121095.111

Kanehisa, M., Goto, S., Sato, Y., Kawashima, M., Furumichi, M. & Tanabe, M. 2014. Data, information, knowledge and principle: back to metabolism in KEGG. *Nucleic acids research,* 42**,** D199-D205.

Karuppusamy, S. 2009. A review on trends in production of secondary metabolites from higher plants by in vitro tissue, organ and cell cultures. *Journal of Medicinal Plants Research,* 3**,** 1222-1239.

Kato, M., Mizuno, K., Crozier, A., Fujimura, T. & Ashihara, H. 2000. Plant biotechnology: Caffeine synthase gene from tea leaves. *nature,* 406**,** 956-957.

Kaul, S., Koo, H. L., Jenkins, J., Rizzo, M., Rooney, T., Tallon, L. J., Feldblyum, T., Nierman, W., Benito, M. I. & Lin, X. 2000. Analysis of the genome sequence of the flowering plant Arabidopsis thaliana. *nature,* 408**,** 796-815.

Kaya, N. & Aki, C. 2013. Effect of plant growth regulators on in vitro biomass changing in Catharanthus roseus (L) G Don. *Annals of Biological Research,* 4**,** 164-168.

Ketchum, R. E., Rithner, C. D., Qiu, D., Kim, Y. S., Williams, R. M. & Croteau, R. B. 2003. Taxus metabolomics: methyl jasmonate preferentially induces production of taxoids oxygenated at C-13 in Taxus x media cell cultures. *Phytochemistry,* 62**,** 901-909.

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. *Frontiers in plant science,* doi: 10.3389/fpls.2016.00225

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.

Kilgore, M. B., Holland, C. K., Jez, J. M. & Kutchan, T. M. 2016b. Identification of a Noroxomaritidine Reductase with Amaryllidaceae Alkaloid Biosynthesis Related Activities. *Journal of Biological Chemistry***,** jbc. M116. 717827.

Kilgore, M. B. & Kutchan, T. M. 2016. The Amaryllidaceae alkaloids: biosynthesis and methods for enzyme discovery. *Phytochemistry Reviews,* 15**,** 317-337.

Kim, H. K., Choi, Y. H., Erkelens, C., Lefeber, A. W. & Verpoorte, R. 2005. Metabolic fingerprinting of Ephedra species using 1H-NMR spectroscopy and principal component analysis. *Chemical and Pharmaceutical Bulletin,* 53**,** 105-109.

Kim, H., Choi, Y. & Verpoorte, R. 2006. Metabolomic analysis of Catharanthus roseus using NMR and principal component analysis. *Plant Metabolomics.* Springer, pp. 261-276

Kim, H. K., Choi, Y. H. & Verpoorte, R. 2010. NMR-based metabolomic analysis of plants. *Nature protocols,* 5**,** 536-549

Kim, H. K., Choi, Y. H. & Verpoorte, R. 2011. NMR-based plant metabolomics: where do we stand, where do we go? *Trends in biotechnology,* 29**,** 267-275.

Kopka, J. 2006. Gas chromatography mass spectrometry. *Plant Metabolomics.* Springer, pp. 3-20.

Kölling, K., Müller, A., Flütsch, P. and Zeeman, S.C., 2013. A device for single leaf labelling with CO 2 isotopes to study carbon allocation and partitioning in Arabidopsis thaliana. *Plant methods*, *9*(1), p.45.

Kornienko, A. & Evidente, A. 2008. Chemistry, biology, and medicinal potential of narciclasine and its congeners. *Chemical reviews,* 108**,** 1982-2014.

Kovačević, N. & Grubišić, D. 2005. In Vitro. Cultures of Plants from the Rhamnaceae: Shoot Propagation and Anthraquinones Production. *Pharmaceutical biology,* 43**,** 420-424.

Kozera, B. & Rapacz, M. 2013. Reference genes in real-time PCR. *Journal of applied genetics,* 54**,** 391-406.

Kreh, M. 2002. *Studies on galanthamine extraction from Narcissus and other Amaryllidaceae* (pp. 256-272), Taylor & Francis: London, UK.

Kumar, S., Banks, T. W. & Cloutier, S. 2012. SNP discovery through next-generation sequencing and its applications. *International journal of plant genomics,* doi:10.1155/2012/831460

Kurbidaeva, A. & Novokreshchenova, M. 2011. Somaclonal variations of Nicotiana tabacum transgenic plants. *Moscow University biological sciences bulletin,* 66**,** 86-90.

Kuroyanagi, M., Arakawa, T., Mikami, Y., Yoshida, K., Kawahar, N., Hayashi, T. & Ishimaru, H. 1998. Phytoalexins from Hairy Roots of Hyoscyamus a lbus Treated with Methyl Jasmonate. *Journal of natural products,* 61**,** 1516-1519.

Kusano, M., Tohge, T., Fukushima, A., Kobayashi, M., Hayashi, N., Otsuki, H., Kondou, Y., Goto, H., Kawashima, M. & Matsuda, F. 2011. Metabolomics reveals comprehensive reprogramming involving two independent metabolic responses of Arabidopsis to UV-B light. *The Plant Journal,* 67**,** 354-369.

Kvam, V. M., Liu, P. & Si, Y. 2012. A comparison of statistical methods for detecting differentially expressed genes from RNA-seq data. *American journal of botany,* 99**,** 248-256.

Lake, J. A., Field, K. J., Davey, M. P., Beerling, D. J. & Lomax, B. H. 2009. Metabolomic and physiological responses reveal multi-phasic acclimation of Arabidopsis thaliana to chronic UV radiation. *Plant, cell & environment,* 32**,** 1377-1389.

Lambert, M., Wolfender, J.-L., Staerk, D., Christensen, S. B., Hostettmann, K. & Jaroszewski, J. W. 2007. Identification of natural products using HPLC-SPE combined with CapNMR. *Analytical chemistry,* 79**,** 727-735.

Lange, B. M. & Ahkami, A. 2013. Metabolic engineering of plant monoterpenes, sesquiterpenes and diterpenes—current status and future opportunities. *Plant biotechnology journal,* 11**,** 169-196.

Langmead, B. & Salzberg, S. L. 2012. Fast gapped-read alignment with Bowtie 2. *Nature methods,* 9**,** 357-359.

Langmead, B., Trapnell, C., Pop, M. & Salzberg, S. L. 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome biology,* 10**,** 1.

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.

Lenz, T. L., Eizaguirre, C., Rotter, B., Kalbe, M. & Milinski, M. 2013. Exploring local immunological adaptation of two stickleback ecotypes by experimental infection and transcriptome-wide digital gene expression analysis. *Molecular Ecology,* 22**,** 774-786.

Leonard, E., Runguphan, W., O'connor, S. & Prather, K. J. 2009. Opportunities in metabolic engineering to facilitate scalable alkaloid production. *Nature Chemical Biology,* 5**,** 292-300.

Li, S., Fan, C., Li, Y., Zhang, J., Sun, J., Chen, Y., Tian, C., Su, X., Lu, M. & Liang, C. 2016. Effects of drought and salt-stresses on gene expression in Caragana korshinskii seedlings revealed by RNA-seq. *BMC Genomics,* 17**,** 1.

Li, H. & Durbin, R. 2010. Fast and accurate long-read alignment with Burrows–Wheeler transform. *Bioinformatics,* 26**,** 589-595.

Li, R., Li, Y., Kristiansen, K. & Wang, J. 2008. SOAP: short oligonucleotide alignment program. *Bioinformatics,* 24**,** 713-714.

Liscombe, D. K., Louie, G. V. & Noel, J. P. 2012. Architectures, mechanisms and molecular evolution of natural product methyltransferases. *Natural product reports,* 29**,** 1238-1250.

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 Journal,* 60**,** 729-743.

Liu, S. H., Edwards, C., Hoch, P. C., Raven, P. H. & Barber, J. C. 2016. Complete plastome sequence of Ludwigia octovalvis (Onagraceae), a globally distributed wetland plant. *Genome announcements,* 4**,** e01274-16.

Llabrés, J. M., Viladomat, F., Bastida, J., Codina, C. & Rubiralta, M. 1986. Phenanthridine alkaloids from Narcissus assoanus. *Phytochemistry,* 25**,** 2637-2638.

López, S., Bastida, J., Viladomat, F. & Codina, C. 2003. Galanthamine pattern in Narcissus confusus plants. *Planta medica,* 69**,** 1166-1168.

Loyola-Vargas, V. M., De-La-Peña, C., Galaz-Avalos, R. & Quiroz-Figueroa, F. R. 2008. Plant tissue culture. *Molecular Biomethods Handbook***,** 875-904.

Lubbe, A., Choi, Y. H., Vreeburg, P. & Verpoorte, R. 2011. Effect of fertilizers on galanthamine and metabolite profiles in Narcissus bulbs by 1H NMR. *Journal of agricultural and food chemistry,* 59**,** 3155-3161.

Lubbe, A., Gude, H., Verpoorte, R. & Choi, Y. H. 2013. Seasonal accumulation of major alkaloids in organs of pharmaceutical crop Narcissus Carlton. *Phytochemistry,* 88**,** 43-53.

Lubbe, A., Pomahačová, B., Choi, Y. H. & Verpoorte, R. 2010. Analysis of metabolic variation and galanthamine content in Narcissus bulbs by 1H NMR. *Phytochemical Analysis,* 21**,** 66-72.

Lubbe, A., Verpoorte, R. & Choi, Y. H. 2012. Effects of fungicides on galanthamine and metabolite profiles in Narcissus bulbs. *Plant physiology and biochemistry,* 58**,** 116-123.

Luk, L. Y., Bunn, S., Liscombe, D. K., Facchini, P. J. & Tanner, M. E. 2007. Mechanistic studies on norcoclaurine synthase of benzylisoquinoline alkaloid biosynthesis: an enzymatic Pictet-Spengler reaction. *Biochemistry,* 46**,** 10153-10161.

M Perez-De-Castro, A., Vilanova, S., Cañizares, J., Pascual, L., M Blanca, J., J Diez, M., Prohens, J. & Picó, B. 2012. Application of genomic tools in plant breeding. *Current genomics,* 13**,** 179-195.

MacKenzie, D. J., Mclean, M. A., Mukerji, S. & Green, M. 1997. Improved RNA extraction from woody plants for the detection of viral pathogens by reverse transcription-polymerase chain reaction. *Plant Disease,* 81**,** 222-226.

MacRae, E. 2007. Extraction of plant RNA. *Protocols for Nucleic Acid Analysis by Nonradioactive Probes***,** 15-24.

Mahmud, I., Thapaliya, M., BoroujerdI, A. & Chowdhury, K. 2014. NMR-based metabolomics study of the biochemical relationship between sugarcane callus tissues and their respective nutrient culture media. *Analytical and bioanalytical chemistry,* 406**,** 5997-6005.

Mahmud, I., Shrestha, B., Boroujerdi, A. & Chowdhury, K. 2015. NMR-based metabolomics profile comparisons to distinguish between embryogenic and non-embryogenic callus tissue of sugarcane at the biochemical level. *In Vitro Cellular & Developmental Biology-Plant,* 51**,** 340-349.

Maheswari, T. & Ravi, M. 2016. Next Generation Sequencing–Technical Report. *LIFE SCIENCE RESEARCH,* 1 (06).

Malik, M. 2008. Comparison of different liquid/solid culture systems in the production of somatic embryos from Narcissus L. ovary explants. *Plant Cell, Tissue and Organ Culture,* 94**,** 337-345.

Marco-Contelles, J., Do Carmo Carreiras, M., Rodríguez, C., Villarroya, M. & García, A. G. 2006. Synthesis and pharmacology of galantamine. *Chemical reviews,* 106**,** 116-133.

Marioni, J. C., Mason, C. E., Mane, S. M., Stephens, M. & Gilad, Y. 2008. RNA-seq: an assessment of technical reproducibility and comparison with gene expression arrays. *Genome research,* 18**,** 1509-1517.

Martin, M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet. journal,* 17**,** pp. 10-12.

Masi, M., Frolova, L. V., Yu, X., Mathieu, V., Cimmino, A., De Carvalho, A., Kiss, R., Rogelj, S., Pertsemlidis, A. & Kornienko, A. 2015. Jonquailine, a new pretazettine-type alkaloid isolated from Narcissus jonquilla quail, with activity against drug-resistant cancer. *Fitoterapia,* 102**,** 41-48.

Matsuda, F., Hirai, M. Y., Sasaki, E., Akiyama, K., Yonekura-Sakakibara, K., Provart, N. J., Sakurai, T., Shimada, Y. & Saito, K. 2010. AtMetExpress development: a phytochemical atlas of Arabidopsis development. *Plant physiology,* 152**,** 566-578.

Mazid, M., Khan, T. & Mohammad, F. 2011. Role of secondary metabolites in defence mechanisms of plants. *Biology and medicine,* 3**,** 232-249.

Menke, F. L., Parchmann, S., Mueller, M. J., Kijne, J. W. & Memelink, J. 1999. Involvement of the octadecanoid pathway and protein phosphorylation in fungal elicitor-induced expression of terpenoid indole alkaloid biosynthetic genes in Catharanthus roseus. *Plant physiology,* 119**,** 1289-1296.

Metzker, M. L. 2010. Sequencing technologies—the next generation. *Nature reviews genetics,* 11**,** 31-46.

Mewis, I., Smetanska, I. M., Müller, C. T. & Ulrichs, C. 2011. Specific poly-phenolic compounds in cell culture of Vitis vinifera L. cv. Gamay Fréaux. *Applied biochemistry and biotechnology,* 164**,** 148-161.

Mizrachi, E., Hefer, C. A., Ranik, M., Joubert, F. & Myburg, A. A. 2010. De novo assembled expressed gene catalog of a fast-growing Eucalyptus tree produced by Illumina mRNA-Seq. *BMC Genomics,* 11**,** 1.

Mizutani, M. & Sato, F. 2011. Unusual P450 reactions in plant secondary metabolism. *Archives of Biochemistry and Biophysics,* 507**,** 194-203.

Molnár, Z., Virág, E. & Ordog, V. 2011. Natural substances in tissue culture media of higher plants. *Acta Biologica Szegediensis,* 55**,** 123-127.

Mortazavi, A., Williams, B. A., Mccue, K., Schaeffer, L. & Wold, B. 2008. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nature methods,* 5**,** 621-628.

Mroczek, T. & Mazurek, J. 2009. Pressurized liquid extraction and anticholinesterase activity-based thin-layer chromatography with bioautography of Amaryllidaceae alkaloids. *Analytica chimica acta,* 633**,** 188-196.

Mu, H. M., Wang, R., Li, X. D., Jiang, Y. M., Wang, C. Y., Quan, J. P., Peng, F. & Xia, B. 2009. Effect of abiotic and biotic elicitors on growth and alkaloid accumulation of Lycoris chinensis seedlings. *Zeitschrift für Naturforschung C,* 64**,** 541-550.

Muhitch, M. J. & Fletcher, J. S. 1985. Influence of culture age and spermidine treatment on the accumulation of phenolic compounds in suspension cultures. *Plant physiology,* 78**,** 25-28.

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.

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

Murthy, H. N., Georgiev, M. I., Park, S. Y., Dandin, V. S. & Paek, K. Y. 2015. The safety assessment of food ingredients derived from plant cell, tissue and organ cultures: A review. *Food Chemistry,* 176**,** 426-432.

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.

Nalawade, S. M. & Tsay, H. S. 2004. In vitro propagation of some important Chinese medicinal plants and their sustainable usage. *In Vitro Cellular & Developmental Biology-Plant,* 40**,** 143-154.

Namdeo, A. 2007. Plant cell elicitation for production of secondary metabolites: a review. *Pharmacogn Rev,* 1**,** 69-79.

Namdeo, A., Patil, S. & Fulzele, D. P. 2002. Influence of fungal elicitors on production of ajmalicine by cell cultures of Catharanthus roseus. *Biotechnology progress,* 18**,** 159-162.

Naoumkina, M., Thyssen, G. N. & Fang, D. D. 2015. RNA-seq analysis of short fiber mutants Ligon-lintless-1 (Li 1) and–2 (Li 2) revealed important role of aquaporins in cotton (Gossypium hirsutum L.) fiber elongation. *BMC plant biology,* 15**,** 1.

Naurin, S., Bensch, S., Hansson, B., Johansson, T., Clayton, D. F., Albrekt, A. S., Von Schantz, T. & Hasselquist, D. 2008. TECHNICAL ADVANCES: A microarray for large-scale genomic and transcriptional analyses of the zebra finch (Taeniopygia guttata) and other passerines. *Molecular ecology resources,* 8**,** 275-281.

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 Reports,* 31**,** 597-620.

Ness, R. W., Siol, M. & Barrett, S. C. 2011. De novo sequence assembly and characterization of the floral transcriptome in cross-and self-fertilizing plants. *BMC Genomics,* 12**,** 1.

Novaes, E., Drost, D. R., Farmerie, W. G., Pappas, G. J., Grattapaglia, D., Sederoff, R. R. & Kirst, M. 2008. High-throughput gene and SNP discovery in Eucalyptus grandis, an uncharacterized genome. *BMC Genomics,* 9**,** 1.

Ober, D. 2005. Seeing double: gene duplication and diversification in plant secondary metabolism. *Trends in plant science,* 10**,** 444-449.

Ogata, H., Goto, S., Sato, K., Fujibuchi, W., Bono, H. & Kanehisa, M. 1999. KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic acids research,* 27**,** 29-34.

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. *Journal of Experimental Botany,* 58**,** 3373-3383.

Oksman-Caldentey, K. M. & Inzé, D. 2004. Plant cell factories in the post-genomic era: new ways to produce designer secondary metabolites. *Trends in plant science,* 9**,** 433-440.

Oksman-Caldentey, K. M. & Saito, K. 2005. Integrating genomics and metabolomics for engineering plant metabolic pathways. *Current opinion in biotechnology,* 16**,** 174-179.

Oliver, S. L., Lenards, A. J., Barthelson, R. A., Merchant, N. & Mckay, S. J. 2013. Using the iPlant collaborative discovery environment. *Current Protocols in Bioinformatics***,** 1.22. 1-1.22. 26.

Osier, T. L. & Lindroth, R. L. 2001. Effects of genotype, nutrient availability, and defoliation on aspen phytochemistry and insect performance. *Journal of chemical ecology,* 27**,** 1289-1313.

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). *Phytochemistry Letters,* 3**,** 161-163.

Osorio, S., Ruan, Y.L. and Fernie, A.R., 2014. An update on source-to-sink carbon partitioning in tomato. *Frontiers in plant science*, *5*, p.516.

Ouzounidou, G. & Ilias, I. 2005. Hormone-induced protection of sunflower photosynthetic apparatus against copper toxicity. *Biologia Plantarum,* 49**,** 223-228.

Ozsolak, F. 2012. Third-generation sequencing techniques and applications to drug discovery. *Expert opinion on drug discovery,* 7**,** 231-243.

Ozsolak, F. & Milos, P. M. 2011. RNA sequencing: advances, challenges and opportunities. *Nature reviews genetics,* 12**,** 87-98.

Pages, H., Aboyoun, P., Gentleman, R. & Debroy, S. 2009. String objects representing biological sequences, and matching algorithms. *R package version,* 2(2).

Palama, T. L., Menard, P., Fock, I., Choi, Y. H., Bourdon, E., Govinden-Soulange, J., Bahut, M., Payet, B., Verpoorte, R. & Kodja, H. 2010. Shoot differentiation from protocorm callus cultures of Vanilla planifolia (Orchidaceae): proteomic and metabolic responses at early stage. *BMC plant biology,* 10**,** 1.

Pant, B. D., Pant, P., Erban, A., Huhman, D., Kopka, J. & Scheible, W. R. 2015. Identification of primary and secondary metabolites with phosphorus status-dependent abundance in Arabidopsis, and of the transcription factor PHR1 as a major regulator of metabolic changes during phosphorus limitation. *Plant, cell & environment,* 38**,** 172-187.

Parida, A.K. and Das, A.B., 2005. Salt tolerance and salinity effects on plants: a review. *Ecotoxicology and environmental safety*, *60*(3), pp.324-349.

Parimalan, R., Giridhar, P. & Ravishankar, G. 2008. Mass multiplication of Bixa orellana L. through tissue culture for commercial propagation. *Industrial Crops and Products,* 28**,** 122-127.

Park, M. R., Wang, Y. H. & Hasenstein, K. H. 2014. Profiling gene expression in germinating Brassica roots. *Plant molecular biology reporter,* 32**,** 541-548.

Paszkiewicz, K. & Studholme, D. J. 2010. De novo assembly of short sequence reads. *Briefings in bioinformatics***,** bbq020.

Pavlov, A., Berkov, S., Courot, E., Gocheva, T., Tuneva, D., Pandova, B., Georgiev, M., Georgiev, V., Yanev, S. & Burrus, M. 2007. Galanthamine production by Leucojum aestivum in vitro systems. *Process Biochemistry,* 42**,** 734-739.

Pearce, G., Strydom, D., Johnson, S. & Ryan, C. A. 1991. A polypeptide from tomato leaves induces wound-inducible proteinase inhibitor proteins. *Science,* 253**,** 895-897.

Pease, J. & Sooknanan, R. 2012. A rapid, directional RNA-seq library preparation workflow for Illumina® sequencing. *Nat Methods Applic Note,* 9**,** i-ii.

Phalaraksh, C., Lenz, E., Lindon, J., Nicholson, J., Farrant, R., Reynolds, S., Wilson, I., Osborn, D. & Weeks, J. 1999. NMR spectroscopic studies on the haemolymph of the tobacco hornworm, Manduca sexta: assignment of 1 H and 13 C NMR spectra. *Insect biochemistry and molecular biology,* 29**,** 795-805.

Phan, H.T., Ellwood, S.R., Adhikari, K., Nelson, M.N. and Oliver, R.P., 2007. The first genetic and comparative map of white lupin (Lupinus albus L.): identification of QTLs for anthracnose resistance and flowering time, and a locus for alkaloid content. *DNA research*, *14*(2), pp.59-70.

Pop, M. & Salzberg, S. L. 2008. Bioinformatics challenges of new sequencing technology. *Trends in Genetics,* 24**,** 142-149.

Prabhudas, S. K., Raju, B., Kannan Thodi, S., Parani, M. & Natarajan, P. 2015. The complete chloroplast genome sequence of Indian mustard (Brassica juncea L.). *Mitochondrial DNA***,** 1-2.

Ptak, A., El Tahchy, A., Dupire, F., Boisbrun, M., Henry, M., Chapleur, Y., Mos, M. & Laurain-Mattar, D. 2008. LCMS and GCMS for the screening of alkaloids in natural and in vitro extracts of Leucojum aestivum. *Journal of natural products,* 72**,** 142-147.

Ptak, A., Tahchy, A. E., Wyżgolik, G., Henry, M. & Laurain-Mattar, D. 2010. Effects of ethylene on somatic embryogenesis and galanthamine content in Leucojum aestivum L. cultures. *Plant Cell, Tissue and Organ Culture (PCTOC),* 102**,** 61-67.

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.

Qing, T., Yu, Y., Du, T. and Shi, L., 2013. mRNA enrichment protocols determine the quantification characteristics of external RNA spike-in controls in RNA-Seq studies. *Science China. Life Sciences*, *56*(2), p.134.

Radman, R., Saez, T., Bucke, C. & Keshavarz, T. 2003. Elicitation of plants and microbial cell systems. *Biotechnology and Applied Biochemistry,* 37**,** 91-102.

Raman, S. B. & Rathinasabapathi, B. 2003. β-Alanine N-methyltransferase of Limonium latifolium. cDNA cloning and functional expression of a novel N-methyltransferase implicated in the synthesis of the osmoprotectant β-alanine betaine. *Plant physiology,* 132**,** 1642-1651.

Rao, S. R. & Ravishankar, G. 2002. Plant cell cultures: chemical factories of secondary metabolites. *Biotechnology advances,* 20**,** 101-153.

Rezaei, M., Naghavi, M. R., Hoseinzade, A. H. & Abbasi, A. 2016. Developmental accumulation of thebaine and some gene transcripts in different organs of Papaver bracteatum. *Industrial Crops and Products,* 80**,** 262-268.

Rhee, S. Y., Wood, V., Dolinski, K. & Draghici, S. 2008. Use and misuse of the gene ontology annotations. *Nature reviews genetics,* 9**,** 509-515.

Rhee, I. K., Appels, N., Hofte, B., Karabatak, B., Erkelens, C., Stark, L. M., Flippin, L. A. & Verpoorte, R. 2004. Isolation of the acetylcholinesterase inhibitor ungeremine from Nerine bowdenii by preparative HPLC coupled on-line to a flow assay system. *Biological and Pharmaceutical Bulletin,* 27**,** 1804-1809.

Rivero, R. M., Mestre, T. C., Mittler, R., Rubio, F., Garcia-Sanchez, F. & Martinez, V. 2014. The combined effect of salinity and heat reveals a specific physiological, biochemical and molecular response in tomato plants. *Plant, cell & environment,* 37**,** 1059-1073.

Robertson, G., Schein, J., Chiu, R., Corbett, R., Field, M., Jackman, S. D., Mungall, K., Lee, S., Okada, H. M. & Qian, J. Q. 2010. De novo assembly and analysis of RNA-seq data. *Nature methods,* 7**,** 909-912.

Roewer, I., Cloutier, N., Nessler, C. & De Luca, V. 1992. Transient induction of tryptophan decarboxylase (TDC) and strictosidine synthase (SS) genes in cell suspension cultures of Catharanthus roseus. *Plant Cell Reports,* 11**,** 86-89.

Rønsted, N., Savolainen, V., Mølgaard, P. & Jäger, A. K. 2008. Phylogenetic selection of Narcissus species for drug discovery. *Biochemical Systematics and Ecology,* 36**,** 417-422.

Safdari, Y. & Kazemitabar, S. K. 2010. Direct shoot regeneration, callus induction and plant regeneration from callus tissue in Mose Rose (Portulaca grandiflora L.). *Plant Omics,* 3**,** 47.

Sage, D. Propagation and protection of flower Bulbs: Current approaches and future prospects, with special reference to Narcissus. IX International Symposium on Flower Bulbs 673, 2004. 323-334.

Sage, D. O., Lynn, J. & Hammatt, N. 2000. Somatic embryogenesis in Narcissus pseudonarcissus cvs. Golden Harvest and St. Keverne. *Plant science,* 150**,** 209-216.

Saito, K. & Matsuda, F. 2010. Metabolomics for functional genomics, systems biology, and biotechnology. *Annual review of plant biology,* 61**,** 463-489.

Saliba, S., Ptak, A., Boisbrun, M., Spina, R., Dupire, F. & Laurain-Mattar, D. 2016. Stimulating effect of both 4'-O-methylnorbelladine feeding and temporary immersion conditions on galanthamine and lycorine production by Leucojum aestivum L. bulblets. *Engineering in Life Sciences*, 00, 1–9

Saliba, S., Ptak, A. & Laurain-Mattar, D. 2015. 4′-O-Methylnorbelladine feeding enhances galanthamine and lycorine production by Leucojum aestivum L. shoot cultures. *Engineering in Life Sciences,* 15**,** 640-645.

Sánchez-Sampedro, M. A., Fernández-Tárrago, J. & Corchete, P. 2005. Yeast extract and methyl jasmonate-induced silymarin production in cell cultures of Silybum marianum (L.) Gaertn. *Journal of biotechnology,* 119**,** 60-69.

Sanger, F. & Coulson, A. R. 1975. A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase. *Journal of molecular biology,* 94**,** 441-448.

Sangwan, R. S., Tripathi, S., Singh, J., Narnoliya, L. K. & Sangwan, N. S. 2013. De novo sequencing and assembly of Centella asiatica leaf transcriptome for mapping of structural, functional and regulatory genes with special reference to secondary metabolism. *Gene,* 525**,** 58-76.

Santos, A., Fidalgo, F., Santos, I. & Salema, R. 2002. In vitro bulb formation of Narcissus asturiensis, a threatened species of the Amaryllidaceae. *The Journal of Horticultural Science and Biotechnology,* 77**,** 149-152.

Satcharoen, V., Mclean, N. J., Kemp, S. C., Camp, N. P. & Brown, R. C. 2007. Stereocontrolled synthesis of (-)-galanthamine. *Organic letters,* 9**,** 1867-1869.

Schäfer, H. & Wink, M. 2009. Medicinally important secondary metabolites in recombinant microorganisms or plants: progress in alkaloid biosynthesis. *Biotechnology journal,* 4**,** 1684-1703.

Schmittgen, T. D. & Zakrajsek, B. A. 2000. Effect of experimental treatment on housekeeping gene expression: validation by real-time, quantitative RT-PCR. *Journal of biochemical and biophysical methods,* 46**,** 69-81.

Schripsema, J., Caprini, G. P., Van Der Heijden, R., Bino, R., De Vos, R. & Dagnino, D. 2007. Iridoids from Pentas lanceolata. *Journal of natural products,* 70**,** 1495-1498.

Schripsema, J. 2010. Application of NMR in plant metabolomics: techniques, problems and prospects. *Phytochemical Analysis,* 21**,** 14-21.

Schulte, J. H., Marschall, T., Martin, M., Rosenstiel, P., Mestdagh, P., Schlierf, S., Thor, T., Vandesompele, J., Eggert, A. & Schreiber, S. 2010. Deep sequencing reveals differential expression of microRNAs in favorable versus unfavorable neuroblastoma. *Nucleic acids research***,** gkq342.

Schumann, A., Berkov, S., Claus, D., Gerth, A., Bastida, J. & Codina, C. 2012. Production of galanthamine by Leucojum aestivum shoots grown in different bioreactor systems. *Applied biochemistry and biotechnology,* 167**,** 1907-1920.

Schumann, A., Torras-Claveria, L., Berkov, S., Claus, D., Gerth, A., Bastida, J. & CODINA, C. 2013. Elicitation of galanthamine production by Leucojum aestivum shoots grown in temporary immersion system. *Biotechnology progress,* 29**,** 311-318.

Scotto-Lavino, E., G. Du and Frohman, M. A. (2006). 5′ end cDNA amplification using classic RACE. *Nature protocols,* 1(6): 2555-2562.

Schuster, S., Fell, D. A. & Dandekar, T. 2000. A general definition of metabolic pathways useful for systematic organization and analysis of complex metabolic networks. *Nature biotechnology,* 18**,** 326-332.

Seabrook, J. E. & Cumming, B. G. 1982. In vitro morphogenesis and growth of Narcissus in response to temperature. *Scientia horticulturae,* 16**,** 185-190.

Sellés, M., Bergoñón, S., Viladomat, F., Bastida, J. & Codina, C. 1997. Effect of sucrose on growth and galanthamine production in shoot-clump cultures of Narcissus confusus in liquid-shake medium. *Plant Cell, Tissue and Organ Culture,* 49**,** 129-136.

Selles, M., Viladomat, F., Bastida, J. & Codina, C. 1999. Callus induction, somatic embryogenesis and organogenesis in Narcissus confusus: correlation between the state of differentiation and the content of galanthamine and related alkaloids. *Plant Cell Reports,* 18**,** 646-651.

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. *Journal of plant physiology,* 179**,** 40-55.

Sharma, Y. & Kanwar, S. B. Studies on micropropagation of tulips and daffodils. XXVI International Horticultural Congress: Elegant Science in Floriculture 624, 2002. 533-540.

Shawky, E., Abou-Donia, A. H., Darwish, F. A., Toaima, S. M., Takla, S. S., Pigni, N. B. & Bastida, J. 2015. HPTLC and GC/MS Study of Amaryllidaceae Alkaloids of Two Narcissus Species. *Chemistry & biodiversity,* 12**,** 1184-1199.

Simoh, S., Quintana, N., Kim, H. K., Choi, Y. H. & Verpoorte, R. 2009. Metabolic changes in Agrobacterium tumefaciens-infected Brassica rapa. *Journal of plant physiology,* 166**,** 1005-1014.

Singh, A. & Desgagné-Penix, I. 2014. Biosynthesis of the Amaryllidaceae alkaloids. *Plant Science Today,* 1**,** 114-120.

Singh, M. & Chaturvedi, R. 2012. Screening and quantification of an antiseptic alkylamide, spilanthol from in vitro cell and tissue cultures of Spilanthes acmella Murr. *Industrial Crops and Products,* 36**,** 321-328.

Skidmore, E., Kim, S. J., Kuchimanchi, S., Singaram, S., Merchant, N. & Stanzione, D. iPlant atmosphere: a gateway to cloud infrastructure for the plant sciences. Proceedings of the 2011 ACM workshop on Gateway computing environments, 2011. ACM, 59-64.

Smetanska, I. 2008. Production of secondary metabolites using plant cell cultures. *Food biotechnology,* (pp. 187-228). Springer Berlin Heidelberg.

Smith, J. E. & Nair, R. 2005. The architecture of virtual machines. *Computer,* 38**,** 32-38.

Sochacki, D. & Orlikowska, T. 2005. The obtaining of narcissus plants free from potyviruses via adventitious shoot regeneration in vitro from infected bulbs. *Scientia horticulturae,* 103**,** 219-225.

Srivastava, N.K. and Srivastava, A.K., 2006. Utilization of exogenously supplied 14 C-saccharose into primary metabolites and alkaloid production in Catharanthus roseus. *Photosynthetica*, *44*(3), pp.478-480.

Stahl, E. 2003. The Secondary Metabolism of Plants: Secondary Defence Compounds; Online article; © Peter v. Sengbusch.

Staikidou, I., Selby, C. & Arvey, B. 1994. Stimulation by auxin and sucrose of bulbil formation in vitro by single leaf cultures of Narcissus. *New phytologist,* 127**,** 315-320.

Staikidou, I., Watson, S., Harvey, B. M. & Selby, C. 2005. Narcissus bulblet formation in vitro: effects of carbohydrate type and osmolarity of the culture medium. *Plant Cell, Tissue and Organ Culture,* 80**,** 313-320.

Stavrinides, A., Tatsis, E. C., Foureau, E., Caputi, L., Kellner, F., Courdavault, V. & O'connor, S. E. 2015. Unlocking the diversity of alkaloids in Catharanthus roseus: nuclear localization suggests metabolic channeling in secondary metabolism. *Chemistry & biology,* 22**,** 336-341.

Stewart, C., 2007. Use of coumarin derivatives in antifungal therapy. U.S. Patent Application 12/303,958.

Stöckigt, J., Barleben, L., Panjikar, S. & Loris, E. A. 2008. 3D-Structure and function of strictosidine synthase–the key enzyme of monoterpenoid indole alkaloid biosynthesis. *Plant physiology and biochemistry,* 46**,** 340-355.

Strickler, S., Bombarely, A. & Mueller, L. 2012. Designing a transcriptome next-generation sequencing project for a nonmodel plant species. *American journal of botany,* 99(2): 257–266.

Sugiyama, A., Linley, P. J., Sasaki, K., Kumano, T., Yamamoto, H., Shitan, N., Ohara, K., Takanashi, K., Harada, E. & Hasegawa, H. 2011. Metabolic engineering for the production of prenylated polyphenols in transgenic legume plants using bacterial and plant prenyltransferases. *Metabolic engineering,* 13**,** 629-637.

Surget-Groba, Y. & Montoya-BurgoS, J. I. 2010. Optimization of de novo transcriptome assembly from next-generation sequencing data. *Genome research,* 20**,** 1432-1440.

Suzuki, Y., K. Yoshitomo-Nakagawa, et al. (1997). Construction and characterization of a full length-enriched and a 5′-end-enriched cDNA library. *Gene,* 200(1–2): 149-156.

Swarbreck, S. M., Lindquist, E. A., Ackerly, D. D. & Andersen, G. L. 2011. Analysis of leaf and root transcriptomes of soil-grown Avena barbata plants. *Plant and cell physiology,* 52**,** 317-332.

Szabo, E., Thelen, A. & Petersen, M. 1999. Fungal elicitor preparations and methyl jasmonate enhance rosmarinic acid accumulation in suspension cultures of Coleus blumei. *Plant Cell Reports,* 18**,** 485-489.

Szlávik, L., Gyuris, Á., MinárovitS, J., Forgo, P., Molnár, J. & Hohmann, J. 2004. Alkaloids from Leucojum vernum and antiretroviral activity of Amaryllidaceae alkaloids. *Planta medica,* 70**,** 871-873.

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.

Takos, A. M. & Rook, F. 2013. Towards a molecular understanding of the biosynthesis of amaryllidaceae alkaloids in support of their expanding medical use. *International journal of molecular sciences,* 14**,** 11713-11741.

Taleb, A. M. A., Hamed, E. R., Shimaa, A. Z. & Adel, B. 2013. Enhancement of alkaloids production in tissue culture of Narcissus tazetta var. italicus I: Effect of growth regulators and fungal elicitors. *Journal of Agricultural Technology,* 9**,** 503-514.

Tanimoto, H., Kato, T. & Chida, N. 2007. Total synthesis of (+)-galanthamine starting from D-glucose. *Tetrahedron letters,* 48**,** 6267-6270.

Tasseva, G., Richard, L. & Zachowski, A. 2004. Regulation of phosphatidylcholine biosynthesis under salt stress involves choline kinases in Arabidopsis thaliana. *FEBS letters,* 566**,** 115-120.

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 research***,** gkv1179.

Tolstikov, V. V. and O. Fiehn (2002). Analysis of highly polar compounds of plant origin: combination of hydrophilic interaction chromatography and electrospray ion trap mass spectrometry. *Analytical biochemistry,* 301(2): 298-307.

Tong, Z., Gao, Z., Wang, F., Zhou, J. & Zhang, Z. 2009. Selection of reliable reference genes for gene expression studies in peach using real-time PCR. *BMC Molecular Biology,* 10**:** 71.

Torras-Claveria, L., Berkov, S., Codina, C., Viladomat, F. & Bastida, J. 2013. Daffodils as potential crops of galanthamine. Assessment of more than 100 ornamental varieties for their alkaloid content and acetylcholinesterase inhibitory activity. *Industrial Crops and Products,* 43**,** 237-244.

Torras-Claveria, L., Berkov, S., Codina, C., Viladomat, F. & Bastida, J. 2014. Metabolomic analysis of bioactive Amaryllidaceae alkaloids of ornamental varieties of Narcissus by GC–MS combined with k-means cluster analysis. *Industrial Crops and Products,* 56**,** 211-222.

Trapnell, C., Pachter, L. & Salzberg, S. L. 2009. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics,* 25**,** 1105-1111.

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. *Nature protocols,* 7**,** 562-578.

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. *Nature biotechnology,* 28**,** 511-515.

Trick, M., Long, Y., Meng, J. & Bancroft, I. 2009. Single nucleotide polymorphism (SNP) discovery in the polyploid Brassica napus using Solexa transcriptome sequencing. *Plant biotechnology journal,* 7**,** 334-346.

Trujillo-Villanueva, K., Rubio-Piña, J., Monforte-González, M. & Vázquez-Flota, F. 2010. Fusarium oxysporum homogenates and jasmonate induce limited sanguinarine accumulation in Argemone mexicana cell cultures. *Biotechnology letters,* 32**,** 1005-1009.

Untergasser, A., Cutcutache, I., Koressaar, T., Ye, J., Faircloth, B. C., Remm, M. & Rozen, S. G. 2012. Primer3—new capabilities and interfaces. *Nucleic acids research,* 40**,** e115-e115.

Van Der Kooy, F., Maltese, F., Choi, Y. H., Kim, H. K. & Verpoorte, R. 2009. Quality control of herbal material and phytopharmaceuticals with MS and NMR based metabolic fingerprinting. *Planta medica,* 75**,** 763-775.

Van Den Berg, R. A., Hoefsloot, H. C., Westerhuis, J. A., Smilde, A. K. & Van Der Werf, M. J. 2006. Centering, scaling, and transformations: improving the biological information content of metabolomics data. *BMC Genomics,* 7**:** 142.

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

Vandesompele, J., De Preter, K., Pattyn, F., Poppe, B., Van Roy, N., De Paepe, A. & Speleman, F. 2002. Accurate normalisation of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome biology, 3*(7), p.1.

Verma, S. K., Sahin, G., Yucesan, B., Eker, I., Sahbaz, N., Gurel, S. & Gurel, E. 2012. Direct somatic embryogenesis from hypocotyl segments of Digitalis trojana Ivan and subsequent plant regeneration. *Industrial Crops and Products,* 40**,** 76-80.

Vasil, I. K. & Thorpe, T. A. eds., 2013. *Plant cell and tissue culture*, Springer Science & Business Media.

Vázquez-Flota, F., Hernández-Domínguez, E., De Lourdes Miranda-Ham, M. & Monforte-González, M. 2009. A differential response to chemical elicitors in Catharanthus roseus in vitro cultures. *Biotechnology letters,* 31**,** 591-595.

Verpoorte, R., Choi, Y. & Kim, H. 2007. NMR-based metabolomics at work in phytochemistry. *Phytochemistry Reviews,* 6**,** 3-14.

Verpoorte, R., Contin, A. & Memelink, J. 2002. Biotechnology for the production of plant secondary metabolites. *Phytochemistry Reviews,* 1**,** 13-25.

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. *Molecular Ecology,* 22**,** 620-634.

Vinterhalter, B., Janković, T., Šavikin, K., Nikolić, R. & Vinterhalter, D. 2008. Propagation and xanthone content of Gentianella austriaca shoot cultures. *Plant Cell, Tissue and Organ Culture,* 94**,** 329-335.

Wall, P. K., Leebens-Mack, J., Chanderbali, A. S., Barakat, A., Wolcott, E., Liang, H., Landherr, L., Tomsho, L. P., Hu, Y. & Carlson, J. E. 2009. Comparison of next generation sequencing technologies for transcriptome characterization. *BMC Genomics, 10* (1), p.1.

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.

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.

Ward, J. A., Ponnala, L. & Weber, C. A. 2012. Strategies for transcriptome analysis in nonmodel plants. *American journal of botany,* 99**,** 267-276.

Ward, J. L., Baker, J. M. & Beale, M. H. 2007. Recent applications of NMR spectroscopy in plant metabolomics. *Febs Journal,* 274**,** 1126-1131.

Weber, A. P., Weber, K. L., Carr, K., Wilkerson, C. & Ohlrogge, J. B. 2007. Sampling the Arabidopsis transcriptome with massively parallel pyrosequencing. *Plant physiology,* 144**,** 32-42.

Weng, R., Shen, S., Tian, Y., Burton, C., Xu, X., Liu, Y., Chang, C., Bai, Y. & Liu, H. 2015. Metabolomics approach reveals integrated metabolic network associated with serotonin deficiency. *Scientific reports,* 5,doi: 10.1038/srep11864.

Wibberley, M. S., Lenton, J. R. & Neill, S. J. 1994. Sesquiterpenoid phytoalexins produced by hairy roots of Nicotiana tabacum. *Phytochemistry,* 37**,** 349-351.

Widarto, H. T., Van Der Meijden, E., Lefeber, A. W., Erkelens, C., Kim, H. K., Choi, Y. H. & Verpoorte, R. 2006. Metabolomic differentiation of Brassica rapa following herbivory by different insect instars using two-dimensional nuclear magnetic resonance spectroscopy. *Journal of chemical ecology,* 32**,** 2417-2428.

Wilson, M. A. 2013. *NMR Techniques & Applications in Geochemistry & Soil Chemistry* (Book), Elsevier.

Wiktorowska, E., Długosz, M. & Janiszowska, W. 2010. Significant enhancement of oleanolic acid accumulation by biotic elicitors in cell suspension cultures of Calendula officinalis L. *Enzyme and Microbial Technology,* 46**,** 14-20.

Wolf, J. B. 2013. Principles of transcriptome analysis and gene expression quantification: an RNA-seq tutorial. *Molecular ecology resources,* 13**,** 559-572.

Wolf, J. B., Bayer, T., Haubold, B., Schilhabel, M., Rosenstiel, P. & Tautz, D. 2010. Nucleotide divergence vs. gene expression differentiation: comparative transcriptome sequencing in natural isolates from the carrion crow and its hybrid zone with the hooded crow. *Molecular Ecology,* 19**,** 162-175.

Wishart, D. S., Jewison, T., Guo, A. C., Wilson, M., Knox, C., Liu, Y., Djoumbou, Y., Mandal, R., Aziat, F. & Dong, E. 2012. HMDB 3.0—the human metabolome database in 2013. *Nucleic acids research***,** gks1065.

Wu, D., Cai, S., Chen, M., Ye, L., Chen, Z., Zhang, H., Dai, F., Wu, F. & Zhang, G. 2013. Tissue metabolic responses to salt stress in wild and cultivated barley. *PloS one,* 8**,** e55431.

Wu, S., Schalk, M., Clark, A., Miles, R. B., Coates, R. & Chappell, J. 2006. Redirection of cytosolic or plastidic isoprenoid precursors elevates terpene production in plants. *Nature biotechnology,* 24**,** 1441-1447.

Xia, J., Sinelnikov, I. V., Han, B. & Wishart, D. S. 2015. MetaboAnalyst 3.0—making metabolomics more meaningful. *Nucleic acids research,* 43**,** W251-W257.

Xia, J. & Wishart, D. S. 2010. MetPA: a web-based metabolomics tool for pathway analysis and visualization. *Bioinformatics,* 26**,** 2342-2344.

Xia, J., Mandal, R., Sinelnikov, I. V., Broadhurst, D. & Wishart, D. S. 2012. MetaboAnalyst 2.0—a comprehensive server for metabolomic data analysis. *Nucleic acids research,* 40**,** W127-W133.

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. *Journal of biotechnology,* 166**,** 122-134.

Xie, Y., Wu, G., Tang, J., Luo, R., Patterson, J., Liu, S., Huang, W., He, G., Gu, S. & Li, S. 2014. SOAPdenovo-Trans: de novo transcriptome assembly with short RNA-Seq reads. *Bioinformatics,* 30**,** 1660-1666.

Yadava, R. & Chawla, H. 2002. Role of genotypes, growth regulators and amino acids on callus induction and plant regeneration from different developmental stages of inflorescence in wheat. *Indian Journal of Genetics and Plant Breeding,* 62**,** 55-60.

Yan, Q., Hu, Z., Tan, R. X. & Wu, J. 2005. Efficient production and recovery of diterpenoid tanshinones in Salvia miltiorrhiza hairy root cultures with in situ adsorption, elicitation and semi-continuous operation. *Journal of biotechnology,* 119**,** 416-424.

Yanagawa, T. Propagation of bulbous ornamentals by simple cultures of bulb-scale segments using plastic vessels. IX International Symposium on Flower Bulbs 673, 2004. 343-348.

Yang, S. O., Kim, S.H., Kim, Y., Kim, H. S., Chun, Y. J. & Choi, H. K. 2009. Metabolic discrimination of Catharanthus roseus calli according to their relative locations using 1H-NMR and principal component analysis. *Bioscience, biotechnology, and biochemistry,* 73**,** 2032-2036.

Yang, S. Y., Kim, H. K., Lefeber, A. W., Erkelens, C., Angelova, N., Choi, Y. H. & Verpoorte, R. 2006. Application of two-dimensional nuclear magnetic resonance spectroscopy to quality control of ginseng commercial products. *Planta medica,* 72**,** 364-369.

Yang, H., Hu, L., Hurek, T. & Reinhold-Hurek, B. 2010. Global characterization of the root transcriptome of a wild species of rice, Oryza longistaminata, by deep sequencing. *BMC Genomics, 11* (1), p.1.

Yang, S. O., Kim, S. H., Kim, Y., Kim, H. S., Chun, Y. J. & Choi, H.-K. 2009. Metabolic discrimination of Catharanthus roseus calli according to their relative locations using 1H-NMR and principal component analysis. *Bioscience, biotechnology, and biochemistry,* 73**,** 2032-2036.

Ye, X., Al-Babili, S., Klöti, A., Zhang, J., Lucca, P., Beyer, P. & Potrykus, I. 2000. Engineering the provitamin A (β-carotene) biosynthetic pathway into (carotenoid-free) rice endosperm. *Science,* 287**,** 303-305.

Yuan, J. S., Reed, A., Chen, F. & Stewart, C. N. 2006. Statistical analysis of real-time PCR data. *BMC bioinformatics, 7* (1), p.1.

Yun, D. J., Hashimoto, T. & Yamada, Y. 1992. Metabolic engineering of medicinal plants: transgenic Atropa belladonna with an improved alkaloid composition. *Proceedings of the National Academy of Sciences,* 89**,** 11799-11803.

Zeng, Y. & Yang, T. 2002. RNA isolation from highly viscous samples rich in polyphenols and polysaccharides. *Plant molecular biology reporter,* 20**,** 417-417.
Zhao, J., Davis, L. C. & Verpoorte, R. 2005. Elicitor signal transduction leading to production of plant secondary metabolites. *Biotechnology advances,* 23**,** 283-333.
Zhao, J., Fujita, K., Yamada, J. & Sakai, K. 2001a. Improved β-thujaplicin production in Cupressus lusitanica suspension cultures by fungal elicitor and methyl jasmonate. *Applied microbiology and biotechnology,* 55**,** 301-305.
Zhao, J., Zhu, W. H. & Hu, Q. 2001b. Enhanced catharanthine production in Catharanthus roseus cell cultures by combined elicitor treatment in shake flasks and bioreactors. *Enzyme and Microbial Technology,* 28**,** 673-681.
Zhao, N., Wang, G., Norris, A., Chen, X. & Chen, F. 2013. Studying plant secondary metabolism in the age of genomics. *Critical Reviews in Plant Sciences,* 32**,** 369-382.
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. *Applied microbiology and biotechnology,* 98**,** 2371-2383.
Zhi, H. J., Qin, X. M., Sun, H. F., Zhang, L. Z., Guo, X. Q. & Li, Z. Y. 2012. Metabolic Fingerprinting of Tussilago farfara L. Using 1H-NMR Spectroscopy and Multivariate Data Analysis. *Phytochemical Analysis,* 23**,** 492-501.
Arabidopsis Genome Initiative, 2000 (www.genomenewsnetwork.org)
Analytik Jena AG 2016 (www.analytik-jena.de)
BD Medical Technology, UK (www.bd.com)
Bhanot, 2014 (lab-training.com)
Beijing Genomics Institute, China (www.genomics.cn)
CyVerse (www.cyverse.org)
Chenomx Software (www.chenomx.com)
Eli Lily and Company Ltd. UK (www.lilly.co.uk)
GSK house, UK (www.gsk.com)
Gene Ontology Consortium, 2004 (www.geneontology.org)
GATC Biotech (www.gatc-biotech.com)
Illumina, 2010 (www.illumina.com)
Illumina, Inc. 2016 (www.illumina.com)
Kyoto Encyclopedia of Genes and Genomes (www.genome.jp/kegg)
MedPlant RNA Seq Database (www.medplantrnaseq.org)
MetaboAnalyst 3.0 (www.metaboanalyst.ca)
NanoDrop (www.nanodrop.com)
NHS, 2016 (www.nhs.uk)
Novocraft, 2010 (www.novocraft.com)
Plant Reactome Pathway Database (plantreactome.gramene.org)
Primer 3 (www.simgene.com/Primer3)
PhytoMetaSyn (www.phytometasyn.ca)
QuickGO-European Bioinformatics Institute (www.ebi.ac.uk/QuickGO-Beta)
Royal Horticultural Society, UK (www.rhs.org.uk)
Royal Botanic Gardens, Kew, UK (www.kew.org)
Ribo-Zero rRNA removal kit, Illumina (www.illumina.com/products/ribo-zero-rrna-removal-plant)
UniProt (www.uniprot.org)
Zhejiang Yixin Pharmaceutical Company Ltd. China (www.herbs-tech.com)
Shams(@MindArtRandom (twitter.com/mindartrandom)
www.3402bioinformaticsgroup.com