Bioinformatic and morphological characterization of *Catharanthus roseus* mutants

By

Graham George Jones

Submitted in partial fulfillment of the requirements for the degree of

Master of Science

Biotechnology

Faculty of Mathematics and Science, Brock University

St. Catharines, Ontario

Abstract

Catharanthus roseus, a member of the Apocynaceae family, has been studied extensively for its valuable chemotherapeutic monoterpenoid indole alkaloids (MIAs). Ethyl methanesuphonate (EMS) mutagenesis is a screening tool that has been used to look for altered MIA profiles in hope of discovering mutations of crucial MIA biosynthetic genes. Without a high-throughput mutation detection screen for *C. roseus* sequencing data, a range of techniques must be used to discover the EMS-induced changes within the plant.

Bioinformatic and morphological analysis revealed the likely alterations leading to unique MIA profiles in two *C. roseus* EMS mutants: the high-ajmalicine accumulating line M2-0754 and the low-MIA accumulating line M2-1582. Expression of geissoschizine synthase (GS) was downregulated almost seven-fold in the leaves of M2-0754, leading to the accumulation of an alternate pathway MIA from the labile intermediate. The low-MIA profile and increased auxin sensitivity of M2-1582 is likely due to the expression of a dysfunctional auxin influx transport protein homologue.

Acknowledgements

I'd like to thank my mother (Nicola Sims-Jones), sister (Alison Jones), extended family (Deb & Jim Milne et al.), business partners (Donald Craig Wiggins & Andrew Peter Udell), lab technique and instrumentation educators (Dr. Yang (Vince) Qu & Dr. Kyung hee Kim) and supervisory committee (Professor Jeffrey Atkinson, Professor Alan Castle, Professor Heather Gordon & Professor Michael Phillips) for your patience and support throughout my graduate studies.

Thank you also to Trevor Kidd and Danielle Williams for helping me improve my tissue culture knowledge and technique.

Table of Contents

List of Figures

Chapter 1: Literature review

Chapter 2: Geissoschizine synthase controls flux in the formation of monoterpenoid indole alkaloids in a *Catharanthus roseus* mutant

Chapter 3: Inter-organ transport of secologanin allows assembly of monoterpenoid indole alkaloids in a *Catharanthus roseus* mutant down-regulated in iridoid biosynthesis

List of Tables

Abbreviations

4CL - 4-coumarate-CoA ligase
7DLS - 7-deoxyloganic acid synthase
8-HGO - Geraniol 8-hydroxygeraniol dehydrogenase
10-HGO - 10-hydroxygeraniol dehydrogenase
16OMT - 16-hydroxytabersonine 16-O-methyltransferase
ACS - 1-aminocyclopropane-1-carboxylate synthase
BIS1 - Basic helix-loop-helix iridoid synthesis 1
BIS2 - Basic helix-loop-helix iridoid synthesis 2
Bp – Base pair
C3'H - p-coumaroyl shikimate 3'-hydroxylase
C4H - Cinnamate 4-hydroxylase
CAD - Cinnamyl alcohol dehydrogenase
CCoAOMT - Caffeoyl-CoA O-methyltransferase
CCR - Cinnamoyl-CoA reductase
CKX - Cytokinin dehydrogenase
COMT - Caffeic acid 3-O-methyltransferase
CPS - Copalyl diphosphate synthase
CS - Catharanthine synthase
Ct - Critical threshold
CYCD2 - cyclin D2
D4H - Desacetoxyvindoline 4-hydroxylase
DAT - Desacetylvindoline O-acetyltransferase
DL7H - 7-deoxyloganic acid 7-hydroxylase
DLGT - 7-deoxyloganetic acid glucosyltransferase
DNA – Deoxyribonucleic acid
DPAS - Dihydroprecodylocarpine acetate synthase
EDTA - Ethylenediaminetetraacetic acid
EIN3 - Ethylene insensitive 3-like
EMS - Ethyl methanesulfonate
F5H - Ferulate 5-hydroxylase
FLP – First leaf pair
FPKM – Fragments per kilobase of transcript per million mapped reads
G8O/G10H – Geraniol 8-oxidase/Geraniol 10-hydroxylase
GA – Gibberellin
GA2ox - Gibberellin 2-β-dioxygenases
GA20-ox - Gibberellin 20-oxidase
GAox – Gibberellin oxidase
Gbp – Giga base pair (1,000,000,000 bp)
GH3.6 – Indole-3-acetic acid-amido synthetase
GFW - Gram fresh weight
GO - Geissoschizine oxidase
GPP- Geraniol pyrophosphate
GS – Geissoschizine synthase
HCT - Hydroxycinnamoyl transferase
HFCA - 2-chloro-9-hydroxyfluorene-9-carboxylic acid

HYS – Heteroyohimbine synthase
IAA - Indole-3-acetic acid
IAM - Indole-3-acetamide
IAOx - Indole-3-acetaldoxime
IAN – Indole-3-acetonitrile
IBA – Indole-3-butyric acid
Ident - Identity
IPA – Indole-3-pyruvic acid
IPAP - Internal phloem-associated parenchyma
IPT - Adenylate isopentenyltransferase
IS - Iridoid synthase
JAR1 - Jasmonic acid-amido synthetase
KAO - Ent-kaurenoic acid oxidase
KD1 - Knotted1-like homeobox protein
KO - Ent-kaurene oxidase
KS - Ent-kaurene synthase
LAMT - Loganic acid O-methyltransferase
MAT - Minovincinine 19-*O*-actyltransferase
Mbp – Mega base pair (1,000,000 bp)
MeJA - Methyl Jasmonate
MIA - Mono-indole Alkaloid
NMT - N-methyltransferase
NPA - 1-N-naphthylphtalamic acid
P5βR – Progesterone 5β-reductase
PAL - Phenylalanine ammonia-lyase
PAS - Precodylocarpine acetate synthase
PIN - PIN-Formed proteins
qPCR - Qualitative polymerase chain reaction
REV – Revoluta
RNA - Ribonucleic acid
SA – Salicylic acid
SAT - Stemmadenine-O-acetyltransferase
SGD - Strictosidine β-glucosidase
SNP - Single nucleotide polymorphism
SLP – Second leaf pai
SLS - Secologanin synthase
STR - Strictosidine synthase
T16H - Tabersonine 16-hydroxylase
T19H - Tabersonine 19-*O*-acetyltransferase
T3O - Tabersonine 3-oxidase
T3R - Tabersonine 3-reductase
TAA - Tryptophan aminotransferase
TDC – Tryptophan decarboxylase
TEM - Tempranillo
THAS – Tetrahydroalstonine synthase
TLC – Thin-layer chromatography

trpA - Tryptophan synthase α-chain
trpB - Tryptophan synthase β-chain
trpD - Anthranilate phosphoribosyltransferase
trpE - Anthranilate synthase
TS - Tabersonine synthase
VIGS - Virus induced gene silencing
WAT1 – Walls are thin 1 protein
WT – Wild-type
YUC/YUCCA - Flavin monooxygenase

Co-Authorship

Chapter 2: Geissoschizine synthase controls flux in the formation of monoterpenoid indole alkaloids in a *Catharanthus roseus* mutant

The research was conceived by Antje M. K. Thamm, Yang Qu, Matthew Czerwinski, Sayaka Masada and Vincenzo De Luca. The experiments were performed by Antje M. K. Thamm, Yang Qu, Matthew Czerwinski and Kyung Hee Kim. The differential gene expression analysis was performed by Graham Jones using the reference transcriptome assembled by Ping Liang. The article was written by Antje M. K. Thamm, Yang Qu and Vincenzo De Luca.

Chapter 3: Inter-organ transport of secologanin allows assembly of monoterpenoid indole alkaloids in a *Catharanthus roseus* mutant down-regulated in iridoid biosynthesis.

The research was conceived by Trevor Kidd, Michael AE Easson, Yang Qu and Vincenzo De Luca. The experiments were performed by Michael AE Easson, Trevor Kidd and Graham Jones. Trevor Kidd wrote the methods, Graham Jones wrote the first draft, Vincenzo De Luca wrote the final draft and edits were make by Yang Qu and Graham Jones. Graham Jones isolated the RNA for Illumina sequencing for differential gene expression using the reference transcriptome assembled by Ping Liang. The data included in the publication from this experiment was moved to Chapter 4 for contextual appropriateness. Conception, implementation and assembly of the morphological and bioinformatic research in Chapter 4 was conducted independently by Graham Jones.

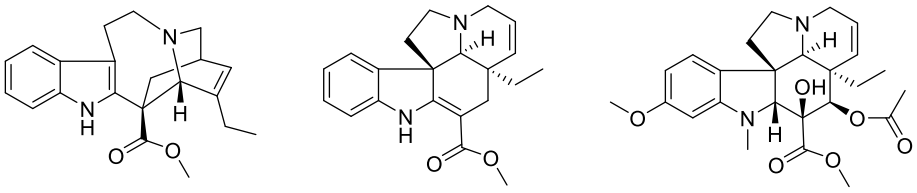
Chapter 1: Literature review

1.1 Iridoid and MIA Biosynthesis

Alkaloids are a diverse class of nitrogen containing specialized metabolites found across most kingdoms of life. In plants, they are believed to play an important role in defence against pathogens and herbivores (Dugé De Bernonville et al. 2017). Monoterpenoid indole alkaloids (MIA) are alkaloids with a backbone formed by the condensation of the secoiridoid secologanin and tryptamine. Vincristine and vinblastine are two essential chemotherapeutic alkaloids that share common precursors synthesised exclusively in *Catharanthus roseus*, a plant in the Apocynaceae family extensively studied for MIA biosynthesis (De Luca et al. 2014). These two MIAs are anti-microtubule agents that inhibit the formation of tubulin, preventing cell division and impacting cellular structure.

The biosynthesis of leaf MIAs in *C. roseus* is compartmentalised in different specialized cell types. Some accumulating MIAs are stored within cells, while others are excreted to the cuticle (Murata et al. 2008). The two precursors to vincristine and vinblastine biosynthesis, vindoline and catharanthine, are specially separated (Figure 1). Catharanthine is transported to the leaf surface while vindoline is stored in idioblasts and laticifers (Yu & De Luca 2013). The removal of physical barriers by herbivory causes the enzymatic dimerization of catharanthine and vindoline in a herbivore, with the physiological impact on the herbivore acting as the deterrent (Dugé De Bernonville et al. 2017). The intricacies of root MIA biosynthesis have not been studied as thoroughly as the leaf biochemistry, transport and regulation in *C. roseus*, yet the roots contain many unique MIAs not found in leaves (Van der Heijden et al. 2004). Ajmalicine, serpentine, lochnericine and hörhamericine are four medicinally important MIAs found mostly in roots; only comparatively small quantities of ajmalicine and serpentine are detected in leaves.

Ajmalicine and serpentine are used to treat hypertension, while lochnericine and hörhamericine demonstrate anti-tumor potential.



catharanthine tabersonine vindoline

Figure 1: Common accumulating MIAs and intermediates found in the leaves of *C. roesus*.

1.2 Secoiridoid biosynthesis and regulation

In leaves, the biosynthesis of the iridoid loganic acid from geraniol occurs in internal phloem-associated parenchyma (IPAP) cells (Murata et al. 2008). The C-8 hydroxylation of geraniol by geraniol 8-oxidase/geraniol 10-hydroxylase (G8O/G10H) catalyses the first step in secoiridoid biosynthesis (Collu et al. 2001; Miettinen et al. 2014). Both hydroxyl groups are oxidized again by the cytochrome P450 monooxygenase 8-hydroxygeraniol dehydrogenase (8-HGO) to form the dialdehyde 8-oxogeranial (Figure 2). Iridoid synthase (IS) cyclises 8-oxogeranial to the bicyclic iridoid nepetalactol, an atypical substrate for a monoterpene cyclase which typically uses geraniol pyrophosphate (GPP) as substrate (Geu-Flores et al. 2012). 7-deoxyloganic acid synthase (7DLS) catalyzes the three-step oxidation of nepetalactol to produce 7-deoxyloganic acid (Salim et al. 2014). Glycosylation of 7-deoxyloganic acid by 7-deoxyloganetic acid glucosyltransferase (DLGT) and subsequent hydroxylation by 7-deoxyloganic acid 7-hydroxylase (DL7H) catalyzes the formation of loganic acid (Salim et al. 2013). An alternative pathway to iridoid formation has been demonstrated, involving the hydroxylation of geraniol by G8O/G10H and subsequent oxidation by 10-hydroxygeraniol dehydrogenase (10-HGO) to form

10-oxogeranial (Krithika et al. 2015). Enzyme assays with 10-HGO and IS reveal the pair can convert 10-hydroxygeraniol to nepetalactol.

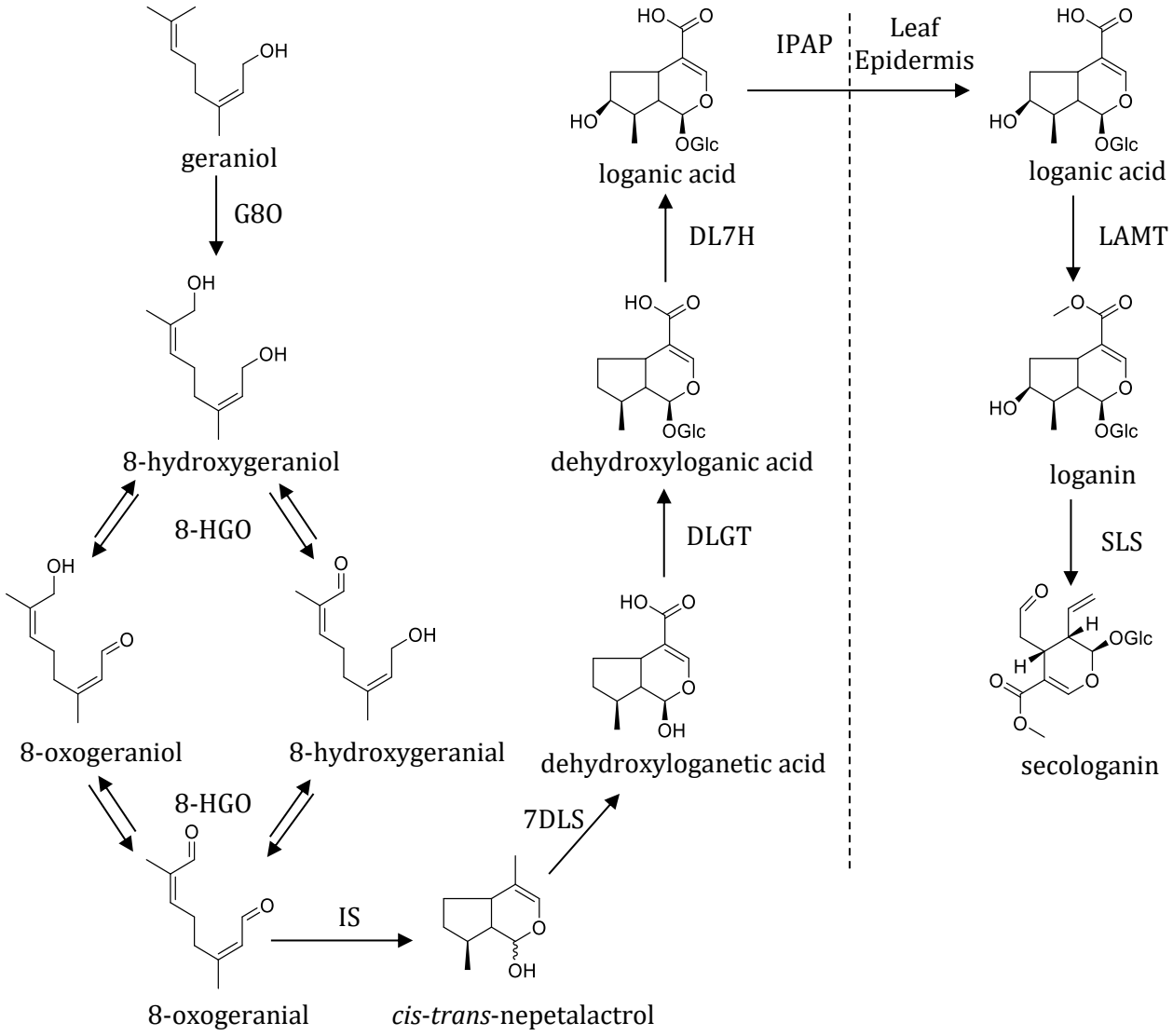


Figure 2: Secoiridoid biosynthesis from geraniol to secologanin occurs in internal phloem-associated parenchyma (IPAP) and leaf epidermal cells in *C. roseus*. The biosynthetic pathway consists of geraniol 8-oxidase (G8O), 8-hydroxygeraniol dehydrogenase (8-HGO), iridoid synthase (IS), 7-deoxyloganic acid synthase (7DLS), 7-deoxyloganetic acid glucosyltransferase (DLGT), 7-deoxyloganic acid 7-hydroxylase (DL7H), loganic acid O-methyltransferase (LAMT) and secologanin synthase (SLS).

After synthesis in IPAP cells, loganic acid is transported to the leaf epidermis via an uncharacterised mechanism then converted to loganin by loganic acid O-methyltransferase (LAMT) (Guirimand et al. 2011; Miettinen et al. 2014). Loginin undergoes oxidative opening of the pentyl ring by secologanin synthase (SLS) to form secologanin (Figure 2) (Irmler et al. 2000). Secologanin and tryptamine are condensed by strictosidine synthase (STR) to form the glycosylated MIA strictosidine (Treimer & Zenk 1979). Deglycosylation by strictosidine β-glucosidase (SGD) creates a labile pool of aglycones used to synthesise over 100 MIAs found in *C. roseus* (Figure 3) (Van der Heijden et al. 2004; Luijendijk et al. 1998).

Several transcription factors have been characterized as regulators of key secoiridioid biosynthetic genes, among them are basic helix-loop-helix iridoid synthesis 1 and 2 (BIS1 & BIS2) (Van Moerkercke et al. 2015, 2016). Both transcription factors are regulated by methyl jasmonate (MeJA) and expressed in IPAP cells. BIS1 and BIS2 have been shown to independently activate G8O, 10-HGO, IS, DLGT and DL7H. ORCA3 is the third characterized transcription factor, also induced by MeJA. It has been shown to upregulate G8O and STR in hairy roots (Zhang et al. 2011).

1.3 MIA Biosynthesis

Cleavage of the glucose moiety from strictosidine by SGD produces the labile resonance intermediates of cathenamine and 4,21-dehydrogeissoschizine (Qu, Easson, et al. 2018). The intermediates can be converted to 19E-geissoschizine via geissoschizine synthase (GS), tetrahydroalstonine via tetrahydroalstonine synthase (THAS) or ajmalicine via an uncharacterized pathway (Figure 3) (Qu, Thamm, et al. 2018). 19E-geissochizine undergoes several successive oxidations by geissoschizine oxidase (GO), Redox1 and Redox2 to generate

the stable MIA stemmadenine. The proposed intermediates between GO, Redox1 and Redox2 are unstable. When Redox1 or Redox2 are silenced by virus induced gene silencing (VIGS), MIAs that resembled oxidative degradation products of the proposed pathway intermediates accumulate, these accumulates were shown to be inactive as substrates to MIA pathway enzymes (Qu, Easson, et al. 2018).

Stemmadenine is acetylated by stemmadenine-O-acetyltransferase (SAT) to form O-acetylstemmadenine (Qu, Easson, et al. 2018). Subsequent successive oxidations by precodylocarpine acetate synthase (PAS) and dihydroprecodylocarpine acetate synthase (DPAS) produces the labile intermediate dihydroxyprecodylocarpine acetate (Caputi et al. 2018). The next biosynthetic step involves catharanthine synthase (CS) or tabersonine synthase (TS), two hydrolases that cleave the acetoxyl moiety and cyclize to the iboga (catharanthine) or aspidosperma (tabersonine) structural classes of MIAs.

After synthesis in epidermal cells, catharanthine is transported to the leaf surface by an ABC transporter (Yu & De Luca 2013). Tabersonine is converted to desacetoxyvindoline in the leaf epidermis by tabersonine 16-hydroxylase (T16H), 16-hydroxytabersonine 16-O-methyltransferase (16OMT), tabersonine 3-oxidase (T3O), tabersonine 3-reductase (T3R) and a N-methyltransferase (NMT) (Qu et al. 2015). Desacetoxyvindoline is then transported to laticifers or idioblasts by an uncharacterized transporter and converted to vindoline by desacetoxyvindoline 4-hydroxylase (D4H) and desacetylvindoline O-acetyltransferase (DAT). Tabersonine produced in the roots of *C. roseus* is converted into different MIA derivatives, the tabersonine to vindoline pathway is exclusively expressed in aerial organs.

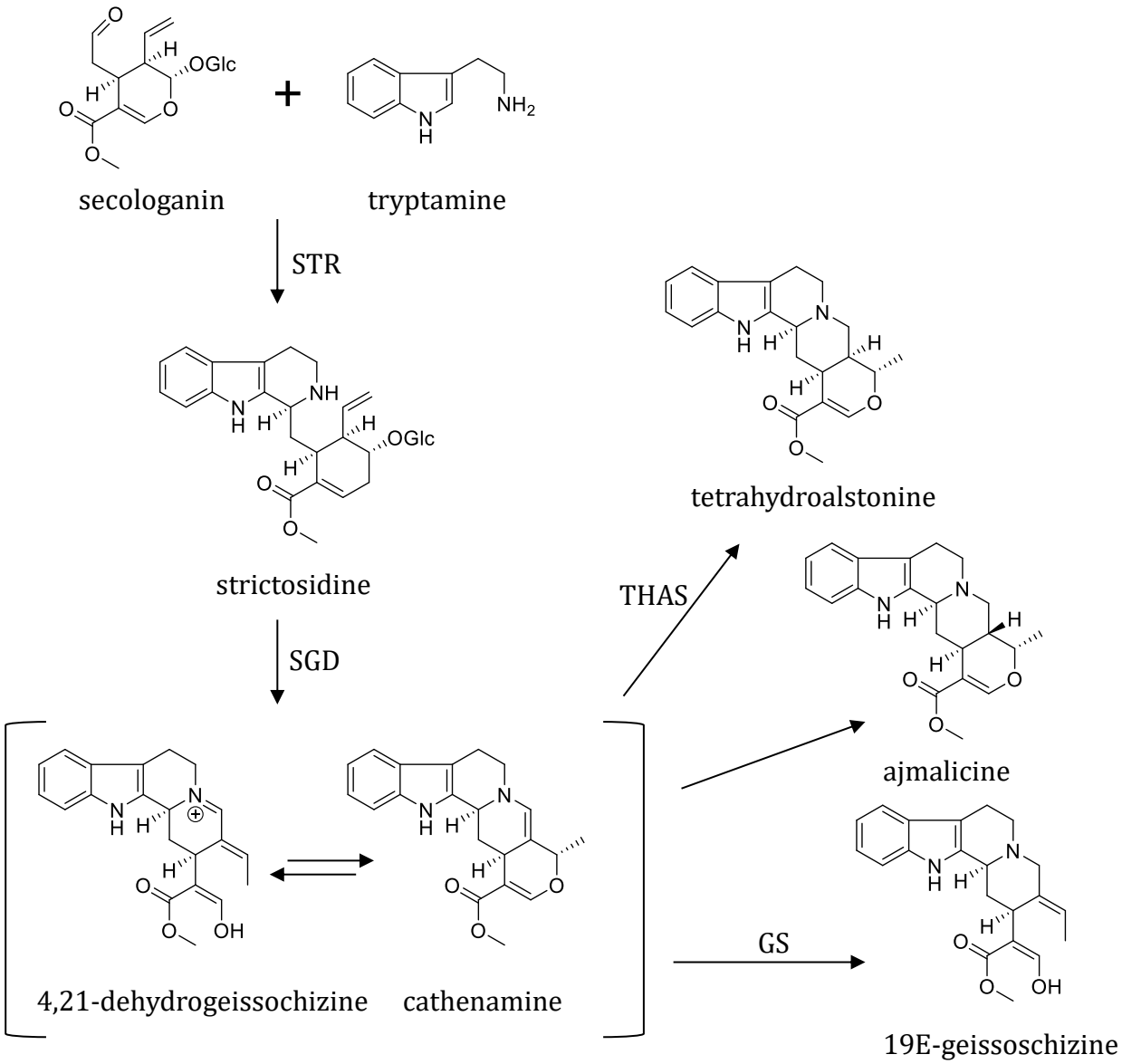


Figure 3: Early MIA biosynthetic pathway in the leaf epidermis of *C. roseus*. The pathway consists of strictosidine synthase (STR), strictosidine β-glucosidase (SGD) and tetrahydroalstonine synthase (THAS), or geissoschizine synthase (GS), or by an uncharacterized pathway to ajmalicine.

1.4 Auxin in Plant Development

Plants utilize several crucial hormones conserved throughout the kingdom for development and regulation. Auxins are a class of hormones involved in almost every major plant process (Aloni 2010; Ma et al. 2015; Olatunji et al. 2017; de Smet 2012). One of the fundamental roles of auxin in development is controlling cell division and elongation through altering localized cellular concentration ratios with its antagonist cytokinin, another class of plant hormone.

Auxin transport in plants is directionally controlled between cells, the process has been termed polar auxin transport. PIN-Formed proteins (PIN) and auxin influx carrier proteins are transporters localized to a portion of the plasma membrane (Adamowski & Friml 2015; Gälweiler et al. 1998). The location of PINs and auxin influx carriers in cells dictates the direction of auxin influx and efflux across membranes. Non-directional long-distance auxin transport occurs in the phloem (Jin et al. 2015). The flow of auxin through a plant depends both on biosynthesis in specific tissues and regulated intercellular and intracellular transport.

Intracellular auxin transport from the cytosol to the vacuole is facilitated by the "walls are thin 1" protein (WAT1) (Ranocha et al. 2013). WAT1 has also shown to play a crucial role in auxin promotion of secondary wall formation in vasculature. *A. thaliana wat1* knockouts displayed reduced fiber content in stems, which was corrected by the exogenous application of auxin. The *wat1* mutants also demonstrated specific resistance to vascular pathogens, with no change in resistance to other sites of infection (Denancé et al. 2013). Promoter-reporter experiments using GUS revealed WAT1 is preferentially expressed in the roots and cotyledons during early plant development. The *wat1* phenotype of fiber content is only apparent in older plants as the vasculature matures. This indicates there is a direct connection between intracellular auxin regulation and stem fiber deposition. Metabolomic analysis of the *wat1* mutant roots

showed lower levels of indole-3-acetic acid (IAA) and indole metabolites derived from auxin, but higher levels of the defence hormone salicylic acid (SA). The resistance of vascular pathogens in *wat1* mutants was attributed to the trade-off of defence for development, with higher levels of SA and lower levels of auxin, a trend widely observed *in planta* (Blomster et al. 2011; Wang et al. 2007; Zhang et al. 2007). This example shows the crucial role auxin plays in xylem differentiation.

1.5 Auxin induces xylem formation

Xylem are water-containing vasculature tissues in plants comprised of lignin polymers, responsible for long-distance root-to-shoot transport of metabolites, proteins and nutrients (Aloni 2010). Xylem are made of dead cells, with thickened cell walls for structural support of the stem. Primary xylem are produced in the primary vascular bundles next to primary phloem and fiber lumen. Secondary xylem distribution extends from proliferation at the secondary vascular cambium towards the pith, a characteristic of all woody plants (Zhang et al. 2011). Auxin has shown to play a role in the initiation of protoxylem through the transport of auxin from newly formed leaves down the procambium by means of PINs, accumulating at the vascular cambium (Aloni 2010; Johnsson et al. 2018). The accumulation of auxin leads to cell expansion, driven by acidification, degradation of the plasma membrane and the uptake of water into the expanding cell.

In the trees *Populus tremula* and *Populus tremuloides* the exogenous application of auxin efflux inhibitors have provided insight into how the transport of auxin through PINs controls leaf abscission, leaf primordia formation and vascular differentiation (Jin et al. 2015; Johnsson et al. 2018; Porth et al. 2014). *P. tremuloides* grown with the efflux inhibitors 1-*N*-naphthylphthalamic

acid (NPA) or 2-chloro-9-hydroxyfluorene-9-carboxylic acid (HFCA) had reduced stem length, number of leaf pairs and overall root growth (Schuetz et al. 2007). Stem cross sections revealed at all concentrations of NPA or HFCA, the number of protoxylem was comparable to controls, but the number of metaxylem had increased. The increased vasculature deposition was attributed to poor vertical drainage of auxin. The change in metaxylem, but not protoxylem, suggests auxin contributes to the formation of metaxylem from protoxylem, in addition to initiating protoxylem proliferation at the vascular cambium.

A plethora of herbaceous plants also form secondary xylem, the model plant *Arabidopsis thaliana* among them. Cross sections of *A. thaliana pin1* mutants displayed increased vascular differentiation below newly developing leaves and an inability to flower (Jones et al. 2005). Though polar auxin transport was inhibited, levels of free IAA in the stem of the *pin1* mutant and wild-type (WT) were similar, suggesting IAA could have been transported from the roots. The increase in vasculature was believed to be associated with poor apical basipetal transport in *pin1* mutants, leading to an uncontrolled accumulation of auxin at the vascular cambium. Exogenous application of the auxin efflux inhibitor NPA on WT *A. thaliana* displayed the same vascular morphology as the *pin1* mutant. Levels of free IAA in the xylem of *Ricinus communis* were measured to be 0.32 ng/mL compared to 13.0 ng/mL in phloem (Baker 2000). Due to endogenous levels of IAA being low in xylem, it is possible xylem can act as an alternative route for long distance IAA transport.

Lignin polymers are mostly comprised of three phenolic monomeric units, H, G and S, condensed by oxidative coupling (Ghaffar & Fan 2013). Lignin composition between plants varies by the percent representation of each monomeric unit in the polymer. The differences in lignin composition vary widely in plants with seemingly similar morphologies, including

common herbaceous fiber crops. For example, lignin in *Linum usitatissimum* is comprised of mostly G monomers (67%), while the lignin of *Corchorus olitorius* contains mostly S monomers (62%).

The first dedicated step in the phenylpropanoid pathway is the deamination of L-phenylalanine by phenylalanine ammonia-lyase (PAL) (Mahroug et al. 2006). The product *trans*-cinnamate is then converted to *p*-coumarate by cinnamate 4-hydroxylase (C4H) and condensed with CoA by 4-coumarate-CoA ligase (4CL) to generate *p*-coumaroyl-CoA (Leple et al. 2007). Hydroxycinnamoyl transferase (HCT) catalyzes the next intermediary transesterification of a shikimate acid moiety in place of CoA, allowing *p*-coumaroyl shikimate 3'-hydroxylase (C3'H) to *meta*-hydroxylate the *p*-coumaloyl moiety, before reversal of the transesterification. This forms caffeoyl-CoA, which undergoes *meta*-O-methylation by caffeoyl-CoA O-methyltransferase (CCoAOMT) to create feruloyl-CoA. The diversity in ratios of the monolignol units is determined by the activity of cinnamoyl-CoA reductase (CCR), caffeic acid 3-O-methyltransferase (COMT), ferulate 5-hydroxylase (F5H) and cinnamyl alcohol dehydrogenase (CAD) on structurally diverse metabolites described (Ghaffar & Fan 2013). H units are derived from CCR catalysing the conversion of *p*-coumaroyl-CoA to *p*-coumaraldehyde. G and S units are derived from CCR accepting feruloyl-CoA or sinapoyl-CoA respectively, with F5H and COMT capable of converting G monomer intermediates to S monomer derivatives. Several of the regulatory factors involved in xylem differentiation induced by auxin are also involved in leaf abscission.

1.6 Auxin regulates formation of leaf primordia and leaf abscission

Auxin plays a crucial role in leaf primordia formation at the shoot apical meristem. Auxin has been shown to induce similar response cascades during leaf abscission as it does in vascular differentiation (Ma et al. 2015; Porth et al. 2014). Leaf abscission initiates by the accumulation of local auxin maxima at the abscission zone between the leaf and stem (Jin et al. 2015). Auxin accumulating at the abscission zone promotes cell expansion and proliferation before leaf separation.

The role of auxin in shade-induced leaf abscission was demonstrated in *P. tremula* using the most common auxin, indole-3-acetic acid (IAA), the auxin efflux inhibitor morphactin and/or ethylene insensitive *P. tremula* mutants (Jin et al. 2015). Ethylene is a fruit ripening and plant senescence hormone; ethylene insensitive mutants allowed for the independence of the ethylene and auxin responses to be evaluated. IAA or morphactin were applied directly to the leaf axil, where the leaf connects to the stem, and showed a delay of abscission by 1 and 7 days respectively. IAA delaying abscission was attributed to disruption of lower concentrations of auxin that could be required in surrounding cells for separation after auxin accumulations along the abscission zone (Sorefan et al. 2009). With higher concentrations of auxins in surrounding cells, a local auxin maxima differential was not achieved. The longer delay in abscission caused by morphactin could have been due to a delay in auxin accumulation at the abscission zone.

Histological sections of β-glucuronidase (GUS) promoter reporter PIN::GUS transformed *P. tremula* were taken showing the abscission zone (Jin et al. 2015). Expression of the PIN homologues in leaves without shading showed unorganized distribution along the axil. Nine days after shading expression of PINs were clustered to the stem side of the abscission zone. This suggests the establishment of high auxin concentrations at the abscission zone are directionally

controlled to flow from the stem, potentially to prevent an accumulation of auxin at the site of separation, next to the abscission zone. Qualitative polymerase chain reaction (qPCR) analysis of the leaf petiole of shaded leaves confirmed expression of PINs were two-fold downregulated compared to the control.

A class of homeobox-leucine zipper transcription factors have been shown to regulate auxin at the abscission zone. Of this family of transcription factors, the knotted1-like homeobox protein (KD1) and revoluta (REV) classes have been characterized (Ma et al. 2015; Porth et al. 2014). KD1 and REV have been shown to regulate auxin signaling in vascular differentiation and proliferation, suggesting conserved mechanisms between leaf abscission and vascular development. The role of auxin transport in roots has shown a connection between pericycle cells, xylem proliferation and lateral root initiation.

1.7 Auxin regulates root formation

Roots contain the largest pools of auxin in plants, despite the aerial organs synthesizing the majority of auxins. Root initiation and early replication from a stem is dictated by localized auxin gradients (Olatunji et al. 2017). Stem cells in the root cap contain the highest concentration of auxin in roots, promoting expansion and division (Li et al. 2014). Auxin in the root cap is also responsible for the root gravitropic response. Positing of auxin transport proteins allows for high concentrations of auxin to accumulate in the most basal root cap stem cells, promoting downwards proliferation of the primary root. *A. thaliana* roots express PIN2, a close homologue of PIN1, which is expressed predominantly in aerial organs (Müller et al. 1998). The differences in cellular localization between PIN1 and PIN2 are specific to the direction of auxin transport from their expressed organs. Auxin transport at the shoot apical and root meristems requires

basipetal transport of auxin in opposite gravitational directions. This is essential in the root cap for eliciting a normal gravitropic response.

Lateral root formation is initiated by auxin in pericycle cells next to the basal end of xylem at the center of the primary root (de Smet 2012). To initiate lateral root formation, pericycle cells use high local auxin concentration gradients to asymmetrically elongate and divide towards the root epidermis. This process creates the lateral root primordia from which the lateral root will proliferate. Control of auxin in pericycle cells is crucial as alterations to native regulation/transport could lead to no division, or uncontrolled division, producing atypical root morphologies (Gou et al. 2010; Khan & Stone 2007; de Smet 2012).

Application of the auxin efflux inhibitor NPA represses lateral root formation off the primary root. Increased expression of genes regulating mitotic division of pericycle cells, specifically D-type cyclin genes, did not lead to increased lateral root formation (Sanz et al. 2011). Instead undirected pericycle proliferation was observed in the root, demonstrating the importance of auxin in directing division. *A. thaliana* overexpressing cyclin D2 (CYCD2) showed an increased sensitivity to auxin in lateral root initiation. In contrast, the *cycd2* knockout mutants demonstrated exogenous auxin insensitivity to lateral root formation. Intriguingly, the impact of endogenous auxin initiation of lateral roots in *cycd2* mutants was minimal comparatively. Cyclin-dependent kinases control cyclin genes by inactivating the protein complex in the nucleus of pericycle cells. Auxin inhibits the expression of these cyclin-dependent kinases, preventing them from inactivating the cyclin protein inducing cell division. For auxin to act as an inhibitor of cell division inhibition, intracellular auxin levels in the nucleus must be high enough to induce the inhibition. The control of specialization cells synthesizing and transporting auxin is important to development, as is understanding the route of auxin synthesis.

1.8 Auxin biosynthesis and regulation

The primary route of auxin biosynthesis in plants is the conversion of L-tryptophan to IAA through the indole-3-pyruvic acid (IPA) pathway (Stepanova et al. 2011; Won et al. 2011). The IPA pathway consists of a tryptophan aminotransferase (TAA) and a flavin monooxygenase (YUC/YUCCA), IPA being the intermediate metabolite (Figure 4).

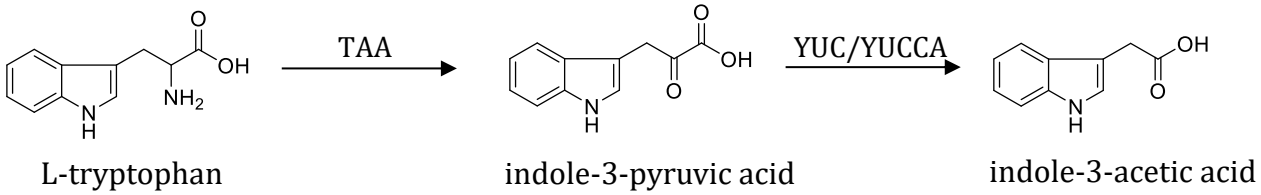


Figure 4: IPA pathway to produce indole-3-acetic acid (IAA) from L-tryptophan. The pathway consists of tryptophan aminotransferase (TAA) and a flavin monooxygenase (YUC/YUCCA).

Apart from the IPA pathway, several other biosynthetic pathways exist in nature to synthesize IAA. Biosynthesis of IAA from tryptamine has been reported in the roots of *Pisum sativum* (Quittenden et al. 2009). Radiolabeled tryptamine and a few suspected biosynthetic pathway intermediates were fed and shown to convert to IAA *in planta*; however, the pathway has yet to be characterized. Another IAA biosynthetic route in plants utilizes a pair of cytochrome P450's, CYP79B2 and CYP79B3, to covert tryptophan to the intermediate indole-3-acetaldoxime (IAOx) (Sugawara et al. 2009). IAOx is subsequently converted to indole-3-acetonitrile (IAN), then IAA by a nitrolase 1-3. Biosynthesis of IAA from the conversion of tryptophan to indole-3-acetamide (IAM) is the primary method for IAA biosynthesis in plant pathenogenic bacteria (Lehmann et al. 2010).

Several YUC homologues from *A. thaliana* have been shown to be biosynthetically active towards endogenous auxin production *in vitro* (Hentrich et al. 2013; Stepanova et al. 2011). These YUC homologues have specific spatial expression in different plant organs. Five of

the YUC homologues have been shown to have high expression levels in roots and phylogenetically cluster in a group distinct from aerial specific YUC homologues. To determine the importance of different YUC homologues in *A. thaliana* roots, root expressing *yuc* mutant knockout lines were generated and grown on media with or without IAA (Chen et al. 2014). Levels of free IAA in the different *yuc* mutants were 55% less than in the control, inhibiting root growth and the gravitropic response of the primary root. Exogenous IAA in the media reversed the mutant phenotype. Overexpression of shoot specific YUC homologues in a root *yuc* mutant background displayed an aerial phenotype characteristic of high levels of auxin and the shriveled root morphology of auxin deficiency. This indicates that auxin overproduction in aerial organs cannot compensate for endogenous auxin deficiency in roots. Several classes of plant hormones directly regulate auxin biosynthesis and transport.

1.9 Jasmonates in Auxin Biosynthesis and Transport

The plant hormone MeJA has been shown both endogenously and by exogenous application to upregulate YUC homologues in *A. thaliana* (Hentrich et al. 2013). Overexpression of two YUC homologues (YUC8 and YUC9) expressed in both roots and shoots of *A. thaliana* displayed similar aerial morphologies to the exogenous application of MeJA on WT plants (Hentrich et al. 2013). These include increased secondary lignification and the upregulation of ethylene responsive genes.

The influence of MeJA on auxin in root development has been discovered by the elucidation of a GH3 protein member, indole-3-acetic acid-amido synthetase (GH3.6), responsible for amino acid conjugated auxin turnover (Khan & Stone 2007). Conjugation with amino acids eliminates the bioactivity of the auxin moiety, removing the molecule from the pool

of active auxin. Another member of the GH3 protein family is the amino acid conjugating protein of jasmonates, jasmonic acid-amido synthetase (JAR1). JAR1 initiates the jasmonate signal cascade by coupling isoleucine to jasmonate, creating the signaling hormone jasmonyl isoleucine (Suza & Staswick 2008). Both *gh3.9* and *jar1 A. thaliana* mutants have shown to be insensitive to MeJA mediated root inhibition and grow longer roots when grown with MeJA than WT (Khan & Stone 2007). The similar root morphology is surprising because *Atjar1* mutants naturally have a short primary root and *Atgh3.9* mutants a long primary root. This suggests that functional jasmonate response and auxin degradation pathways are both necessary for MeJA induced inhibition of root growth. Gibberellins are another class of plant hormone shown to regulate auxin transport.

1.10 Gibberellins in Auxin Transport

Gibberellins (GA) are a class of plant hormones, with a tetracyclic diterpene backbone, involved with auxin in a range of developmental processes. These include floral development (Osnato et al. 2012), vascular differentiation (Johnsson et al. 2018), root signaling (Gou et al. 2010; Li et al. 2014) and PIN regulation (Willige et al. 2011). Removing the apical meristem in a variety of plant species has been shown to reduce endogenous levels of GA (Johnsson et al. 2018). Normal GA levels were restored after decapitation with the exogenous application of IAA. Auxin has shown to induce GA biosynthesis to regulate intercellular auxin transport (Ulmasov et al. 1997).

The first step in GA biosynthesis is the cyclization of geranylgeranyl diphosphate by two terpene cyclases, copalyl diphosphate synthase (CPS) and ent-kaurene synthase (KS) (Weiss & Ori 2007; Yamaguchi & Kamiya 2000). The subsequent conversion of ent-kaurene to GA_{12} is catalysed by two cytochrome P450 monooxygenases, ent-kaurene oxidase (KO) and ent-

kaurenoic acid oxidase (KAO) (Figure 5). GA_{12} is converted to the diverse range of GA hormones by different GA oxidases (GAox). Several of these GAox enzymes can catalyse consecutive oxidations on a variety of GA structural substrates. Turnover of active GA is mediated by gibberellin 2-β-dioxygenases (GA2ox). GAox biosynthetic genes have been shown to be directly downregulated by the transcription factor tempranillo (TEM) (Osnato et al. 2012). *A. thaliana tem* mutants displayed elongated hypocotyls and early flowering, while overexpression of TEM led to dwarfism, loss of apical dominance and late flowering. Promoter reporter experiments using GUS of GAox genes in TEM overexpressing lines showed a decrease in GAox expression at the apical meristem and an increased expression in roots.

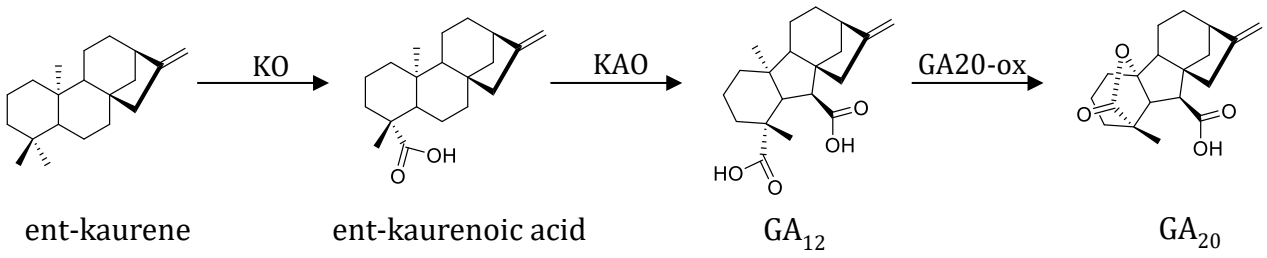


Figure 5: Part of the gibberellin biosynthetic pathway from ent-kaurene to GA_{20}. The pathway consists of ent-kaurene oxidase (KO), ent-kaurenoic acid oxidase (KAO) and gibberellin 20-oxidase (GA20-ox).

The influence of GA on auxin in roots has been observed in transgenic *Populus* with a GA deficiency or insensitivity (Gou et al. 2010). GA deficient lines were created by overexpressing GA2ox, increasing catabolism of biologically active GA. GA insensitive lines were generated by overexpressing a GA induced homologue lacking a DELLA protein domain required for degradation, blocking the signal cascade. Both GA insensitive and deficient transgenic lines displayed degrees of aerial dwarfism and increased levels of root biomass. GA

insensitive lines showed an increase in GA biosynthetic genes and a reduction in GA2ox, suggesting a negative feedback loop regulates the response. Additionally, auxin biosynthetic genes and PINs were upregulated, abscisic acid biosynthetic genes were downregulated, while ethylene biosynthetic and responsive genes showed mixed differential expression.

GA deficient and insensitive lines displayed an increase in lateral root primordia formation compared to WT *Populus* (Gou et al. 2010). In contrast, exogenous application of GA on WT, GA deficient and GA insensitive lines reduced the number of lateral root primordia formed compared to controls. This suggests the inhibition of lateral root primordia by GA occurs early in lateral root initiation.

The crosstalk and impact of IAA and GA on roots has been explored in *A. thaliana*. Exogenous application of GA_3 with IAA showed reduced root elongation compared to IAA alone at IAA concentrations of 5 nM (Li et al. 2014). Promoter-reporter experiments using GUS revealed the application of GA enhanced the expression of the auxin response proteins IAA2 and DR5 in their normative root cell types (Ulmasov et al. 1997). Auxin transport and signaling mutants were used to understand how GA enhances the root inhibition of exogenous IAA. Auxin signaling mutants of the TIR1 auxin receptor and AXR1 involved in the auxin signal cascade both showed reduced sensitivity to auxin and the increased responsiveness to exogenous GA_3 with IAA was not observed. The auxin influx and efflux transport mutants also displayed reduced sensitivity to IAA and did not have increased root inhibition with the additional application of GA_3. This suggests that the role of GA in increasing the exogenous auxin response involves auxin transport.

GA has been shown to mediate auxin transport in roots using gravitropism experiments (Willige et al. 2011). It plays a role in promoting the clustering of PINs at the basal end of the

root to induce asymmetric division. Discovering biosynthetic genes, transporters and regulators in non-model plant homologous to the elucidated genes, often from *A. thaliana*, requires sequencing and assembly of data to make the comparison.

1.11 Bioinformatic Analysis of Non-Model Plants

Higher plants have a large variation in genome size and complexity, partly due to a larger number of transposable elements, or transposons, on average compared to the genomes of animals or microorganisms (Claros et al. 2014; Treangen & Salzberg 2012). These complex genomes have allowed plants to display a wide range of phenotypes across, and sometimes within, species with variation increasing when grown under different environmental conditions. The size and complexity of plant genomes make it difficult for researchers to create assembled, curated and annotated plant genomes at the same rate as other organisms. For example, the model plant *A. thaliana* has a genome size of approximately 135 Mbp (Arabidopsis Genome Initiative, 2000), while our commercially cultivated crops have a median genome size of 766 Mbp (Feuillet et al. 2011). The relatively small genome of *A. thaliana* is one quality that makes it an attractive model plant, however, this also contributes to its limitations as a model for all plants. By comparison, the average size of known land plant genomes is approximately 6 Gbp (Claros et al. 2014; Garcia et al. 2014).

Plant genomes are capable of undergoing ploidy and gene duplication events to create variation with less issue of offspring viability than animals or microorganisms. The variation in potential expression patterns is controlled by a high percentage of transposons in the genome of some plants (Claros et al. 2014; Feuillet et al. 2011). The genome of *A. thaliana* is comprised of 5.6% transposable elements compared to 85% in the genome of *Zea mays*. Assessing the quality

of assembly of a genome or transcriptome involves using the metric n50, which is the minimum contig length of half the total number of nucleotides. That is, lining up the entire sequenced dataset from longest to shortest contig, the n50 is the length of the contig physically in the middle of the dataset.

Without a curated genome, creating transcriptomes of non-model plants has become a popular method of studying unique traits found in some plants. To identify genes producing valuable medicinal compounds in non-model plants, a large-scale transcriptomic analysis project was conducted under the PhytoMetaSyn project (Xiao et al. 2013). The project involved exploiting divergent specialized metabolic pathways only conserved in families producing similar classes of metabolites. Using transcriptomic data for biochemical elucidation of pathways can be approached in a number of ways. Differential transcript expression analysis of specialized cells producing the metabolites of interest is one method used to enrich the presence of the unidentified genes. This is complemented with annotation of conserved domains to sort protein families and predict potential function. When searching for MIA biosynthetic genes, the candidate genes should be similar to genes found in other MIA producing Apocynaceae, yet dissimilar from genes in non-MIA producing plants like *A. thaliana*. The approach has led to the discovery of MIA biosynthetic genes in *C. roseus* (Murata et al. 2008), cannabinoid biosynthetic genes in *Cannabis sativa* (Stout et al. 2012) and benzylisoquinoline alkaloid biosynthetic genes in *Papaver somniferum* (Desgagné-Penix et al. 2010). Differential gene expression datasets are quantified by fragments per kilobase of transcript per million mapped reads (FPKM). Plants contain many specialized cells with different gene expression profiles, enriching specialized tissues in a ribonucleic acid (RNA) isolation increases the FPKM of genes preferentially expressed in those cell types (Murata et al. 2008). In contrast, if a whole root RNA extract is

taken for sequencing, a gene only expressed in the root cap will have a comparatively low FPKM compared to genes expressed in more abundant root cells.

Elucidation and analysis of primary metabolic networks conserved in all higher plants is another use for transcriptomic data from non-model plants. A high degree of sequence conservation in genes involved with primary metabolism is typically necessary for normal development (Aharoni & Galili 2011). The ability to compare genes to a model organism allows for a more robust functional analysis of pathways based solely on sequencing data. The metric used in the comparison is identity (ident), a percentage of identical amino acids in aligned positioning (Johnson et al. 2008). Conserved primary metabolic pathways are involved in anabolism and catabolism of the monomeric and polymeric units of life, carbohydrates, amino acids, nucleotides and lipids. They are also involved with transcriptional and hormonal regulation in response to stimuli and environmental conditions (Aharoni & Galili 2011; Weiss & Ori 2007). The diversity of specialized metabolites has evolved in pockets as variant pathways of conserved metabolic networks found in all plants. One of the tools scientists use to discover novel genes in specialization biosynthetic pathways is mutagenesis.

1.12 EMS Mutagenesis in *C. roseus*

Ethyl methanesuphonate (EMS) mutagenesis has provided a useful tool to create and study variable traits in plants (Chen et al. 2016; Cooper et al. 2008). It induces a predictable guanine to adenine or cytosine to thymine substitution pattern, using standard methodology, in approximately 1 of every 200 kbps of the genome. The randomness of the base pair substitutions is the greatest weakness of using EMS mutagenesis because there is no method of detecting these mutations in high-throughput sequencing data of non-model plants. Several EMS mutagenized

C. roseus plants with unique MIA profiles have been identified and published. One is a low vindoline mutant with a single amino acid substitution on T3R, reducing its activity by 95% (Edge et al. 2017). The reduced enzyme activity caused accumulation of the intermediate 2,3-epoxytabersonine on the leaf surface. The second *C. roseus* EMS mutant M2-0754 contained elevated levels of ajmalicine and reduced levels of catharanthine and vindoline (Qu et al. 2018). Several differentially regulated transcription factors were identified as potential candidates for the shift in flux of the common metabolite. The discovery of the cause behind the unique chemotypes of these two mutants was through screening an epidermal enriched database unrelated to the mutated plants. A third EMS mutant M2-1582 contained low levels of alkaloids and had to be grown in tissue culture conditions due to its susceptibility to infection (Kidd et al. *in press*). Under *in vitro* conditions the plant displayed unique morphologies suggestive of hormone imbalances. Characterizing the alterations in the mutagenized plant may offer insights into cross-talk between primary and specialized metabolism.

1.13 Objective of Research

Bioinformatic analysis to understand *C. roseus* focuses on elucidating biosynthetic genes and regulatory elements of MIAs. Hairy root and callus cultures are typically used to study the impact of hormones on MIA production. The characterization of a *C. roseus* EMS-mutant with complex morphological and biochemical challenges has never been attempted. The model plant *A. thaliana* is typically used to study plant hormones. This study attempts to use our knowledge gained from *A. thaliana, C. roseus* and the greater plantae kingdom to identify the underlying cause of the unique physical morphologies and low-MIA content of M2-1582 using transcriptomic and physiological analysis of alterations (Chapters 3 & 4). This study also proposes the basis of the unique MIA profile observed in M2-0754 (Chapter 2).

Chapter 2: Geissoschizine synthase controls flux in the formation of monoterpenoid indole alkaloids in a *Catharanthus roseus* mutant

Authors: Yang Qu, Antje M. K. Thamm, Matthew Czerwinski, Sayaka Masada, Kyung Hee Kim, Graham Jones, Ping Liang, Vincenzo De Luca

***Planta*, 2018, 247:625-634**

2.1 Introduction

Catharanthus roseus is an important medicinal plant that is the source of the dimeric monoterpenoid indole alkaloid (MIA) anhydrovinblastine and its derivatives (Van der Heijden et al. 2004). The coupling of vindoline and catharanthine monomers leads to the formation of this naturally occurring dimer and further chemical modifications have led to several synthetic derivatives used as chemotherapeutic agents to treat cancers such as Hodgkin's disease and early childhood leukemia. These leaf- and stem-derived *C. roseus* MIAs remain the only commercial source of these chemotherapy agents, while the roots accumulate ajmalicine that is useful in the treatment of hypertension and its oxidized derivative serpentine (Van der Heijden et al. 2004).

MIAs are derived from the assembly of a secoiridoid, secologanin and tryptamine to form the central intermediate strictosidine, the precursor of thousands of MIAs in several plant families (Thamm et al., 2016). Rapid progress has recently been made in completing our understanding of the remaining steps in secoiridoid biosynthesis (Asada et al. 2013; Geu-Flores et al. 2012; Miettinen et al. 2014; Salim et al. 2013, 2014; Simkin et al. 2013) and on the complete elucidation of the seven-step pathway conversion of tabersonine to vindoline (Kellner et al. 2015; Liscombe et al. 2010; Qu et al. 2015). In addition, the recent functional identification of several medium chain reductases that convert strictosidine aglycones to ajmalicine and/or its 20-epimer, tetrahydroalstonine, may explain how these different gene products mediate the

assembly of corynanthe alkaloids in *C. roseus* (Stavrinides et al. 2015, 2016). These developments have been made possible with the availability of high throughput sequencing, bioinformatics-based candidate gene selection, virus-induced gene silencing (VIGS) of selected genes combined with targeted MIA profiling (De Luca et al. 2012) and Agrobacterium co-infiltration methods in *Nicotiana benthamiana* (Miettinen et al. 2014) for functional characterization of selected genes. The characterization of iridoid and MIA pathways led to the creation of prototype strictosidine- and vindoline-producing yeast strains that, respectively, expressed 13 genes for conversion of geranyl diphosphate to strictosidine (Brown et al. 2015) and 7 genes for conversion of tabersonine to vindoline (Qu et al. 2015).

While transcriptomics and bioinformatics have been essential tools for identifying missing steps in the assembly of secoiridoids and for the conversion of tabersonine to vindoline, the biochemistry involved for the elaboration of the iboga and aspidosperma ring structures that lead to the formation of catharanthine and tabersonine, respectively, remains to be identified. The development of additional tools to facilitate discovery of novel MIA genes include the development of *C. roseus* mutants. Mutant collections have been used successfully in crop breeding or in studying basic cellular or biochemical processes in model plants such as *Arabidopsis thaliana* or rice. Despite its common use for crop breeding, development of mutant populations of *C. roseus* using ethyl methyl sulfonate (EMS) mutagenesis has not been widely used or reported except in India where three mutant *C. roseus* strains with high contents of leaf or root MIAs were identified (Kulkarni et al. 1999; Kulkarni and Baskaran 2014), but these strains are not available for basic research because of their purported commercial value.

The present study used a simple thin-layer chromatography (TLC) to screen 3600 EMS-mutagenized *C. roseus* plants that identified a single high ajmalicine, low

catharanthine/vindoline mutant (M2-0754). The MIA phenotype was confirmed by structure elucidation of ajmalicine and by documenting the MIA profiles at different stages of plant growth and development by UPLC-MS. Genetic crosses demonstrated a Mendelian inheritance pattern for the high ajmalicine phenotype. High-throughput sequencing combined with sequence annotation and bioinformatic analyses of the mutant compared to the WT revealed that the mutant expressed increased levels of strictosidine-β-glucosidase (*Sgd*) and heteroyohimbine synthase (*Hys*) transcripts and decreased level of geissoschizine synthase (*Gs*) that explains the mutant phenotype. These differences were confirmed by qRT-PCR analyses of the respective transcripts and by performing biochemical assays in mutant and WT leaf extracts. The characterization of GS revealed its critical role in controlling the flux for the biosynthesis of corynanthe (ajmalicine, tetrahydroalstonine) and iboga/aspidosperma (catharanthine, tabersonine) type MIAs.

2.2 Results

A TLC-based screen has been used to identify MIA mutants. EMS mutagenesis introduces random mutations throughout a target genome that can be used to identify loss- or gain-offunction mutants with appropriate screening methods. The discovery that catharanthine occurs on the surface of leaves (Roepke et al. 2010) was used to develop a simple TLC based screen for identifying MIA mutants with altered MIA profiles. Fresh *C. roseus* third leaf pairs were harvested from each mutant and dipped in chloroform for 1 h to extract leaf surface MIAs (Roepke et al. 2010). After evaporation of the chloroform, the MIA profiles of approximately 250 mutants per week were screened by TLC, leading to the testing of 3600 EMS-mutagenized plants over a 16-week period. This TLC screen led to the identification of mutant line M2-0754

containing low levels of catharanthine and two new MIAs in leaf surface extracts (Fig. S1). The most abundant MIA (Fig. S1) was purified by TLC from M2-0754, and the NMR profile (Fig. S2) corresponded to the chemical shifts of the corynanthe MIA, ajmalicine (Hong et al. 2010). The less abundant MIA was purified by TLC from in vitro assays described later, and its NMR profile (Fig. S3) corresponded to tetrahydroalstonine (Stavrinides et al. 2016).

The *C. roseus* mutant M2-0754 accumulates high levels of ajmalicine at the expense of aspidosperma and iboga MIAs. Further characterization of the M2-0754 mutant line was performed by extracting leaf surface MIAs as well as MIAs remaining within leaves after chloroform extraction (Murata et al. 2008; Roepke et al. 2010). UPLC–MS analyses confirmed the reduction (less than 50%) of catharanthine and vindoline, and the increase (more than fivefold) of ajmalicine and tetrahydroalstonine (Fig. 1 and Fig. S4). The high ajmalicine/tetrahydroalstonine leaf content was maintained in older leaves (Fig. S4, leaf pairs 2–5) suggesting that the biosynthesis of corynanthe MIAs was enhanced in the mutant background throughout leaf development.

Further comparative analyses of mutant M2-0754 involved observing whole plants under UV (254 nm) light. Remarkably, the leaves in the mutant background were highly florescent compared to the WT that appeared to only have limited florescence in the leaf veins (Fig. S4G). While the metabolite(s) responsible for the florescence was not in the chloroform dip (Fig. S1), a highly florescent MIA was extracted together with vindoline in methanol washes of leaves after chloroform dipping. The florescent compound was purified by preparative TLC and was analyzed by UPLC–MS. The compound with a retention time of 2.63 min (*m*/*z* 349.3) showed characteristic absorbance maxima at 249, 306 and 359 nm that have been attributed to serpentine,

an oxidation product of ajmalicine. Mutant leaves of different ages (Fig. S4E, leaf pairs 1–5) accumulated five to eight times more serpentine than WT leaves.

Transcriptomic analysis of the *C. roseus* mutant M2-0754 revealed a series of up-/down-regulated genes. The high ajmalicine phenotype of mutant M2-0754 was further analyzed by performing large-scale comparative transcriptome analysis of mutant and WT transcripts isolated from leaf pair 1 using Illumina sequencing. Mutant and WT databases were assembled separately, mapped to a pre-annotated *C. roseus* transcriptome, and combined in a database to allow comparison of transcript read counts of selected genes (Facchini et al. 2012; Xiao et al. 2013). Inspection of the methylerythritol pathway, the secologanin pathway, and the known MIA pathway genes revealed that only *Sgd* and *Hys* from the MIA pathway out of 34 genes queried were represented approximately twice as many times in the mutant background compared to WT (Table S1). In addition to known MIA-related genes, several others were up- or down-regulated in the mutant (Table S1, S2).

Silencing GS by VIGS resulted in the same chemical phenotype as mutant M2-0754. To investigate the gene responsible for the phenotype of mutant M2-0754, candidates were selected from genes with altered transcription level (Table S1, S2), based on their presence in leaf epidermis enriched express sequence tag (EST) dataset (Facchini et al. 2012) and with their possible enzyme functions. The candidate genes were transiently silenced by VIGS. Remarkably, silencing a cinnamyl alcohol dehydrogenase-type gene that was reduced to 14% in M2-0754 transcriptome (Table S2) to 24%, compared to the empty vector (EV) controls, resulted in significant decrease of catharanthine and vindoline to 26 and 38%, respectively, concurrent with a significant increase of ajmalicine (785%) and tetrahydroalstonine (Fig. 1 and Fig. S5). The similarity between the VIGS phenotype and that of M2-0754 (Fig. 1) strongly suggested that the

reduction of this gene accounted for the phenotype seen in M2-0754. The decrease of iboga and aspidosperma MIAs and the increase in corynanthe MIAs suggested that this enzyme might compete with ajmalicine synthase for strictosidine aglycones.

Biochemical analyses revealed the identity of geissoschizine synthase (GS). After cloning this cinnamyl alcohol dehydrogenase-type cDNA, it was expressed as a His-tagged protein in *E. coli*. The affinity-purified enzyme (Fig. S6) was incubated with strictosidine aglycone (*m*/*z* 351) and NADPH cofactor to yield two closely migrated MIAs of the same UV and mass (*m*/*z* 353), and a third minor product (Fig. 2). The reaction products were purified by preparative TLC and subjected to NMR analyses. The two closely migrating MIAs were identified as 19*E*and 19*Z*-geissoschizine (Fig. 2, Fig. S7, Fig. S8, Fig. S9, Table S3, Table S4). The minor product was identified as tetrahydroalstonine (Stavrinides et al. 2016). Based on the result, this enzyme was named geissoschizine synthase (GS). When NADPH was replaced with NADH in enzyme assays, no reaction product was detected. The activity of GS in the reverse direction with affinity-purified recombinant enzyme and NADP+ converted approximately 31% of the 19*E*-geissocshicine to strictosidine aglycones, two unidentified MIA suspected to be re-arranged forms of strictosidine and to a small amount of tetrahydroalstonine (Fig. S10). In contrast, no product was formed when GS did not catalyze the reverse reaction when incubated with NADP+ and 19*Z*-geissocshicine (Fig S10).

Enzyme activities of SGD and HYS were up-regulated in M2-0754 plants. The greater representation of *Sgd* and *Hys* transcripts in the mutant background (Table S3) was further investigated by measuring their transcripts by qRT-PCR and enzyme activities in leaves of different ages (Fig. 3). The *Sgd* transcripts were > 12-fold higher in mutant leaf pair 1 compared to WT (Fig. 3, leaf pair 1), and the level of this transcript was several-fold higher in all mutant

leaves of different ages than in WT (Fig. 3, leaf pairs 2–5). The increased *Sgd* transcript levels were corroborated by enzyme activities that were highest in leaf pair 1 in the mutant (Fig. 3 leaf pair 1), and were consistently double those of WT in older leaves (Fig. 3, leaf pair 1–5).

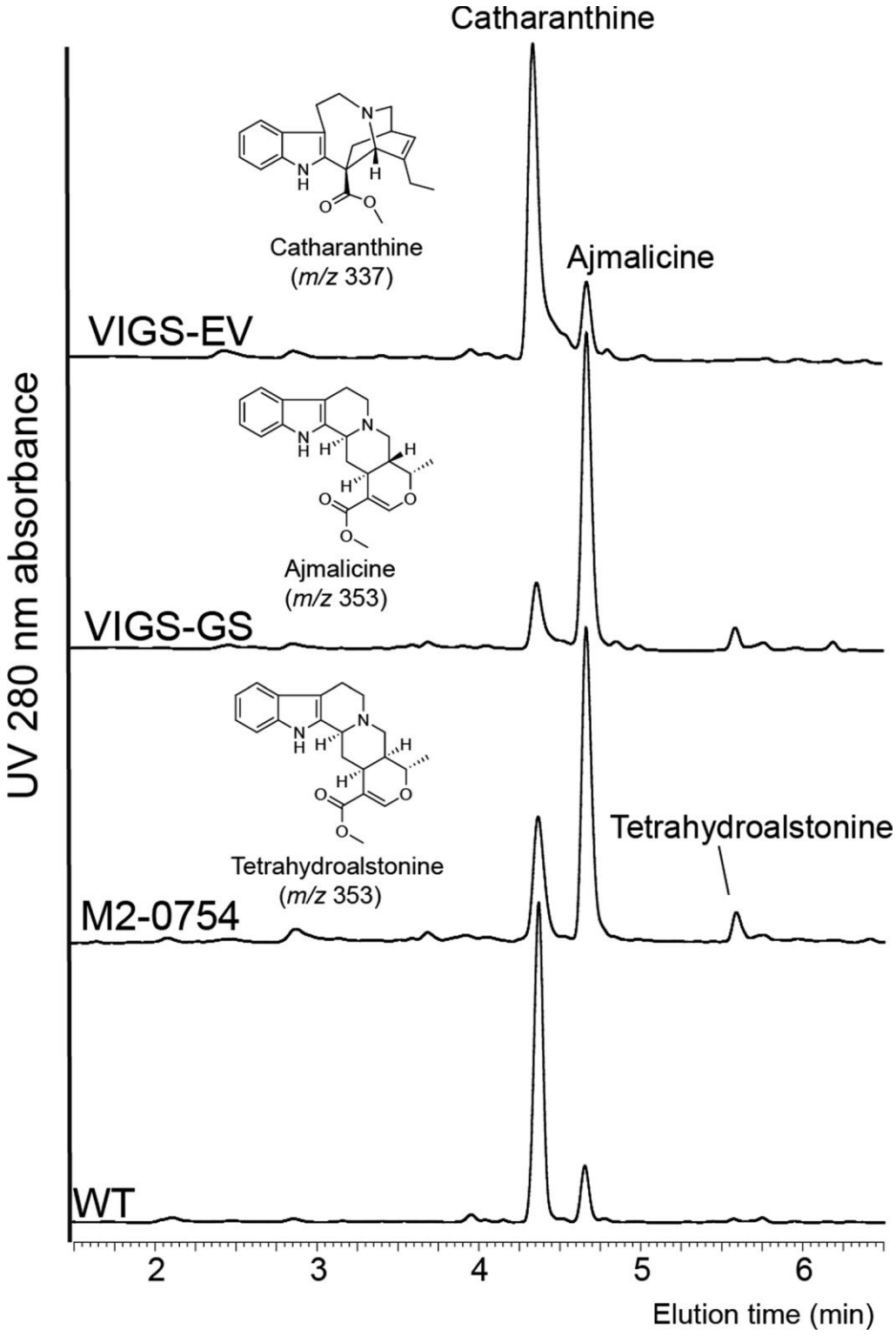


Fig. 1 UPLC–MS chromatograms of leaf surface extract of wild type (WT), M2-0754, VIGS-GS, and VIGS-empty vector (EV) control plants at 280 nm. Each trace represents an injection of total alkaloids extracted from 0.0.25 mg of fresh WT, M2-0754, VIGS-GS, or VIGS-EV leaves.

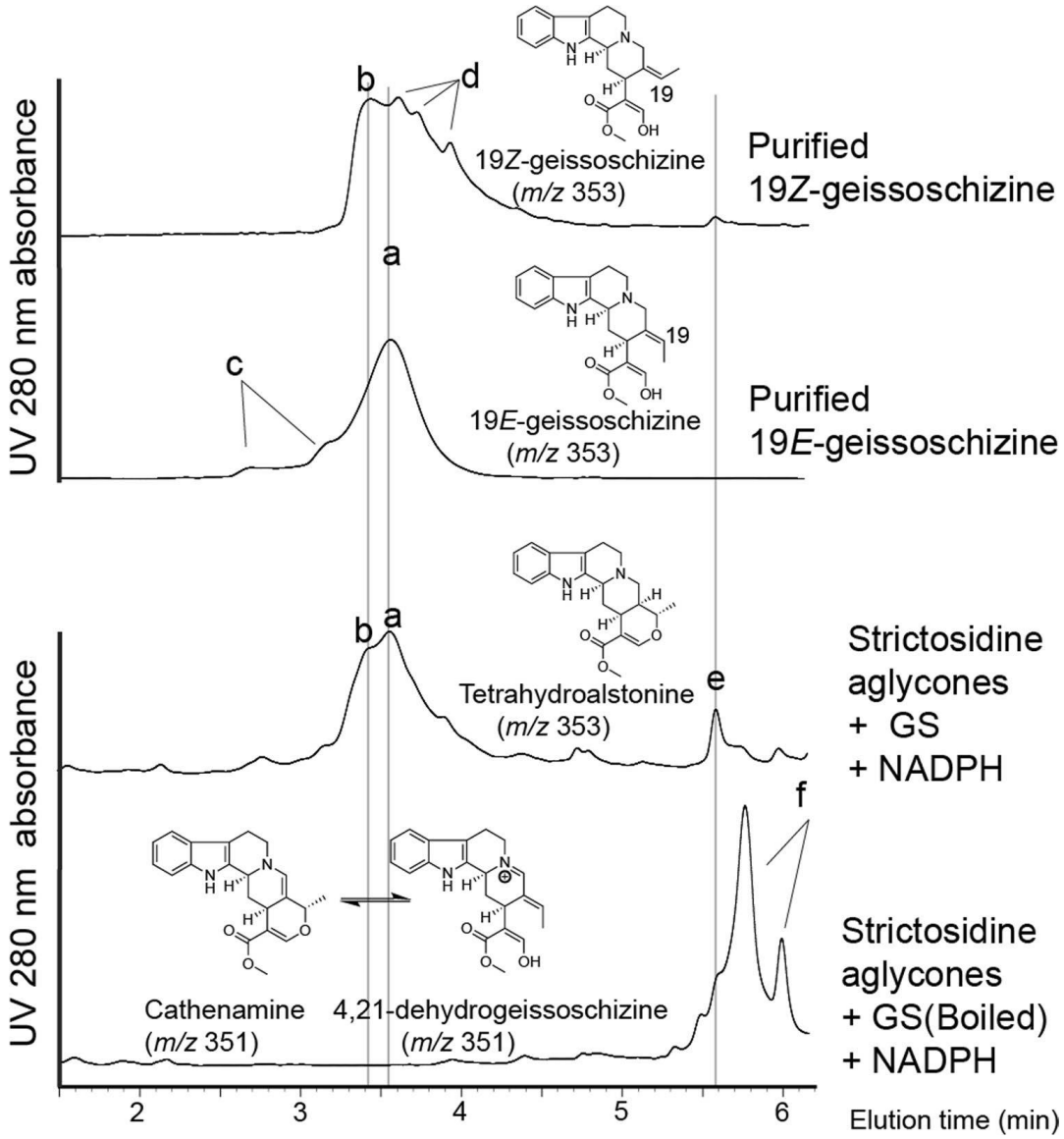


Fig. 2 Recombinant GS purified from *E. coli* converts strictosidine aglycones to both 19*E*- and 19*Z*-geissoschizine and small amount of THA in vitro with NADPH. Reaction products were analyzed by UPLC–MS at 280 nm. a, 19*E*-geissoschizine; b, 19*Z*-geissoschizine; c, 19*E*-geissoschizine methanol adducts; d, 19*Z*-geissoschizinemethanol adducts; e, tetrahydroalstonine; f, strictosidine aglycones. Both 19E and 19Z-geissoschizine form methanol adducts (*m*/*z* 385, gain of mass 32) when dissolved in methanol. Dissolving the standards in acetonitrile eliminates these adducts. Strictosidine aglycones occur as an equilibrium mixture of cathenamine, epicathenamine, and 4,21-dehydrogeissoschizine. Identification of the components of this equilibrium mixture (f) in the bottom trace is not possible. The structure of cathenamine that leads to the formation of tetrahydroalstonine/ajmalicine, and the structure of 4,21-dehydrogeissoschizine that leads to the formation of geissoschizine are shown. The bottom two traces are under the same scale, whereas the top two traces are under the same scale too." For better clarity.

Similarly, *Hys* transcripts were approximately twofold higher in mutant throughout the leaf developmental stages (Fig. 3). Unlike SGD, several HYS/THAS enzymes exists in *C. roseus* leaves (De Luca et al. 2012), which reduces strictosidine aglycones to ajmalicine and tetrahydroalstonine Remarkably, HYS/THAS enzyme assays with mutant and WT leaf extracts showed that mutant leaf pair 1 (Fig. 3, leaf pair 1) was > 5- and 9-fold more active in the formation of ajmalicine and tetrahydroalstonine, respectively, compared to WT and was several times higher in the mutant in leaves of different ages (Fig. 3, leaf pairs 2–5). Together, the increased SGD and HYS/THAS activities in the mutant background also partly accounted for the increased ajmalicine production, in addition to the reduction of GS in the mutant. However, it was not possible to determine GS activity in leaf extracts, possibly caused by weak GS activity or due to reaction product lability.

The inheritance of MIA phenotypes in mutant M2-0754 is recessive. The altered phenotype of mutant M2-0754 was further investigated by determining the inheritance pattern of these traits. After confirmation that the mutant was homozygous for the high ajmalicine phenotype, reciprocal crosses were performed between the mutant (M2-0754) and WT. All F1 plants showed a WT phenotype (Fig. 4), suggesting that inheritance for the phenotype was controlled by one recessive nuclear gene.

To further analyze the segregation pattern, four F1 generation plants from each reciprocal crossing were randomly chosen and self-fertilized to generate seeds of the F2 generation. In contrast to the F1 generation, the 54 F2 plants analyzed showed larger variation in the MIA phenotypes (Fig. S11) and plants were categorized as mutants if ajmalicine/tetrahydroalstonine contents were within those of M2-0754 and 1 standard deviation. Using these criteria, these two MIAs were inherited as a recessive trait (14/54 = 0.26; Fig. S11).

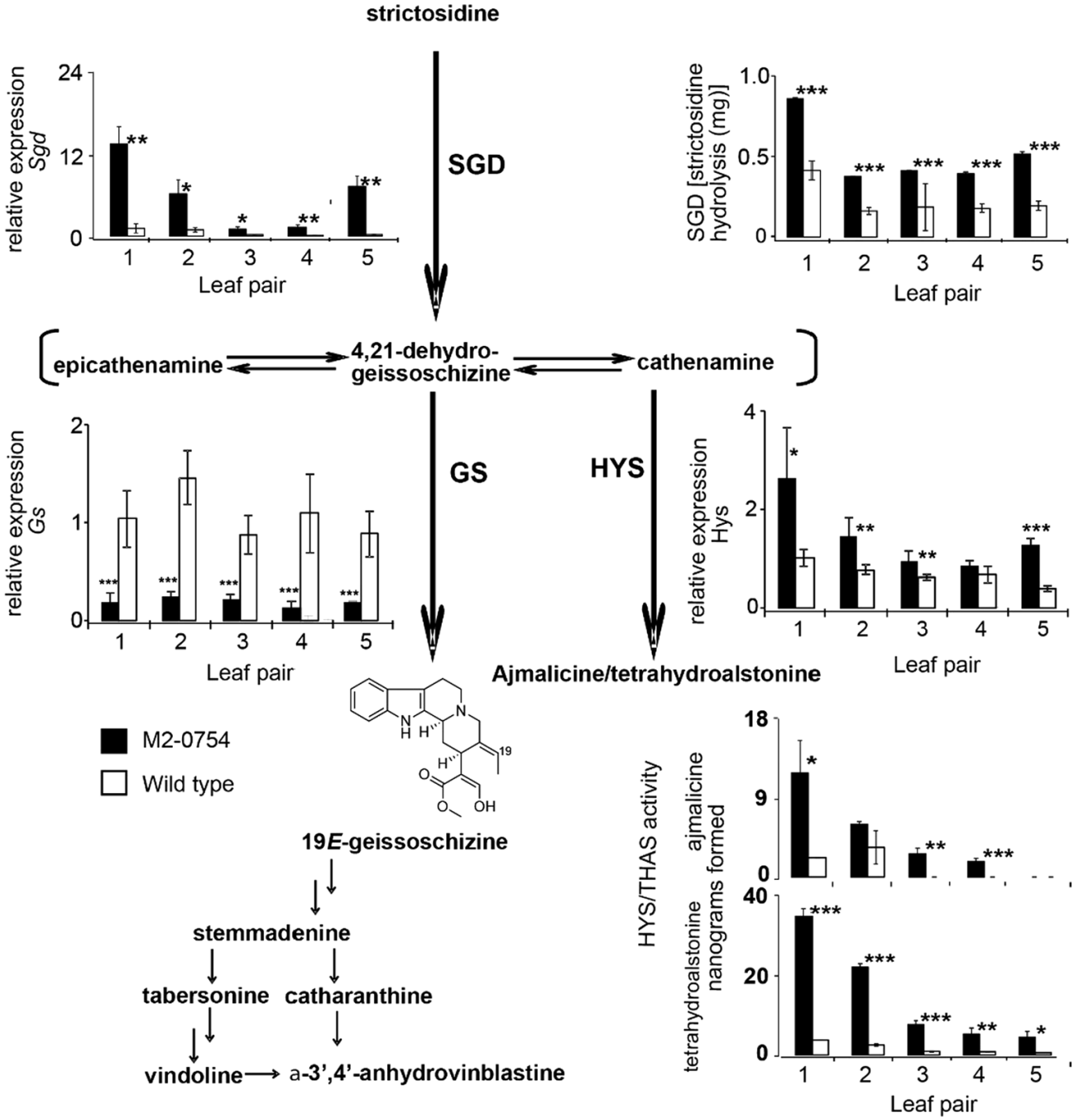


Fig. 3 Transcript level of *Sgd*, *Hys*, and *Gs*, and enzyme activity of SGD, HYS/THAS in M2-0754 plants compared to wild type. Transcript levels and enzyme activities were measured from the youngest developing leaf (leaf pair 1) to older leaves (leaf pair 2–5). The qRT-PCR data represent three biological replicates with three technical replicates for transcript analysis of each leaf pair, while in vitro enzyme assays for SGD and for HYS/THAS represent three biological replicates without any technical replicates for each leaf pair. Error bars represent the standard deviation. *$P < 0.05$; **$P < 0.01$; ***$P < 0.001$.

2.3 Discussion

A simple TLC procedure is useful for screening an EMS *C. roseus* population to identify MIA-accumulating mutants. Despite the value of mutagenesis for improvement of crops, there appear to be few examples (Kulkarni et al. 1999; Kulkarni and Baskaran 2015) of successful use of this technique for producing medicinal plants that accumulate altered or novel MIAs. It is possible that the lack of simple MIA screening protocols has hampered the discovery process (Chaudhary et al. 2011). The present study shows that 3600 EMS lines of *C. roseus* could be screened over a 16-week period using a simple TLC method to identify a single line that accumulated low levels of the iboga MIA, catharanthine, in favor of corynanthe MIA, ajmalicine (Fig. 1). This mutant accumulates five times more ajmalicine than the WT, while levels of both catharanthine and vindoline were reduced by 50%. This screen has also been useful in the recent identification of a tabersonine 3-reductase mutant that accumulates MIA epoxides rather than vindoline (Edge et al. 2017).

While most of the ajmalicine was recovered in chloroform, the oxidized derivative serpentine that also increased five- to eightfold compared to WT was recovered in the extracts of leaves after the chloroform treatment (Fig. S4). Previous studies with plant cell suspension cultures have shown that ajmalicine may be transported into vacuoles against a concentration gradient by a specific MIA uptake system (Deus-Neumann and Zenk 1984) that may involve a proton-driven antiporter (I. Carqueijeiro et al. 2013). Vacuolar peroxidases then convert ajmalicine to positively N-charged serpentine that may hinder it from diffusing across the tonoplast (Blom et al. 1991). This possible mechanism may help to explain why serpentine is not extracted by chloroform treatment (Figs. 1, 2), while the majority of the ajmalicine is. It would be interesting if the catharanthine transporter CrTPT2 (Yu & De Luca 2013) responsible for

secretion of this MIA to the leaf surface may also be responsible for the transport and presence of ajmalicine in the chloroform fraction. Since CrTPT2 is preferentially expressed in younger leaves (Yu & De Luca 2013), it is possible that declines of CrTPT2 expression with leaf age is responsible for shifting the accumulation of ajmalicine to inside the cell and retention inside vacuoles after this MIA is converted to serpentine.

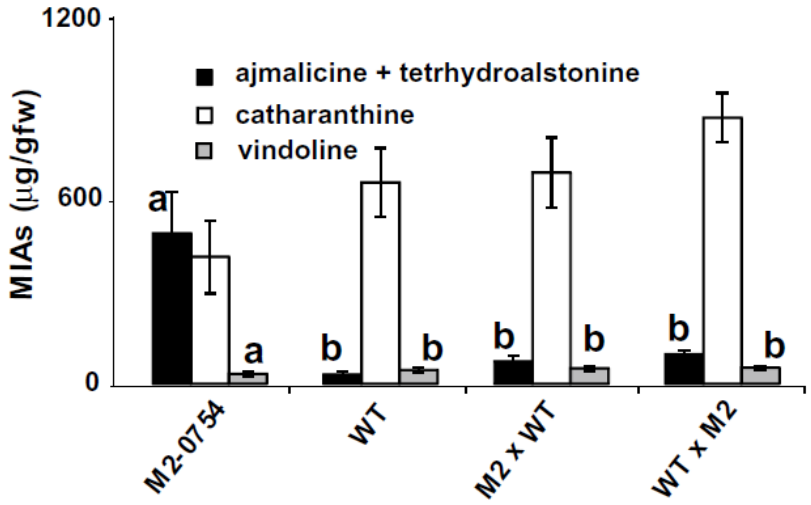


Fig. 4: The M2-0754 high ajmalicine mutant follows a recessive inheritance pattern. Reciprocal crosses of mutant with WT generate a hybrid where WT MIA profiles are re-established. Total MIAs were extracted from leaf pair 2 and were analyzed by UPLC–MS. Data points for the same MIA with different letters are statistically different (ANOVA, $n = 9$, $df = 3$). Error bars represent standard error.

Comparative bioinformatics of mutant and WT transcriptomes show increased expression of *Sgd* and *Hys* combined with decreased expression of *Gs* in *C. roseus* leaves of different ages in the mutant. The increased ajmalicine, tetrahydroalstonine, and serpentine levels found in the mutant were accompanied by 50% declines of catharanthine and vindoline compared to the WT (Fig. S4). Comparative bioinformatic analyses of transcript abundance of all known MIA pathway enzymes showed that *Sgd* and *Hys* transcripts were 3.61- and 1.86-fold more represented (Table S1), while Gs was almost 7-fold less represented in the mutant compared to

the WT (Table S2). Analysis by qRT-PCR showed that *Sgd, Hys, and Gs* had similar representative expression profiles (Fig. 3) to those of comparative bioinformatics (Tables S1, S2) in mutant leaves compared to the WT.

The combined use of bioinformatics and VIGS identified GS, the gateway enzyme that controls the flux of strictosidine aglycones to iboga/aspidosperma type MIAs (catharanthine/tabersonine). Silencing of *Gs* by VIGS alone suppressed the accumulation of catharanthine and vindoline (Fig. 1) as also observed in the M2-0754 mutant (Fig. 3), suggesting that reduction of GS largely accounted for the chemical phenotype. A similar screening approach identified a *C. roseus* mutant (M2-1865) with reduced levels of vindoline and high levels of tabersonine-2,3-epoxide and 16-methoxytabersonine-2,3-epoxide (Edge et al. 2017). This MIA phenotype was caused by a single amino acid substitution (H189Y) of tabersonine-3-reductase that diminished its biochemical activity by 95% and blocked a coupled tabersonine-3-oxidase/tabersonine-3-reductase conversion of these tabersonine or 16-methoxytabersonine to intermediates required for the vindorosine and vindoline pathways, respectively. Further studies from our laboratory identified several subsequent enzymatic steps that convert the critical precursor 19*E*-geissoschizine to the final product catharanthine and tabersonine, confirming the importance of *Gs* and the 19*E*-geissoschizine as an intermediate in the formation of iboga and aspidosperma MIAs (data not shown). During the preparation of this report, *Gs* (genbank: KF302079.1) was also described in another study (Tatsis et al. 2017). In contrast to the present study, assays with recombinant GS did not result in the production of geissoschizine but of two further reduced products, the 19-epimers of 16*R*-isositsirikine. In the same study, silencing of *Gs* by VIGS did not result in a significant change in the MIA profile (Tatsis et al. 2017). In the present study, the formation of isositsirikine by GS was not detected, possibly caused by

differences in the expression vector used or in the in vitro reaction conditions used in the two studies. Instead, the formation of both 19*E*and 19*Z*-geissoschizine by GS in this study was evident by the mass of the reaction products obtained and by confirmation of the structures by NMR analyses (Figs. S7, S8, S9; Table S3, S4). We conclude that the name geissoschizine synthase is appropriate for the gene described here and by Tatsis et al. (2017). While two 19-epimers of geissoschizine were produced in vitro, only 19*E*-geissoschizine accumulated when geissoschizine oxidase (GO) was silenced in plants by VIGS and the 19*Z*-epimer could not be converted to stemmadenine (data not shown). In addition, tetrahydroalstonine was also produced by GS in vitro (Fig. 2); however, the tetrahydroalstonine level was elevated in M2-0754 and VIGS-GS plant (Fig. 1, Fig. S4). The results suggest that both 19*Z*-geissoschizine and tetrahydroalstonine could be an artifact when conducting in vitro assays, since only 19*E*-geissoschizine is produced *in planta*.

Comparative biochemical analysis of mutant and WT show that SGD and HYS/THAS enzyme activities are increased in *C. roseus* leaves of different ages in the mutant. The conversion of strictosidine to the aglycone generates MIA intermediates that may be channeled by different enzyme activities to form corynanthe or iboga/aspidosperma MIAs. The reduction of cathenamine to ajmalicine/tetrahydroalstonine was shown with partially purified *C. roseus* HYS/THAS activities isolated from MIA-accumulating cell suspension cultures (Hemscheidt and Zenk 1985). The increased expression of *Sgd* and *Hys* transcripts in the mutant background correlated with increased activities of the corresponding enzymes in mutant leaves of different ages compared to those of WT (Fig. 3). Biochemical analysis of M2-0754 showed very high levels of HYS/THAS activity in leaves of all ages compared to those of WT (Fig. 3) and may in part explain why strictosidine aglycones might be preferentially channeled to form corynanthe

MIAs (ajmalicine, tetrahydroalstonine and serpentine) rather than to iboga or aspidosperma MIAs.

Four *C. roseus Thas* genes (*Thas*1–4) have recently been cloned and functionally characterized as tetrahydroalstonine synthases (Stavrinides et al. 2015, 2016), generating mostly tetrahydroalstonine together with minor amounts of mayumbine. A fifth *Thas* gene product named heteroyohimbine synthase *Hys*, (Stavrinides et al. 2016) displayed a different enzyme activity profile to generate mostly ajmalicine (ajmalicine:tetrahydroalstonine:mayumbine; 55:27:15). Reciprocal crosses suggest that that high ajmalicine phenotype is inherited as a Mendelian recessive trait. Ultimately, the concurrent increase of *Sgd/Hys* transcripts and decrease of *Gs* transcripts may be caused by altered transcription regulation. Several putative transcription factors were found up-regulated in M2-0754 transcriptome (Table S3). Further investigation is necessary for the complete elucidation of the mutation in M2-0754.

In conclusion, the present study describes a simple targeted TLC-based screen for identifying MIA mutants in *C. roseus* that has led to the identification and characterization of a high ajmalicine mutant, and the gateway enzyme geissoschizine synthase for the formation iboga/aspidosperma MIA. Since ajmalicine is an antihypertensive drug used for treatment of high blood pressure, this mutant may represent an alternative inexpensive commercial source of this MIA.

2.4 Methods and Materials:

2.4a Plant materials

Control (WT) and mutagenized seeds of *Catharanthus roseus* (L.) G. don were provided by PanAmerican Seed Co. (panamseed.com). Mutagenesis and mutant seed production were conducted as described previously (Edge et al. 2017). Non-mutagenized *C. roseus* control seeds

and ~ 4000 M2 seeds were germinated in soil and plants were grown in a greenhouse under long-day photoperiod (16/8 h light/dark) at 30 °C.

2.4b Leaf MIA extraction

Fresh *C. roseus* third leaf pairs were harvested and fresh weights were recorded. To obtain extracts containing leaf surface MIAs, leaves were dipped in 3 mL chloroform in 15 mL conical sterile capped polypropylene tubes for 1 h at room temperature. The surface wax-stripped leaves were transferred to fresh tubes, dried by evaporation of the chloroform and extracted in 2 mL methanol for 1 h to obtain MIAs not harvested by the chloroform treatment. The resulting extracts were analyzed by TLC and UPLC–MS analyses. For the VIGS experiment, leaf pair 1 was used because silencing is only maintained in the young developing leaf tissue.

2.4c Ajmalicine purification

Young leaves (50 g of second and third leaf pairs) were dipped in 1 L of chloroform overnight. After evaporating the chloroform, dried surface extract was re-suspended in 5 mL methanol and purified by TLC using ethyl acetate:methanol 9:1 (v/v). Ajmalicine (2 mg) Rf was 0.62.

2.4d Geissoschizine synthase (GS) enzyme assay and geissoschizine/tetrahydroalstonine purification

Escherichia coli BL21-DE3 harboring pET30b + GS was grown to OD 0.8 induced with 0.1 mM IPTG at 15 °C overnight. 6X-his-tagged GS was purified by Ni–NTA affinity chromatography from sonicated lysate in buffer (20 mM Tris pH 7.5, 100 mM NaCl, 1 mM dithiothreitol, 1 mM phenylmethanesulfonyl fluoride, and 10% v/v glycerol). Enzyme assay (100 μL) contains 20 mM Tris pH 7.5, 2 μg recombinant GS, and 5 μg strictosidine aglycones, and were incubated at 30 °C for 1 h. The reaction products were extracted by ethyl acetate, dried, dissolved in methanol, and analyzed by UPLC–MS. For large-scale products purification, 30 mg

secologanin and 18.6 mg tryptamine were reacted with 1.35 mg strictosidine synthase for 1 h in buffer (50 mM Hepes pH 7.5). Then, 5.2 mg SGD was added. After 1 h, the aglycones were collected by centrifugation at 10,000*g* for 20 min. The collected aglycones were reacted with 40 mg purified recombinant GS and 0.3 g NADPH in buffer (Tris pH 7.5) at 30 °C for 6 h. The products were extracted with ethyl acetate, reduced to small volume, and resolved by TLC in solvent combination of toluene:ethyl acetate:methanol 15:4:1 (by vol.). Rf of 19*E*-geissoschizine (1.5 mg): 0.17; Rf of 19*Z*-geissoschizine (1.5 mg): 0.11; Rf of tetrahydroalstonine (1.8 mg), 0.62.

2.4e SGD/HYS/THAS enzyme assays with leaf extracts

Leaf tissue was ground in 50 mM Na2HPO4 buffer, pH 6.6. After collection of cell debris by centrifugation, the supernatant was fractionated with 20–50% ammonium sulfate precipitation and the 50% protein pellet was dissolved in 50 mM Na2HPO4 buffer, pH 6.6, and desalted by gel filtration. Enzyme assays (200 µL) for SGD activity contained 200 µg protein and 1 mM strictosidine and was incubated for 1 h at 30 °C. Enzyme assays (200 µL) for HYS/THAS activities contained 50 µg protein, 50 µg strictosidine aglycones, and 0.5 mM NADPH, and were incubated for 1 h at 30 °C. The reactions were stopped by the addition of base (NaOH) to pH 10 and by extraction of MIAs into ethyl acetate.

2.4f Plant crosses and segregation of MIA phenotype

Reciprocal crosses were made between parental line (WT) and M2-0754 plant lines as described previously (Edge et al. 2017) to generate the F1 generation. Phenotypic analysis involved determination of Ajmalicine/tetrahydroalstonine and catharanthine in leaf surface extracts and vindoline in leaves after chloroform extraction. Leaf pairs 2 (LP2) and 3 (LP3) were used for phenotypic analysis of the F1 and F2 generations, respectively.

2.4g UPLC-MS

Samples were analyzed using Acquity UPLC systems (Waters, Milford, MA, USA) equipped with BEH C18 column (2.1 × 50 mm, particle size 1.7 μm), a photodiode array detector and a mass spectrometer. The solvent systems for alkaloid analysis were as follows: solvent A, methanol: acetonitrile: 5 mM ammonium acetate at 6:14:80; solvent B, methanol: acetonitrile: 5 mM ammonium acetate at 24:64:10. The following linear elution gradient was used: 0–0.5 min 99% A, 1% B at 0.3 mL min^{-1}; 0.5–0.6 min 99% A, 1% B at 0.4 mL min^{-1}; 0.6–8.0 min 1% A, 99% B at 0.4 mL min^{-1}; 8.0–8.3 min 99% A, 1% B at 0.4 mL min^{-1}; 8.3–10.0 min 99% A, 1% B at 0.3 mL min^{-1}. The mass spectrometer was operated as capillary voltage 3.1 kV, cone voltage 48 V, desolvation gas flow 600 L h^{-1}, desolvation temperature 350°C, source temperature 150°C, and positive ion mode.

2.4h RNA extraction, cDNA synthesis and qRT-PCR

Leaf pairs of different ages and stages of development (Leaf pairs 1 to 5) were harvested from mutant and WT lines, were ground in liquid nitrogen to a fine powder and extracted with Trizol® reagent (Invitrogen, Carlsbad, CA, USA) according to manufacturer's instructions with minor modifications. After precipitation with isopropanol, RNA was treated with DNAse I to remove genomic DNA. First-strand cDNA was synthesized using SuperScript II reverse transcriptase (Invitrogen) and oligo (dT) 12-18 primer (Invitrogen) using 1-5 μg total RNA according to the manufacturer's protocol. qRT-PCR was performed (CFX96^{TM} Real-Time system, Bio-Rad, Hercules, CA, USA) using iTaq^{TM} Universal SYBR® Green Supermix (Bio-Rad), 5 μM primers and cDNA template (equivalent to 5 ng total RNA) in a reaction volume of 10 μL. The reaction conditions for qRT-PCR included, 1 cycle of 95 °C for 1 min and 40 cycles of 95 °C for 15 s and 58 °C for 1 min. The Critical Threshold (Ct) values were used to calculate the relative transcript

abundance with 60S ribosome RNA as the internal control. The primer efficiency was calculated from qRT-PCR of the serial dilution of total cDNA, and the specificity of the primers was confirmed by the dissociation curve for each primer set. The qRT-PCR primers are listed in (Table S5).

2.4i Cloning

Full-length *Sgd* cDNA was amplified by primers (1/2) and cloned in pCRT7 vector within NcoI/XhoI sites. Partial *Str* cDNA (deletion of amino acid 1-19 that contains a signal peptide) was amplified by primers (3/4) and cloned in pET30b+ vector within NcoI/XhoI sites. Full-length GS cDNA was amplified by primers (5/6) and cloned in pET30b+ vector within BamHI/SalI sites.

2.4j NMR instrumental

Spectra were recorded on a Bruker Avance AV I 600 Digital NMR spectrometer with a 14.1 Tesla Ultrashield Plus magnet using TOPSPIN 2.1 software for data acquisition and analysis on a Windows 7 workstation. The 1D spectra were acquired with a sweep width of 20.5 ppm with an FID size of 32k points for proton. Deuterated acetone (99.8% pure, Cambridge Isotope Laboratories) was used as the solvent using the internal reference of (1H = 2.05 ppm). The 2D spectra were acquired with 2048 points and 256 increments and they were processed with 1024 × 1024 points. The HC-coupling constant in the HSQC, edHSQC and HMBC were set to 145 Hz and long-range coupling constant in the HMBC was set to 10 or 6 Hz. Majority of samples were run in standard 5mm NMR tubes. Most of samples were run on a BBO probe. The most dilute sample was run on a TXI probe to benefit from increased signal-to-noise ratio.

2.4k Identification of 19E-/19Z-geissoschizine

TLC purified respective geissoschizine epimes (1.5 mg each) were analyzed by ^{1}H, ^{13}C (only for 19*E*-geissoschizine), edHSQC and HMBC to resolve the structures. edHSQC and HMBC cross peaks clearly revealed that both are geissoschizine (Takayama et al. 1992; Tirkkonen et al. 2994). H(18) of 19*E*-geissoschizine (1.76 dd, Table S7) has a small coupling due to the correlation to one of H(21) as seen in a number of MIA with 19*E*-configuration. Instead, H(18) in 19*Z*-geissoschizine (1.62 d, 1.56 d, Fig. S9, Table S3) is only a doublet, demonstrating a 19*Z*-configuration. 19*Z*-geissoschizine is seen as a single compound when analyzed on TLC and UPLC-MS. When examined by NMR in acetone-d_6, the apparent pure compound behaved as two tautomers with several peaks of overlapping chemical shifts [e.g. H(9)-H(10)], and a number of peaks with slightly different chemical shifts [e.g. H(18), Table S3]. The ^{13}C chemical shifts deduced from edHSQC and HMBC of both tautomers largely overlapped, except for C(3, 14, 15, 21), C=O, and O-CH_3 (Table S4). The broad peaks in ^{1}H spectrum (Fig. S9) for H(14) and H(15) also indicate fluctuation in the molecules that may explain the formation of the tautomers. The tautomers of geissoschizine were also documented previously (Takayama et al. 1992).

2.4l Differential Gene Expression of M2-0754

The RNA-seq data for the wild-type C. roseus ssp. Pacific Peach and M2-0754 mutant were generated as part of a previously conducted transcriptomic analysis of 75 non-model, medicinally relevant plants (Xiao et al. 2013). The wild-type *C. roseus* and M2-0754 RNA-seq files were aligned to the indexed reference transcriptome using STAR (Dobin & Gingeras 2015). Cufflinks suite (Trapnell et al. 2012) was used to assemble the indexed sequencing data to the indexed transcriptome. Cuffcompare was used to combine the assembled datasets for the differential gene expression analysis using Cuffdiff. The cut-off was set to FPKM<0.01 as the

minimum quantified integer, anything below is set to 0. Genes outlined were manually screened for protein length greater than 80 amino acids and sequence homology above 65% indent to characterized proteins.

Annotation of Upregulated Genes	WT FPKM	Mut FPKM	Mut/WT FPKM
WD-40 repeat protein	0.34	43.33	126.71
Salicylate carboxymethyltransferase	0.19	10.35	54.80
Cysteine protease	0.32	16.67	51.49
Peroxidase N1	1.20	38.87	32.48
Salicylate carboxymethyltransferase	0.59	17.92	30.38
Terpene synthase	1.16	34.98	30.18
Tryptophan synthase beta chain 2	4.74	91.58	19.33
Glutamate receptor 2.7	1.04	13.24	12.74
Epoxide hydrolase 3	1.17	14.39	12.28
Outer arm dynein light chain 1 protein	1.60	13.69	8.57
Arginine decarboxylase-like protein	31.53	266.18	8.44
Short-chain dehydrogenase/reductase	3.29	27.26	8.28
Cytokinin dehydrogenase 5	1.94	14.67	7.57
Kinesin heavy chain-like protein	4.14	29.63	7.15
Terpene synthase 10-like protein	15.62	111.55	7.14
Presenilin	3.37	20.34	6.03
Oxidation resistance protein 1	3.19	19.04	5.97
Syntaxin-132	5.56	31.69	5.70
Cysteine proteinase RD21a	8.12	37.61	4.63
Jasmonate O-methyltransferase-like protein	3.47	15.50	4.47
Xyloglucan endotransglucosylase	2.37	10.26	4.34
Mediator of RNA polymerase II transcription subunit 14	3.41	14.02	4.12
WRKY18 Transcription factor	2.64	10.87	4.11
WRKY32 Transcription factor	2.69	10.65	3.96
GDSL esterase-like protein	5.91	22.66	3.84
Salicylic acid-binding protein 2	17.39	65.88	3.79
Protein Stay-Green like	8.68	32.18	3.71
Strictosidine beta-glucosidase	60.21	217.05	3.61
Aspartate aminotransferase	5.97	21.29	3.57
Probable receptor-like protein kinase	5.95	20.75	3.49
Respiratory burst oxidase-like protein	3.54	11.80	3.33
Early light-induced protein 2	25.88	75.50	2.92
Senescence-specific cysteine protease SAG39	7.09	20.62	2.91
Endochitinase EP3	15.77	45.82	2.91

Polyneuridine-aldehyde esterase	9.67	28.07	2.90
Secologanin synthase	24.46	69.96	2.86
Cationic peroxidase 1	4.17	11.90	2.86
BTB/POZ domain-containing protein FBL11	4.43	12.58	2.84
UDP-glycosyltransferase 91A1	12.75	36.18	2.84
Jasmonate ZIM domain 2	77.40	204.03	2.64
Scopoletin glucosyltransferase	6.36	16.80	2.64
Aspartic proteinase	7.61	20.02	2.63
Helicase and polymerase-containing protein TEBICHI	3.93	10.21	2.60
Eugenol synthase 1	60.39	152.45	2.52
Protein detoxification 21	15.19	36.91	2.43
Carbonyl reductase	22.55	53.05	2.35
Auxin response factor 19	7.11	16.55	2.33
Nucleoside diphosphate kinase 1	5.01	11.67	2.33
Anaphase-promoting complex subunit 4	5.23	11.96	2.28
Kinesin-like protein KIN12A	15.46	35.32	2.28
Ureidoglycolate hydrolase	5.36	11.66	2.18
Lipoxygenase	85.78	186.18	2.17
Vicianin hydrolase-like protein	255.81	554.95	2.17
DSPc domain-containing protein	9.05	19.48	2.15
Phragmoplast orienting kinesin 2	5.83	12.49	2.14
Wall-associated receptor kinase 5	17.27	36.96	2.14
Nuclear intron maturase 2	6.54	13.93	2.13
Mitotic-spindle organizing protein 1B	9.38	19.96	2.13
PRp27-like protein	6.92	14.69	2.12
ZCT1	7.36	15.57	2.11
Serine carboxypeptidase-like 35	15.93	33.14	2.08
Tyrosine ligase-like protein 12	6.82	14.17	2.08
ABC transporter C family member 4	7.19	14.85	2.07
Methanol O-anthraniloyltransferase	28.31	58.22	2.06
Alpha,alpha-trehalose-phosphate synthase 9	5.13	10.53	2.05
Deacetoxyvindoline 4-hydroxylase-like protein	5.68	11.53	2.03
Reticulata-related protein 5	5.89	11.94	2.03
Beta-fructofuranosidase	5.93	12.02	2.02
NBS-LRR class resistance protein Fy2-Ry2	8.81	17.69	2.01
Serine protease	8.41	16.87	2.01
Heteroyohimbine synthase	205.82	383.79	1.86
Loganic acid methyltransferase	61.64	80.69	1.31

Table S1 List of up-regulated genes in M2-0754 compared to wild type *C. roseus*

Annotation of downregulated genes	WT FPKM	Mut FPKM	WT/Mut FPKM
ABC transporter I family member 1	11.13	0.18	62.72
Protein FAF-like, chloroplastic	23.45	0.46	51.13
Putative phospholipid-transporting ATPase 9	24.86	1.32	18.86
Putative pentatricopeptide repeat-containing protein	12.88	0.72	18.00
Putative phospholipid-transporting ATPase 12	213.93	13.98	15.30
Retrovirus-related Pol polyprotein from transposon TNT 1-94	448.28	35.02	12.80
Aspartyl-tRNA synthetase	226.82	20.11	11.28
Aldose reductase	13.27	1.29	10.26
Autophagy-related protein 9	12.43	1.25	9.91
Geissoschizine synthase	313.38	45.04	6.96
MADS-box protein SVP	21.11	3.09	6.83
WD-40 repeat protein	51.64	7.82	6.60
Serine protease SPPA, chloroplastic	54.95	9.29	5.91
Presenilin-like protein	29.01	5.06	5.73
Syntaxin-132	35.70	6.69	5.34
DNA-directed RNA polymerase V subunit 5A	51.85	10.20	5.08
Laccase-1	10.24	2.12	4.84
Myosin heavy chain kinase C	18.88	4.02	4.69
Probable serine/threonine-protein kinase	26.54	6.41	4.14
Mediator of RNA polymerase II transcription subunit 14	15.80	3.96	3.99
Caffeoylshikimate esterase	34.07	8.63	3.95
Actin-related protein 5	11.32	2.98	3.80
Ubiquitin-conjugating enzyme E2	13.42	3.57	3.76
Kirola-like protein	40.77	11.04	3.69
Ferredoxin-3	11.83	3.67	3.23
Small heat shock protein, chloroplastic	11.53	3.79	3.04
Pathogenesis-related protein PR-1	28.87	9.73	2.97
Glutaredoxin	230.81	79.46	2.90
Galactinol synthase	647.85	231.00	2.80
Ammonium transporter 1 member 2	18.31	6.59	2.78
Haloacid dehalogenase-like hydrolase	98.80	37.00	2.67
Lysine-specific histone demethylase	11.70	4.47	2.62
CYC02 protein	93.97	37.46	2.51
Caffeic acid 3-O-methyltransferase	11.44	4.71	2.43
HVA22-like protein a	32.18	13.30	2.42
Ferredoxin	14.93	6.21	2.41
THO complex subunit 5B	31.27	13.13	2.38
Glyceraldehyde-3-phosphate dehydrogenase	29.68	12.78	2.32
Early nodulin-93	12.47	5.43	2.29
Auxin-binding protein ABP19a	10.73	4.70	2.28

Dimethylallyl pyrophosphate isomerase isoform 1	217.48	95.54	2.28
Ultraviolet-B receptor UVR8	49.24	21.66	2.27
Pectinesterase	11.46	5.12	2.24
Highly ABA-induced PP2C gene 2 isoform 1	14.83	6.79	2.18
Chlorophyll a-b binding protein 21	214.14	98.45	2.18
Carotenoid cleavage dioxygenase 4	133.43	61.74	2.16
Zinc finger, CCCH-type	47.74	22.20	2.15
Vacuolar sorting protein 9	17.33	8.07	2.15
Heavy metal-associated isoprenylated plant protein 26	16.74	7.91	2.12
Derlin-1.1	20.49	9.82	2.09
Vomilenine reductase 2	11.69	5.61	2.08
Transcriptional activator FHA1	20.85	10.10	2.06
Peroxisomal membrane protein 2	72.61	35.39	2.05
Glycosyl transferase	73.21	35.79	2.05
Flavonoid 4'-O-methyltransferase homologue 1	22.73	11.21	2.03
TRAF-like superfamily protein	18.85	9.30	2.03
Oligouridylate-binding protein 1	110.81	54.78	2.02
Flavonoid 4'-O-methyltransferase homologue 2	25.11	12.42	2.02

Table S2 List of down-regulated genes in M2-0754 compared to wild type *C. roseus*

Chapter 3: Inter-organ transport of secologanin allows assembly of monoterpenoid indole alkaloids in a *Catharanthus roseus* mutant down-regulated in iridoid biosynthesis

Authors: Trevor Kidd, Michael AE Easson, Yang Qu, Graham Jones, Vincenzo De Luca

Phytochemistry, in press

3.1 Introduction

The chemotherapeutic value of vinblastine and vincristine, two monoterpene indole alkaloids (MIAs) isolated exclusively in *Catharanthus roseus* are derived from the oxidative coupling of catharanthine and vindoline monomers and have led to extensive studies on the biosynthesis of these MIAs over the past 50 years (De Luca et al. 2012). These studies have revealed that formation of catharanthine and vindoline from geraniol and tryptophan involve 28 separate biochemical reactions that have been characterized at the biochemical and molecular level (Qu, Thamm, et al. 2018; Qu, Easson, et al. 2018; Qu, Safonova, et al. 2018; Caputi et al. 2018).

The biochemical and molecular characterization of the 9-step conversion of geraniol to secologanin (Collu et al. 2001; Geu-Flores et al. 2012; Salim et al. 2013 & 2014; Asada et al. 2013; Simkin et al. 2013, Miettinen et al. 2014; Brown et al. 2015), the decarboxylase-mediated formation of tryptamine from tryptophan (De Luca et al. 1989) and the strictosidine synthase (Kutchan et al. 1988) mediated conversion of secologanin and tryptamine to form the central intermediate strictosidine has been used to create the 1st recombinant yeast strains that accumulate this precursor (Brown et al. 2015) of several thousand known MIAs (Szabó 2008). The hydrolysis of strictosidine by strictosidine β-glucosidase (Geerlings et al. 2000) generates reactive aglycones that undergo 9 biochemical reactions that have been characterized at the

molecular level (Qu, Thamm, et al. 2018; Qu, Easson, et al. 2018; Qu, Safonova, et al. 2018; Caputi et al. 2018) to generate the iboga MIA, catharanthine and aspidosperma MIA, tabersonine. The expression of 7 other genes involved in vindoline biosynthesis in prototype recombinant yeast strains have been shown to convert tabersonine to vindoline (Qu et al. 2015). The discovery of the complete pathways for the formation of catharanthine and vindoline from geraniol and tryptophan provide remarkable tools for functional expression of their pathways in yeast or in other heterologous systems.

Hairy roots of *C. roseus* contain detectable levels of catharanthine (Parr et al. 1988) and oxidized tabersonine derivatives such as lochnericine, hörhammericine and echitovenine rather than vindoline. The restricted occurrence of vindoline and its post-tabersonine pathway to above ground photosynthetic tissues has helped to explain why the dimeric anticancer MIAs have not been consistently observed in roots and plant cell cultures (Aerts & De Luca 1992). The root-specific conversion of tabersonine to lochnericine involves a cytochrome P450 (Rodriguez et al. 2003) that has been described at the molecular level by the functional identification of 2 individual CYPs that both catalyze the same reaction (Carqueijeiro et al. 2018), while another CYP catalyses the 19-hydroxylation of lochnericine to form hörhammericine (Giddings et al. 2011). Recently tabersonine 19-*O*-acetyltransferase (T19H) (Carqueijeiro et al. 2018) rather than minovincinine 19-*O*-actyltransferase (MAT) (Laflamme 2001) has been suggested to catalyse the terminal reaction in echitovenine biosynthesis.

The biosynthesis of leaf MIAs in *C. roseus* is compartmentalized in different specialized cell types, with the assembly of the iridoid, loganic acid, from geraniol occurring in specialized internal phloem associated parenchyma (IPAP) cells, followed by its transport to leaf epidermis where it is converted to secologanin for the assembly of MIAs (Courdavault et al. 2014; Murata

et al. 2008). While the cell compartmentation of root iridoid and MIA biosynthesis has yet to be studied thoroughly, it is possible that the iridoid and MIA pathways occur in the same cells (Laflamme et al. 2001). Alternatively transport of iridoid substrates from other organs may be possible as found for the phloem mediated transport of the iridoid glycoside, antirrhinoside in *Asarina scandens* (Gowan et al. 1995) and in *Antirrhinum majus* (Beninger et al. 2007).

This report describes a *Catharanthus roseus* mutant that does not accumulate MIAs which was identified using a simple thin layer chromatography screen of over 3000 mutants (Edge et al. 2018) generated by ethyl methanesulphonate (EMS) mutagenesis. This screen was successfully used to identify mutant lines containing high levels of ajmalicine (Qu et al. 2018), O-acetylstemmadenine (Qu, Safonova, et al. 2018) and tabersonine epoxide (Edge et al. 2018) that have been important in elucidating the biosynthetic pathway for the formation of catharanthine and vindoline from strictosidine (Qu et al. 2015; Qu, Easson, et al. 2018). The lack of MIAs in the mutant are attributed to the loss of secologanin accumulation that is required together with tryptamine for assembly of strictosidine. This study shows that two transcription factors, BIS 1 and BIS2 (**B**asic helix-loop-helix **i**ridoid **s**ynthesis 1 and 2) are down regulated in the mutant. As a result, the mutant does not express several iridoid pathway genes that are required for the formation of secologanin. The study demonstrates that supplying secologanin to the mutant through the roots, re-establishes the formation of catharanthine and vindoline in the leaves of the mutant. This suggests that secologanin is mobile between *C. roseus* roots and shoots.

3.2 Results

Mutant line M2-1582 accumulates almost no iridoids or MIAs. Several *C. roseus* mutants have recently been identified (Edge et al. 2018; Qu et al. 2015; Qu, Thamm, et al. 2018; Qu, Easson,

et al. 2018; Qu, Safonova, et al. 2018) by screening an EMS population using chloroform dipping to harvest MIAs associated with leaf surfaces (Roepke et al. 2010) and monitoring altered MIA profiles by thin layer chromatography (TLC). Among more than 3000 mutants screened, line M2-1582 did not to accumulate any MIAs. This line was cultivated in the greenhouse for approximately 3 months when it began to display disease symptoms and the mutant was rescued by harvesting cuttings that were surface sterilized and transferred to *in vitro* cultivation and propagation before the plant died. The *in vitro* mutant line survived, and several hundred plants were clonally propagated together with the *in vitro* WT plants to investigate the mutant phenotype. The mutant displayed several physiological abnormalities including slow growth, short internodes, undifferentiated root architecture, premature leaf abscission, a rigid stem due to increased lignification and slow growth compared to WT plants cultivated under the same conditions. Measurements of gram fresh weights (GFW) revealed that while mutant GFWs resembled closely to the WT, the GFWs of the mutant were redistributed between above and below ground organs (Table S1). The GFWs of wild type roots was 16.5 % of the total, while mutant roots represented 69.3% of the GFW of the entire plant. (Table S1). The extremely different phenotype of the mutant was reflected by comparative transcriptome analysis of the mutant (Fig. S1) that confirmed that numerous biological processes had been affected compared to WT.

Leaves from the WT and the M2-1582 mutant were extracted and analyzed for iridoid and MIA content. The major secoiridoid, secologanin, was reduced 97% from 207.5 ± 37 µg/gfw in the WT to 5.8 ± 4.3 µg/gfw in the mutant (Fig. 1). The major MIAs catharanthine and vindoline were reduced 95% and 97%, respectively, from 203.1 ± 47 µg/gfw and 32.1 ± 7 µg/gfw in the WT to 5.6 ± 4.2 µg/gfw and 1.7 ± 1 µg/gfw in the mutant (Fig. 1). Minor leaf

MIAs including serpentine, anhydrovinblastine, vindolidine, deacetylvindoline, tabersonine and 16-methoxytabersonine detected in WT leaves (Fig. 2) were undetectable in the mutant background. Furthermore, precursors of secologanin such 7-deoxyloganic acid, loganic acid and loganin were not detected in the mutant background suggesting that the very low iridoid/MIA phenotype might be caused by greatly reduced expression of iridoid pathways genes.

Secologanin feeding enhances the accumulation of catharanthine and vindoline in M2-1582. Mutant and WT plants were tested for their ability to produce MIAs when their roots were supplied with secologanin (Fig. 1). Provision of secologanin to WT roots over a 48hr period suppressed its accumulation in leaves by over 50 % (Fig. 1) compared with controls growing in media alone. The inhibition observed was perhaps due to inhibition of secologanin pathway enzymes or to other non-specific effects that remain to be discovered. In contrast, mutant roots provided with secologanin increased leaf secologanin levels over 13-fold compared to those growing in media alone. This large increase in secologanin levels in mutant leaves was accompanied by over 8.6- and 4.5-fold increase in catharanthine and vindoline levels, respectively compared to control mutant roots (Fig. 1). These results show that the M2-1582 plants that accumulate little or no secologanin, appear to transport this iridoid from roots to leaves where it was accumulated and was partly transformed into the major MIAs, catharanthine and vindoline.

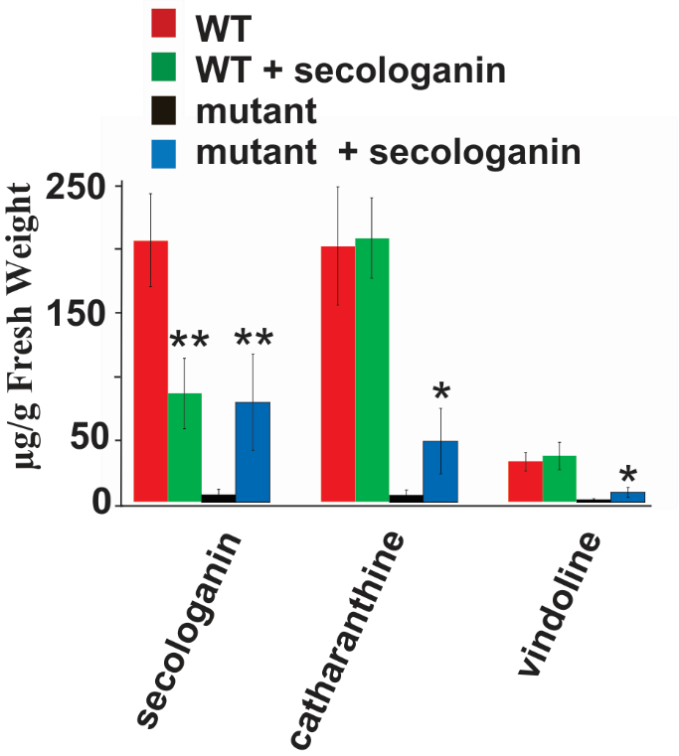


Fig. 1 A very low secologanin, catharanthine, and vindoline mutant line accumulates MIAs when roots are provided with 1 mM secologanin. The error bars represent standard deviations of measurements of three biological replicates with each line. Two-tailed student t-tests were calculated and (***) denotes <0.001 significance, (**) <0.01 significance and (*) <0.05 significance. Significance was compared between feeding/no-feeding secologanin for either WT or mutant plants.

Grafting of mutant shoots onto wild-type roots re-establishes MIA biosynthesis and accumulation in the mutant. The Top Wedge Grafting method (Kimura and Sinha 2008) was used for all grafting of the mutant scion onto WT stalk, where the stalk plant was unable to support a scion with more than one leaf pair. A horizontal cut made with a scalpel was used to separate and remove the terminal portion of the stock plant. The scalpel was then used to split the stem of the stock plant, cutting vertically approximately 1 cm. The scion was cut horizontally around the second node, and then the lower 1 cm was tapered using the scalpel into a “V’ shaped

wedge. The scion was inserted into the stock, and grafting wax is applied to reduce desiccation and to provide support.

Once grafting had been completed, the youngest mutant leaf pairs were harvested after 1- and 3-weeks, respectively, for extraction of iridoids and MIAs. Remarkably secologanin levels increased over 27-fold in leaves of the mutant after 1 week of grafting to reach secologanin levels found in WT leaves (Fig. 2). This substantial movement of secologanin through the graft union and accumulation in mutant leaves was accompanied with >11- and 27-fold increase in catharanthine and vindoline, respectively in mutant leaves compared to those found in un-grafted mutant leaves. The levels of secologanin declined by 56.5% in mutant leaves after three weeks of grafting, while catharanthine and vindoline increased over 2- and 1.7-fold over the levels found in mutant leaves at the 1-week graft stage (Fig. 2). Other MIAs (serpentine, 3',4'-anhydrovinblastine, vindorosine, deacetylvindoline, tabersonine and 16-methoxytabersonine) that were not detected in mutant leaves of un-grafted plants, all increased significantly (Fig. 3) in mutant leaves of the grafted plant, suggesting that the downstream MIA pathway is active in the mutant. These grafting experiments strongly support the results obtained from the secologanin feeding studies (Fig. 1) to show that the no MIA phenotype is dependent on the inability of the mutant to produce secologanin substrate for the assembly of MIAs. The results of both experiments strongly imply that secologanin is a mobile molecule that can be transported from underground roots to above ground shoots where it is assembled into MIAs that normally accumulate in WT leaves.

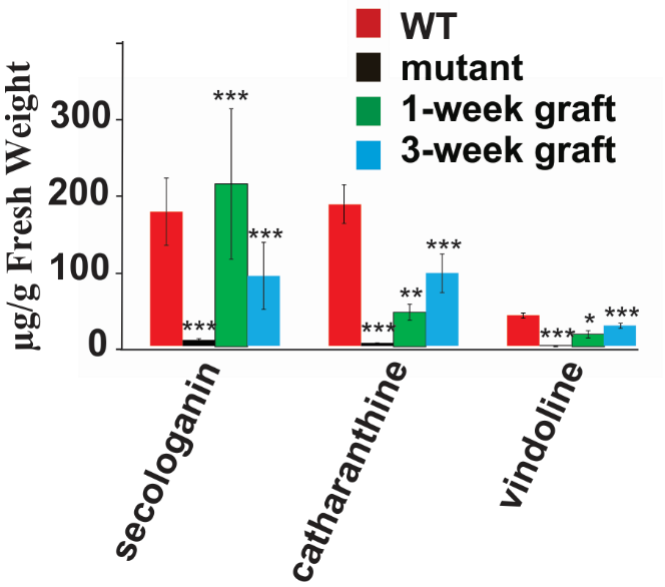


Fig. 2 A very low secologanin, catharanthine, and vindoline mutant line accumulates MIAs in leaves when grafted on to WT roots. Leaves from the scion were harvested 1- and 3-weeks after grafting onto WT roots. The error bars represent standard deviations of four biological replicates with each line. Two-tailed student t-tests were calculated and (***) denotes <0.001 significance, (**) <0.01 significance and (*) <0.05 significance. Significance was compared between mutant/WT, mutant/mutant 1-week grafting, or mutant/mutant 3-week grafting.

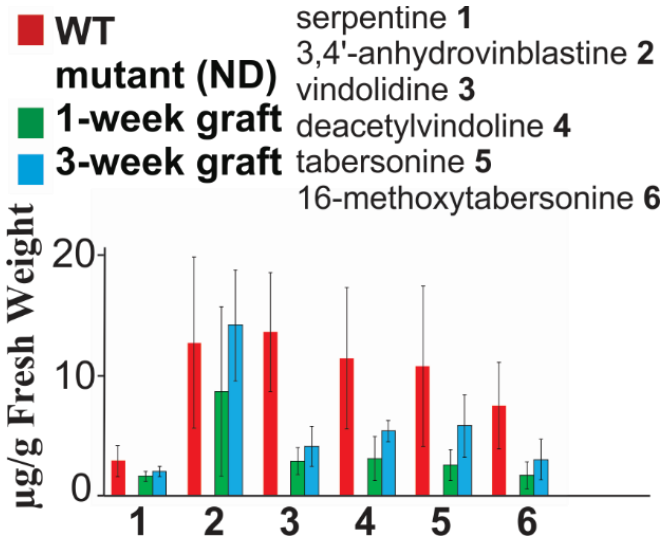


Fig. 3 A very low secologanin, catharanthine, and vindoline mutant line also accumulates other minor MIAs in leaves when grafted on to WT roots. Leaves from the scion were harvested 1- and 3-weeks after grafting onto WT roots. The error bars represent standard deviations of measurements of 4 biological replicates with each line. These MIAs were not detected in the mutant as indicated (ND).

Low secologanin and MIA levels in the M2-1582 line is caused by iridoid pathway down-regulation. It has been proposed based on studies with *C. roseus* cell suspension and hairy root cultures, that the transcription factors BIS1 and/or BIS2 activate expression of 7 genes involved in conversion of geraniol to secologanin (Van Moerkercke et al. 2016). Since secologanin levels were very low in the mutant, the relative expression of *C. roseus* iridoid pathway transcription factors and seven secologanin biosynthesis genes were monitored by real-time PCR (qPCR) mutant and WT leaves. Remarkably, both *BIS1* and *BIS2* transcript levels were 3- and 13-fold lower in the M2-1582 line, respectively in mutant leaf pair 1 than in that of WT (Fig. 4). Expression levels of *BIS1* and *BIS2* were similar in mutant and WT leaf pair 2 while both were several for higher in mutant leaf pair 3 compared to that of WT.

The effects of decreased *BIS1* and *BIS2* expression observed in the mutant leaf pair 1 background were further investigated by observing the expression levels of several different iridoid and MIA pathway genes. Geraniol 10-hydroxylase (*G10H*) that catalyses the conversion of geraniol to form the iridoid precursor 10-hydroxygeraniol was increased 8-fold in the mutant compared to WT (Fig. 5). Two separate genes, 10-hydroxygeraniol oxidoreductase (*10HGO*) and 8-hydroxygeraniol oxidoreductase (*10HGO*) catalyse similar reactions, but only *8HGO* appeared to be expressed in IPAP cells and was suggested to be directly involved in iridoid biosynthesis (Miettinen et al. 2014). Expression of *10HGO* increased 10-fold in the mutant compared to WT, while that of *8HGO* declined almost 200-fold in the mutant compared to the WT. This result provides complementary evidence supporting the direct role of *8HGO* in precursor supply in the secologanin pathway.

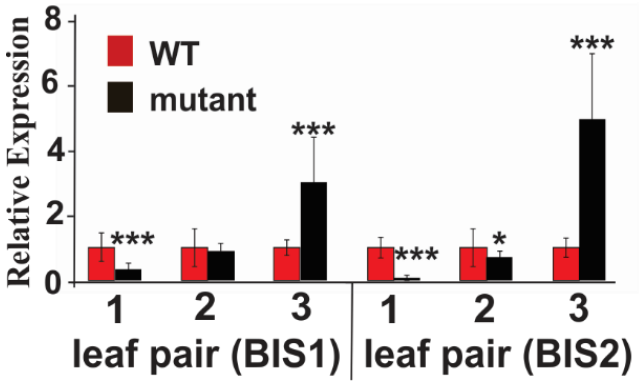


Fig. 4 The expression of transcription factors *BIS1* and *BIS2* is very low in leaf pair 1 of the mutant when compared to those found in the WT. The expression levels of *BIS1* and *BIS2* in the youngest (leaf pair 1) and in developmentally older leaves (leaf pairs 2 and 3) of the mutant were compared to those in WT. The error bars represent standard deviations of four biological and four technical replicates, normalized to the WT profiles in each set of leaf pairs. Two-tailed student t-tests were calculated and (***) denotes <0.001 significance, (**) <0.01 significance and (*) <0.05 significance.

Iridoid synthase (*IS*) (Geu-Flores et al. 2012), originally classified as a progesterone 5β – reductase (*P5βR*), is part of a six-member family found in *C. roseus* (Munkert et al. 2015). Previous studies (Munkert et al. 2015) showed that 6 progesterone 5β-reductase-like genes isolated from *C. roseus* (*CrP5βR1, CrP5βR2, CrP5βR3, CrP5βR4, CrP5βR4*5 and *CrP5βR6)* could functionally reduce 8-oxogeranial to iridoidial, raising the possibility that more than one of these genes might participate in iridoid biosynthesis in addition to the *C. roseus* iridoid synthase (*IS is CrP5βR*5*)* originally identified by Geu-Flores et al. (2012). Their studies concluded that *CrP5βR4* in addition to *CrP5βR5* might also be involved in iridoid biosynthesis. Expression analysis of *IS* showed that an almost 48-fold decline in line M2-1582 compared to WT (*IS* in Fig. 4, *CrP5βR4* in Fig. 6). Remakably *CrP5βR4* also declined over 6-fold in the mutant compared to WT, while gene expression for the other 4 members of this gene family was relatively unaffected (Fig. 6). Expression of 7-deoxyloganetic acid synthase (*7DLS*) and deoxyloganic acid glucosyltransferase (*DLGT*) declined 5- and 3-fold in line M2-1582, respectively, compared to

WT. These results clearly suggested that the trace MIA phenotype of the mutant was likely caused by a mutation that alters expression of *BIS1* and *BIS*2 that would normally activate multiple steps in iridoid biosynthesis (*8HGO, IS (CrP5βR5), 7DLS, DLGT*) and are consistent with results reported in previous studies (Van Moerkercke et al. 2015, 2016).

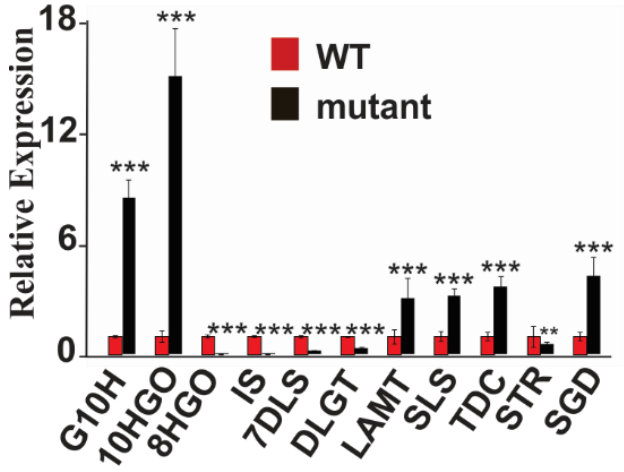


Fig. 5 Comparison of secoiridoid and early stage MIA pathway gene expression profiles in leaf pair 1 of the mutant and WT plants. The error bars represent standard deviations of four biological and four technical replicates, normalized to the WT profiles. Abbreviations: *BIS1* and *BIS2* (bHLH transcription factor 1 and 2); *G10H* (geraniol 10-hydroxylase); *10HGO* (10-hydroxygeraniol oxidase), *8HGO* (8-hydroxygeraniol oxidase), *IS* (iridoid synthase), *7DLS* (7-deoxyloganetic acid synthase), *DLGT* (7-deoxyloganetic acid glucosyltransferase), *LAMT* (loganic acid O-methyltransferase), *SLS* (secologanin synthase), *SGD* (strictosidine-β-glucosidase), *TDC* (tryptophan decarboxylase), *STR* (strictosidine synthase). Two-tailed student t-tests were calculated and (***) denotes <0.001 significance, (**) <0.01 significance and (*) <0.05 significance.

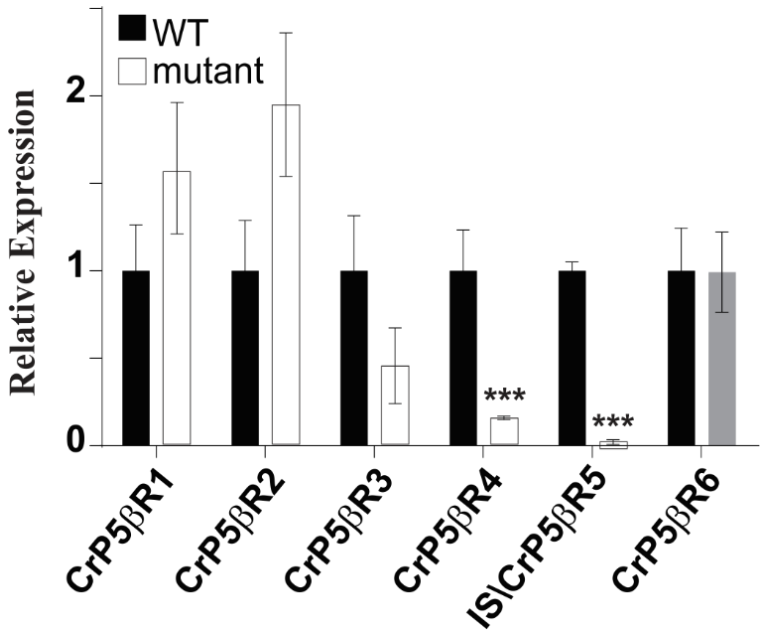


Fig. 6 Comparison of gene expression profiles in leaf pair 1 of six members to (*CrP5βR1* to *6*) of the progesterone 5β-reductase family found in *C. roseus*. The error bars represent *standard deviations* of four biological and four technical replicates normalized to the WT profiles. Iridoid synthase (Geu-Flores et al., 2012) corresponds to *CrP5βR5*.

Expression of *LAMT*, *SLS*, *TDC* and *SGD* are not down-regulated in line M2-1582. LAMT and SLS that catalyze the last two steps in the assembly of secologanin are expressed in leaf epidermal cells (Murata et al. 2008; Guirimand et al. 2011), unlike all previous pathway steps that occur in IPAP cells. Previous studies with *BIS1* expressing hairy root cultures (Van Moerkercke et al. 2015) showed that expression of *LAMT, SLS, TDC* and *SGD* genes were suppressed. Based on these earlier studies, the decline in *BIS1/BIS2* expression observed in the mutant (Fig. 4) released the suppression and resulted in the several-fold increase in *LAMT* (3-fold), *SLS* (3-fold) *TDC* (4-fold) and SGD (4-fold) expression of *STR* was slightly decreased while in the mutant (Fig. 5). Together these results provide significant insights on the different regulation patterns of IPAP localized parts of iridoid biosynthesis compared to the last two steps in iridoid biosynthesis by *LAMT* and *SLS* together with *TDC* and *SGD* in the assembly of strictosidine and downstream MIAs occurring in the leaf epidermis.

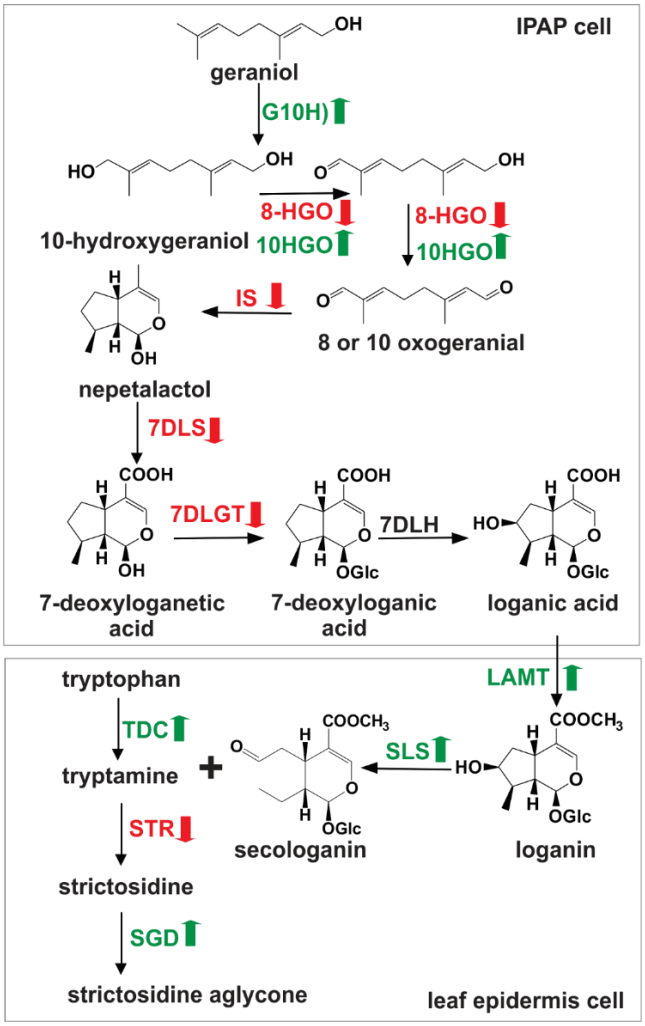


Fig. 7 Summary of gene expression analysis responsible for the low secologanin and MIA phenotype of mutant line M-1582. The iridoid pathway located in IPAP cells is down regulated in several steps highlighted in red. *G10H* and *10HGO* are upregulated in the mutant (green), but it is not clear if this increased expression occurs in IPAP cells. Leaf epidermis reaction steps in iridoid and MIA biosynthesis (*LAMT*, *SLS*, *TDC* and *SGD*) are increased in the mutant (green) while *STR* expression declined significantly, Abbreviations: *G10H* (geraniol 10-hydroxylase; CYP76B6); *10HGO* (10-hydroxygeraniol oxidase), *8HGO* (8-hydroxygeraniol oxidase), *IS* (iridoid synthase), *7DLS* (7-deoxyloganetic acid synthase; CYP76A26), *DLGT* (7-deoxyloganetic acid glucosyltransferase), *7DLH* (7-deoxyloganic acid hydroxylase; CYP72A224), *LAMT* (loganic acid O-methyltransferase), *SLS* (secologanin synthase; CYP72A1); *SGD* (strictosidine-β-glucosidase), *TDC* (tryptophan decarboxylase), *STR* (strictosidine synthase).

3.3 Discussion

A simple TLC screening procedure has been used to identify several EMS-mutagenized *C. roseus* plants with altered levels of MIAs (Qu, Thamm, et al. 2018; Qu, Easson, et al. 2018; Edge et al. 2018) that have been successfully characterized. The present identification of a very low iridoid and MIA accumulating mutant provided an interesting experimental system to study the possible molecular basis for the phenotype and to determine if iridoids are mobile molecules in an MIA accumulating plant such as *C. roseus*. The mobility of iridoids has been documented in iridoid accumulating plant species such as *Asarina scandens* (Gowan et al. 1995) and in *Antirrhinum majus* (Beninger et al. 2007). When leaves of *A scandens* were exposed to $^{14}CO_2$, the iridoid antirrhinoside was labelled within 20 min and the iridoid was transported together with labelled sucrose via the petiole through the phloem (Gowan et al. 1995). Similar labelling studies with snapdragon leaves rapidly labelled the major iridoids, antirrhinoside and antirrhide but only antirrhinoside and sucrose appeared to be phloem mobile (Beninger et al. 2007). These studies both suggested that formation and transport of iridoids in the phloem might deter feeding by phloem feeders such as aphids (Gowan et al. 1995; Beninger et al. 2007), while others have speculated that iridoids may also function to maintain osmoregulatory water balance within the plant (Voitsekhovskaja 2005). The transport of secologanin in the *C. roseus* mutant demonstrated by secologanin uptake and transport experiments (Fig. 1 and 2) suggest that iridoid transport may be quite common and is not restricted to members of the iridoid accumulating Scrophulariaceae family such as in the genus *Asarina* and *Antirrhinum*.

Exogenous provision of secologanin to mutant roots or grafting of WT rootstock to the mutant resulted in transport of secologanin to mutant leaves where the secologanin was transformed into the major leaf MIAs catharanthine and vindoline (Fig. 1, 2). Gene expression

analysis of the mutant revealed that the transcription factors *BIS1* and *BIS2* were considerably diminished at certain stages of development in young *C. roseus* leaves of the mutant, while they were more highly expressed than WT in older leaves (Fig. 4). Since both *BIS1* and *BIS2* transcription factors function as heterodimers to activate several steps in iridoid biosynthesis in IPAP cells (Van Moerkercke et al. 2015, 2016), it is not surprising that IPAP localized expression of *8HGO, IS (CrP5βR5), 7DLS* and *DLGT* was greatly decreased, while leaf epidermis localized expression of *LAMT, SLS, TDC* and *SGD* was increased in the *C. roseus* mutant (Fig. 5). The significantly increased expression of he G10H and 10HGO in young leaves of the mutant, suggest that regulation of these early iridoid biosynthesis steps are also respond to other regulators in addition to *BIS1* and *BIS2*, perhaps leading to formation other 10-hydroxygeraniol derivatives, unrelated to secologanin that could only be observed in the mutant background.

While progesterone 5β –reductase was first characterized in the formation of 5β-pregnane-3,20-dione involved in cardenolide biosynthesis, these enzyme appear to be widely spread among angiosperms and that the IS activity of these enzymes appears to be intrinsic to this family. (Munkert et al. 2015). While all 6 members of the progesterone 5β –reductase family in *C. roseus* catalyse the formation of iridoidial only *CrP5βR4* and IS/*CrP5βR*5 are IPAP localized and are highly expressed compared to the other 4 members of the family. The 6- and 48-fold declines in *CrP5βR4* and IS/*CrP5βR*5 gene expression in the mutant background provide strong support for this conclusion.

In summary EMS-mutagenesis has provided a useful tool to create and study variable traits in plants (Chen et al. 2016; Cooper et al. 2008). Recently, several EMS-mutagenized *C. roseus* plants with altered MIA profiles have been characterised. One low vindoline mutant

contained a single amino acid substitution that reduced the activity of 17-*O*-acetylstemmadenine oxidase by over 97% and resulted in mutant plants that accumulated 17-*O*-acetylstemmadenine instead of catharanthine, tabersonine or vindoline in *C. roseus* leaves (Qu, Safonova, et al. 2018). Another low vindoline mutant with a single amino acid substitution that reduced the activity of tabersonine 3-reductase by 95% (Edge et al. 2017) resulted in mutant plants that accumulate 2,3-epoxytabersonine rather than vindoline. A third *C. roseus* EMS-mutant contained elevated levels of ajmalicine and reduced levels of catharanthine and vindoline (Qu, Thamm, et al., 2018). The present study corroborates the important role(s) of *BIS1* and *BIS2* in activating the iridoid pathway involved in loganic acid biosynthesis in IPAP cells. A major finding shows that secologanin is transported by unknown mechanisms from sites of biosynthesis to other cells or organs where the iridoid may also participate in MIA biosynthesis or in role(s) that remain to be established.

3.4 Methods and Materials

3.4a Chemicals

All reagents were purchased from Sigma-Aldrich (Oakville, Canada) in ≥90% purity unless otherwise stated. Secologanin was purchased from Sigma-Aldrich (Oakville, Canada). Loganin was obtained from Wako Pure Chemicals (Osaka, Japan). 7-Deoxyloganic acid was prepared from 7-deoxyloganin tetraacetate as described by Salim (2013) and Nagatoshi et al. (2011). Loganic acid was obtained from Extrasynthese (Genay, France). The standard curves for Catharanthine (y=0.0394x), Vindoline (y=0.0155x+3.226), Secologanin (y=0.018x-19.925), 3',4'-Anhydrovinblastine (y=0.0286+10.567), 16-methoxytabersonine (y=0.0182+0.1555), Deacetylvindoline (y=0.0504+3.2245), Serpentine (y=0.0233x), Tabersonine

(y=0.0106+3.8138), and Vindolidine (y=0.0155x+3.226) quantification were obtained from previous studies (Qu et al. 2015; Salim et al. 2013). X represents peak area. Y is nanograms of alkaloid analyzed.

3.4b Plant Materials

Seeds (WT and mutagenized) of Catharanthus roseus (L.) G. don were donated by PanAmerican Seed Co. (panamseed.com). The methods used for mutagenesis and for seed production were described previously (Edge et al. 2018). Control (WT) and mutant M2 seeds (~4000) were cultivated in soil in a greenhouse (16/8 h light/dark) at 30 ˚C.

3.4c *In vitro* Plant Cultivation

C. roseus variety Pacifica Peach (wild-type) and EMS mutant (M2-1582) were grown in sterile magenta boxes containing plant regenerating media composed of autoclaved Woody plant medium salts (2.3g/L), indole butyric acid (0.2 mg/L), glycine (2.0 mg/L), myo-inositol (100 mg/L), nicotinic acid (0.5 mg/L), pyridoxine (0.5 mg/L), thiamine-HCl (0.1 mg/L) and sucrose (30 g/L), activated charcoal (2.5 g/L), gelrite (2.6 g/L) adjusted to pH 5.8 with ~4 drops of 1M KOH. Plants were propagated by taking cuttings and inserting them into new medium.

3.4d Feeding Experiments

Several feeding experiments were attempted. Mature plants, both wild-type and mutant lines were transferred from the magenta box to 50 ml falcon tubes with 3 – 5 ml of Murashige and Skoog (MS) fluid for the control and the MS fluid plus the addition of 1 mM of either secologanin, loganin, loganic acid or 7-deoxyloganic acid (Murashige & Skoog 1962). Leaves were removed after 48 hours for alkaloid and iridoid quantification by UPLC-MS.

3.4e Grafting

The Top Wedge Grafting method was used for all grafting (Kimura & Sinha 2008). Experimentation found that the stock plant was unable to support a scion with more than one leaf pair. A horizontal cut made with a scalpel was used to separate and remove the terminal portion of the stock plant. The scalpel was then used to split the stem of the stock plant, cutting vertically approximately 1 cm. The scion is cut horizontally around the second node, and then the lower 1 cm is tapered using the scalpel into a "V' shaped wedge. The scion is inserted into the stock, and grafting wax is applied to reduce desiccation and provide support.

3.4f MIA Screening and UPLC-MS Quantification

EMS-mutagenized *C. roseus* leaves were dipped in chloroform to harvest MIAs secreted to the leaf surface, followed by methanol extractions of leaves to harvest alkaloids found within the leaf. Samples were dried by vacuum centrifugation and suspended in a small amount of methanol in preparation for screening by thin layer chromatography (TLC silica gel 60 F_{254}). After TLC separation using ethylacetate:methanol (9:1), the plates were exposed to short (252 nm) and long (365 nm) wave ultraviolet light to visualize MIAs and were documented by photography.

Samples for UPLC-MS analysis were prepared identically to TLC preparation. Analytes were separated using an Aquity UPLC BEH C18 column with particle size 1.7 µm and dimensions of 1.0-50 mm. Alkaloids were quantified using peaks measured by a photodiode array at 280 nm or 305 nm.

3.4g RNA extraction and cDNA synthesis

Each leaf (leaf pair 1, leaf pair 2 or leaf pair 3) stem and root was ground in liquid nitrogen to a fine powder and immediately mixed with 0.8 mL of Trizol reagent (Life Technologies) to extract the RNA. 0.16 mL of chloroform was added, the aqueous phase was removed. After precipitation

with 0.4 mL of isopropanol and 0.8 mL of ethanol, the RNA was suspended in 25 µl of DEPC H_2O. The RNA was subsequently treated with DNAase I to remove genomic DNA. DNAase I was in turn inactivated with 1 mM EDTA for 15 min at 75°C. Quality and quantity of RNA were determined visually by gel electrophoresis and nanophotometry. First-strand cDNA was synthesized using SuperScript II reverse transcriptase (Invitrogen) and oligo (dT) 12-18 primer (Invitrogen) using 1-4 µg total RNA according to manufacturer's protocol.

3.4h Quantitative Real-time PCR for iridoid expression analysis

qRT-PCR was performed (CFX96^{TM} Real-Time system, Bio-Rad, Hercules, CA and CFX96 Connect^{TM} Real-Time system, Bio-Rad, Hercules, CA) using 5 µL iTaq^{TM} Universal SYBR® Green Supermix (Bio-Rad, Hercules, CA) 0.25 µL primer and cDNA template at 5 ng total RNA per 10 µL reaction volume.

The reaction conditions for qRT-PCR included 1 cycle of 95°C for 1 min and 40 cycles of 95°C for 15s and 58°C for 1 min. Critical Threshold (Ct) values were used to calculate the relative transcript abundance with the housekeeping gene 60S ribosome cDNA as the internal control. Expression levels of each target gene were analyzed with the Bio-Rad CFX Manager Software (Bio-Rad) and normalized to the 60S ribosomal cDNA (Table 2).

	Wild-type	Mutant
Whole Plant (GFW)	2.27 ± 0.38	2.21 ± 0.56
Leaves/stems (GFW)	1.89 ± 0.26	0.69 ± 0.17
Roots (GFW)	0.38 ± 0.11	1.52 ± 0.49
% distribution of GFW	16.5%	69.3%

Table S1 GFW of WT and M2-1582 mutant. GFW = Gram fresh weight. Error represents standard deviation obtained for independent *in vitro* grown plants. (n=6).

Chapter 4: Bioinformatic and morphological characterization of a low iridoid and MIA accumulating *Catharanthus roseus* mutant

Unpublished

4.1 Results

4.1a M2-1582 morphologies are likely related to auxin homeostasis

The low iridoid and MIA accumulating mutant M2-1582 grown *in vitro* displayed several physiological abnormalities including small leaves, slow growth, short internode spaces, undifferentiated root architecture, lack of flowering, premature leaf abscission, and a ridged stem (Figure 6). In addition to altered root architecture, the roots contributed to 69.3% of total plant mass compared to the 16.5% of the *C. roseus* WT var. Pacifica Peach (Kidd, 2016).

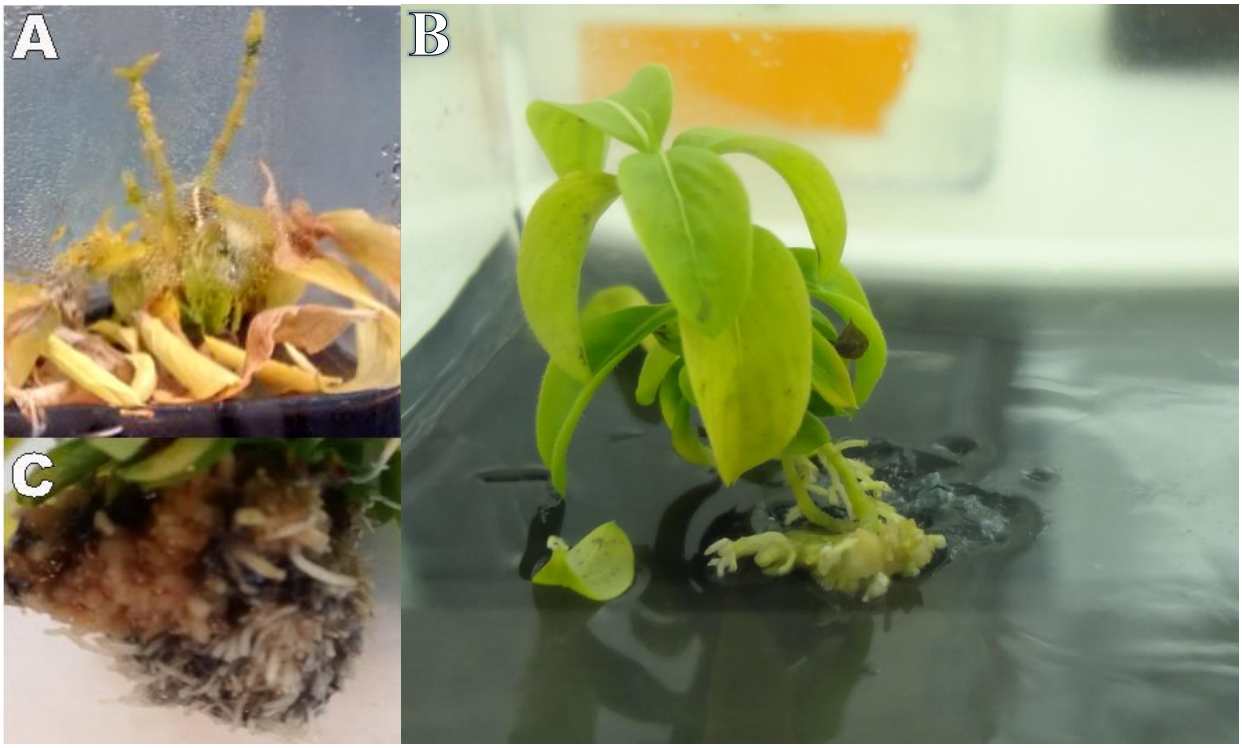


Figure 6: Old (A and C) M2-1582 (Kidd, 2016) and a younger (B) M2-1582 with green roots.

These morphologies had previously been attributed to ethylene or jasmonate excess due to the early leaf senescence, low-iridoid production and lack of flowering (Kidd, 2016). After a more granular assessment of the plant and the literature, a disruption in auxin homeostasis was determined to be a likely candidate for inducing most, if not all the morphologies. Cross sections of the *in vitro* grown M2-1582 and WT revealed M2-1582 had significantly altered secondary vascular deposition. The WT cross section was taken from the most heavily lignified part of the stem at the base, while the M2-1582 cross section was taken towards the stem apex where the vasculature is actively differentiating. Cross sections imaged together allows for relative measurements to be compared between pith and vasculature (Figure 7). Approximate representative radii of vasculature and pith diameter were taken to accommodate for the warping of the cross section caused by the razor blade. The ratio of pith/vasculature for M2-1582 was 1.45 while the pith/vasculature ratio for the WT was 3.85. This represents a 2.66-fold decease in the ratio of pith to vasculature (Appendix). Secondary vasculature expands towards the pith, therefore the ratio of pith/vasculature does not get larger in older sections of stem (Zhang et al. 2011).

One M2-1582 *in vitro* cultivated plant displayed a unique morphology after two years of uniform propagation. It displayed relatively normal root architecture, but the roots were green (Figure 6). Green roots are characteristic of high cytokinin levels inducing chloroplast production in roots (Kobayashi et al. 2012). This unique M2-1582 plant displaying relatively normal green roots was suggestive of increased cytokinin production, counteracting the typical root morphology of M2-1582. Cytokinin's antagonistic relationship with auxin responses are well-established (Kobayashi et al. 2012; Marhavý et al. 2014; Weiss & Ori 2007).

Figure 7: Cross section of wild-type *C. roseus* (WT) (right) taken at the base of the stem and M2-1582 stem cross section (left) taken between the first and second leaf pairs. Diameter of pith (yellow) and radii of vasculature (white) are taken as approximates.

4.1b M2-1582 is sensitive to exogenous auxin

M2-1582 and WT cuttings were taken and put into the standard plant regeneration media with different concentrations of the auxin indole-3-butyric acid (IBA). After 90 days, M2-1582 with the standard 200 ug/L IBA displayed the normal morphology of M2-1582 (Figure 8). 75% of the M2-1582 cuttings died when grown with 0 ug/L IBA, the lone survivor displayed a short, green primary root. Some of the M2-1582 grown in 100 ug/L IBA displayed gravitropic response issues in the root one month into the experiment. After 90 days, M2-1582 grown at 100 ug/L were more similar in morphology to WT than M2-1582 grown in 200 ug/L IBA (Figures 9 & 10). The roots were distinguished and had similar patterning to WT with several primary roots resonating from the stem base. The aerial part of the plant was also more reminiscent of WT, with large, dark green leaves and no abscission. In contrast, the WT never displayed obvious changes in morphologies during the over 2 years of *in vitro* cultivation and under the different concentrations of IBA. The extreme morphological differences of M2-1582 under these relatively small changes in auxin supports the impact of the EMS mutation on the plant involving auxin biosynthesis and/or regulation.

Figure 8: 1-month old gravitropism confused roots of M2-1582 grown in media containing 100 ug/L IBA (top), 90 day old roots of M2-1582 grown in media containing 200 ug/L IBA (middle) and 90 day old roots of M2-1582 grown in media containing 0 ug/L IBA (bottom).

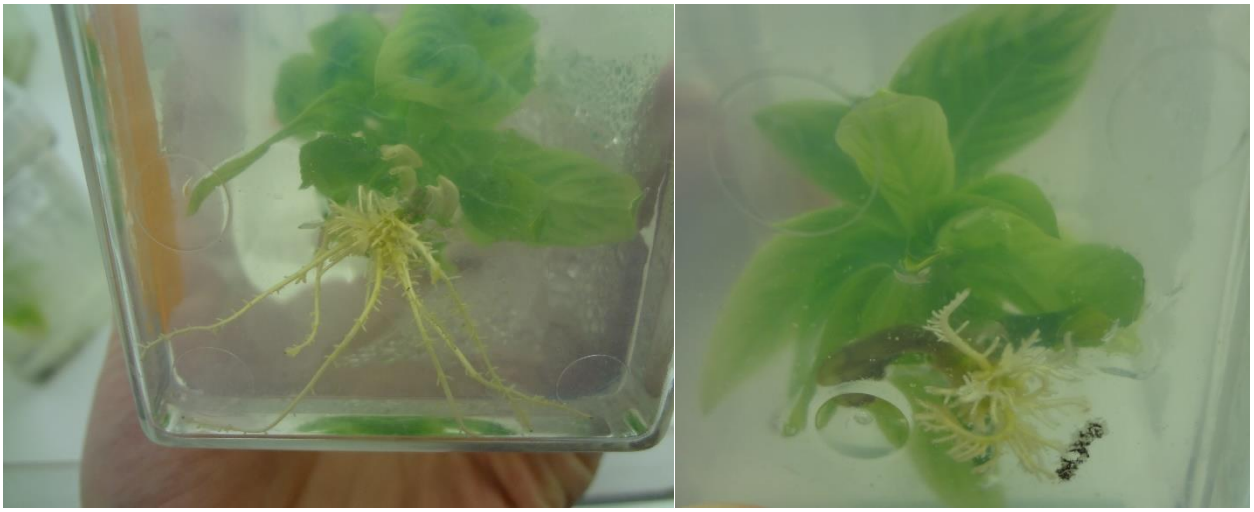


Figure 9: Roots of wild-type *C. roseus* (WT) (left) and M2-1582 (right) 90 days after cutting and transplanting to 100 ug/L IBA containing media.

Figure 10: Wild-type *C. roseus* (WT) (left) and M2-1582 (right) 90 days after cutting and transplanting to 100 ug/L IBA containing media.

4.1c Trend summary of differential gene expression and example mutation detection analysis

The completed *C. roseus* reference transcriptome has over 200,000 sequences with an average length of 889 bp and a N50 of 1256. Some of these sequences are real homologues from duplication events or are created by single nucleotide polymorphism (SNP) variants between *C. roseus* cultivars. In the assembly some of the real homologues are combined due to the lack of sequence consensus in assembly typically provided by an assembled and curated genome (Claros et al. 2014). Differential gene expression analysis of the first leaf, second leaf, and roots between M2-1582 and WT showed a total of 52,019 sequences in the reference transcriptome were expressed across these organs. Of those, roots had the most upregulated genes, with 2,979 over two-fold upregulated and 1,095 over two-fold downregulated in M2-1582. The second leaf pair had the most downregulated genes with 2,838 two-fold downregulated and 1,840 upregulated. The first leaf pair of M2-1582 had 978 genes two-fold downregulated and 1,343 over two-fold upregulated.

Clustering of two-fold increase or decrease in expected annotated gene representation in M2-1582 by gene ontology terms shows which cellular components are most impacted (Figure 11) (Mi et al. 2017). Genes with two-fold increase in roots annotated as responsive to hydrogen peroxide were represented almost 16-fold higher than the expected representation of all cellular biological processes. The most impacted cellular components in the roots were the mitochondria. Genes with an over two-fold decrease in expression involved with the mitochondrial intermembrane space and respiratory chain complex were over 12- and 8-fold increased respectively in expected representation.

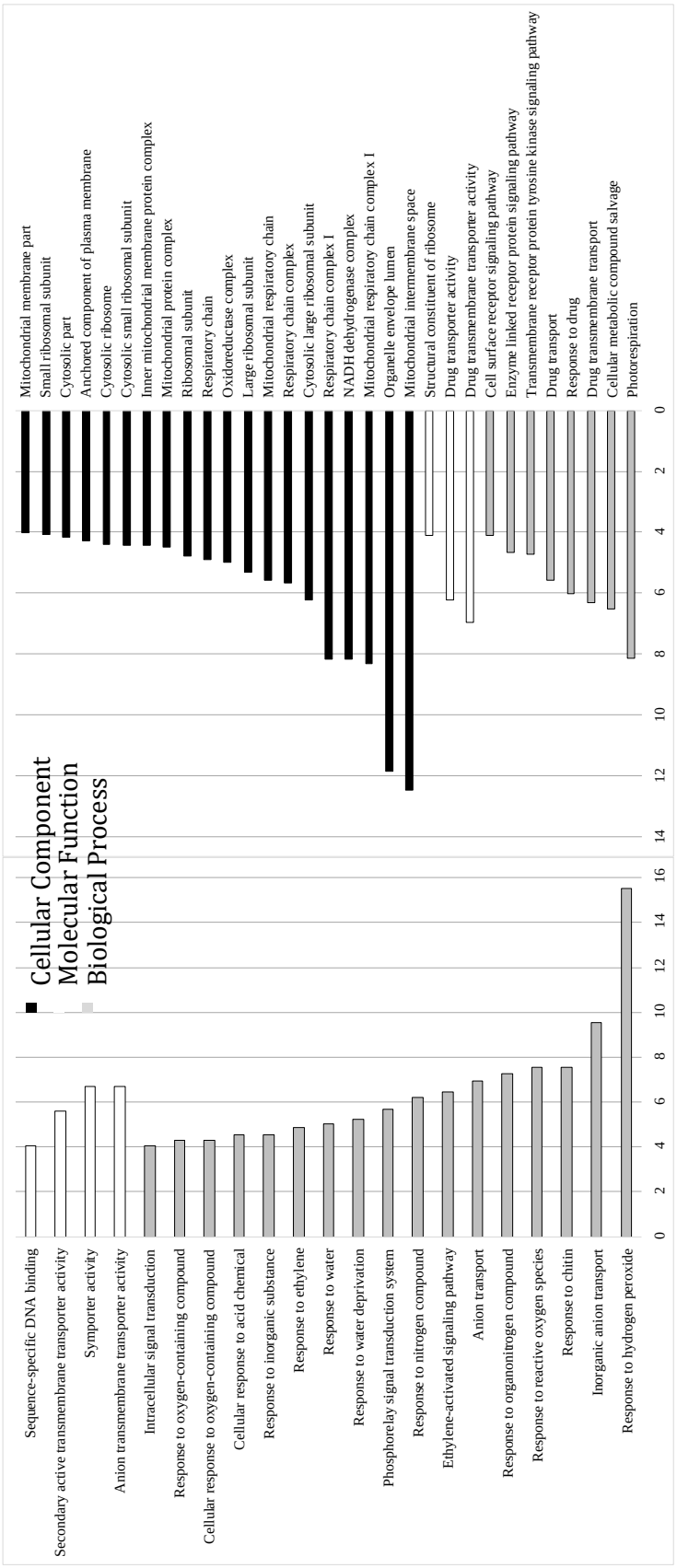


Figure 11: Two-fold increase (left) or decrease (right) in relative transcript abundance in roots between M2-1582 and wild-type *C. roseus* (WT) in expected annotated gene representation of gene ontology terms. Gene ontology terms are divided into cellular component (black), molecular function (white) and biological process (grey).

String (Szklarczyk et al. 2015) was used to look at which processes in cellular function had the most impacted known protein-protein interactions (Figures 12 & 13). The example mutated genes from the trial mutation detection run shows how the physical interactions between mutated genes and differentially regulated genes, as a direct or indirect consequence of the mutations, can be visualized. A homologue of a lignin biosynthetic gene is increased in the roots of M2-1582 and interacts with a downregulated homologue involved in amino acid processing. Several other proteins involved in protein processing known to interact with each other were also downregulated in roots. The greatest clustering of upregulated interacting protein in roots are involved with translation. The same trend of upregulated interacting proteins involved in translation and lignin biosynthesis is found in the leaves of M2-1582. The greatest clustering of interacting downregulated proteins in leaves are involved with protein and amino acid processing.

High-throughput annotation and visualization of data has its uses in communication and initial exploration. Manual confirmation and reconstruction of pathways is the most robust method of sorting data from non-model organisms.

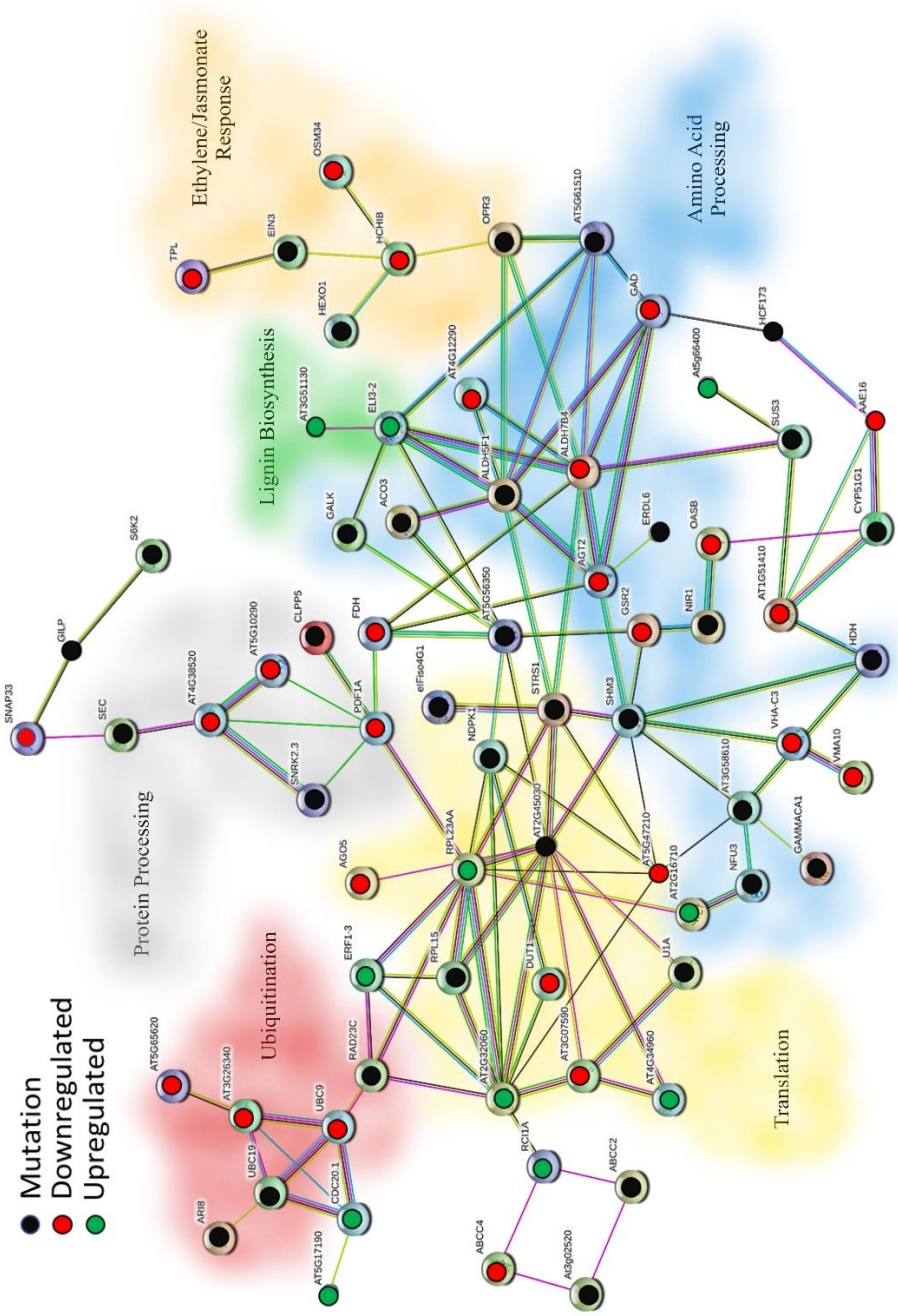


Figure 12: Protein-protein interaction network of genes 2-fold differentially expressed in both first and second leaf pairs of M2-1582. The network shows upregulated (green), downregulated, (red) and predicted to contain a nonsense or start codon EMS mutation (black) in the example analysis. The lines between genes represent different lines of evidence: fusion (red), neighborhood (green), cooccurrence (blue), experimental (purple), textmining (yellow), database (light blue) and coexpression (black). The genes are clustered by colour coding to represent ubiquitination (red), translation (yellow), protein processing (grey), amino acid processing (blue), lignin biosynthesis (green) and ethylene/jasmonate response (orange).

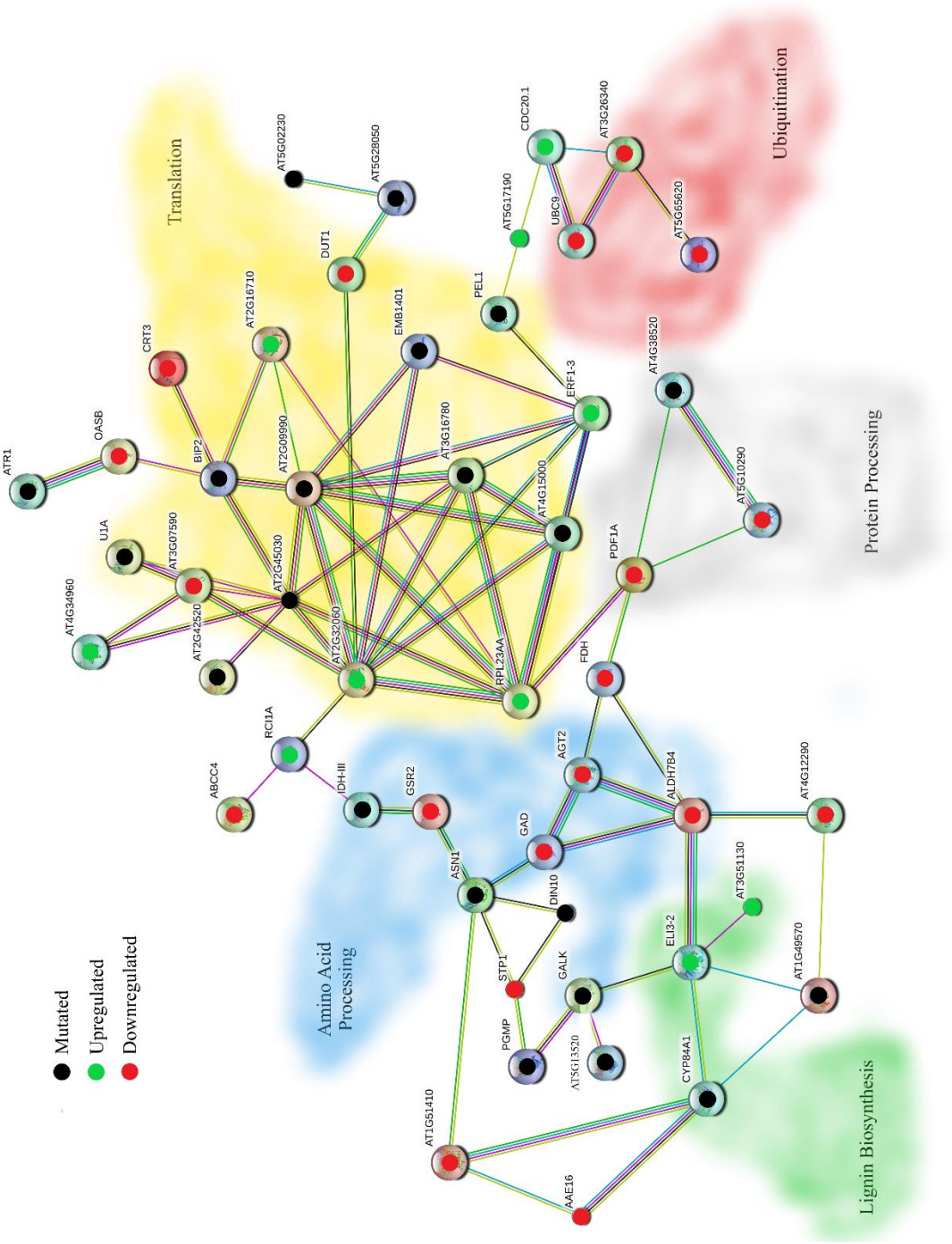


Figure 13: Protein-protein interaction network of genes 2-fold differentially expressed in roots of M2-1582. The network shows upregulated (green), downregulated, (red) and predicted to contain a nonsense or start codon EMS mutation (black) in the example analysis. The lines between genes represent different lines of evidence: fusion (red), neighborhood (green), cooccurrence (blue), experimental (purple), textmining (yellow), database (light blue) and coexpression (black). The genes are clustered by colour coding to represent ubiquitination (red), translation (yellow), protein processing (grey), amino acid processing (blue) and lignin biosynthesis (green).

4.1d Differential gene expression of different organs reveals auxin biosynthetic, regulatory and auxin interacting pathways are altered in M2-1582

All specific genes were manually confirmed to separate duplicates created by the assembly from real homologues. Assessment of differentially expressed iridoid and MIA biosynthetic genes and regulators showed the analysis was mostly in agreement with the qPCR data conducted a year before the *in vitro* RNA isolation for Illumina sequencing. BIS1 and BIS2 both showed downregulation in the first leaf pair (FLP), along with most of the iridoid biosynthetic genes (Kidd et al. *in press*). The analyses disagree on the differential expression of G8O, LAMT and SLS (Table 1). The sequencing data suggests these three genes are all downregulated over two-fold in the FLP of M2-1582 while the qPCR data shows all three at least three-fold upregulated. Otherwise the two analyses agreed on the downregulation of the other iridoid biosynthetic genes and the upregulation of SGD in M2-1582. OCRA3, a regulator of iridoid biosynthesis responsive to MeJA that was not investigated by qPCR, was shown to be significantly downregulated in the FLP and second leaf pair (SLP) but upregulated in the roots.

Protein	FLP MT/WT	SLP MT/WT	Root MT/WT
Geraniol 8-oxidase	0.05	0.02	0.05
8-hydroxygeraniol dehydrogenase	0.03	0.04	0.41
Iridoid synthase	0.11	< 0.00	*
7-deoxyloganic acid synthase	0.58	0.52	1.60
7-deoxyloganetic acid glucosyltransferase	1.15	1.45	2.74
Loganic acid O-methyltransferase	0.16	0.47	32.61
Secologanin synthase	0.38	1.96	8.78
Tryptophan decarboxylase	0.54	0.78	5.38
Strictosidine synthase	0.93	1.15	1.73
Strictosidine β-glucosidase	1.21	1.96	19.96
10-hydroxygeraniol dehydrogenase	4.23	3.74	0.78
BIS1	0.33	0.31	0.11
BIS2	0.49	0.32	0.78
ORCA3	*	*	**

Table 1: Differential expression of secoiridoid biosynthetic genes and regulators between M2-1582 (MT) and wild-type *C. roseus* (WT). (*) Denotes presence in WT but not MT, while (**) represents a presence in MT but not WT. Supporting information in Appendix: Table 5.

To explore auxin biosynthesis TAA, YUC1 and a YUC homologue (84% ident) were identified, both were downregulated in leaves and upregulated in roots. FPKMs of TAA in the leaves of M2-1582 were below the threshold of 0.01 FPKM (Table 2). The root expression of TAA goes from below the threshold in the WT to 7.61 FPKM in M2-1582. Expression of auxin biosynthesis is expected to be low in total leaf and root RNA extract due to the localized expression in specific cell types (Chen et al. 2014). The only PIN homologue (87% ident) identified was over two-fold downregulated in both leaves and over three-fold upregulated in roots. The homologue of an auxin influx protein (84% ident) showed the greatest differential expression in the second leaf pair, a decrease by over two-fold. The auxin vacuole transporter homologue of WAT1 (84% ident) was upregulated in all sampled organs, with the most drastic being over 8-fold in roots. A GH3.6 homologue (88% indent), responsible for the turnover of biologically active auxin, was downregulated 100-fold in the first leaf pair and almost 22-fold upregulated in roots. Two potential auxin response factor homologues (80% and 75% ident) were identified; weighing the relative difference in FPKM between the homologues, there is a general increase in expression across all sampled organs of M2-1582. Increase in root auxin biosynthetic, transport and turnover genes indicates the roots could be producing more auxin than normal. Other auxin biosynthetic pathways were investigated but no close homologues to known biosynthetic genes were found.

Auxin	FLP MT/WT	SLP MT/WT	Root MT/WT
Tryptophan aminotransferase	*	*	**
Flavin monooxygenase 1	0.05	0.04	3.73
Flavin monooxygenase 2	< 0.00	0.12	8.18
Indole-3-acetic acid-amido synthetase	0.01	1.02	21.98
Auxin efflux carrier component	0.38	0.48	3.51
Auxin influx carrier protein	0.83	0.39	1.15
Walls are thin protein	1.34	2.33	8.11
Auxin reponse factor 1	1.57	1.74	1.25
Auxin reponse factor 2	0.64	0.91	1.15
Cytokinin			
Adenosine phosphate-isopentenyltransferase	0.09	*	1.81
Cytokinin dehydrogenase	3.51	12.46	0.36
Gibberellin			
Gibberellin 20-oxidase	1.18	0.05	15.56
Ent-kaurene oxidase	1.07	4.28	3.31
Ent-kaurenoic acid oxidase	0.79	1.86	2.67
GA-2 oxidase	4.20	6.39	5.14
Tempranillo	1.54	0.56	8.31
Ethylene			
1-aminocyclopropane-1-carboxylate synthase 1	< 0.00	*	0.24
1-aminocyclopropane-1-carboxylate synthase 2	< 0.00	5.87	0.04
1-aminocyclopropane-1-carboxylate synthase 3	1.62	57.20	0.30
Ethylene insensitive 3-like 1	1.08	2.77	0.50
Jasmonate			
Jasmonic acid-amido synthetase	2.77	5.11	1.51

Table 2: Differential expression of plant hormone anabolic and catabolic biosynthetic genes and regulators between M2-1582 and wild-type *C. roseus* (WT). (*) Denotes presence in WT but not M2-1582, while (**) represents a presence in M2-1582 but not WT. Supporting information in Tables 6 & 9.

Cytokinin biosynthetic and turnover genes were also impacted in leaves and roots of M2-1582 (Table 2). An adenylate isopentenyltransferase (IPT) homologue (70% ident), potentially involved in cytokinin biosynthesis, was downregulated over 10-fold in leaves and upregulated

80% in roots (Marhavý et al. 2014). A cytokinin dehydrogenase (CKX) homologue (80% ident), involved in cytokinin turnover, was upregulated over three-fold in leaves and downregulated over two-fold in roots.

GA biosynthetic, turnover and response homologues were all upregulated in roots with mixed differences of expression in the leaves of M2-1582 (Table 2). These include the potential biosynthetic gene homologues of GA-20ox (88% ident), KO (71% ident) and KAO (79% ident), each at least two-fold upregulated in roots. The GA turnover gene GA2ox homologue (92% ident) was upregulated in each sampled organ at least four-fold in M2-1582. The potential GA biosynthesis negative regulator TEM homologue (76% ident) was most differentially expressed in the roots with an 8-fold increase.

Three potential ethylene biosynthetic gene homologues of 1-aminocyclopropane-1-carboxylate synthase (ACS) (85%, 81% and 77% ident) and one ethylene response transcription factor homologue (88% ident), ethylene insensitive 3-like (EIN3), were identified (Wang et al. 2002). All were downregulated in the roots of the M2-1582 (Table 2). Two of the ethylene biosynthetic homologues and the response factor are all upregulated over two-fold in the second leaf pair, with little difference in the first leaf pair. The jasmonate biosynthetic gene JAR1 homologue (80% ident) was upregulated in all sampled organs, with a 50% increase in root and over two-fold increase in the leaves of M2-1582.

Protein	FLP MT/WT	SLP MT/WT	Root MT/WT
Anthranilate synthase alpha-subunit	0.07	0.02	1.82
Anthranilate synthase beta-subunit	1.79	0.96	2.93
Anthranilate phosphoribosyltransferase 1	1.78	2.64	1.82
Anthranilate phosphoribosyltransferase 2	1.77	2.24	2.65
Tryptophan synthase alpha chain	1.22	0.73	2.55
Tryptophan synthase beta chain	0.28	1.20	4.74

Table 3: Differential expression of tryptophan biosynthetic genes between M2-1582 (MT) and wild-type *C. roseus* (WT). Supporting information in Tables 7 & 10.

Tryptophan biosynthetic gene homologues were all upregulated in the roots of M2-1582 (Table 3) (Radwanski 1995). Two potential homologues of the alpha and beta chain (82% and 73% ident) of anthranilate synthase (trpE) were identified (Figure 14). The alpha subunit was downregulated over 10-fold in both sampled leaf organs. Two homologues of the second tryptophan biosynthetic pathway gene (85% and 83% ident) anthranilate phosphoribosyltransferase (trpD) were identified and shown to be upregulated in all sampled organs. The last two steps in tryptophan biosynthesis are catalysed by tryptophan synthase alpha (trpA) and beta (trpB) chains, a potential homologue of each was identified (77% and 84% ident respectively). Apart from their upregulation in roots, the only large difference in expression is the over three-fold downregulation of trpB in the first leaf pair of M2-1582.

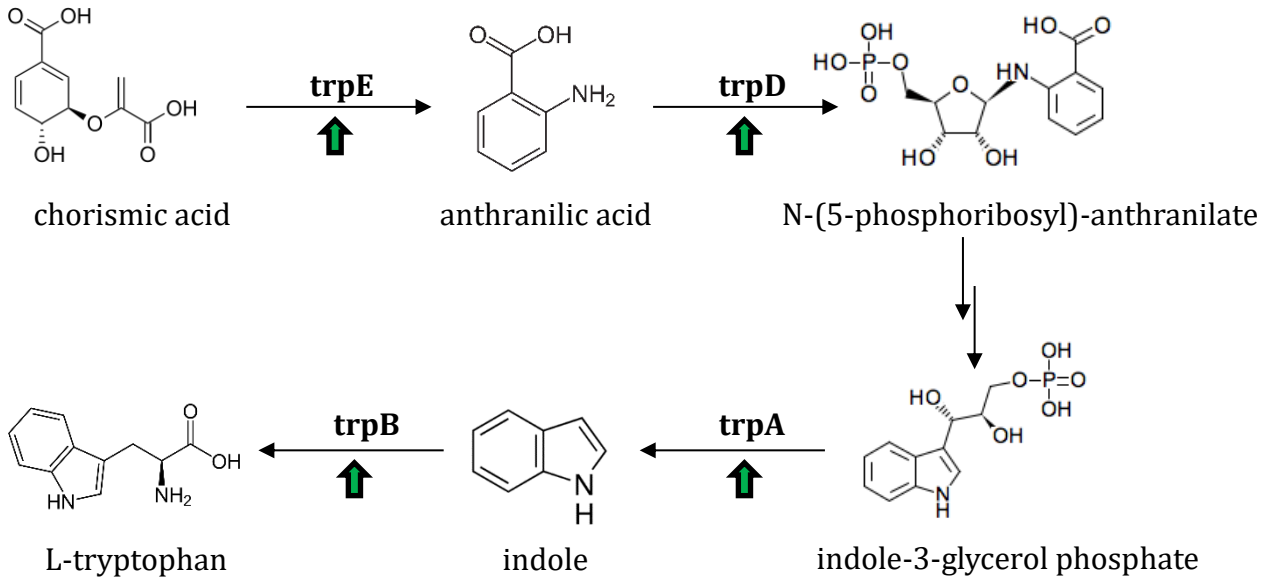


Figure 14: L-tryptophan biosynthesis showing upregulation of all identified genes in the roots of M2-1582. The outlined biosynthetic pathway consists of anthranilate synthase α+β (trpE), anthranilate phosphoribosyltransferase (trpD), tryptophan synthase α (trpA) and tryptophan synthase β (trpB).

4.1e Lignin biosynthetic genes and regulators were mostly upregulated in the roots of M2-1582

All lignin biosynthetic genes and potential homologues in the phenylpropanoid pathway were upregulated in the roots of M2-1582, except C4H and a F5H homologue (Figure 15). PAL was upregulated in all sampled organs by at least 60%, while C4H was downregulated in all organs by at least 50% (Table 4). Homologues of 4CL and HCT (99% and 90% ident) were upregulated over 60% in roots but only slightly differentially expressed in leaves. A C3'H homologue (84% ident) was upregulated in all organs, with the roots showing the least increase of expression of 38%. Conversely, COMT showed the greatest increase in expression in roots of 68%, with a minimal increase in the FLP. A potential homologue of CCR (79% ident) was upregulated in all sampled organs, with the greatest increase in roots by over two-fold. CAD was downregulated in leaves and upregulated over two-fold in the roots of M2-1582.

Protein	FLP MT/WT	SLP MT/WT	Root MT/WT
Phenylalanine ammonia-lyase	3.86	1.71	1.66
Cinnamate 4-hydroxylase	0.42	0.40	0.65
4-coumarate-CoA ligase	1.18	1.11	2.56
Hydroxycinnamoyl transferase	0.95	0.83	1.60
p-coumaroyl shikimate 3'-hydroxylase	1.89	2.62	1.38
Caffeic acid 3-O-methyltransferase	1.22	1.01	1.68
Cinnamoyl-CoA reductase 1	1.50	1.40	2.36
Cinnamyl alcohol dehydrogenase	0.78	0.21	2.29
Ferulate 5-hydroxylase	1.11	2.26	0.42
Homeobox-leucine zipper 1	0.47	0.50	1.56
Homeobox-leucine zipper 2	0.48	0.44	1.19
Homeobox-leucine zipper 3	0.62	0.43	1.62

Table 4: Differential expression of lignin biosynthetic genes between M2-1582 (MT) and wild-type *C. roseus* (WT). Supporting information in Tables 8 & 11.

A F5H homologue (82% ident) that converts G monomers to S monomers was upregulated in leaves but downregulated over two-fold in roots (Table 4) (Fraser & Chapple 2011). Three KD1 and REV homologues (88%, 87% and 87% ident) involved in auxin signaling

of secondary xylem differentiation and leaf abscission were all downregulated in leaves and upregulated in roots of M2-1582 (Ma et al. 2015; Porth et al. 2014). All three homologues show an over two-fold reduced expression in both leaf organ samples, except for one homologue in the first leaf pair which has a 61% reduction in expression. All three homologues had increased expression patterns in the roots of 19%, 56% and 62%.

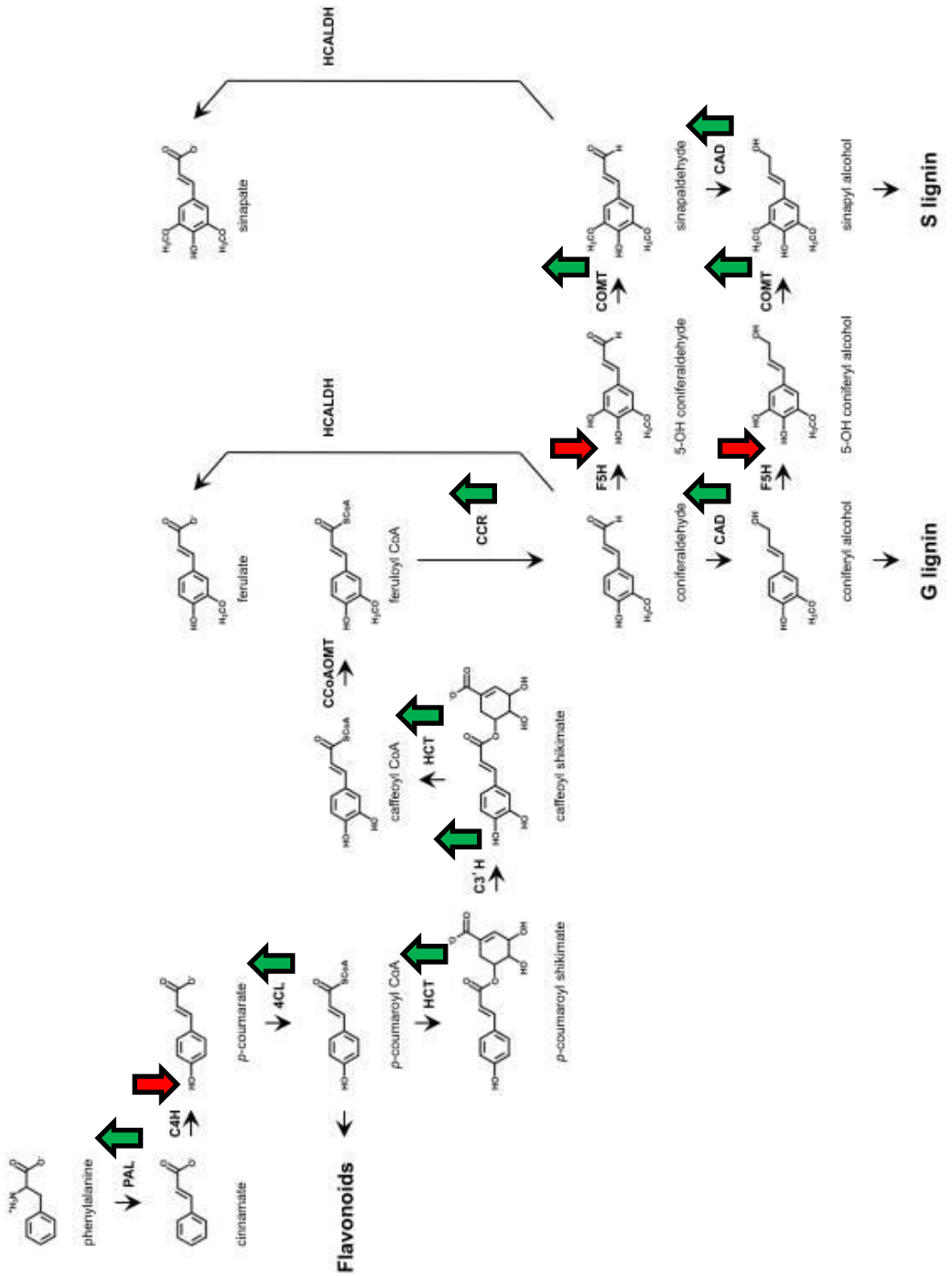


Figure 15: Phenylpropanoid biosynthetic pathway leading to flavonoids and lignin monomers showing differential expression in the roots of M2-1582, figure appropriated from Fraser & Chapple (2011). The pathways consist of phenylalanine ammonia-lyase (PAL), cinnamate 4-hydroxylase (C4H), 4-coumarate-CoA ligase (4CL), hydroxycinnamoyl transferase (HCT), p-coumaroyl shikimate 3'-hydroxylase (C3'H), caffeoyl-CoA O-methyltransferase (CCoAOMT), cinnamoyl-CoA reductase (CCR), caffeic acid 3-O-methyltransferase (COMT), ferulate 5-hydroxylase (F5H) and cinnamyl alcohol dehydrogenase (CAD).

4.1f Comparison of transcriptomes reveals expression of an alternate auxin influx transport protein in M2-1582

C. roseus var. Pasifica Peach and M2-1582 specific transcriptomes were generated to identify potential polymorphisms, EMS induced mutations and expressed sequences. The var. Pasifica Peach transcriptome is comprised of 145,601 sequences with an average length of 435 bps and a N50 of 940. The M2-1582 transcriptome comprised 82,888 sequences with an average length of 820 bps and a N50 of 1981. Identified potential auxin transporters and GH3.6 retrieved from the reference, var. Pasifica Peach reference and M2-1582 transcriptomes revealed differences only in the auxin influx transporter (Figure 16 & Appendix: Figures 17, 18 & 19). M2-1582 was shown to express an 85% indent homologue of the auxin influx transporter found in the reference transcriptomes (Figure 16). Sequences of both homologues are represented in the *C. roseus* draft genome (Appendix: Table 12).

```
Reference       MASEKVETMIAGNYIEMEREGEDGHSSAAAGGSSKTRLSKLFWHGGSVYDAWFSCASNQVAQ 62
PP_Ref          ------------------------------------------------------------ 0
M2-1582         MASEKVETMIAGNYIEMEREGEDGHSSAAAGGSSKTRLSKLFWHGGSVYDAWFSCASNQVAQ 62


Reference       VLLTLPYSFSQLGMLSGVIFQLFYGLMGSWTAYLICILYIEYRTRKEKENVSFKNHVIQWFE 124
PP_Ref          ------------------------------------------------------------ 0
M2-1582         VLLTLPYSFSQLGMLSGVIFQLFYGLMGSWTAYLICILYIEYRTRKEREKVDFRNHVIQWFE 124


Reference       VLDGLLGPQWKAVGLAFNCTFLLFGSVIQLIACASNIYYINDRLDKRTWTYIFGACCATTVF 186
PP_Ref          ----------------------------------SNIYYINDRLDKRTWTYIFGACCATTVF 28
M2-1582         VLDGLLGKHWRNIGLFFNCTFLLFGSVIQLIACASNIYYINDSLDKRSWTYIFGACCATTVF 186
                                                 ******** ****:**************


Reference       IPSFHNYRIWSFLGLGMTTYTAWYMTIAAFVHGQVDGVTHSGPTKLALYFTGATNILYTFGG 248
PP_Ref          IPSFHNYRIWSFLGLGMTTYTAWYMTIAAFVHGQVDGVTHSGPTKLALYFTGATNILYTFGG 90
M2-1582         IPSFHNYRIWSFLGLIMTTYTAWYLTIASLIHGQIEGVKHSGPTKIVLYFTGATNILYTFGG 248
                *************** ********:***:::***::**.******:.***************


Reference       HAVTVEIMHAMWKPQKFKYIYLYATFYVFTLTIPSASAVYWAFGDQLLDHANAFSLLPRTGF 310
PP_Ref          HAVTVEIMHAMWKPQKFKYIYLYATFYVFTLTIPSASAVYWAFGDQLLDHANAFSLLPRTGF 152
M2-1582         HAVTVEIMHAMWKPQKFKLIYLFATVYVLTLTLPSASAVYWAFGDLLLNHSNALSLLPRTGF 310
                ****************** ***:**.**:***:************ **:*:**:********


Reference       RDAAVILMLIHQFITFGFACTPLYFVWEKVIGVHDTSSICLRALARLPVVIPIWFLAIIFPF 372
PP_Ref          RDAAVILMLIHQFITFGFACTPLYFVWEKVIGVHDTSSICLRALARLPVVIPIWFLAIIFPF 214
M2-1582         RDTAVILMLIHQFITFGFACTPLYFVWEKLIGVHETKSLFKRALARLPVVIPIWFLAIIFPF 372
                **:****************************:****:*.*:  **********************


Reference       FGPINSAVGALLVSFTVYIIPALAHMLTYRKASARKNAAEKPPFFLPSWTAMYAVNIFVVAW 434
PP_Ref          FGPINSAVGALLVSFTVYIIPALAHMLTYRKASARKNAAEKPPFFLPSWTAMYAVNIFVVAW 276
M2-1582         FGPINSTVGSLLVSFTVYIIPALAHMITFASSSARENAVERPPKILGGWVGLYSMNVFVVVW 434
                ******:**:*****************:*: .:***:**.*:** :* .*..:*::*:***.*


Reference       VFVVGFGFGGWASMSNFIKQVDTFGLFAKCYQCKPPQHQHH 475
PP_Ref          VFVVGFGFGGWASMSNFIKQVDTFGLFAKCYQCKPPQHQHH 317
M2-1582         IFVIGFGFGGWASMVNFVRQINTFGLFTKCYQCPPKK---A 472
                :**:********** **::*::*****:***** * :
```

Figure 16: Multiple sequence alignment of auxin influx carrier protein homologues from the *C. roseus* reference transcriptome (Reference), *C. roseus* var. Pacifica Peach reference transcriptome (PP_Ref) and M2-1582 transcriptome (M2-1582).

4.2 Discussion

4.2a M2-1582 morphology is altered by exogenous auxin concentrations

Identification of M2-1582, a low iridoid and MIA EMS *C. roesus* mutant, provided the opportunity to study root to leaf iridoid transport (Kidd et al. *in press*). The unique morphologies of M2-1582 under different concentrations of auxin provided the opportunity to gain insight into plant development through a non-model organism. The typical phenotype of M2-1582 grown under 200ug/L IBA included early leaf abscission, root callus and slow growth. These phenotypes changed to produce a plant morphologically similar to WT when M2-1582 was grown in media containing 100 ug/L IBA. Auxin induces leaf abscission by forming a local auxin maximum at the abscission zone (Jin et al. 2015; Ma et al. 2015). Auxin also plays a role in lateral root initiation in the pericycle cells next to the basal end of the xylem (Gou et al. 2010; Khan & Stone 2007). Most of the M2-1582 grown under 0 ug/L IBA died, with only a single plant able to produce a stunted primary root with no visible lateral roots. This suggests M2-1582 has an issue with auxin transport, not accumulation through uncontrolled biosynthesis and turnover.

A single M2-1581 plant grown under 200 ug/L IBA displaying a relatively normal green root suggests the root morphology can be partially reverted by increased cytokinin levels, as indicated by the green chloroplast induction in roots (Kobayashi et al. 2012). This outlier seems to be reacting to the exogenous auxin by abusing the antagonistic relationship between auxin and cytokinin. While the roots had a distinguished primary root, the aerial part of the plant still displayed small, yellowing leaves periodically undergoing abscission.

4.2b Stem cross-sections reveal increased vascular proliferation in M2-1582

Cross-sections from *in vitro* WT *C. roseus* and M2-1582 were taken at points of physical necessity for the experiment and visualized together for direct relative comparison (Figure 7). The difference in locations on the stem make the vasculature to pith ratio differences less drastic than if they were taken at the same point. The decrease in pith to vasculature ratio from 3.85 in the WT to 1.45 in M2-1582 demonstrates the increased vasculature proliferation is the cause of the ridged stem (Appendix). The phenotype of increased secondary vascular deposition could be due to poor auxin drainage from the vascular cambium when M2-1582 is grown media containing 200 ug/L IBA (Jones et al. 2005).

4.2c Auxin related gene expression profiles of multiple organs in M2-1582 are altered

The preliminary high-throughput analysis using Panther and String visualized several crucial biological processes were over two-fold differentially regulated in the first leaf pair, second leaf pair and roots of M2-1582 (Figures 11, 12 & 13). Closer inspection of specific biosynthetic and regulatory networks related to auxin revealed complex organ expression differences that can be attributed to the unique morphologies observed.

Differential gene expression analysis revealed almost 3000 more than two-fold upregulated genes in the roots of M2-1582. Of those, both TAA and YUCCA were upregulated in the roots but downregulated in the leaves. Auxin biosynthetic genes would not typically represent a large portion of the total root transcriptome as root auxin biosynthesis is limited to a few specialized cells, like pericycle cells or the root cap (Olatunji et al. 2017; de Smet 2012). In the roots of M2-1582 expression levels of TAA are 7.61 FPKM, while transcript representation is below the 0.01 FPKM threshold in WT. TAA is the first dedicated step in IAA biosynthesis

from tryptophan in the IPA pathway, indicating an increase in representativeness of specialized root cells synthesizing IAA (Won et al. 2011). Transcript levels of tryptophan biosynthetic gene homologues were all upregulated in the roots of M2-1582, potentially to meet the increased demand from the IPA pathway. The directional proliferation of cells involved in primary and lateral root growth is controlled by the accumulation of local auxin maxima. The increased representation of cells synthesizing IAA could be due to poor regulation of local auxin maxima, potentially causing the callus-like root morphology of M2-1582 grown in 200 ug/L IBA. The almost full reversion of the root morphology in 100 ug/L IBA to WT and lethality of 0 ug/L IBA adds support to the hypothesis that this EMS mutation impacts auxin transport, and not uncontrolled auxin biosynthesis or turnover. The almost 22-fold increase in the IAA turnover protein GH3.6 in the roots of M2-1582 could be induced to regulate intracellular auxin levels. Surprisingly, the two auxin response protein homologues did not show a large increase in differential expression in the roots.

Both the PIN homologue and the auxin influx carrier protein were upregulated in the roots of M2-1582 and downregulated in both sampled leaves. The vacuole auxin transport protein WAT1 was increased in all sampled tissues of M2-1582, with the greatest increase by over 8-fold in roots. The increased storage of auxin in the vacuole has likely contributed to the increased secondary vasculature in the stem (Ranocha et al. 2013). Interestingly, the expression data suggests the role of WAT1 in secondary vascular proliferation may involve the root system. The reversion of M2-1582 root morphology to almost normal in M2-1582 suggests the transport issue is intercellular. The increase in WAT1 expression and deposition of IAA into the vacuole could have been an attempt to correct for poor intercellular transport.

Expression of an alternate homologue of the potential auxin influx transporter could be the cause of the M2-1582 mutant's sensitivity to exogenous auxin. The 85% indent homologue was not represented in either reference transcriptome generated but is found in the draft genome (Appendix: Table 12). It could be the remnant of a duplication event that is not expressed under normal conditions. The EMS mutation in M2-1582 could have knocked out the typically expressed auxin influx transporter, causing the induction of its dysfunctional homologue. It is also possible the EMS induced mutation could have induced the homologue to be preferentially expressed over the typical auxin influx transporter.

4.2d Gibberillin, cytokinin and jasmonate biosynthetic, regulatory and turnover genes are mostly upregulated in the roots of M2-1582

Expression of the cytokinin biosynthetic gene IPT homologue increased 80% in the roots of M2-1582 and the cytokinin turnover gene CKX homologue decreased over two-fold. The increase in cytokinin biosynthesis and reduction of turnover could be to combat high levels of auxin in roots due to poor intercellular transport. The partial correction of root morphology in the M2-1582 outlier grown on 200 ug/L IBA displaying green roots supports the connection (Kobayashi et al. 2012).

The several fold increase in all identified GA biosynthetic, turnover and regulatory genes in the roots of M2-1582 suggests GA levels could have increased to mediate auxin transport. GA deficient *Populus* lines displayed aerial dwarfism and increased root mass due to an increase in lateral root primordia (Gou et al. 2010). These *Populus* lines are deficient in their ability to increase PIN expression and promoting clustering in cell wall deposition mediated by GA. M2-1582 grown in 200 ug/L IBA is displaying similar morphological features to *pin1* mutants or

plants deficient in GA (Gou et al. 2010; Jones et al. 2005). Differential expression of GA biosynthetic, turnover and regulatory genes in the leaves of M2-1582 displayed mixed patterning, with the most noticeable being the over four-fold increase of the GA turnover gene GA2ox homologue in both sampled leaves.

The three ethylene biosynthetic gene ACS homologues and the ethylene response factor EIN3 were all two-fold downregulated or more in the roots of M2-1582. Two of the ACS homologues and EIN3 are increased over two-fold in the second leaf pair of M2-1582. These data suggest ethylene levels could be reduced in roots and increased in older leaves. The increase in older leaves could be explained by the routine leaf abscission of M2-1582 grown in 200 ug/L IBA. It could also be explained by issues with auxin transport, as ethylene biosynthetic genes and response factors were upregulated *P. tremula* after the application of a PIN inhibitor (Jin et al. 2015). Ethylene levels may be reduced in roots due to their antagonistic relationship with GA in mature plants (Weiss & Ori 2007). If GA biosynthesis was upregulated to accommodate for an auxin transport issue, ethylene biosynthesis could be downregulated as a consequence.

The jasmonate biosynthetic gene JAR1 homologue showed increased expression in all sampled tissues of M2-1582. The increase in expression of JAR1 is potentially related to the increase in expression of ORCA3, an iridioid biosynthetic gene regulator upregulated by MeJA. Expression levels of ORCA3 in WT roots is below the FPKM threshold, while expression levels in the roots of M2-1582 was 2.22 FPKM. MeJA is derived from the same common precursor used as substrate for JAR1, jasmonic acid. An increase in JAR1 and OCRA3 in the roots suggests there is an increase in jasmonates in the roots of M2-1582, potentially due to impaired auxin transport (Khan & Stone 2007).

4.2e Auxin regulation of lignin biosynthesis is altered in M2-1582

The aerial section of M2-1582 grown in media containing 200 ug/L IBA displayed two distinctive phenotypes related to the accumulation of local auxin maxima, induced by the same transcription factors, which are increased secondary vasculature and early leaf abscission. Both biological processes are regulated by KD1 and REV homologues. Three homologues of these two types of homeobox-leucine zipper proteins were identified to be upregulated in roots and downregulated in both sampled leaves. M2-1582 grown in 100 ug/L IBA did not display early leaf abscission, demonstrating one of these biological processes was corrected by reducing exogenous application of auxins to the roots. This raises questions about the diversity of roles roots can play in signaling and regulation of aerial organs.

Lignin biosynthetic genes showed the greatest differential regulation in the roots of M2-1582, with most having a 60% or more increase in transcript abundance. The exceptions were C4H, which was downregulated in all sampled organs, and F5H, which was downregulated in roots and upregulated in both leaves. The exogenous application of cinnamic acid and its derivatives to the roots of *Glycine max* upregulated lignin biosynthetic genes in roots, reduced root growth and increased lignin content. The increased lignin content disproportionately favored G monomers with H:G:S ratios changing from 10:55:35 to 14:72:14 (Lima et al. 2013). It is possible the increased lignin biosynthetic genes in M2-1582 is showing a similar preference for increased G unit lignin composition by downregulating F5H.

The EMS mutation on M2-1582 severely impacted hormone homeostasis, however the link to the low alkaloid and iridoid content is not clear. ORCA3 was upregulated in the roots of M2-1582, yet the ORCA3 upregulated iridoid biosynthetic gene G8O was downregulated 20-fold. Since BIS1 and BIS2 do not upregulate G8O, another factor is influencing transcript

abundance of G8O in M2-1582. The early iridoid biosynthetic genes are the most impacted with transcript representation of IS in the roots dropping from 18.95 FPKM in WT to below threshold in M2-1582. IS cyclizes the monoterpene 8-oxogenaniol, dedicating the terpene to a specific chemical backbone. The oxidation of the diol to the dialdehyde by 8-HGO is reversable. It is possible the iridoid pathway is downregulated because the labile terpene precursor is being used for another biosynthetic pathway, GA biosynthesis for example.

4.3 Methods and Materials

4.3a Plant materials

Seeds (WT and EMS mutagenized) of *Catharanthus roseus* (L.) G. don were donated by PanAmerican Seed Co. (panamseed.com). The methods used for mutagenesis and for seed production were described previously (Edge et al. 2018). Control (WT) and mutant M2 seeds (~4000) were cultivated in soil in a greenhouse (16/8 h light/dark) at 30 ˚C.

4.3b *In vitro* plant cultivation

C. roseus variety Pacifica Peach (WT) and EMS mutant (M2-1582) were grown in sterile magenta boxes containing plant regenerating media composed of autoclaved Woody plant medium salts (2.3g/L), indole butyric acid (0.2 mg/L), glycine (2.0 mg/L), myo-inositol (100 mg/L), nicotinic acid (0.5 mg/L), pyridoxine (0.5 mg/L), thiamine-HCl (0.1 mg/L) and sucrose (30 g/L), activated charcoal (2.5 g/L), gelrite (2.6 g/L) adjusted to pH 5.8 with ~4 drops of 1M KOH. Plants were propagated by taking cuttings and inserting them into new medium.

4.3c *In vitro* exogenous auxin gradients

WT and M2-1582 were grown in sterile magenta boxes containing plant regenerating media composed of autoclaved woody plant medium salts (2.3g/L), glycine (2.0 mg/L), myo-inositol (100 mg/L), nicotinic acid (0.5 mg/L), pyridoxine (0.5 mg/L), thiamine-HCl (0.1 mg/L) and sucrose (30 g/L), gelrite (2.6 g/L) adjusted to pH 5.8 with ~4 drops of 1M KOH. Four M2-1582 and WT plants were each grown in indole butyric acid concentrations of 0 mg/L, 0.1mg/L and 0.2 mg/L for 90 days. Activated charcoal was removed to allow for improved observation of root morphologies. Activated charcoal was primarily being used to make the alkaloids excreted by the roots inert (Thomas 2008), M2-1582 doesn't produce root alkaloids (unpublished data) and the WT displayed no obvious physical differences without.

4.3d RNA extraction and cDNA synthesis

Each leaf (leaf pair 1 or leaf pair 2) stem and root from *in vitro* cultivated M2-1582 and WT were pooled separately and ground in liquid nitrogen to a fine powder and immediately mixed with 0.8 mL of Trizol reagent (Life Technologies) to extract the RNA. 0.16 mL of chloroform was added, the aqueous phase was removed. After precipitation with 0.4 mL of isopropanol and 0.8 mL of ethanol, the RNA was suspended in 25 µl of DEPC H_2O. The RNA was subsequently treated with DNAase I to remove genomic DNA. DNAase I was in turn inactivated with 1 mM EDTA for 15 min at 75°C. Quality and quantity of RNA were determined visually by gel electrophoresis and nanophotometry. First-strand cDNA was synthesized using SuperScript II reverse transcriptase (Invitrogen) and oligo (dT) 12-18 primer (Invitrogen, Carlsbad, CA) using 1-4 µg total RNA according to manufacturer's protocol.

RNA extraction and cDNA synthesis destined for qPCR was conducted approximately a year before the same methodology was applied for Illumina sequencing. RNA samples of first and second leaf pairs and roots from WT *C. roseus ssp.* Pacific Peach and the M2-1582 mutant were sequenced on an Illumina NextSeq500 with mid output 150bp paired-end reads. A minimum of 26M clusters/sample was generated.

4.3e Carborundum-assisted isolation of epidermal enriched RNA from *C. roseus* var. Pacifica Peach

2.5g of fresh *C. roseus* var. Pacifica Peach leaves were added to a 50 mL falcon tube with 8 mL of PureLink Plant RNA Reagent (Invitrogen) and 2.5g of carborundum (SiC) (Fisher Scientific). The falcon tube was vortexed for 1 minute then left to settle for 5 minutes at room temperature. Leaves were removed with tweezers and the previous step repeated with an additional 2.5g of fresh leaves in the same falcon tube. The leaves were removed and the falcon tube centrifuges at 2600 g for 5 minutes at 4°C. The top phase was removed, and 100 µl of chloroform was added

per 500 µl of the remaining bottom phase. The falcon tube was inverted for 30 seconds and incubated at room temperature for 2 minutes. The falcon tube was centrifuged at 12,000 g for 15 minutes at 4°C. The colourless top phase was transferred to a fresh 50 mL falcon tube and 250 µl of isopropanol was added. The solution was incubated at room temperature for 10 minutes then centrifuged at 12,000 g for 10 minutes at 4°C. The supernatant was removed, and pellet washed with 70% ethanol by votexing. The solution was centrifuged at 7,500 g for 5 minutes at 4°C, supernatant removed, and RNA pellet left to dry to almost completeness. The RNA was resuspended in 40 µl of DEPC H_2O. The RNA was subsequently treated with DNAase I to remove genomic DNA. DNAase I was in turn inactivated with 1 mM EDTA for 15 min at 75°C. Quality and quantity of RNA were determined visually by gel electrophoresis and nanophotometry. First-strand cDNA was synthesized using SuperScript II reverse transcriptase (Invitrogen) and oligo (dT) 12-18 primer (Invitrogen) using 1-4 µg total RNA according to manufacturer's protocol. The sample was sequenced on an Illumina NextSeq500 with mid output 150bp paired-end reads. A minimum of 25M clusters/sample was generated.

4.3f Quantitative PCR to test for enrichment of carborundum-assisted RNA isolation

qRT-PCR was performed (CFX96™ Real-Time system, Bio-Rad, Hercules, CA and CFX96 Connect™ Real-Time system, Bio-Rad, Hercules, CA) using 5 µL iTaq™ Universal SYBR® Green Supermix (Bio-Rad, Hercules, CA) 0.25 µL primer and cDNA template at 5 ng total RNA per 10 µL reaction volume. cDNA samples from wild-type *C. roseus* var. Pacifica Peach isolated as part of the M2-1582 sequencing analyses was used as a control to test for enrichment.
The reaction conditions for qRT-PCR included 1 cycle of 95°C for 1 min and 40 cycles of 95°C for 15s and 58°C for 1 min. Critical Threshold (Ct) values were used to calculate the relative transcript abundance with the housekeeping gene 60S ribosome cDNA as the internal control.

Expression levels of each target gene were analyzed with the Bio-Rad CFX Manager Software (Bio-Rad) and normalized to the 60S ribosomal cDNA (Figure 20).

4.3g Assembly of reference transcriptome

All *C. roseus* pair-ended RNA-seq data sets in the European Nucleotide Archive (https://www.ebi.ac.uk/ena/about) available before June, 2016 were assembled together with the three WT *C. roseus* biological samples sequenced alongside the M2-1582 mutant samples using Trinity's *de novo* suite with the default settings (Haas et al. 2013). The resulting file had a total of 290,033,768 bps from 292,193 sequences with a length average: 992 bps; N50: 1453.

CD-HIT (Fu et al. 2012) was used to cluster nucleotide sequences in the combined transcriptome and resulted in a file with a total of 180,106,363 bps from 203,358 sequences with a length average: 885 bps; N50: 1240. SOAP *de novo*'s gap closer was applied to close sequence gaps and resulted in a file with a total of 178,354,257 bps from 200,255 sequences with a length average: 890 bps; N50: 1259.

86,956 mRNA entries from *C. roseus* were obtained from GenBank and assembled using CAP3 (Huang & Madan 1999). Creating a transcriptome file with a total of 107,471,342 bps from 86,956 sequences with a length average: 1235 bps; N50 1680.

CAP3 (Huang & Madan 1999) was used to assemble both the transcriptome files created from mRNA data and the combined RNA-seq data assembled with Trinity, then processed with CD-HIT (Fu et al. 2012) and SOAPdenovo2's gap closer (Luo et al. 2012).
Sequences in the resulting transcriptome shorter than 200 base pairs were removed resulting in a total of 182,562,426 bps from 205,237 sequences with a length average: 889 bps; N50: 1256. This file is the reference transcriptome used for all downstream processing and is hereby referred

to as the ‘reference transcriptome’. The refence transcriptome was indexed using STAR for downstream processing (Dobin & Gingeras 2015).

4.3h Annotation of reference transcriptome

The reference transcriptome was annotated using the trinotate suite (Grabherr et al. 2011). Trinotate requires the output of complementary and redundant programs to be loaded into the suite before processing and annotation results can be retrieved.

SwissProt (Bairoch 2000) and Pfam (Finn et al. 2014) databases were retrieved and converted into a protein Blast database using NCBI’s Blast+ (Camacho et al. 2009). TransDecoder (Haas et al. 2013) was used to find the longest open reading frames in the reference transcriptome with a minimum length of 50 amino acids and separately predict the likely coding regions. HMMscan (Finn et al. 2011) was used with the Pfam database on the TransDecoder predicted longest open reading frame amino acid sequences to predict likely protein sequences. BlastX (Camacho et al. 2009) was used to align the reference transcriptome against the created SwissProt and Pfam protein database using Blast+ to retrieve annotation results of the closest aligned sequence. BlastP was used to blast the reference transcriptome’s longest open reading frame amino acid sequences against the created SwissProt and Pfam protein database using Blast+ to retrieve annotation results of the closest aligned sequence. SignalP (Petersen et al. 2011) was used to predict signal peptide cleavage sites on the reference transcriptome’s longest open reading frame amino acid sequences. TmHMM (Krogh et al. 2001) was used to predict transmembrane domains in the reference transcriptome’s longest open reading frame amino acid sequences. All output files created from these analyses were loaded into the Trinotate.sqlite database and exported as the newly annotated reference transcriptome (Haas et al. 2013).

4.3i Assembly of *C. roseus* var. Pacifica Peach reference transcriptome

CAP3 (Huang & Madan 1999) was used to assemble all the wild-type *C. roseus* var. Pacifica Peach RNA sequencing data isolated as a control with the high ajmalicine mutant (Qu et al. 2018), high O-acetylstemmadenine mutant (M2-10698) and low-alkaloid mutant (M2-1582). The assembled data was processed with CD-HIT (Fu et al. 2012) and SOAPdenovo2's gap closer (Luo et al. 2012).

Sequences in the resulting transcriptome shorter than 100 base pairs were removed resulting in a total of 63,423,297 bps from 145,601 sequences with a length average: 435 bps; N50: 940.

4.3j Assembly of *C. roseus* M2-1582 transcriptome and multiple sequence alignment

CAP3 (Huang & Madan 1999) was used to assemble the M2-1582 *C. roseus* RNA sequencing data isolated from 3 different . The assembled data was processed with CD-HIT (Fu et al. 2012) and SOAPdenovo2's gap closer (Luo et al. 2012).

Sequences in the resulting transcriptome shorter than 100 base pairs were removed resulting in a total of 68,011,986 bps from 82,888 sequences with a length average: 820 bps; N50: 1981.

Identical translated sequences, or the closest homologues of select auxin transport genes from the reference transcriptome, *C. roseus* var. Pacifica Peach reference transcriptome and M2-1582 transcriptome were aligned using T-Coffee (Notredame et al. 2000).

4.3k Differential gene expression of multiple organs in M2-1582

The WT *C. roseus* and M2-1582 RNA-seq files were aligned to the indexed reference transcriptome using STAR (Dobin & Gingeras 2015) with MAPQ integer set to 0to255 for downstream processing. Cufflinks suite (Trapnell et al. 2012) was used to assemble the indexed

sequencing data to the indexed transcriptome. Cuffcompare was used to combine the assembled datasets for the differential gene expression analysis using Cuffdiff. The cut-off was set to FPKM<0.01 as the minimum quantified integer, anything below is set to 0. Genes outlined were manually screened and duplicate transcripts in the reference caused by native SNPs were combined. Secoirioid biosynthetic genes and regulators all aligned to their known sequences and were not given a table of closest homologues.

4.3l Exploratory mutation detection analysis of M2-1582

GATK is typically used to detect SNPs in RNA-seq datasets using a curated genome as a reference (McKenna et al. 2010). The tool was appropriated for use with a transcriptome reference to show the potential it has to aid in modeling EMS mutagenized plants, when curated genomes for these plant species become available.

GATK requires several pre-processing steps before the variant caller can be run. The indexed RNA-Seq files generated for differential gene expression analysis were run through picard tools (Ebbert et al. 2016) for AddOrReplaceReadGroups and MarkDuplicates to sort the data. The resulting files were further pre-processed by GATK to split N cigar reads. GATK uses a default map quality integer of 255, coinciding with the map quality integer set when indexing the RNA-seq datasets with STAR.

GATK (McKenna et al. 2010) was run with the older variant caller 'UnifiedGenotyper' as the new 'HaplotypeCaller' was unable to process the data using a transcriptome reference, regardless of the pre-processing steps applied to the datasets. Varients were called with interval padding set to 100, minimum confidence threshold for calling set to 20 and minimum base calling set to 1. Setting the minimum score to call the base as a variant higher than 1 resulted in 0 variants being called under all tested parameters. Variant caller files of the WT *C. roseus* and

M2-1582 mutant were combined based on each organ type (first leaf pair, second leaf pair and roots). Combining the mutant and WT variant calls allows the native SNP variants from the Pacific Peach cultivar to be separated. The cultivar specific SNP's are identified as variants of the reference transcriptome, which contains RNA-seq datasets from different *C. roseus* cultivars. The variant caller files were also filtered for EMS-specific mutations, which occur in a predictable guanine to adenine or cytosine to thymine substitution pattern.

GATK's variant annotator, SnpEff (Cingolani et al. 2012), was used to identify the impact of mutations on coding regions. Sequences flagged as 'LowQual' were removed; nonsense mutations and start codon substitutions were selected to represent 'high-impact' mutations that were later included in the String interactions analysis. String is a protein-protein interactions database that allows for the visualization of protein interaction networks and lists the varying strength of support for the interactions based on the source data.

It's important to note that the mutation detection data generated is not accurate. UnifiedGenotyper generated over 100-fold the expected number of mutations on the expressed genome that should be produced by EMS under standard methodology. High-fidelity PCR amplification and sanger sequencing confirmed the false positives. It's being used here to show how modeling of EMS-mutagenized plants can be improved with a curated genome.

4.3m Bioinformatic analysis of pathway alterations in M2-1582

Panther was used to identify pathways overrepresented in the root and combined first and second leaf pair differential gene expression analysis (Mi et al. 2017). Gene ontology terms for genes showing a two-fold difference in expression were used with an *A. thaliana* background and create the gene ontology annotated datasets. Pathways with statistically significant

overrepresentation ($P<0.05$) using Bonferroni correction were sorted and graphed with the gene ontology categories.

To show the protein-protein interactions of highly impacted genes in M2-1582, interactions maps were generated with String using an *A. thaliana* background (Szklarczyk et al. 2015). Protein sequences for genes showing a two-fold difference in expression in both leaf samples, or the root sample, or containing high-impact mutations were uploaded to String. Proteins with less than 65% identity to *A. thaliana* proteins were removed. The leaf and root interaction maps generated were manually curated to exclude any non-interacting proteins to the major cluster of interactions.

4.3n Compound microscopy

Cross sections of *in vitro* WT *C. roseus* was taken at the base of the stem due to poor lignification making razorblade cuttings at other points difficult. M2-1582 stem cross sections were taken between the first and second leaf pairs due to the intense lignification making razorblade cuttings at other points difficult. The cross sections were visualized together using a Zeiss phase contrast microscope. Relative measurements between the piths and vasculature were taken.

Future Directions

The unique morphologies displayed by M2-1582 grown *in vitro* with 200 ug/L IBA has provided an opportunity to gain insight into the plant development through a non-model organism. It has also provided an opportunity to look at the cross-regulation between primary and specialized metabolism in plants.

Discovery of M2-1582 expressing an 85% homologue to the auxin influx transporter represented in the reference transcriptome allows for a simple experiment to be designed for confirmation of the observed phenotype. Expression of both homologues in *Saccharomyces cerevisiae* and collection of auxin transport kinetics across the membrane would demonstrate the potential differences in transport capabilities.

Several factors would typically make M2-1582 a poor choice for studying primary metabolism. *C. roseus* is not a model plant and mandatory tissue culture propagation of the M2-1582 were the most significant. However, studying M2-1582 has provided insight into how roots can regulate aerial parts of the plant and produce a range of unique morphologies with relatively minor changes in concentrations of exogenous auxin. M2-1582 also brings into question the potential contribution of the roots in the production of lignin monomers.

The high ajmalicine *C. roseus* EMS mutant M2-0754 will provide a valuable source of ajmalicine if there is ever demand for production.

Curation of the assembled *C. roseus* genome would allow for the sequencing data to be used for a more in-depth analysis of M2-1582 and M2-0754, performing mutation detection analysis, for example.

Summary and Conclusions

The benefits of using EMS mutagenesis and specialized cell enriched transcriptomes for elucidation of MIA biosynthetic genes in *C. roseus* separately are regularly echoed. The continued heavy reliance on manual screening of genes selected from specialized cell enriched transcriptomes to discover the basis of the unique MIA profiles in EMS mutants is rarely discussed. The partial bioinformatic and morphological characterization of the EMS mutant M2-1582 demonstrates the numerous challenges of discovering specific causes of EMS induced phenotypes in non-model plants.

After the discovery of GS by manual gene selection from the epidermal enriched transcriptome and screening by VIGS, differential gene expression analysis revealed GS was almost 7-fold downregulated in M2-0754 compared to WT.

The most severe impact of EMS mutagenesis on M2-1582 was likely related to auxin transport. The change in morphologies when M2-1582 was grown in media containing different concentrations of IBA, the differences in vascular deposition revealed in the cross sections and the differential expression of crucial biological processes in different organs all support auxin transport as having the largest impact on the unique morphologies observed. The unique morphology of M2-1582 and identification of differentially expressed pathways would make a good starting point for further investigation of how auxin transport impacts plant development in *A. thaliana*.

References

Adamowski, M., & J. Friml. (2015). PIN-Dependent Auxin Transport: Action, Regulation, and Evolution. *The Plant Cell Online*.

Aerts, R. J., & V. De Luca. (1992). Phytochrome is Involved in the Light-Regulation of Vindoline Biosynthesis in Catharanthus. *Plant physiology*. **100**:1029-1032.

Aharoni, A., & G. Galili. (2011). Metabolic Engineering of the Plant Primary-Secondary Metabolism Interface. *Current Opinion in Biotechnology*. **22**:239-244.

Aloni, R.. (2010). The Induction of Vascular Tissues by Auxin. *Plant Hormones*. 485-518.

Arabidopsis Genome Initiative. (2000). Analysis of the Genome Sequence of the Flowering Plant *Arabidopsis thaliana*. *Nature*. **408**:769-815.

Asada, K., Salim, V., Masada-Atsumi, S., Edmunds, E., Nagatoshi, M., Terasaka, K., Mizukami, H. & V. De Luca. (2013). A 7-Deoxyloganetic Acid Glucosyltransferase Contributes a Key Step in Secologanin Biosynthesis in Madagascar Periwinkle. *The Plant cell.* **25**:4123-4134.

Bairoch, A. (2000). The SWISS-PROT Protein Sequence Database and its Supplement TrEMBL in 2000. *Nucleic Acids Research*. **28**:45-48.

Baker, D.A.. (2000). Long-distance Vascular Transport of Endogenous Hormones in Plants and their Role in Source:Sink Regulation. *Israel Journal of Plant Sciences*. **3**:199-203.

Beninger, C.W., Cloutier, R.R., Monteiro, M.A., & B. Grodzinski. (2007). The Distribution of Two Major Iridoids in Different Organs of *Antirrhinum majus* L. at Selected Stages of Development. *Journal of Chemical Ecology*. **33**:731-747.

Blom, T.J.M., Sierra, M., van Vliet, T.B., Franke-van Dijk, M.E.I., de Koning, P., van Iren, F., Verpoorte, R. & K.R. Libbenga. (1991). Uptake and Accumulation of Ajmalicine into Isolated Vacuoles of Cultured Cells of *Catharanthus roseus* (L.) G. Don. and Its Conversion into Serpentine. *Planta*. **183**:170-177.

Blomster, T. Salojärvi, J., Sipari, N., Brosché, M., Ahlfors, R., Keinänen, M., Overmyer, K. & J. Kangasjärvi. (2011). Apoplastic Reactive Oxygen Species Transiently Decrease Auxin Signaling and Cause Stress-Induced Morphogenic Response in Arabidopsis. *Plant Physiology*. **157**:1866-1883.

Brown, S., Clastre, M., Courdavault, V. & S.E. O'Connor. (2015). De Novo Production of the Plant-Derived Alkaloid Strictosidine in Yeast. *Proceedings of the National Academy of Sciences of the United States of America.* **112**:3205-3210.

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K. & T.L. Madden. (2009). BLAST+: Architecture and Applications. *BMC Bioinformatics*. **10**:421.

Caputi, L., Franke, J., Farrow, S.C., Chung, K., Payne, R.M.E., Nguyen, T.D., Dang, T.T., Soares Teto Carqueijeiro, I., Koudounas, K., Dugé de Bernonville, T., Ameyaw, B., Jones D.M., Vieira, I.J.C., Courdavault, V. & S.E. O'Connor. (2018). Missing Enzymes in the Biosynthesis of the Anticancer Drug Vinblastine in Madagascar Periwinkle. *Science*. **360**:1235-1239.

Carqueijeiro, I., Noronha, H., Duarte, P., Gerós, H. & M. Sottomayor. (2013). Vacuolar Transport of the Medicinal Alkaloids from *Catharanthus roseus* is Mediated by a Proton-Driven Antiport. *Plant Physiology*. **162**:1486-1496.

Carqueijeiro, I., Brown, S., Chung, K., Dang, T.T., Walia, M., Besseau, S., Dugé de Bernonville, T., Oudin, A., Lanoue, A., Billet, K., Munsch, T., Koudounas, K., Melin, C., Godon, C., Razafimandimby, B., de Craene, J.O., Glévarec, G., Marc, J., Giglioli-Guivarc'h, N., Clastre, M., St-Pierre, B., Papon, N., Andrade, R.B., O'Connor, S.E. & V. Courdavault. (2018). Two Tabersonine 6,7-Epoxidases Initiate Lochnericine-Derived Alkaloid Biosynthesis in *Catharanthus roseus*. *Plant Physiology*. **177**:1473-1486.

Chaudhary, S., Sharma, V., Prasad, M., Bhatia, S., Tripathi, B.N., Yadav, G. & S. Kumar. (2011). Characterization and Genetic Linkage Mapping of the Horticulturally Important Mutation Leafless Inflorescence (Lli) in Periwinkle *Catharanthus roseus*. *Scientia Horticulturae*. **129**:142-153.

Chen, J., Xie, J., Duan, Y., Hu, H., Hu, Y. & W. Li. (2016). Genome-Wide Identification and Expression Profiling Reveal Tissue-Specific Expression and Differentially-Regulated Genes Involved in Gibberellin Metabolism between Williams Banana and its Dwarf Mutant. *BMC plant biology.* **16**:123.

Chen, Q., Dai, X., De-Paoli, H., Cheng, Y., Takebayashi, Y., Kasahara, H., Kamiya, Y. & Y. Zhao. (2014). Auxin Overproduction in Shoots Cannot Rescue Auxin Deficiencies in Arabidopsis Roots. *Plant and Cell Physiology*. **55**:1072-1079.

Cingolani, P., Platts, A., Wang le, L., Coon, M., Nguyen, T., Wang, L., Land, S.J., Lu, X. & D.M. Ruden. (2012). A Program for Annotating and Predicting the Effects of Single Nucleotide Polymorphisms, SnpEff. *Fly.* **6**:80-92.

Claros, M.G., Bautista, R., Guerrero-Fernández, D., Benzerki, H., Seoane, P. & N. Fernández-Pozo. (2014). Why Assembling Plant Genome Sequences is so Challenging. *The Role of Bioinformatics in Agriculture*. **1**:439-459.

Collu, G., Unver, N., Peltenburg-Looman, A.M., van der Heijden, R., Verpoorte, R. & J. Memelink. (2001). Geraniol 10-Hydroxylase, a Cytochrome P450 Enzyme Involved in Terpenoid Indole Alkaloid Biosynthesis. *FEBS letters.* **508**:215-220.

Cooper, J., Till B.J., Laport, R.G., Darlow, M.C., Kleffner, J.M., Jamai, A., El-Mellouki, T., Liu S., Ritchie, R., Nielsen, N., Bilyeu, K.D., Meksem, K., Comai, L. & S. Henikoff (2008). TILLING to Detect Induced Mutations in Soybean. *BMC Plant Biology* **8**:9.

Courdavault, V., Papon, N., Clastre, M., Giglioli-Guivarc'h, N., St-Pierre, B. & V. Burlat. (2014). A Look inside an Alkaloid Multisite Plant: The Catharanthus Logistics. *Current Opinion in Plant Biology*. **19**:43-50.

De Luca, V., Salim, V., Thamm, A., Masada, S.A. & F. Yu. (2014). Making Iridoids/Secoiridoids and Monoterpenoid Indole Alkaloids: Progress on Pathway Elucidation. *Current Opinion in Plant Biology.* **19**:35-42.

De Luca, V., Salim, V., Atsumi, S.A & F. Yu. (2012). Mining the Biodiversity of Plants: A Revolution in the Making. *Science*. **19**:35-42.

Denancé, N., Ranocha, P., Oria, N., Barlet, X., Rivière, M.P., Yadeta, K.A., Hoffmann, L., Perreau, F., Clément, G., Maia-Grondard, A., van den Berg, G.C., Savelli, B., Fournier, S., Aubert, Y., Pelletier, S., Thomma, B.P., Molina, A., Jouanin, L., Marco, Y. & D. Goffner. (2013). Arabidopsis Wat1 (Walls Are Thin1)-Mediated Resistance to the Bacterial Vascular Pathogen, *Ralstonia solanacearum*, Is Accompanied by Cross-Regulation of Salicylic Acid and Tryptophan Metabolism. *Plant Journal*. **73**:225-239.

Desgagné-Penix, I., Khan, M.F., Schriemer, D.C., Cram, D., Nowak, J. & P.J. Facchini. (2010). Integration of Deep Transcriptome and Proteome Analyses Reveals the Components of Alkaloid Metabolism in Opium Poppy Cell Cultures. *BMC plant biology*. **10**:252.

Deus-Neumann, B. & M.H. Zenk. (1984). A Highly Selective Alkaloid Uptake System in Vacuoles of Higher Plants. *Planta*. **162**: 250-260.

Dobin, A. & T.R. Gingeras. (2015). Mapping RNA-Seq Reads with STAR. *Current Protocols in Bioinformatics*. **51**:1-11.

Dugé de Bernonville, T., Carqueijeiro, I., Lanoue, A., Lafontaine, F., Sánchez Bel, P., Liesecke, F., Musset, K., Oudin, A., Glévarec, G., Pichon, O., Besseau, S., Clastre, M., St-Pierre, B., Flors, V., Maury, S., Huguet, E., O'Connor, S.E. & V. Courdavault. (2017). Folivory Elicits a Strong Defense Reaction in *Catharanthus roseus*: Metabolomic and Transcriptomic Analyses Reveal Distinct Local and Systemic Responses. *Scientific Reports*. **7**:40453.

Ebbert, M., Wadsworth, M.E., Staley, L.A., Hoyt, K.L., Pickett, B., Miller, J., Duce, J.; Alzheimer's Disease Neuroimaging Initiative, Kauwe, J.S. & P.G. Ridge. (2016). Evaluating the Necessity of PCR Duplicate Removal from Next-Generation Sequencing Data and a Comparison of Approaches. *BMC Bioinformatics*. **17**:239.

Facchini, P.J., Bohlmann, J., Covello, P.S., De Luca, V., Mahadevan, R., Page, J.E., Ro, D.K., Sensen, C.W., Storms, R. & V.J. Martin. (2012). Synthetic Biosystems for the Production of High-Value Plant Metabolites. *Trends in Biotechnology*. **30**:127-131.

Feuillet, C., Leach, J.E., Rogers, J., Schnable, P.S. & K. Eversole. (2011). Crop Genome Sequencing: Lessons and Rationales. *Trends in Plant Science*. **16**:77-88.

Finn, R.D., Bateman, A., Clements, J., Coggill, P., Eberhardt, R.Y., Eddy, S.R., Heger, A., Hetherington, K., Holm, L., Mistry, J., Sonnhammer, E.L., Tate, J. & M. Punt. (2014). Pfam: The Protein Families Database. *Nucleic Acids Research*. **42**.

Finn, R.D., Clements, J. & S.R. Eddy. (2011). HMMER Web Server: Interactive Sequence Similarity Searching. *Nucleic Acids Research*. **39**.

Fraser, C.M. & C. Chapple. (2011). The Phenylpropanoid Pathway in Arabidopsis. *The Arabidopsis Book*. **9**:e0152.

Fu, L., Nju, B., Zhu, Z., Wu, S. & W. Li. (2012). CD-HIT: Accelerated for Clustering the next-Generation Sequencing Data. *Bioinformatics*. **28**:3150-3152.

Gälweiler, L., Guan, C., Müller, A., Wisman, E., Mendgen, K., Yephremov, A. & K. Palme. (1998). Regulation of Polar Auxin Transport by AtPIN1 in Arabidopsis Vascular Tissue. *Science*. **282**:2226-2230.

Garcia, S., Leitch, I.J., Anadon-Rosell, A., Canela, M.Á., Gálvez, F., Garnatje, T., Gras, A., Hidalgo, O., Johnston, E., Mas de Xaxars, G., Pellicer, J., Siljak-Yakovlev, S., Vallès, J., Vitales, D. & M.D. Bennett. (2014). Recent Updates and Developments to Plant Genome Size Databases. *Nucleic Acids Research*. **42**.

Geerlings, A. Ibañez, M.M., Memelink, J., van Der Heijden, R. & R. Verpoorte. (2000). Molecular Cloning and Analysis of Strictosidine β-D-Glucosidase, an Enzyme in Terpenoid Indole Alkaloid Biosynthesis in *Catharanthus roseus*. *Journal of Biological Chemistry*. **275**:3051-3056.

Geu-Flores, F., Sherden, N.H., Courdavault, V., Burlat, V., Glenn, W.S., Wu, C., Nims E., Cui, Y. & S.E. O'Connor. (2012). An Alternative Route to Cyclic Terpenes by Reductive Cyclization in Iridoid Biosynthesis. *Nature*. **492**:138-142.

Ghaffar, S.H. & M. Fan. (2013). Structural Analysis for Lignin Characteristics in Biomass Straw. *Biomass and Bioenergy*. **57**:264-279.

Giddings, L.A., Liscombe, D.K., Hamilton, J.P., Childs, K.L., Della Penna, D., Buell, C.R. & S.E. O'Connor. (2011). A Stereoselective Hydroxylation Step of Alkaloid Biosynthesis by a Unique Cytochrome P450 in *Catharanthus roseus*. *Journal of Biological Chemistry*. **286**:16751-16757.

Gou, J., Strauss, S.H., Tsai, C.J., Fang, K., Chen, Y., Jiang, X. & V.B. Busov. (2010). Gibberellins Regulate Lateral Root Formation in *Populus* through Interactions with Auxin and Other Hormones. *The Plant Cell*. **22**:623-639.

Gowan, E.B., Lewis, A. & R. Turgeon. (1995). Phloem Transport of Antirrhinoside, an Iridoid Glycoside, in *Asarina scandens* (Scrophulariaceae). *Journal of Chemical Ecology.* **21**: 1781-1788.

Grabherr, M.G., Haas, B.J., Yassour, M., Levin, J.Z., Thompson, D.A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., Chen, Z., Mauceli, E., Hacohen, N., Gnirke, A., Rhind, N., di Palma, F., Birren, B.W., Nusbaum, C., Lindblad-Toh, K., Friedman, N. & A. Regev. (2011). Full-Length Transcriptome Assembly from RNA-Seq Data without a Reference Genome. *Nature Biotechnology*. **29**:644-652.

Guirimand, G., Guihur, A., Poutrain, P., Héricourt, F., Mahroug, S., St-Pierre, B., Burlat, V. & V. Courdavault. (2011). Spatial Organization of the Vindoline Biosynthetic Pathway in *Catharanthus roseus*. *Journal of Plant Physiology*. **168**:549-557.

Guirimand, Grégory, Guihur, A., Ginis, O., Poutrain, P., Héricourt, F., Oudin, A., Lanoue, A., St-Pierre, B., Burlat, V. & V. Courdavault. (2011). The Subcellular Organization of Strictosidine Biosynthesis in *Catharanthus roseus* Epidermis Highlights Several Trans-Tonoplast Translocations of Intermediate Metabolites. *FEBS Journal.* **278**:749-763.

Haas, B.J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P.D., Bowden, J., Couger, M.B., Eccles, D., Li B., Lieber, M., MacManes, M.D., Ott, M., Orvis, J., Pochet, N., Strozzi, F., Weeks, N., Westerman, R., William, T., Dewey, C.N., Henschel, R., LeDuc, R.D., Friedman, N. & A. Regev. (2013). De Novo Transcript Sequence Reconstruction from RNA-Seq Using the Trinity Platform for Reference Generation and Analysis. *Nature Protocols*. **8**:1494-1512.

van Der Heijden, R., Jacobs, D.I., Snoeijer, W., Hallard, D. & R. Verpoorte. (2004). The Catharanthus Alkaloids: Pharmacognosy and Biotechnology. *Current medicinal chemistry.* **11**:607-628.

Hentrich M., Sánchez-Parra, B., Pérez Alonso, M.M., Carrasco Loba, V., Carrillo, L., Vicente-Carbajosa, J., Medina, J. & S. Pollmann. (2013). YUCCA8 and YUCCA9 Overexpression Reveals a Link between Auxin Signaling and Lignification through the Induction of Ethylene Biosynthesis. *Plant signaling & behavior*. **8**:e26363.

Hemscheidt, T. & M.H. Zenk. (1985). Partial Purification and Characterization of a NADPH Dependent Tetrahydroalstonine Synthase from *Catharanthus roseus* Cell Suspension Cultures. *Plant Cell Reports*. **4**:216-219.

Huang, X. & A. Madan. (1999). CAP3: A DNA Sequence Assembly Program. *Genome Research*. **9**:868-877.

Irmler, S. Schröder, G., St-Pierre, B., Crouch, N.P., Hotze, M., Schmidt, J., Strack, D., Matern, U. & J. Schröder. (2000). Indole Alkaloid Biosynthesis in *Catharanthus roseus*: New Enzyme Activities and Identification of Cytochrome P450 CYP72A1 as Secologanin Synthase. *Plant Journal.* **24**:797-804.

Jin, X., Zimmermann, J., Polle, A. & U. Fischer. (2015). Auxin is a Long-Range Signal That Acts Independently of Ethylene Signaling on Leaf Abscission in *Populus*. *Frontiers in Plant Science*. **6**:634.

Johnson, M., Zaretskaya, I., Raytselis, Y., Merezhuk, Y., McGinnis, S. & T.L. Madden. (2008). NCBI BLAST: A Better Web Interface. *Nucleic acids research*. **36**.

Johnsson, C., Jin, X., Xue, W., Dubreuil, C., Lezhneva, L. & U. Fischer. (2018). The Plant Hormone Auxin Directs Timing of Xylem Development by Inhibition of Secondary Cell Wall Deposition through Repression of Secondary Wall NAC-Domain Transcription Factors. *Physiologia Plantarum*.

Jones, S.E., Demeo, J.S., Davies, N.W., Noonan, S.E. & J.J. Ross. (2005). Stems of the Arabidopsis Pin1-1 Mutant Are Not Deficient in Free Indole-3-Acetic Acid. *Planta*. **222**:530-534.

Kellner, F. Geu-Flores, F., Sherden, N.H., Brown, S., Foureau, E., Courdavault, V. & S.E. O'Connor. (2015). Discovery of a P450-Catalyzed Step in Vindoline Biosynthesis: A Link between the Aspidosperma and Eburnamine Alkaloids. *Chemical Communications*. **51**:7626-7628.

Kellner, F., Kim, J., Clavijo, B.J., Hamilton, J.P., Childs, K.L., Vaillancourt, B., Cepela, J., Habermann, M., Steuernagel, B., Clissold, L., McLay, K., Buell, C.R. & S.E. O'Conner. (2015). Genome-guided investigation of plant natural product biosynthesis. *The Plant Journal*. **82**:680-692.

Khan, S. & J.M. Stone. (2007). *Arabidopsis thaliana* GH3.9 in Auxin and Jasmonate Cross Talk. *Plant Signaling and Behavior*. **2**:483-485.

Kidd, T., Easson, M.A.E., Qu, Y., Jones, G. & V. De Luca. (*in press*) Inter-organ transport of secologanin allows assembly of monoterpenoid indole alkaloids in a *Catharanthus roseus* mutant down-regulated in irioid biosynthesis. *Phytochemistry*.

Kidd, T. (2016). A *Catharanthus roseus* mutant with trace levels of secologanin and monoterpenoid indole alkaloids does not express BIS1/BIS2 transcription factors and fails to activate iridoid biosynthesis. Unpublished masters thesis. Brock University. St. Catharines, Ontario.

Kimura, S. & N. Sinha. (2008). Grafting Tomato Plants. *Cold Spring Harbor Protocols.* 3.

Kobayashi, K., Baba, S., Obayashi, T., Sato, M., Toyooka, K., Keränen, M., Aro, E.M., Fukaki, H., Ohta, H., Sugimoto, K. & T. Masuda. (2012). Regulation of Root Greening by Light and Auxin/Cytokinin Signaling in Arabidopsis. *The Plant Cell*. **24**:1081-1095.

Krithika, R., Lal Srivastava, P., Rani, B., Kolet, S.P., Chopade, M., Soniya, M. & H.V. Thulasiram. (2015). Characterization of 10-Hydroxygeraniol Dehydrogenase from *Catharanthus roseus* Reveals Cascaded Enzymatic Activity in Iridoid Biosynthesis. *Scientific reports.* **5**:8258.

Krogh, A., Larsson, B., von Heijne, G. & E.L.L. Sonnhammer. (2001). Predicting Transmembrane Protein Topology with a Hidden Markov Model: Application to Complete Genomes. *Journal of Molecular Biology*. **305**:567-580.

Kulkarni, R. N., Baskaran, K., Chandrashekara, R.S. & S. Kumar. (1999). Inheritance of Morphological Traits of Periwinkle Mutants with Modified Contents and Yields of Leaf and Root Alkaloids. *Plant Breeding*. **118**:71-74.

Kulkarni, R.N. & S. Baskaran. (2014). Increasing Total Leaf Alkaloid Concentrations in Periwinkle (*Catharanthus roseus*) by Combining the Macro-Mutant Traits of Two Induced Leaf Mutants ('Necrotic Leaf' and 'Nerium Leaf')." *Journal of Horticultural Science and Biotechnology*. **89**:513-518.

Kutchan, T.M., Hampp, N., Lottspeich, F., Beyreuther, K. & M.H. Zenk. (1988). The cDNA Clone for Strictosidine Synthase from *Rauvolfia serpentina* DNA Sequence Determination and Expression in *Escherichia coli*. *FEBS Letters*. **237**:40-44.

Laflamme, P., St-Pierre, B. & V. De Luca. (2001). Molecular and Biochemical Analysis of a Madagascar Periwinkle Root-Specific Minovincinine-19-Hydroxy-O-Acetyltransferase. *Plant Physiology*. **125**:189-198.

Lehmann, T., Hoffmann, M., Hentrich, M. & S. Pollmann. (2010). Indole-3-Acetamide-Dependent Auxin Biosynthesis: A Widely Distributed Way of Indole-3-Acetic Acid Production? *European Journal of Cell Biology*. **89**:895-905.

Leple, J.C., Dauwe, R., Morreel, K., Storme, V., Lapierre, C., Pollet, B., Naumann, A., Kang, K.Y., Kim, H., Ruel, K., Lefèbvre, A., Joseleau, J.P., Grima-Pettenati, J., De Rycke, R., Andersson-Gunnerås, S., Erban, A., Fehrle, I., Petit-Conil, M., Kopka, J., Polle A., Messens, E., Sundberg, B., Mansfield, S.D., Ralph, J., Pilate, G. & W. Boerjan. (2007). Downregulation of Cinnamoyl-Coenzyme A Reductase in *Poplar*: Multiple-Level Phenotyping Reveals Effects on Cell Wall Polymer Metabolism and Structure. *The Plant Cell Online*. **19**:3669-3691.

Li, G., Zhu, C., Gan, L., Ng, D. & K. Xia. (2014). GA3 Enhances Root Responsiveness to Exogenous IAA by Modulating Auxin Transport and Signalling in Arabidopsis. *Plant Cell Reports*. **34**:483-494.

Lima, R.B., Salvador, V.H., dos Santos, W.D., Bubna, G.A., Finger-Teixeira, A., Soares, A.R., Marchiosi, R., Ferrarese, Mde L. & O. Ferrarese-Filho. (2013). Enhanced Lignin Monomer Production Caused by Cinnamic Acid and Its Hydroxylated Derivatives Inhibits Soybean Root Growth. *PLoS ONE*. **8**:e80542.

Liscombe, D.K., Usera, A.R. & S.E. O'Connor. (2010). Homolog of Tocopherol C Methyltransferases Catalyzes N Methylation in Anticancer Alkaloid Biosynthesis. *Proceedings of the National Academy of Sciences*. **107**:18793-18798.

Luijendijk, T.J.C., Stevens, L.H. & R. Verpoorte. (1998). Purification and Characterisation of Strictosidine β-D-Glucosidase from *Catharanthus roseus* Cell Suspension Cultures. *Plant Physiology and Biochemistry*. **36**:419-425.

Luo, R., Liu, B., Xie, Y., Li, Z., Huang, W., Yuan, J., He, G., Chen, Y., Pan, Q., Liu, Y., Tang, J., Wu, G., Zhang, H., Shi, Y., Liu, Y., Yu, C., Wang, B., Lu, Y., Han, C., Cheung, D.W., Yiu, S.M., Peng, S., Xiaoqian, Z., Liu, G., Liao, X., Li, Y., Yang, H., Wang, J., Lam, T.W. & J. Wang. (2012). SOAPdenovo2: An Empirically Improved Memory-Efficient Short-Read de Novo Assembler. *GigaScience*.

Ma, C., Meir, S., Xiao, L., Tong, J., Liu, Q., Reid, M.S. & C.Z. Jiang. (2015). A KNOTTED1-LIKE HOMEOBOX Protein Regulates Abscission in Tomato by Modulating the Auxin Pathway. *Plant Physiology*. **167**:844-853.

Mahroug, S., Courdavault, V., Thiersault, M., St-Pierre, B. & V. Burlat. (2006). Epidermis is a Pivotal Site of at Least Four Secondary Metabolic Pathways in *Catharanthus roseus* Aerial Organs. *Planta*.

Marhavý, P., Duclercq, J., Weller, B., Feraru, E., Bielach, A., Offringa, R., Friml, J., Schwechheimer, C., Murphy, A. & E. Benková. (2014). Cytokinin Controls Polarity of PIN1-Dependent Auxin Transport during Lateral Root Organogenesis. *Current Biology*. **24**:1031-1037.

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M. & M.A. DePristo. (2010). The Genome Analysis Toolkit: A MapReduce Framework for Analyzing next-Generation DNA Sequencing Data. *Genome Research.* **20**:1297-1303.

Mi, H., Huang, X., Muruganujan, A., Tang, H., Mills, C., Kang, D. & P.D. Thomas. (2017). PANTHER Version 11: Expanded Annotation Data from Gene Ontology and Reactome Pathways, and Data Analysis Tool Enhancements. *Nucleic Acids Research.* **45**:183-189.

Miettinen, K., Dong, L., Navrot, N., Schneider, T., Burlat, V., Pollier, J., Woittiez, L., van der Krol, S., Lugan, R., Ilc, T., Verpoorte, R., Oksman-Caldentey, K.M., Martinoia, E., Bouwmeester, H., Goossens, A., Memelink, J. & D. Werck-Reichhart. (2014). The Seco-Iridoid Pathway from *Catharanthus roseus*. *Nature Communications*. **5**:3606.

van Moerkercke, A., Steensma, P., Schweizer, F., Pollier, J., Gariboldi, I., Payne, R., Vanden, Bossche, R., Miettinen, K., Espoz, J., Purnama, P.C., Kellner, F., Seppänen-Laakso, T., O'Connor, S.E., Rischer, H., Memelink, J. & A. Goossens. (2015). The BHLH Transcription Factor BIS1 Controls the Iridoid Branch of the Monoterpenoid Indole Alkaloid Pathway in *Catharanthus roseus*. *Proceedings of the National Academy of Sciences of the United States of America.* **112**:8130-8135.

van Moerkercke, A., Steensma, P., Gariboldi, I., Espoz J., Purnama, P.C., Schweizer, F., Miettinen, K., Vanden Bossche, R., De Clercq, R., Memelink, J. & A. Goossens. (2016). The Basic Helix-Loop-Helix Transcription Factor BIS2 is Essential for Monoterpenoid Indole Alkaloid Production in the Medicinal Plant *Catharanthus roseus*. *Plant Journal.* **88**:3-12.

Müller, A., Guan, C., Gälweiler, L., Tänzler, P., Huijser, P., Marchant, A., Parry, G., Bennett, M., Wisman, E. & K. Palme. (1998). AtPIN2 Defines a Locus of Arabidopsis for Root Gravitropism Control. *EMBO Journal*. **17**:6903-6911.

Munkert, J., Pollier, J., Miettinen, K., Van Moerkercke, A., Payne, R., Müller-Uri, F., Burlat, V., O'Connor, S.E., Memelink, J., Kreis, W. & A. Goossens. (2015). Iridoid Synthase Activity is Common among the Plant Progesterone 5β-Reductase Family. *Molecular Plant* **8**:136-152.

Murashige, T. & F. Skoog. (1962). A Revised Medium for Rapid Growth and Bio Assays with Tobacco Tissue Cultures. *Physiologia Plantarum* **15**:473-497.

Murata, J., Roepke, J., Gordon, H. & V. De Luca. (2008). The Leaf Epidermome of *Catharanthus roseus* Reveals its Biochemical Specialization. *The Plant cell* **20**:524-542.

Notredame, C., Higgins D.G. & J. Heringa. (2000). T-Coffee: A novel method for fast and accurate multiple sequence alignment. *Journal of Molecular Biology*. **302**:205-217.

Olatunji, D., Geelen, D. & I. Verstraeten. (2017). Control of Endogenous Auxin Levels in Plant Root Development. *International Journal of Molecular Sciences*. **18**:2587.

Osnato, M., Castillejo, C., Matías-Hernández, L. & S. Pelaz. (2012). TEMPRANILLO Genes Link Photoperiod and Gibberellin Pathways to Control Flowering in Arabidopsis. *Nature Communications*. **3**:808.

Parr, A.J., Peerless, A.C., Hamill, J.D., Walton, N.J., Robins, R.J. & M.J. Rhodes. (1988). Alkaloid Production by Transformed Root Cultures of *Catharanthus roseus*. *Plant Cell Reports*. **7**:309-312.

Peebles, C.A.M., Hughes, E.H. Shanks, J.V. & K. Yiu San. (2009). Transcriptional Response of the Terpenoid Indole Alkaloid Pathway to the Overexpression of ORCA3 along with Jasmonic Acid Elicitation of *Catharanthus roseus* Hairy Roots over Time. *Metabolic Engineering*. **11**:76-86.

Petersen, T.N., Brunak, S., von Heijne, G. & H. Nielsen. (2011). SignalP 4.0: Discriminating Signal Peptides from Transmembrane Regions. *Nature Methods*. **8**:785-786.

Porth, I., Klápste, J., McKown, A.D., La Mantia, J., Hamelin, R.C., Skyba, O., Unda, F., Friedmann, M.C., Cronk, Q.C., Ehlting, J., Guy, R.D., Mansfield, S.D., El-Kassaby, Y.A. & C.J. Douglas. (2014). Extensive Functional Pleiotropy of REVOLUTA Substantiated through Forward Genetics. *Plant Physiology*. **164**:548-554.

Qu, Y., Easson, M.L., Froese, J., Simionescu, R., Hudlicky, T. & V. De Luca. (2015). Completion of the Seven-Step Pathway from Tabersonine to the Anticancer Drug Precursor Vindoline and Its Assembly in Yeast. *Proceedings of the National Academy of Sciences*. **112**:6224-6229.

Qu, Y., Thamm, A.M.K., Czerwinski, M., Masada, S., Kim, K.H., Jones, G., Liang, P. & V. De Luca. (2018). Geissoschizine Synthase Controls Flux in the Formation of Monoterpenoid Indole Alkaloids in a *Catharanthus roseus* Mutant. *Planta*. **247**:625-634.

Qu, Y., Easson, M.E.A., Simionescu, R., Hajicek, J., Thamm, A.M.K., Salim, V. & V. De Luca. (2018). Solution of the Multistep Pathway for Assembly of Corynanthean, Strychnos, Iboga, and Aspidosperma Monoterpenoid Indole Alkaloids from 19 *E* -Geissoschizine. *Proceedings of the National Academy of Sciences*. **115**:3180-3185.

Qu, Y., Safonova, O. & V. De Luca. (2018). Completion of the Canonical Pathway for Assembly of Anticancer Drugs Vincristine/Vinblastine in *Catharanthus roseus*. *Plant Journal*.

Quittenden, L.J., Davies, N.W., Smith, J.A., Molesworth, P.P., Tivendale, N.D. & J.J. Ross. (2009). Auxin Biosynthesis in Pea: Characterization of the Tryptamine Pathway. *Plant Physiology*. **151**:11130-1138.

Radwanski, E.R. & R.L Last. (1995). Tryptophan Biosynthesis and Metabolism: Biochemical and Molecular Genetics. *The Plant Cell Online*. **7**:921-934.

Ranocha, P., Dima, O., Nagy, R., Felten, J., Corratgé-Faillie, C., Novák, O., Morreel, K., Lacombe, B., Martinez, Y., Pfrunder, S., Jin, X., Renou, J.P., Thibaud, J.B., Ljung, K., Fischer, U., Martinoia, E., Boerjan, W. & D. Goffner. (2013). Arabidopsis WAT1 is a Vacuolar Auxin Transport Facilitator Required for Auxin Homoeostasis. *Nature Communications*. **4**:2625.

Rodriguez, S., Compagnon, V., Crouch, N.P., St-Pierre, B. & V. De Luca. (2003). Jasmonate-Induced Epoxidation of Tabersonine by a Cytochrome P-450 in Hairy Root Cultures of *Catharanthus roseus*. *Phytochemistry*. **64**:401-409.

Roepke, J. Salim, V., Wu, M., Thamm, A.M., Murata, J., Ploss, K., Boland, W. & V. De Luca. (2010). Vinca Drug Components Accumulate Exclusively in Leaf Exudates of Madagascar Periwinkle. *Proceedings of the National Academy of Sciences*. **107**:15287-15292.

Salim, V., Wiens, B., Masada-Atsumi, S., Yu, F. & V. De Luca. (2014). 7-Deoxyloganetic Acid Synthase Catalyzes a Key 3 Step Oxidation to Form 7-Deoxyloganetic Acid in *Catharanthus roseus* Iridoid Biosynthesis. *Phytochemistry* **101**:23-31.

Salim, V., Yu, F., Altarejos, J. & V. De Luca. (2013). Virus-Induced Gene Silencing Identifies *Catharanthus roseus* 7-Deoxyloganic Acid-7-Hydroxylase, a Step in Iridoid and Monoterpene Indole Alkaloid Biosynthesis. *Plant Journal*. **76**:754-765.

Sanz, L., Dewitte, W., Forzani, C., Patell, F., Nieuwland, J., Wen, B., Quelhas, P., De Jager, S., Titmus, C., Campilho, A., Ren, H., Estelle, M., Wang, H. & J.A. Murray. (2011). The Arabidopsis D-Type Cyclin CYCD2;1 and the Inhibitor ICK2/KRP2 Modulate Auxin-Induced Lateral Root Formation. *The Plant Cell*. **23**:641-660.

Simkin, A.J., Miettinen, K., Claudel, P., Burlat, V., Guirimand, G., Courdavault, V., Papon, N., Meyer, S., Godet, S., St-Pierre, B., Giglioli-Guivarc'h, N., Fischer, M.J., Memelink, J. & M. Clastre. (2013). Characterization of the Plastidial Geraniol Synthase from Madagascar Periwinkle Which Initiates the Monoterpenoid Branch of the Alkaloid Pathway in Internal Phloem Associated Parenchyma. *Phytochemistry*. **85**:36-43.

de Smet, I.. (2012). Lateral Root Initiation: One Step at a Time. *New Phytologist*. **193**:867-873.

Sorefan, K., Girin, T., Liljegren, S.J., Ljung, K., Robles, P., Galván-Ampudia, C.S., Offringa, R., Friml, J., Yanofsky, M.F. & L. Østergaard. (2009). A Regulated Auxin Minimum is Required for Seed Dispersal in Arabidopsis. *Nature*. **459**:583-586.

Stavrinides, A., Tatsis, E.C., Foureau, E., Caputi, L., Kellner, F., Courdavault, V. & S.E. O'Connor. (2015). Unlocking the Diversity of Alkaloids in *Catharanthus roseus*: Nuclear Localization Suggests Metabolic Channeling in Secondary Metabolism. *Chemistry and Biology*. **22**:336-341.

Stavrinides, A., Tatsis, E.C., Caputi, L., Foureau, E., Stevenson, C.E., Lawson, D.M., Courdavault, V. & S.E. O'Connor. (2016). Structural Investigation of Heteroyohimbine Alkaloid Synthesis Reveals Active Site Elements That Control Stereoselectivity. *Nature Communications*. **7**:12116.

Stepanova, A.N., Yun, J., Robles, L.M., Novak, O., He, W., Guo, H., Ljung, K. & J.M. Alonso. (2011). The Arabidopsis YUCCA1 Flavin Monooxygenase Functions in the Indole-3-Pyruvic Acid Branch of Auxin Biosynthesis. *The Plant Cell*. **23**:3961-3973.

Stout, J.M., Boubakir, Z., Ambrose, S.J., Purves, R.W. & J.E. Page. (2012). The Hexanoyl-CoA Precursor for Cannabinoid Biosynthesis is Formed by an Acyl-Activating Enzyme in *Cannabis sativa* Trichomes. *Plant Journal*. **71**:353-365.

Sugawara, S., Hishiyama, S., Jikumaru, Y., Hanada, A., Nishimura, T., Koshiba, T., Zhao, Y., Kamiya, Y. & H. Kasahara. (2009). Biochemical Analyses of Indole-3-Acetaldoxime-Dependent Auxin Biosynthesis in Arabidopsis. *Proceedings of the National Academy of Sciences*. **106**:5430-5435.

Suza, W.P. & P.E. Staswick. (2008). The Role of JAR1 in Jasmonoyl-l-Isoleucine Production during Arabidopsis Wound Response. *Planta*. **227**:1221-1232.

Szabó, L.F.. (2008). Rigorous Biogenetic Network for a Group of Indole Alkaloids Derived from Strictosidine. *Molecules*. **13**:1875-1896.

Szklarczyk, D., Franceschini, A., Wyder, S., Forslund, K., Heller, D., Huerta-Cepas, J., Simonovic, M., Roth, A., Santos, A., Tsafou, K.P., Kuhn, M., Bork, P., Jensen, L.J. & C. von Mering. (2015). STRING V10: Protein-Protein Interaction Networks, Integrated over the Tree of Life. *Nucleic Acids Research*. **43**:447-452.

Tatsis, E.C., Dugé de Bernonville, T., Franke, J., Dang, T.T., Oudin, A., Lanoue, A., Lafontaine, F., Stavrinides, A.K., Clastre, M., Courdavault, V. & S.E. O'Connor. (2017). A Three Enzyme System to Generate the Strychnos Alkaloid Scaffold from a Central Biosynthetic Intermediate. *Nature Communications*. **8**:316.

Thamm, A.M.K., Qu, Y. & V. De Luca. (2016). Discovery and Metabolic Engineering of Iridoid/Secoiridoid and Monoterpenoid Indole Alkaloid Biosynthesis. *Phytochemistry Reviews*. **15**:339-361.

Thomas, D.T.. (2008). The Role of Activated Charcoal in Plant Tissue Culture. *Biotechnology Advances*. **26**:618-631.

Trapnell, C., Roberts, A., Goff, L., Pertea, G., Kim, D., Kelley, D.R., Pimentel, H., Salzberg, S.L., Rinn, J.L. & L. Pachter. (2012). Differential Gene and Transcript Expression Analysis of RNA-Seq Experiments with TopHat and Cufflinks. *Nature Protocols*. **7**:562-578.

Treangen, T.J. & S.L. Salzberg. (2012). Repetitive DNA and Next-Generation Sequencing: Computational Challenges and Solutions. *Nature reviews. Genetics.* **13**:36-46.

Treimer, J.F. & M.H. Zenk. (1979). Purification and Properties of Strictosidine Synthase, the Key Enzyme in Indole Alkaloid Formation. *European Journal of Biochemistry.* **101**:225-233.

Ulmasov, T. Murfett, J., Hagen, G. & T.J. Guilfoyle. (1997). Aux/IAA Proteins Repress Expression of Reporter Genes Containing Natural and Highly Active Synthetic Auxin Response Elements. *The Plant Cell Online*. **9**:1963-1971.

Voitsekhovskaja, O.V., Koroleva, O.A., Batashev, D.R., Knop, C., Tomos, A.D., Gamalei, Y.V., Heldt, H.W. & G. Lohaus. (2005). Phloem Loading in Two Scrophulariaceae Species. What Can Drive Symplastic Flow via Plasmodesmata? *Plant Physiology*. **140**:383-395.

Wang, D., Pajerowska-Mukhtar, K., Culler, A.H. & X. Dong. (2007). Salicylic Acid Inhibits Pathogen Growth in Plants through Repression of the Auxin Signaling Pathway. *Current Biology*. **17**:1784-1790.

Wang, K.L.C., Li, H. & J.R. Ecker. (2002). Ethylene Biosynthesis and Signaling Networks. *The Plant Cell*. **14**:131-151.

Weiss, D., & N. Ori. (2007). Mechanisms of Cross Talk between Gibberellin and Other Hormones. *Plant Physiology*. **144**:2024.

Willige, B.C., Isono, E., Richter, R., Zourelidou, M. & C. Schwechheimer. (2011). Gibberellin Regulates PIN-FORMED Abundance and Is Required for Auxin Transport–Dependent Growth and Development in *Arabidopsis Thaliana*. *The Plant Cell*. **23**:2184-2195.

Won, C., Xiangling, S., Mashiguchi, K., Zheng, Z., Dia, X., Cheng, Y., Kasahara, H., Kamiya, Y., Chory, J. & Y. Zhao. (2011). Conversion of Tryptophan to Indole-3-Acetic Acid by TRYPTOPHAN AMINOTRANSFERASES OF ARABIDOPSIS and YUCCAs in Arabidopsis. *Proceedings of the National Academy of Sciences*. **108**:18518-18523.

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., MacNevin, G., Masada-Atsumi, S., Reed, D.W., Stout, J.M., Zerbe, P., Zhang, Y., Bohlmann, J., Covello, P.S., De Luca, V., Page, J.E., Ro, D.K., Martin, V.J., Facchini, P.J. & C.W. Sensen. (2013). Transcriptome Analysis Based on Next-Generation Sequencing of Non-Model Plants Producing Specialized Metabolites of Biotechnological Interest. *Journal of Biotechnology*.**166**:122-134.

Yamaguchi, S. & Y. Kamiya. (2000). Gibberellin Biosynthesis: Its Regulation by Endogenous and Environmental Signals. *Plant and Cell Physiology*. **41**:251-257.

Yu, F. & V. De Luca. (2013). ATP-Binding Cassette Transporter Controls Leaf Surface Secretion of Anticancer Drug Components in *Catharanthus roseus*. *Proceedings of the National Academy of Sciences*. **110**:15830-15835.

Zhang, H., Hedhili, S., Montiel, G., Zhang, Y., Chatel, G., Pré, M., Gantet, P. & J. Memelink, (2011). The Basic Helix-Loop-Helix Transcription Factor CrMYC2 Controls the Jasmonate-Responsive Expression of the ORCA Genes That Regulate Alkaloid Biosynthesis in *Catharanthus roseus*. *Plant Journal.* **67**:61-71.

Zhang, J., Elo, A. & Y. Helariutta. (2011). Arabidopsis as a Model for Wood Formation. *Current Opinion in Biotechnology*. **22**:293-299.

Zhang, Z., Li, Q., Li, Z., Staswick, P.E., Wang, M., Zhu, Y. & Z. He. (2007). Dual Regulation Role of GH3.5 in Salicylic Acid and Auxin Signaling during Arabidopsis-Pseudomonas Syringae Interaction. *Plant Physiology*. **145**:450-464.

Appendix

Measurements and Calculations:
Mutant Pith: 2.90 cm
Wild-type Pith: 5.00 cm
Mutant Vasculature: 2.00 cm
Wild-type Vasculature: 1.30 cm
WT Pith/Vasculature: 3.85
MT Pith/Vasculature: 1.45
Difference: 2.66x

```
Reference       MITLSDFYHVMTAVVPLYVAMILAYGSVRWWKIFTPDQCSGINRFVALFAVPLLSFHFIASN 62
PP_Ref          MITLSDFYHVMTAVVPLYVAMILAYGSVRWWKIFTPDQCSGINRFVALFAVPLLSFHFIASN 62
M2-1582         MITLSDFYHVMTAVVPLYVAMILAYGSVRWWKIFTPDQCSGINRFVALFAVPLLSFHFIASN 62
                ************************************************************

Reference       NPYAMNLRFIAADTLQKIIVLTVLFVWSKVSKRGCLEWTITLFSLSTLPNTLVMGIPLLKGM 124
PP_Ref          NPYAMNLRFIAADTLQKIIVLTVLFVWSKVSKRGCLEWTITLFSLSTLPNTLVMGIPLLKGM 124
M2-1582         NPYAMNLRFIAADTLQKIIVLTVLFVWSKVSKRGCLEWTITLFSLSTLPNTLVMGIPLLKGM 124
                ************************************************************

Reference       YGAFSGSLMVQIVVLQCIIWYTLMLFMFEFRGARLLISEQFPDTAGSIVSIHVDSDVVSLDG 186
PP_Ref          YGAFSGSLMVQIVVLQCIIWYTLMLFMFEFRGARLLISEQFPDTAGSIVSIHVDSDVVSLDG 186
M2-1582         YGAFSGSLMVQIVVLQCIIWYTLMLFMFEFRGARLLISEQFPDTAGSIVSIHVDSDVVSLDG 186
                ************************************************************

Reference       RQPLETEAEIKEDGKLHVTVRKSNASRSDIFSRRSQGLSSTTPRPSNLTNAEIYSLQSSRNP 248
PP_Ref          RQPLETEAEIKEDGKLHVTVRKSNASRSDIFSRRSQGLSSTTPRPSNLTNAEIYSLQSSRNP 248
M2-1582         RQPLETEAEIKEDGKLHVTVRKSNASRSDIFSRRSQGLSSTTPRPSNLTNAEIYSLQSSRNP 248
                ************************************************************

Reference       TPRGSSFNHTDFYSMVAAGRNSNFGANDVYGLSASRGPTPRPSNYEEDNAKPRFHYPAGGPG 310
PP_Ref          TPRGSSFNHTDFYSMVAAGRNSNFGANDVYGLSASRGPTPRPSNYEEDNAKPRFHYPAGGPG 310
M2-1582         TPRGSSFNHTDFYSMVAAGRNSNFGANDVYGLSASRGPTPRPSNYEEDNAKPRFHYPAGGPG 310
                ************************************************************

Reference       NNNTQYPAPNPGMFSPTGSKAMGTNPNKKPNGQHKEEGAKDLHMFVWSSSASPVSDVFGGGH 372
PP_Ref          NNNTQYPAPNPGMFSPTGSKAMGTNPNKKPNGQHKEEGAKDLHMFVWSSSASPVSDVFGGGH 372
M2-1582         NNNTQYPAPNPGMFSPTGSKAMGTNPNKKPNGQHKEEGAKDLHMFVWSSSASPVSDVFGGGH 372
                ************************************************************

Reference       DYGALDQPAKEVRVAVSPGKVDTHRDHNQEGYMEREDFTFGNKDMNNHEGGDKVGDNKVKVM 434
PP_Ref          DYGALDQPAKEVRVAVSPGKVDTHRDHNQEGYMEREDFTFGNKDMNNHEGGDKVGDNKVKVM 434
M2-1582         DYGALDQPAKEVRVAVSPGKVDTHRDHNQEGYMEREDFTFGNKDMNNHEGGDKVGDNKVKVM 434
                ************************************************************

Reference       PPTSVMTRLILIMVWRKLIRNPNTYSSLFGIIWSLISFRWNVQMPAIIAQSISILSDAGLGM 496
PP_Ref          PPTSVMTRLILIMVWRKLIRNPNTYSSLFGIIWSLISFRWNVQMPAIIAQSISILSDAGLGM 496
M2-1582         PPTSVMTRLILIMVWRKLIRNPNTYSSLFGIIWSLISFRWNVQMPAIIAQSISILSDAGLGM 496
                ************************************************************

Reference       AMFSLGLFMALQPRIIACGNSIAAFAMAVRFLTGPAVMAAASIAVGLRGVLLHVAIVQAALP 558
PP_Ref          AMFSLGLFMALQPRIIACGNSIAAFAMAVRFLTGPAVMAAASIAVGLRGVLLHVAIVQAALP 558
M2-1582         AMFSLGLFMALQPRIIACGNSIAAFAMAVRFLTGPAVMAAASIAVGLRGVLLHVAIVQAALP 558
                ************************************************************

Reference       QGIVPFVFAKEYNVHPDILSTAVIFGMLIALPITLVYYILVGL 601
PP_Ref          QGIVPFVFAKEYNVHPDILSTAVIFGMLIALPITLVYYILVGL 601
M2-1582         QGIVPFVFAKEYNVHPDILSTAVIFGMLIALPITLVYYILVGL 601
                *********************************************
```

Figure 17: Multiple sequence alignment of PIN homologues from the *C. roseus* reference transcriptome (Reference), *C. roseus* var. Pacifica Peach reference transcriptome (PP_Ref) and M2-1582 transcriptome (M2-1582).

```
Reference       MADTSSSSVVGIKRMCSIPEKFQLHLAMLLLQFGYAGFHVVSRAALNMGISKIVFPVYRNIL 62
PP_Ref          MADTSSSSVVGIKRMCSIPEKFQLHLAMLLLQFGYAGFHVVSRAALNMGISKIVFPVYRNIL 62
M2-1582         MADTSSSSVVGIKRMCSIPEKFQLHLAMLLLQFGYAGFHVVSRAALNMGISKIVFPVYRNIL 62
                ************************************************************

Reference       ALILLLPFAYFLEKKERPPLNWSFMLQFFLLAIVGITANQGFYLLGLENTSPTFASAIQNSV 124
PP_Ref          ALILLLPFAYFLEKKERPPLNWSFMLQFFLLAIVGITANQGFYLLGLENTSPTFASAIQNSV 124
M2-1582         ALILLLPFAYFLEKKERPPLNWSFMLQFFLLAIVGITANQGFYLLGLENTSPTFASAIQNSV 124
                ************************************************************

Reference       PAITFLMAAILRIEQVRLDRKDGIGKMVGTLLCVAGASVITLYKGPIIYSPAPPLQTTTSAA 186
PP_Ref          PAITFLMAAILRIEQVRLDRKDGIGKMVGTLLCVAGASVITLYKGPIIYSPAPPLQTTTSAA 186
M2-1582         PAITFLMAAILRIEQVRLDRKDGIGKMVGTLLCVAGASVITLYKGPIIYSPAPPLQTTTSAA 186
                ************************************************************

Reference       SSAPMLLNLGDANGKNWTLGCIYLIGHCLSWSGWLVLQKPVLKNYPARLSFTSYQCFFGVIQ 248
PP_Ref          SSAPMLLNLGDANGKNWTLGCIYLIGHCLSWSGWLVLQKPVLKNYPARLSFTSYQCFFGVIQ 248
M2-1582         SSAPMLLNLGDANGKNWTLGCIYLIGHCLSWSGWLVLQKPVLKNYPARLSFTSYQCFFGVIQ 248
                ************************************************************

Reference       FLIIALFLERDAEAWLIHSGGELFSVFYAGVVASGIAFAVQVWCIDRGGPVFVAVYQPVQTL 310
PP_Ref          FLIIALFLERDAEAWLIHSGGELFSVFYAGVVASGIAFAVQVWCIDRGGPVFVAVYQPVQTL 310
M2-1582         FLIIALFLERDAEAWLIHSGGELFSVFYAGVVASGIAFAVQVWCIDRGGPVFVAVYQPVQTL 310
                ************************************************************

Reference       VVAIMASIALGEQFYLGGIIGAVLIISGLYLVLWGKNEEQKFAMLQKAAIIQPPADHGRATT 372
PP_Ref          VVAIMASIALGEQFYLGGIIGAVLIISGLYLVLWGKNEEQKFAMLQKAAIIQPPADHGRATT 372
M2-1582         VVAIMASIALGEQFYLGGIIGAVLIISGLYLVLWGKNEEQKFAMLQKAAIIQPPADHGRATT 372
                ************************************************************

Reference       HIKSSSLAQPLLTQSTENV 391
PP_Ref          HIKSSSLAQPLLTQSTENV 391
M2-1582         HIKSSSLAQPLLTQSTENV 391
                ******************
```

Figure 18: Multiple sequence alignment of WAT1 homologues from the *C. roseus* reference transcriptome (Reference), *C. roseus* var. Pacifica Peach reference transcriptome (PP_Ref) and M2-1582 transcriptome (M2-1582).

```
Refernce        MPEAPNEAAGSNYSLAEKNKKALQFIEDVTTKADEVQQRVLSEILSKNADVEYLKRHGLNGH 62
PP_Ref          MPEAPNEAAGSNYSLAEKNKKALQFIEDVTTKADEVQQRVLSEILSKNADVEYLKRHGLNGH 62
M2-1582         MPEAPNEAAGSNYSLAEKNKKALQFIEDVTTKADEVQQRVLSEILSKNADVEYLKRHGLNGH 62
                ************************************************************

Refernce        TDRETFKKIMPVITYEDLQPDINRIANGDKSPIICSQPISEFLTSSGTSGGERKLMPTIEEE 124
PP_Ref          TDRETFKKIMPVITYEDLQPDINRIANGDKSPIICSQPISEFLTSSGTSGGERKLMPTIEEE 124
M2-1582         TDRETFKKIMPVITYEDLQPDINRIANGDKSPIICSQPISEFLTSSGTSGGERKLMPTIEEE 124
                ************************************************************

Refernce        LGRRSLLYSLLMPVMNQSVPGLDKGKGMYFLFIKSEAKTPGGLVARPVLTSYYKSSHFKERP 186
PP_Ref          LGRRSLLYSLLMPVMNQSVPGLDKGKGMYFLFIKSEAKTPGGLVARPVLTSYYKSSHFKERP 186
M2-1582         LGRRSLLYSLLMPVMNQSVPGLDKGKGMYFLFIKSEAKTPGGLVARPVLTSYYKSSHFKERP 186
                ************************************************************

Refernce        FDPYTNYTSPNETILCPDSYQSMYSQMLCGLCLNKEVLRVGAVFASGFIRAIRFLEKHWSLL 248
PP_Ref          FDPYTNYTSPNETILCPDSYQSMYSQMLCGLCLNKEVLRVGAVFASGFIRAIRFLEKHWSLL 248
M2-1582         FDPYTNYTSPNETILCPDSYQSMYSQMLCGLCLNKEVLRVGAVFASGFIRAIRFLEKHWSLL 248
                ************************************************************

Refernce        CNDIRTGTLNPEITDQSVRDAVMKILKPDLQLAEFIEKECSRESWQGIITRLWPNTKYIDVI 310
PP_Ref          CNDIRTGTLNPEITDQSVRDAVMKILKPDLQLAEFIEKECSRESWQGIITRLWPNTKYIDVI 310
M2-1582         CNDIRTGTLNPEITDQSVRDAVMKILKPDLQLAEFIEKECSRESWQGIITRLWPNTKYIDVI 310
                ************************************************************

Refernce        VTGTMSQYIPTLDYYSNGLPLVCTMYASSECYFGVNLNPLCKPSEVAYTLIPTMAYFEFLPV 372
PP_Ref          VTGTMSQYIPTLDYYSNGLPLVCTMYASSECYFGVNLNPLCKPSEVAYTLIPTMAYFEFLPV 372
M2-1582         VTGTMSQYIPTLDYYSNGLPLVCTMYASSECYFGVNLNPLCKPSEVAYTLIPTMAYFEFLPV 372
                ************************************************************

Refernce        HRNNGVNNSISMPKSLNEKEQQELVDLADVKLGQEYELVVTTYAGLYRYRVGDVLRVAGFKN 434
PP_Ref          HRNNGVNNSISMPKSLNEKEQQELVDLADVKLGQEYELVVTTYAGLYRYRVGDVLRVAGFKN 434
M2-1582         HRNNGVNNSISMPKSLNEKEQQELVDLADVKLGQEYELVVTTYAGLYRYRVGDVLRVAGFKN 434
                ************************************************************

Refernce        NAPQFNFICRKNVVLSIDSDKTDEVELQNAVKNAVTHLLPFDAHLTEYTSYADTTTNPGHYV 496
PP_Ref          NAPQFNFICRKNVVLSIDSDKTDEVELQNAVKNAVTHLLPFDAHLTEYTSYADTTTNPGHYV 496
M2-1582         NAPQFNFICRKNVVLSIDSDKTDEVELQNAVKNAVTHLLPFDAHLTEYTSYADTTTNPGHYV 496
                ************************************************************

Refernce        LFWELSLNGSNPPIPPSVFEDCCLSIEESLNSVYRQGRVSDKSIGPLEIKIVETGTFDKLMD 558
PP_Ref          LFWELSLNGSNPPIPPSVFEDCCLSIEESLNSVYRQGRVSDKSIGPLEIKIVETGTFDKLMD 558
M2-1582         LFWELSLNGSNPPIPPSVFEDCCLSIEESLNSVYRQGRVSDKSIGPLEIKIVETGTFDKLMD 558
                ************************************************************

Refernce        YAISLGASINQYKTPRCVKFPPIVELLNSRVVSSYLSPKCPKWAPGHKQWNTINL 613
PP_Ref          YAISLGASINQYKTPRCVKFPPIVELLNSRVVSSYLSPKCPKWAPGHKQWNTINL 613
M2-1582         YAISLGASINQYKTPRCVKFPPIVELLNSRVVSSYLSPKCPKWAPGHKQWNTINL 613
                *********************************************************
```

Figure 19: Multiple sequence alignment of GH3.6 homologues from the *C. roseus* reference transcriptome (Reference), *C. roseus* var. Pacifica Peach reference transcriptome (PP_Ref) and M2-1582 transcriptome (M2-1582).

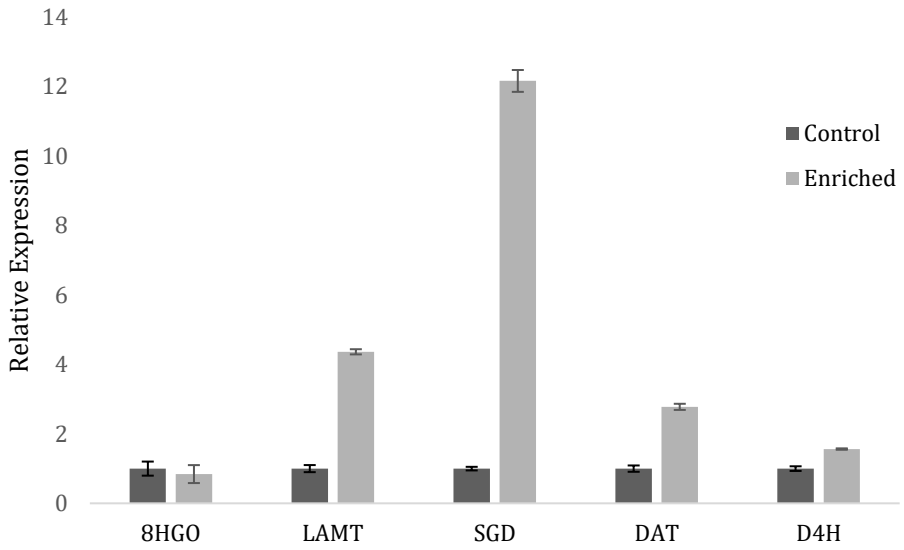


Figure 20: Relative expression of MIA biosynthetic genes showing enrichment of epidermal, idioblast and laticifer transcripts in carborundum-assisted RNA extraction (Enriched). 8HGO is preferentially expressed in IPAP cells, LAMT and SGD in epidermal cells, DAT and D4H in idioblasts and laticifers.

Protein	FLP-WT FPKM	FLP-MT FPKM	SLP-WT FPKM	SLP-MT FPKM	Root-WT FPKM	Root-MT FPKM
Geraniol 8-oxidase	295.46	15.80	148.30	2.43	83.50	4.41
8-hydroxygeraniol dehydrogenase	111.37	3.48	34.36	1.39	44.84	18.29
Iridoid synthase	44.73	5.06	19.34	0.04	18.95	0.00
7-deoxyloganic acid synthase	116.34	67.08	100.48	52.35	32.37	51.70
7-deoxyloganetic acid glucosyltransferase	21.35	24.66	14.45	21.01	12.91	35.34
Loganic acid O-methyltransferase	43.71	7.17	51.12	23.89	8.30	270.80
Secologanin synthase	306.38	116.26	277.57	545.18	127.25	1117.17
Tryptophan decarboxylase	14.07	7.58	9.85	7.67	11.28	60.69
Strictosidine synthase	49.76	46.38	46.23	53.12	88.03	151.96
Strictosidine β-glucosidase	24.39	29.46	17.45	34.21	11.37	226.95
10-hydroxygeraniol dehydrogenase	84.62	358.18	40.48	151.30	995.14	775.23
BIS1	21.36	7.13	2.29	0.72	19.07	2.15
BIS2	7.61	3.76	11.48	3.70	3.29	2.56
ORCA3	3.51	0.00	0.56	0.00	0.00	2.22

Table 5: FPKMs of secoiridoid biosynthetic genes and regulators of M2-1582 (MT) and wild-type *C. roseus* (WT).

Auxin	FLP-WT FPKM	FLP-MT FPKM	SLP-WT FPKM	SLP-MT FPKM	Root-WT FPKM	Root-MT FPKM
Tryptophan aminotransferase	0.13	0.00	0.10	0.00	0.00	7.61
Flavin monooxygenase 1	1.99	0.09	1.76	0.06	0.03	0.12
Flavin monooxygenase 2	0.62	0.00	3.89	0.46	0.74	6.06
Indole-3-acetic acid-amido synthetase	5.35	0.07	5.71	5.85	1.19	26.08
Auxin efflux carrier component	16.70	6.26	10.32	5.00	6.01	21.10
Auxin influx carrier protein	4.42	3.65	10.35	4.06	6.75	7.78
Walls are thin protein	65.05	86.88	35.51	82.80	9.27	75.14
Auxin response factor 1	88.84	139.57	86.30	150.10	49.70	62.12
Auxin response factor 2	5.56	3.54	2.99	2.72	2.74	3.17
Cytokinin						
Adenosine phosphate-isopentenyltransferase	5.79	0.52	0.35	0.00	13.67	24.69
Cytokinin dehydrogenase	0.58	2.04	0.05	0.61	17.92	6.51
Gibberellin						
Gibberellin 20-oxidase	0.98	1.15	2.41	0.13	3.33	51.81
Ent-kaurene oxidase	7.14	7.66	1.85	7.93	16.18	53.55
Ent-kaurenoic acid oxidase	14.49	11.38	4.66	8.67	51.10	136.56
GA-2 oxidase	0.65	2.71	0.50	3.20	1.11	5.73
Tempranillo	33.65	51.87	152.99	85.89	13.90	115.46
Ethylene						
1-aminocyclopropane-1-carboxylate synthase 1	0.00	0.00	0.05	0.00	5.67	1.35
1-aminocyclopropane-1-carboxylate synthase 2	0.00	0.00	0.05	0.30	3.02	0.13
1-aminocyclopropane-1-carboxylate synthase 3	2.39	3.87	0.13	7.18	7.69	2.30
Ethylene insensitive 3-like 1	85.33	91.98	35.72	99.12	75.55	37.81
Jasmonate						
Jasmonic acid-amido synthetase	61.66	171.00	35.26	180.15	63.47	96.13

Table 6: FPKMs of plant hormone anabolic and catabolic biosynthetic genes and regulators of M2-1582 (MT) and wild-type *C. roseus* (WT).

Protein	FLP-WT FPKM	FLP-MT FPKM	SLP-WT FPKM	SLP-MT FPKM	Root-WT FPKM	Root-MT FPKM
Anthranilate synthase alpha-subunit	9.36	0.61	21.38	0.33	0.20	0.36
Anthranilate synthase beta-subunit	9.92	17.75	20.81	19.88	31.34	91.83
Anthranilate phosphoribosyltransferase 1	18.92	33.67	17.32	45.77	47.62	86.68
Anthranilate phosphoribosyltransferase 2	37.21	65.69	37.07	83.00	90.04	238.70
Tryptophan synthase alpha chain	42.23	51.66	72.49	52.90	100.39	256.07
Tryptophan synthase beta chain	106.21	29.87	54.97	65.72	134.25	635.70

Table 7: FPKMs of tryptophan biosynthetic genes of M2-1582 (MT) and wild-type *C. roseus* (WT).

Protein	FLP-WT FPKM	FLP-MT FPKM	SLP-WT FPKM	SLP-MT FPKM	Root-WT FPKM	Root-MT FPKM
Phenylalanine ammonia-lyase	115.17	443.96	193.54	330.72	1021.41	1691.92
Cinnamate 4-hydroxylase	37.27	15.50	14.19	5.65	220.02	143.59
4-coumarate-CoA ligase	30.81	36.37	14.31	15.83	314.21	805.36
Hydroxycinnamoyl transferase	11.25	10.63	5.77	4.80	60.46	96.89
p-coumaroyl shikimate 3'-hydroxylase	17.37	32.90	20.61	53.91	17.99	24.73
Caffeic acid 3-O-methyltransferase	84.46	102.85	101.11	101.94	327.31	550.76
Cinnamoyl-CoA reductase	50.72	76.33	47.26	66.06	31.95	75.39
Cinnamyl alcohol dehydrogenase	29.66	23.00	62.59	13.42	6.13	14.03
Ferulate 5-hydroxylase	11.34	12.58	6.57	14.88	280.27	118.99
Homeobox-leucine zipper 1	27.44	12.85	16.54	8.29	16.74	26.12
Homeobox-leucine zipper 2	8.97	4.29	7.00	3.10	3.28	3.91
Homeobox-leucine zipper 3	9.22	5.75	10.20	4.41	4.35	7.05

Table 8: FPKMs of lignin biosynthetic genes and regulatory elements of M2-1582 (MT) and wild-type *C. roseus* (WT).

Auxin Reference ID	Protein	Query Cover	Ident	Plant Species
trinity4all_125263	Tryptophan aminotransferase 1	100%	98%	*Catharanthus roseus*
gbmRNA_000018	Flavin monooxygenase 1	100%	100%	*Catharanthus roseus*
trinity4all_032139	Indole-3-pyruvate monooxygenase	100%	84%	*Actinidia chinensis*
CAP3contig_002379	Indole-3-acetic acid-amido synthetase GH3.6	99%	88%	*Ricinus communis*
trinity4all_001853	Auxin efflux carrier component 1	100%	87%	*Herrania umbratica*
trinity4all_124220	Auxin influx carrier protein	99%	84%	*Zinnia violacea*
trinity4all_026942	Protein Walls Are Thin 1	96%	83%	*Sesamum indicum*
CAP3contig_003685	Auxin response factor 23	99%	75%	*Capsicum baccatum*
trinity4all_002811	Auxin response factor 8	99%	80%	*Capsicum baccatum*
Cytokinin Reference ID				
CAP3contig_006810	Adenylate isopentenyltransferase	99%	70%	*Actinidia chinensis*
trinity4all_028521	Cytokinin dehydrogenase	99%	80%	*Actinidia chinensis*
Gibberellin Reference ID				
CAP3contig_003779	GA 20-oxidase 1	97%	88%	*Nerium oleander*
CAP3contig_004402	Ent-kaurene oxidase	100%	71%	*Actinidia chinensis*
gbmRNA_009359 |GACD01018458.1|	Ent-kaurenoic acid oxidase 1	100%	79%	*Capsicum baccatum*
trinity4all_052697	GA 2-oxidase 2	100%	92%	*Nerium oleander*
trinity4all_000701	Tempranillo	100%	76%	*Olea europaea*
Ethylene Reference ID				
trinity4all_080218	1-aminocyclopropane-1-carboxylate synthase 7	98%	81%	*Morus notabilis*
trinity4all_124784	1-aminocyclopropane-1-carboxylate synthase 3	97%	85%	*Petunia x hybrida*
trinity4all_126269	1-aminocyclopropane-1-carboxylate synthase	98%	77%	*Handroanthus impetiginosus*
trinity4all_001654	EIN3-like protein EIL1	100%	88%	*Actinidia chinensis*
Jasmonate Reference ID				
trinity4all_000472	JAR1	100%	80%	*Gentiana rigescens*

Table 9: Closest characterized protein to identified homologue of plant hormone anabolic and catabolic biosynthetic genes and regulators of M2-1582 and wild-type *C. roseus* (WT), plant species the protein is from and the contig ID in the reference transcriptome.

Tryptophan Reference ID	Protein	Query Cover	Ident	Plant Species
trinity4all_027070	Anthranilate synthase alpha-subunit 1	94%	82%	*Mitragyna speciosa*
trinity4all_078132	Anthranilate synthase, beta chain	96%	73%	*Handroanthus impetiginosus*
trinity4all_000971	Anthranilate phosphoribosyltransferase	97%	83%	*Capsicum baccatum*
gbmRNA_024001 |GACD01048044.1|	Anthranilate phosphoribosyltransferase	86%	85%	*Solanum pennellii*
trinity4all_052641	Tryptophan synthase alpha chain	99%	77%	*Solanum lycopersicum*
CAP3contig_006354	Tryptophan synthase beta chain 1	100%	84%	*Helianthus annuus*

Table 10: Closest characterized protein to identified homologue of tryptophan biosynthetic genes of M2-1582 and wild-type *C. roseus* (WT), plant species the protein is from and the contig ID in the reference transcriptome.

Lignin Reference ID	Protein	Query Cover	Ident	Plant Species
CAP3contig_000238	Phenylalanine ammonia lyase	100%	99%	*Catharanthus roseus*
gbEST_012317 |gb|EG559401.1|	**mRNA for cinnamate 4-hydroxylase (CYP73)	91%	96%	*Catharanthus roseus*
trinity4all_000719	4-coumarate--CoA ligase	100%	99%	*Vanilla planifolia*
CAP3contig_010067	Hydroxycinnamoyl transferase	100%	90%	*Coffea arabica*
CAP3contig_006154	4-coumaric acid 3`-hydroxylase 25	99%	84%	*Coffea arabica*
trinity4all_026944	Caffeic acid 3-O-methyltransferase	100%	100%	*Catharanthus roseus*
trinity4all_026772	Cinnamoyl-CoA reductase 1	98%	79%	*Sesamum indicum*
trinity4all_027355	CAD4 protein	100%	99%	*Catharanthus roseus*
gbmRNA_023690	Ferulate 5-hydroxylase	96%	82%	*Camptotheca acuminata*
CAP3contig_000921	Homeobox-leucine zipper family protein isoform 5	100%	87%	*Theobroma cacao*
gbmRNA_029912	Transcription factor HEX	100%	88%	*Handroanthus impetiginosus*
trinity4all_001548	Homeobox-leucine zipper protein REVOLUTA	100%	87%	*Jatropha curcas*

Table 11: Closest characterized protein to identified homologue of lignin biosynthetic genes and regulators of M2-1582 and wild-type *C. roseus* (WT), plant species the protein is from and the contig ID in the reference transcriptome. **Contig missing start codon, alignment is based on a mostly complete nucleotide sequence.

Protein	Reference	PP_Ref	M2-1582	Genome
Auxin Influx Carrier Protein 1	trinity4all_124220	C467646	-	cro_scaffold_3010661
Auxin Influx Carrier Protein 2	-	-	scaffold14770	cro_scaffold_3047681
Auxin efflux carrier component	trinity4all_001853	C481950	C366999	cro_scaffold_3059189
Walls are thin protein	trinity4all_026942	C468096	scaffold13427	cro_scaffold_3059041
Indole-3-acetic acid-amido synthetase	CAP3contig_002379	C480398	C365221	cro_scaffold_2992911

Table 12: Unique identifiers of auxin transporters and regulator in different references. Reference representation includes *C. roseus* reference transcriptome (Reference), *C. roseus* var. Pacifica Peach reference transcriptome (PP_Ref), M2-1582 transcriptome (M2-1582) and *C. roseus* genome (Kellner et al. 2015). (-) denotes the lack of presence of an identical sequence.