

Cover Page



Universiteit Leiden



The handle <http://hdl.handle.net/1887/22039> holds various files of this Leiden University dissertation.

**Author:** Miettinen, Karel

**Title:** Elucidation of the secoiridoid pathway in *Catharanthus roseus*

**Issue Date:** 2013-10-24

**Elucidation of the secoiridoid pathway in**  
*Catharanthus roseus*

**Karel Miettinen**

Karel Miettinen

Elucidation of the secoiridoid pathway in *Catharanthus roseus*

Printed by Smart Printing Solutions

Cover design painted by Teresa Capell

**Elucidation of the secoiridoid pathway in**  
*Catharanthus roseus*

Proefschrift

ter verkrijging van  
de graad van Doctor aan de Universiteit Leiden,  
op gezag van Rector Magnificus prof.mr. C.J.J.M. Stolker,  
volgens besluit van het College voor Promoties  
te verdedigen op Donderdag 24 oktober 2013  
klokke 13:45 uur

door

**Karel Miettinen**

geboren te Lappeenranta (Finland)  
in 1982

Promotiecommissie

Promotor: Prof.dr. J.Memelink  
Overige leden: Prof.dr. C.J. ten Cate  
Prof.dr. R. Verpoorte  
Dr.D. Werck-Reichhardt (Université de Strasbourg)  
Prof.dr. H. J. Bouwmeester (Wageningen University)

The research leading to these results has received funding from the European Union Seventh Framework Programme FP7/2007-2013 under grant agreement number 222716 - SMARTCELL.

*To Minna, my love and inspiration*



## Contents

	Page
Chapter 1      Monoterpenoid indole alkaloid and iridoid biosynthesis in <i>Catharanthus roseus</i>	9
Chapter 2      Characterization of the plastidial geraniol synthase from Madagascar periwinkle which initiates the monoterpenoid branch of the alkaloid pathway in internal phloem associated parenchyma	35
Chapter 3      Characterization of two geraniol synthases from <i>Valeriana officinalis</i> and <i>Lippia dulcis</i> : similar activity but difference in subcellular localization	65
Chapter 4      Discovery and reconstitution of the secoiridoid pathway from <i>Catharanthus roseus</i>	105
Chapter 5      Summary	143
Chapter 6      Samenvatting	151
Acknowledgements	158
<i>Curriculum vitae</i>	159





## **Monoterpenoid indole alkaloid and iridoid biosynthesis in *Catharanthus roseus***

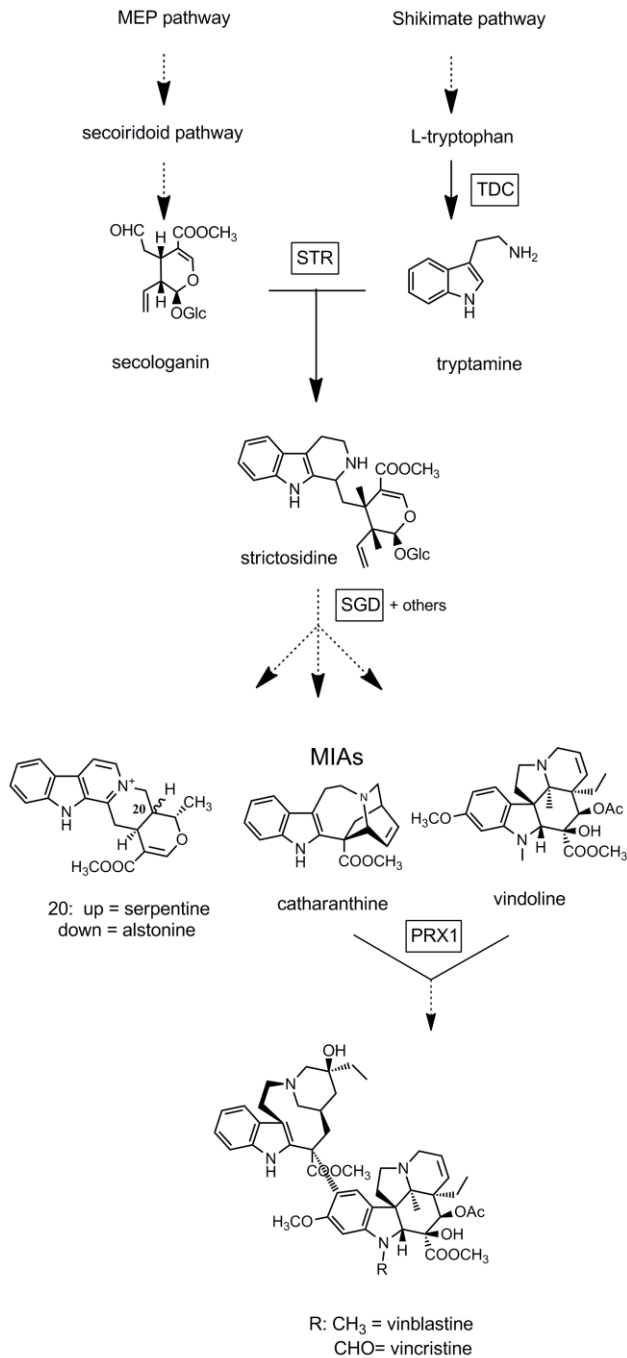
### Introduction

Plants produce tens of thousands of different secondary metabolites, compounds with important roles in defence against herbivores, pests and pathogens or interaction with surroundings but not required for plant growth, development or reproduction in controlled conditions. These compounds have a diversity of traditional and novel applications for example as medicines, flavourings, colorants, or in agriculture. *Catharanthus roseus* (Madagascar periwinkle) is one of the most studied medicinal plants and it has been the object of interest of scientists for several decades. It produces the important class of secondary metabolites the monoterpenoid indole alkaloids (MIA) and their precursors the (seco)iridoids. From the over 130 known *C. roseus* MIAs several have pharmaceutical applications such as the anti-hypertensive drugs serpentine and ajmalicine and the potent antitumor agents bisindole alkaloids vincristine and vinblastine that are widely used to treat several types of cancer such as Hodgkins lymphoma, Kaposi's sarcoma, breast cancer, bladder cancer and testicular cancer. Most MIAs are present in plants in very small amounts making their extraction for pharmacological use uneconomical. Because of their complex structures total synthesis is unfeasible (van der Heijden et al., 2004). A lot of effort has been put into biotechnological ways to produce higher amounts of MIAs such as metabolic engineering of cell cultures (Canel et al., 1997), hairy roots (Jaggi et al., 2011; Hughes et al., 2004; Hong et al., 2006; Magnotta et al., 2007; Peebles et al. 2009, 2011; Wang et al., 2010; Pomachova et al., 2009; Zhou et al., 2011; Liu et al., 2011), and even whole plants (Pan et al., 2012) with the help of hormone treatments and intermediate metabolite feeding (Morgan and Shanks, 1999; El-Sayed and Verpoorte, 2002; El-Sayed et al., 2004; Lee-Parsons and Royce, 2006), but a breakthrough still awaits. To date only parts of the biosynthetic pathway leading to MIAs are known. Knowledge of the biosynthesis is essential to biotechnological production of MIAs and (seco)iridoids.

## MIA biosynthesis

The *C. roseus* MIA biosynthetic pathway is extremely complex with tens of hypothetical steps and different branches. The universal precursor of all MIAs, strictosidine, is produced by condensation of the monoterpenoid secologanin and the indole compound tryptamine (Fig. 1) by the enzyme strictosidine synthase (STR) (Mizukami et al., 1979; Pfitzner and Zenk 1989; McKnight et al 1990; Pasquali et al., 1992; de Waal et al., 1995). Tryptamine originates from the shikimate pathway and is synthesized from tryptophan by the enzyme tryptophan decarboxylase (TDC) (Pennings et al., 1989; De Luca et al., 1989), whereas secologanin originates from the methyl erythritol phosphate (MEP) pathway via the iridoid pathway (Contin et al., 1998; Veau et al., 2000; Chahed et al., 2000; Hong et al., 2003). The iridoid pathway is considered to be the rate-limiting step in MIA biosynthesis (Morgan and Shanks, 1999) and engineering of this part of the MIA biosynthesis pathway may be the key to successfully improve MIA production. All MIAs from plants, including the anti-cancer drug camptothecin from *Camptotheca acuminata* and the antimalarial drug quinine from the *Cinchona* tree, are produced from tryptamine and secologanin (O'Connor and Maresh, 2006). This stresses the important role of the MEP and iridoid pathways in engineering of the MIA pathway.

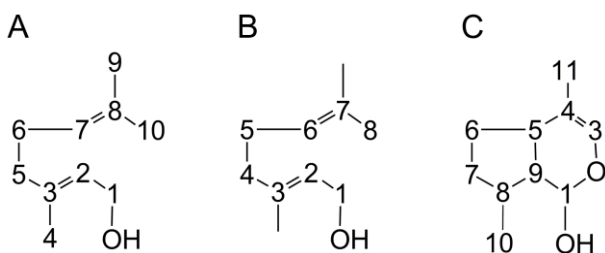
In *C. roseus* MIA biosynthesis occurs via the removal of the glucose moiety from strictosidine by strictosidine- $\beta$ -D-glucosidase (SGD) yielding strictosidine aglucone giving rise to the highly reactive dialdehyde that can go through a number of molecular rearrangements. This is the first branching point of MIA biosynthesis leading to the tens of *C. roseus* monomeric MIAs (Scott et al., 1977; Luiendijk et al., 1998; Geerlings et al., 2000). Dimeric bisindole alkaloids are formed by condensation of catharanthine and vindoline to form 3',4'-anhydrovinblastine by the peroxidase enzyme  $\alpha$ -3',4'-anhydrovinblastine synthase (PRX1) (Costa et al., 2008) further leading to vinblastine and ultimately to vincristine (O'Connor and Maresh, 2006).



**Fig.1 Overview of MIA biosynthesis.** Solid arrows indicate single enzymatic conversions, dashed arrows multiple reactions. MEP-pathway: methyl erythritol phosphate pathway, TDC: tryptophan decarboxylase, STR: strictosidine synthase, SGD: strictosidine-β-D-glucosidase, PRX1: α-3',4'-anhydrovinblastine synthase.

## Numbering of carbon atoms in terpenoid and iridoid structures

In literature there exist two different ways of numbering carbon atoms in linear terpenoids. Both are used in this thesis in names of compounds and enzymes (Fig. 1B). One is used mostly in *C. roseus* and MIA biosynthesis related publications (Fig. 2A)(Uesato et al., 1984). The second one used in most of the monoterpene related literature conforms to the naming standards of the international union of pure and applied chemistry (IUPAC), the world authority on chemical nomenclature (IUPAC-IUB, 1986)(Fig. 2B). This nomenclature will be used in the rest of this chapter. As an example 10-OH-geraniol as named according to the non-IUPAC nomenclature is 8-OH-geraniol in the IUPAC nomenclature. The cyclic iridoids are usually numbered in a non-IUPAC way (Fig. 2C) (Dinda et al., 2007). Whereas other numbering schemes exist this is the most widely used and is used here for clarity's sake.

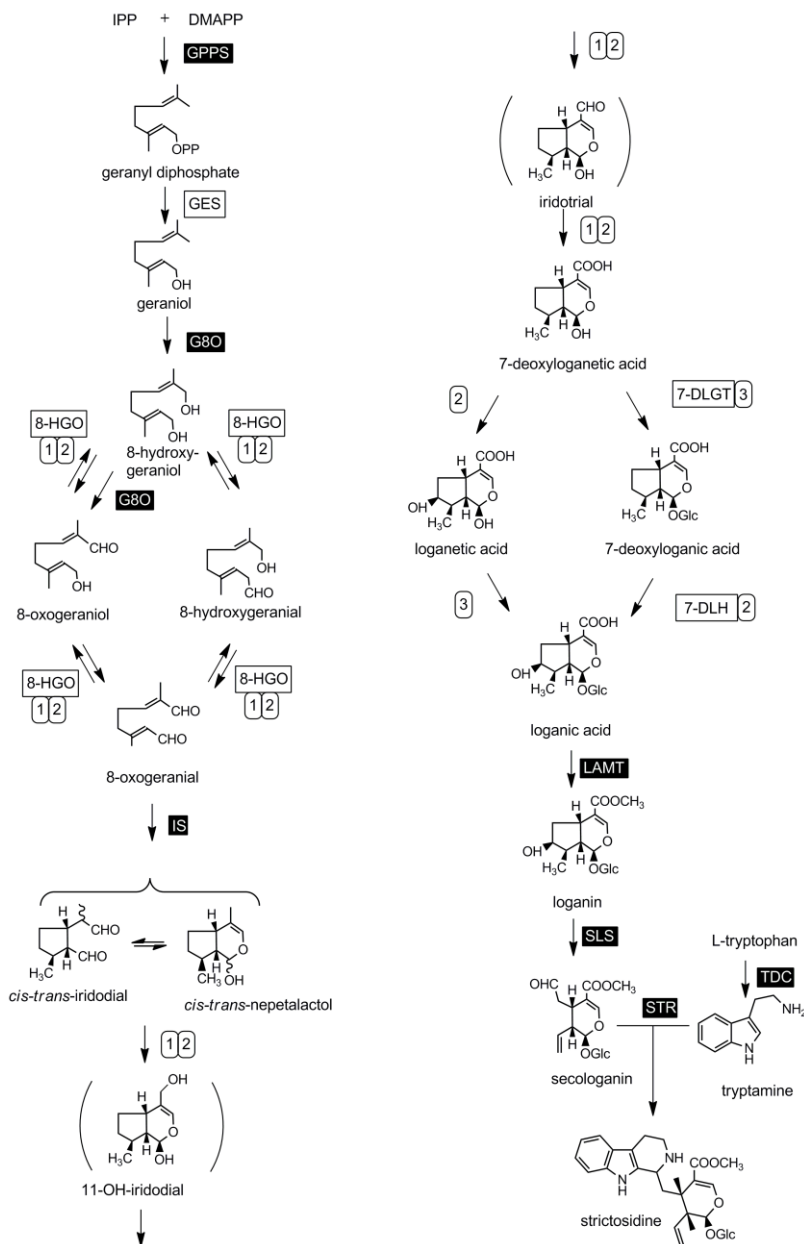


**Fig. 2** Numbering of terpenoid and iridoid carbon atoms. A: non-IUPAC numbering, B: IUPAC compliant numbering, C: typical numbering of iridoid carbons.

## Monoterpene and iridoid biosynthesis

Biosynthesis of the terpenoid moiety of MIAs (Fig. 3) starts from the head to tail condensation of dimethylallyl diphosphate (DMAPP) and isopentenyl diphosphate (IPP)(McGarvey and Croteau, 1995), the basic building blocks of all terpenes, by the enzyme geranyl diphosphate synthase (GPPS) to form geranyl diphosphate (GPP). *C. roseus* has two different kinds of holoenzymes capable of making GPP, the mitochondrial homomeric CrGPPS and the plastidial heteromeric enzyme consisting of a small- and large subunit (CrGPPS-SSU and CrGPPS-LSU). The heteromeric form is thought to supply GPP for MIA production (Rai et al., 2013). The next step is generally assumed to be the synthesis of the acyclic monoterpene alcohol geraniol (Oudin et al., 2007a). While a Mn<sup>2+</sup> dependent terpene synthase called geraniol synthase (GES) is known to produce geraniol from GPP in *Ocimum basilicum* (Iijima et al., 2004), *Cinnamomum tenuipilum* (Yang et al., 2005), and *Perilla citriodora* (Ito and Honda, 2007) no GES from an iridoid producing plant species such as *C. roseus* had been cloned at the start of the present study.

The first dedicated step in (seco)iridoid biosynthesis is the hydroxylation of geraniol (and its *cis*-isomer nerol) at its 8-position by the cytochrome P450 (CYP) enzyme geraniol-8-oxidase (G8O)(also known as geraniol-8-hydroxylase) (Meehan and Coscia, 1973; Collu et al., 2001) to form 8-hydroxygeraniol (and 8-hydroxynerol respectively). This enzyme was shown to also be able to oxidize the 8-position further to 8-oxogeraniol (Höfer et al., 2013). G8O needs the accessory protein cytochrome P450 reductase (CPR) to function (Madyasha and Coscia, 1979; Meijer et al., 1993). The following steps are less well known, though feeding experiments with (radio)isotopes of intermediates and assays with enzyme preparations from *C. roseus* and other plants suggests oxidation of 8-OH-geraniol/nerol to both 8-oxogeraniol and 8-hydroxygeraniol and further oxidation into 8-oxo-geraniol by 8-hydroxygeraniol oxidoreductase (8-HGO) (Uesato et al., 1984, 1986a, 1986b and 1987; Ikeda et al., 1991; Hasnain, 2010). Recently it was found that 8-oxo-geraniol is turned into the first iridoid, *cis-trans*-nepetalactol (also known as iridodial hemiacetal form), by the iridoid synthase (IS) enzyme in an NADPH dependent reductive cyclization reaction forming the first compound



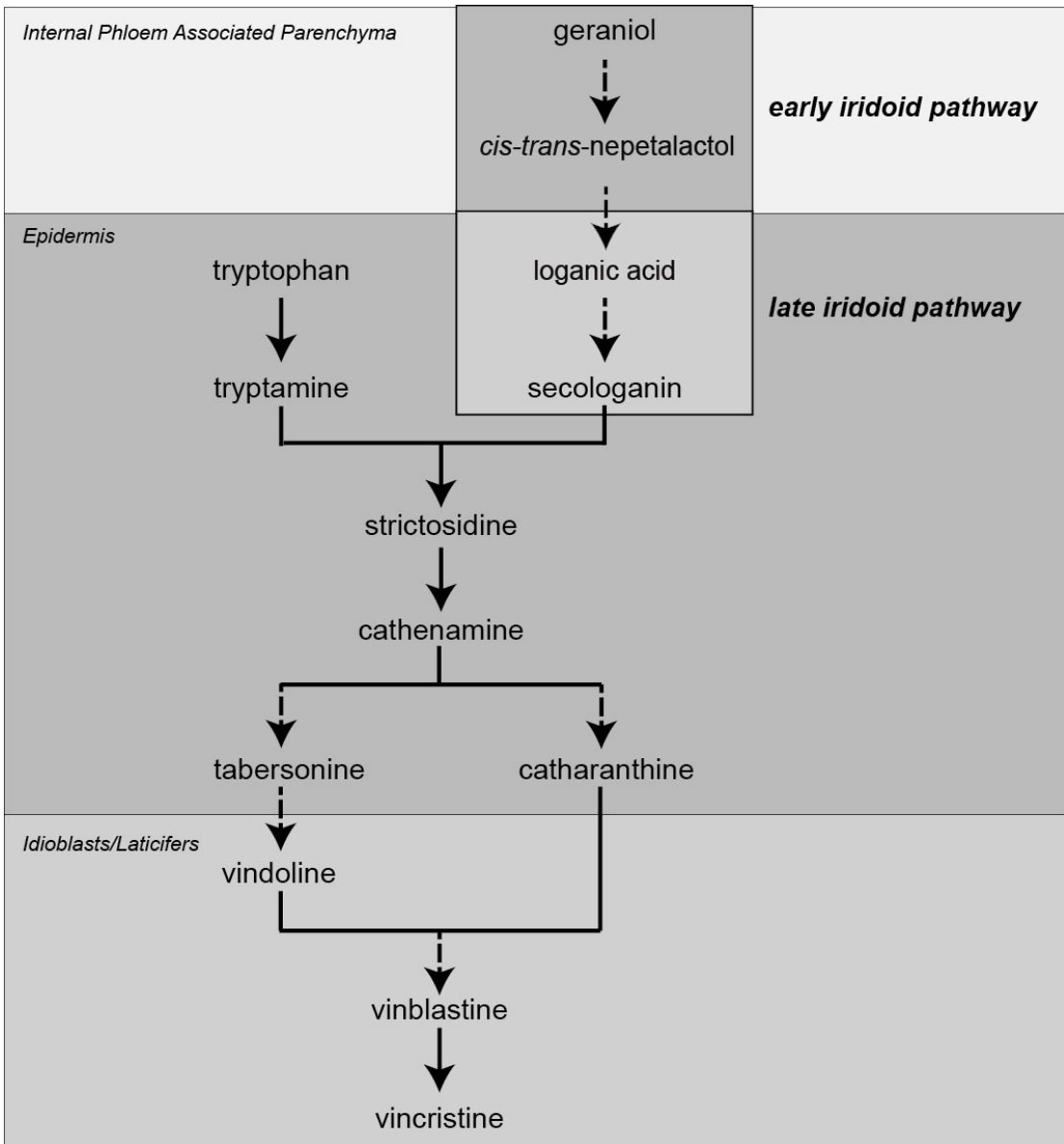
**Fig. 3 Hypothetical model of iridoid biosynthesis.** Solid arrows indicate single enzymatic conversions, dashed arrows multiple reactions. Reactions for which the corresponding *C.roseus* gene has been cloned have a black background, reactions for which the corresponding gene was not cloned but that are described in literature or for which the corresponding gene has been cloned from another organism have a white background. Numbers are for putative enzyme type, 1: oxidoreductase, 2: cytochrome P450, 3: UDP-glucose-glucosyltransferase.



with an iridoid backbone. The compound is considered to exist in equilibrium between the single ring structure *cis-trans*-iridodial and the double ring hemiacetal structure *cis-trans*-nepetalactone (Geu-Flores et al., 2012). Even though a reaction oxidizing *cis-trans*-nepetalactol at carbon-11 was never described, iridotrial was found to be incorporated into MIAs in isotope feeding experiments making it a plausible intermediate. Genes coding for enzymes catalyzing the later steps have not been cloned but some of the reactions have been characterized from protein extracts. A partially purified enzyme from *Lonicera japonica* cell cultures showed UDP-glucose dependent glucosyltransferase activity by glucosylating 7-deoxyloganetic acid and similar compounds into 7-deoxyloganic and other glucosides, 7-deoxyloganic acid glucosyl transferase (7-DLGT) activity being the most specific (Yamamoto et al., 2002). Nagatoshi et al. (2011) described a UDP-glucose:glucosyltransferase from *Gardenia jasminoides* that glucosylates 7-deoxyloganetin but not 7-deoxyloganetic acid. They also proposed that UGT85A23 (AB591741.1) from *C. roseus* could be the enzyme responsible for iridoid glucosylation. 7-deoxyloganic acid hydroxylase activity (7-DLH), leading to loganin, was observed in an incubation of a microsome preparation from a *Lonicera japonica* cell culture. According to inhibition assays the enzyme is likely to be a CYP (Katano et al., 2001). Loganin acid methyl transferase (LAMT) was cloned from *C. roseus* by Murata et al. (2008). It uses S-adenosyl-methionine (SAM) as co-substrate. Loganin is known to be turned to secologanin by secologanin synthase (SLS), a CYP catalyzing a peculiar ring opening reaction (Yamamoto et al., 1999, 2000; Irmeler et al., 2000).

## Tissue and subcellular organization of iridoid biosynthesis

Many different subcellular compartments in different cell types are required in MIA biosynthesis. This is proposed to minimize the toxic effects of the compounds (Facchini, 2001) but is also seen as a multilayered defence strategy (Guirimand et al., 2010; Roepke et al., 2010). It is speculated that the "Early Steps of Monoterpenoid Biosynthesis" (ESMB), at least until G8O, occur in vascular tissue cells called internal phloem associated parenchyma (IPAP) cells (Burlat et al., 2004; Oudin et al., 2007b). In *C. roseus* the known MEP pathway genes are expressed in IPAP cells and the enzymes located in plastids (Oudin et al., 2007b; Ginis et al., 2012; Burlat et al., 2004; Guirimand et al., 2009). It is not yet known in which cell type the genes coding for the next enzymes, the plastid-located CrGPPS-SSU and CrGPPS-LSU, are expressed but the idea of a co-localized early pathway/ESMB suggests localization in the IPAP cells. Tissue and subcellular localization of GES is also expected to be the same. This would mean that an intermediate metabolite must then be transported to the cytosol since the next enzyme G8O is located on the ER facing to the cytosol, again in the IPAP cells (Burlat et al., 2004; Guirimand et al., 2009). IS reported to be located in the cytosol of IPAP cells (Geu-Flores et al., 2012) suggesting similar localization of the in-between step(s) (8-HGO). Localization of the next enzymes is not known. The later steps in the pathway, LAMT, SLS, TDC and STR, dubbed intermediate pathway, are expressed in epidermal cells (Irmeler et al., 2000; Murata et al., 2008; Guirimand et al., 2011b). These enzymes are all localized in the cytosol except STR that localizes to the vacuole. This proposed to have a function in plant defence against herbivores; the glucoside strictosidine is separated from the  $\beta$ -glucosidase SGD as long as subcellular compartments are intact. When cells are disrupted by herbivory, substrate and enzyme mix producing highly reactive and hypothetically toxic protein crosslinking compounds (Guirimand et al., 2010). Dimeric bisindole MIAs are produced in another distinct cell type, i.e. the idioblast/laticifer cells (Guirimand 2011b).



**Fig. 4 Tissue localization of MIA biosynthesis.** Solid arrows indicate single enzymatic conversions, dashed arrows multiple reactions.

## Regulation of MIA and iridoid biosynthesis genes

MIAs and iridoids are secondary metabolites, compounds that are not regarded as essential for the growth and development of an organism but rather for interaction with its surroundings including as mentioned earlier, plant defence. This is reflected in the way their production is regulated. MIA biosynthetic genes are tightly regulated by developmental cues and external stress signals (Aerts et al., 1994; St. Pierre et al., 1999). A striking mechanism of developmental regulation is that each of the known MIA biosynthetic genes is strictly expressed in one of the three distinct cell types in *C. roseus* aerial organs (IPAP cells, epidermis, idioblasts/laticifers). MIA production is also dependent on the developmental stage of leaves. All known MIA pathway genes are regulated by the defence hormone jasmonate (van der Fits and Memelink 2000; Giddings et al., 2011; Geu-Flores et al., 2012; Oudin et al., 2007; He et al., 2011) orchestrating defence against herbivores, necrotrophic pathogens and biotrophic pathogens. Also the plant hormones auxin, ethylene, cytokinins and the signaling molecule nitric oxide are known to affect the expression of MIA biosynthesis genes and MIA production (Ginis et al., 2012; Yahia et al., 1998; Zhou et al., 2010). Many genes, in the precursor (AS $\alpha$ ), intermediate (STR, SLS, LAMT, TDC) and late (D4H) pathway, are regulated by octadecanoid-derivative responsive *Catharanthus* AP2-domain transcription factors (ORCA2, ORCA3) (Menke et al., 1999; van der Fits and Memelink 2000; Hasnain 2010). These jasmonate inducible regulators are known to activate *STR* expression by binding to a GCC box in the jasmonate and elicitor responsive element (JERE) in the *STR* promoter (Menke et al., 1999; van der Fits and Memelink, 2000; van der Fits et al., 2001). The expression of the ORCA3 gene is controlled by the basic helix-loop-helix (bHLH) transcription factor CrMYC2 that binds an element called a G-box (T/G-box) in the ORCA3 promoter (Zhang et al., 2011). ORCA3 expression is induced by jasmonate via jasmonate dependent degradation of the jasmonate ZIM domain (JAZ) repressors, which bind CrMYC2 in the absence of jasmonate, allowing CrMYC2 to activate the ORCA3 promoter. (Zhang, 2008). The *STR* promoter also contains a G-box but CrMYC2 cannot activate the promoter directly (Zhang et al., 2011). On the other hand the G-box binding factors CrGBF1 and CrGBF2 act as transcriptional repressors of the *STR* promoter (Siberil et

al., 2001). A second set of transcription factors, the repressors ZCT1, ZCT2 and ZCT3, bind elsewhere in the STR promoter (Pauw et al., 2004). A similar transcriptional regulation system has been found in tobacco where homologues of ORCA3, the jasmonate inducible NIC2 locus ERF transcription factors activate nicotine biosynthesis gene expression by binding to a GCC box (Shoji et al 2010). ORC1 (ERF221) and ERF189, activate nicotine biosynthesis gene expression but in this case bHLH transcription factors were found to be additionally needed to fully activate expression (Shoji et al., 2011; De Boer et al., 2011). The CrMYC2 homolog NtMYC2 was found to boost activation of nicotine biosynthesis genes by ERF189 by binding a G-box (Shoji et al., 2011) and NbbHLH1 from *Nicotiana benthamiana* was also found to boost activation by ORC1 (De Boer et al., 2011). RNAi silencing of NtMYC2 dramatically reduced the expression of nicotine biosynthesis genes but also the expression of the NIC2 locus ERF genes suggesting that they are regulated by NtMYC2 and leading to a model where nicotine biosynthesis genes are induced by NtMYC2 directly and through ERF transcription factors (Shoji et al., 2011). NtMYC2 itself was found to bind JAZ repressors and be activated by jasmonate via the degradation of the JAZ repressors (Shoji et al 2008). Regulation of *C. roseus* early pathway genes such as G8O seems to employ a different mechanism and they are known not to be regulated by the ORCAs (Hasnain, 2010). While the G8O promoter does not contain a JERE it also has a G-box sequence (Suttipanta et al., 2007). No transcription factor(s) binding to the G-box in the G8O promoter are known yet but it is likely to be a bHLH, possibly CrMYC2 or a related transcription factor.

### **Research strategy and available resources**

This work was part of the SmartCell project, a EU seventh framework funded collaboration between several European groups, where the aim was the elucidation of the secoiridoid pathway and the engineering of the pathway in several organisms for optimized and *de novo* production of iridoids, MIAs and related compounds. A candidate based approach was taken to find enzymes catalyzing both hypothetical and completely unknown intermediate reactions in iridoid biosynthesis and to find suitable enzymes for biotechnical applications. The candidates were picked based on amino acid and nucleotide sequence homology with known enzymes in Leiden and the screening refined by gathering additional information

such as gene expression pattern in different conditions by SmartCell partner Vlaams Instituut voor Biotechnologie, Gent, Belgium, tissue localization of proteins by partners University of Zurich, Switzerland and Université Catholique de Louvain, Belgium and tissue localization of transcripts by the external collaborator Vincent Burlat from University of Toulouse, France. Functional screening of the chosen soluble enzyme candidates was performed in Leiden and P450s were screened by the SmartCell partner Centre national de Recherche Scientifique, Strasbourg, France. Work concerning the *C. roseus* GES was a non-Smart Cell collaboration with Marc Clastre and Andrew Simkin from Tours University, France and collaborators from University of Toulouse, France.

Today's high throughput data mining methods enable the problem to be addressed from multiple angles yielding a broad range of candidates but also the possibility for finer selection and more significant results. Based on the evidence at the onset of these studies the pathway was missing 4-6 enzymes with completely new functionalities. As proposed before (Ikeda et al., 1991; Hasnain, 2010) oxidation of 8-OH geraniol to 8-oxogeraniol should be carried out by a soluble enzyme using NAD(P)<sup>+</sup> as a co-substrate by an alcohol dehydrogenase subtype of oxidoreductases expected to be expressed in IPAP cells. Similar reactions have been showed to be catalyzed also by other kinds of oxygenase enzymes such as cytochrome P450s (Höfer et al., 2013), therefore also other types of candidates were taken into account. In order to get the carboxyl group at the 11-position as it is present in the later metabolites, for example in loganic acid, the iridoid intermediate has to be oxidized to form either the hypothetical intermediate 11-OH-iridodial, iridotrial or directly 7-deoxyloganic acid. While no enzyme types for the reaction have been proposed in literature the hypothetical reaction of 11-hydroxylation is similar to the one of G8O and thus a cytochrome of the CYP76 family seemed a good candidate. In the case of a hydroxylated intermediate an oxidoreductase might be needed to yield iridotrial that can further be oxidized to the carboxylic acid by the same or a different P450. To end up with loganin or any iridoid later in the pathway glucosylation has to happen. This reaction is likely catalyzed by a Plant Secondary Product Glucosyltransferase (PSPG) a subclass of UDP-dependent glucosyltransferases (Vogt et al., 2000; Nagatoshi 2011). Hydroxylation of 7-deoxyloganic acid was previously described as probably done by a P450 (Katano et al., 2001). Candidates

for the *C. roseus* GES were expected to belong to the class of monoterpene synthases identified in several plant species

Several *C. roseus* transcriptomics data sets created with techniques such as cDNA-AFLP (Rischer et al., 2006; Hasnain 2010), random cDNA library sequencing (Murata et al., 2006 and 2008) or next generation sequencing (Van Moerkercke et al., 2013; Gongora-Castillo et al., 2013; Xiao et al., 2013) are currently publicly available. They have been generated from different types of plant material such as different organs/tissues of whole plants, wild type and transgenic cell culture lines and hairy roots in different conditions such as different developmental stages, and with different hormone treatments. This not only gives access to a large amount of sequence data with good coverage of the transcriptome in desired conditions (for example jasmonate treatment) yielding full sequences of candidates, but also enables comparison between datasets and candidate selection based on differential expression. cDNA AFLP data from ORCA2 and ORCA3 inducible overexpression cell culture lines made it possible to screen for ORCA2 and ORCA3 regulated candidates (Hasnain, 2010). Recent proteomics data (Champagne et al., 2012) from MIA producing plant material gives differential data about enzyme distribution over different cell culture lines. As only indirect evidence of the existence of most of the intermediate compounds was available different pathway models were taken into account and the candidate enzymes were assayed with hypothetical intermediates according to the different models. Most compounds to be used in these functional assays were not commercially available so a strategy to obtain such compounds was designed. Some literature on the synthesis exists (Jensen et al., 1987 and 1989; Hasnain 2010) so semisynthesis from natural products according to the literature was chosen as a production method. *Aucuba japonica* leaf material was collected on the Leiden University campus. Aucubine (an iridoid glucoside) was purified by charcoal adsorption followed by silica gel chromatography and compound purity was measured with NMR in Leiden. Iridotrial was then produced by Chiralix BV. (Nijmegen, NL) from aucubine by HCOOH, Pd/C reduction and Vilsmeier formylation. Three other glucosides were produced from iridotrial by organic reduction and oxidation reactions by the same company. Aglucones for enzyme assays were produced by  $\beta$ -glucosidase treatment in Leiden.

In vivo experiments to test candidate genes and known terpenoid and iridoid biosynthesis genes for *in vivo* activity and biotechnological properties in *N. benthamiana*, *N.*

*tabacum* and *C. roseus*, were performed by SmartCell partners University of Wageningen (WUR), University of Lleida, Spain, VTT Technical Research Centre of Finland and IME Fraunhofer, Aachen Germany. Reconstitution of the pathway in *N. benthamiana* by transient transformation and metabolite analysis of products was conducted at WUR.

## Outline of the thesis

**Chapter 2** presents cloning and functional characterization of the *C. roseus* geraniol synthase (CrGES). The enzyme expressed in *E. coli* was functionally characterized *in vitro*. Also tissue-specific and subcellular localization of GES were determined. Proof of enzyme in function *in vivo* was given by recombinant expression of CrGES in *Saccharomyces cerevisiae* followed by metabolite analysis. Expression analysis showed it has a jasmonate inducible early monoterpenoid pathway expression pattern.

**Chapter 3** describes cloning of the geraniol synthase from *Valeriana officinalis* (VoGES) and its comparison with the geraniol synthase from *Lippia dulcis* (LdGES). The enzymes expressed in *E. coli* were functionally characterized *in vitro*. GESs were expressed in stable transgenic tobacco lines and transiently expressed in *N. benthamiana* as full length plastidial and N-terminally truncated cytosolic versions and resulting changes in soluble and volatile metabolites were analyzed in detail.

**Chapter 4** presents transcriptomics and proteomics screening of candidates for the missing iridoid pathway enzymes. It describes the characterization of the genes corresponding to all the missing iridoid pathway enzymes, 8-hydroxygeraniol oxidoreductase (8-HGO), 7-deoxyloganic acid glucosyltransferase (7-DLGT), iridoid oxidase (IO) and 7-deoxyloganic acid hydroxylase (7-DLH) and their functional characterization via *in vitro* enzyme assays. Tissue and subcellular localization is described and proof of function *in vivo* is given by reconstitution of the whole pathway by *Agrobacterium* mediated transient expression in *N. benthamiana*. A detailed metabolite analysis leads to a complete model of the (seco)iridoid biosynthetic pathway.

**Chapter 5** contains a summary and discussion of chapters 2-4



## References

- Aerts RJ, Gisi D, De Carolis E, De Luca V, Baumann TW** (1994) Methyl jasmonate vapor increases the developmentally controlled synthesis of alkaloids in *Catharanthus* and *Cinchona* seedlings. *Plant J* **5**: 635-643
- Burlat V, Oudin A, Courtois M, Rideau M, St-Pierre B** (2004) Co-expression of three MEP pathway genes and geraniol 10-hydroxylase in internal phloem parenchyma of *Catharanthus roseus* implicates multicellular translocation of intermediates during the biosynthesis of monoterpene indole alkaloids and isoprenoid-derived primary metabolites. *Plant J* **38**: 131-141
- Canel C, Lopes-Cardoso MI, Whitmer S, van der Fits L, Pasquali G, van der Heijden R, Hoge JHC, Verpoorte R** (1998) Effects of over-expression of strictosidine synthase and tryptophan decarboxylase on alkaloid production by cell cultures of *Catharanthus roseus*. *Planta* **205**: 414-419
- Chahed K, Oudin A, Guivarc'h N, Hamdi S, Chénieux J-C, Rideau M, Clastre M** (2000) 1-Deoxy-D-xylulose 5-phosphate synthase from periwinkle: cDNA identification and induced gene expression in terpenoid indole alkaloid-producing cells. *Plant Physiol Biochem* **38**: 559-566
- Champagne A, Rischer H, Oksman-Caldentey KM, Boutry M** (2012) In-depth proteome mining of cultured *Catharanthus roseus* cells identifies candidate proteins involved in the synthesis and transport of secondary metabolites. *Proteomics* **12**: 3536-3547
- Collu G, Unver N, Peltenburg-Looman AM, van der Heijden R, Verpoorte R, Memelink J** (2001) Geraniol 10-hydroxylase, a cytochrome P450 enzyme involved in terpenoid indole alkaloid biosynthesis. *FEBS Lett* **508**: 215-220
- Contin A, van der Heijden R, Lefeber AWM, Verpoorte R** (1998) The iridoid glucoside secologanin is derived from the novel triosephosphate/pyruvate pathway in a *Catharanthus roseus* cell culture. *FEBS Lett* **434**: 413-416
- Costa MMR, Hilliou F, Duarte P, Pereira LG, Almeida I, Leech M, Memelink J, Barcelo AR, Sottomayor M** (2008) Molecular cloning and characterization of a vacuolar class

- III peroxidase involved in the metabolism of anticancer alkaloids in *Catharanthus roseus*. *Plant Physiol* **146**: 403-417
- De Boer K, Tilleman S, Pauwels L, van den Bossche R, De Sutter V, Vanderhaeghen R, Hilson P, Hamill J, Goossens A** (2011) APETALA2/ETHYLENE RESPONSE FACTOR and basic helix-loop-helix tobacco transcription factors cooperatively mediate jasmonate-elicited nicotine biosynthesis. *Plant J* **66**: 1053-1065
- De Luca V, Marineau C, Brisson N** (1989) Molecular cloning and analysis of cDNA encoding a plant tryptophan decarboxylase: comparison with animal dopa decarboxylases. *Proc Natl Acad Sci USA*. **86**: 2582-2586
- de Waal A, Meijer AH, Verpoorte R** (1995) Strictosidine synthase from *Catharanthus roseus* - Purification and characterization of multiple forms. *Biochem J* **306**: 571-580
- Dinda B, Debnath S, Harigaya Y** (2007) Naturally Occurring Iridoids. A Review, Part 1. *Chem Pharm Bull* **55**: 159-222
- El-Sayed M, Choi YH, Frederich M, Roytrakul S, Verpoorte R** (2004) Alkaloid accumulation in *Catharanthus roseus* cell suspension cultures fed with stemmadenine. *Biotech Lett* **26**: 793-798
- El-Sayed M, Verpoorte R** (2002) Effect of phytohormones on growth and alkaloid accumulation by a *Catharanthus roseus* cell suspension cultures fed with alkaloid precursors tryptamine and loganin. *Plant Cell Tiss Org* **68**: 265-270
- Facchini PJ** (2001) Alkaloid biosynthesis in plants: Biochemistry, cell biology, molecular regulation, and metabolic engineering applications. *Ann Rev Plant Phys* **52**: 29
- Geu-Flores F, Sherden NH, Courdavault V, Burlat V, Glenn WS, Wu C, Nims E, Cui Y, O'Connor SE** (2012) An alternative route to cyclic terpenes by reductive cyclization in iridoid biosynthesis. *Nature* **492**: 138-142
- Geerlings A, Ibanez MML, Memelink J, van der Heijden R, Verpoorte R** (2000) Molecular cloning and analysis of strictosidine beta-D-glucosidase, an enzyme in terpenoid indole alkaloid biosynthesis in *Catharanthus roseus*. *J Biol Chem* **275**: 3051-3056
- Ginis O, Courdavault V, Melin C, Lanoue A, Giglioli-Guivarc'h N, St-Pierre B, Courtois M, Oudin A** (2012) Molecular cloning and functional characterization of *Catharanthus roseus* hydroxymethylbutenyl 4-diphosphate synthase gene promoter from the methyl erythritol phosphate pathway *Mol Biol Rep* **39**: 5433-5447

- Gongora-Castillo E, Childs KL, Fedewa G, Hamilton JP, Liscombe DK, Magallanes-Lundback M, Mandadi KK, Nims E, Rungtuphan W, Vaillancourt B, Varbanova-Herde M, DellaPenna D, McKnight TD, O'Connor SE, Buell R** (2012) Development of Transcriptomic Resources for Interrogating the Biosynthesis of Monoterpene Indole Alkaloids in Medicinal Plant Species. *PLOS ONE* **7**: e52506
- Guirimand G, Burlat V, Oudin A, Lanoue A, St-Pierre B, Courdavault V** (2009) Optimization of the transient transformation of *Catharanthus roseus* cells by particle bombardment and its application to the subcellular localization of hydroxymethylbutenyl 4-diphosphate synthase and geraniol 10-hydroxylase. *Plant Cell Rep* **28**: 1215-1234
- Guirimand G, Courdavault V, Lanoue A, Mahroug S, Guihur A, Blanc N, Giglioli-Guivarc'h N, St-Pierre B, Burlat V** (2010) Strictosidine activation in *Apocynaceae*: towards a "nuclear time bomb"? *BMC Plant Biol* **10**: 182
- Guirimand G, Guihur A, Ginis O, Poutrain P, Hericourt F, Oudin A, Lanoue A, St-Pierre B, Burlat V, Courdavault V** (2011a) The subcellular organization of strictosidine biosynthesis in *Catharanthus roseus* epidermis highlights several trans-tonoplast translocations of intermediate metabolites. *FEBS J* **278**: 749-763
- Guirimand G, Guihur A, Poutrain P, Héricourt F, Mahroug S, St-Pierre B, Burlat V, Courdavault V** (2011b) Spatial organization of the vindoline biosynthetic pathway in *Catharanthus roseus*. *J Plant Physiol* **168**: 549-557
- Hasnain G** (2010) the ORCA-ome as a key to understanding alkaloid biosynthesis in *Catharanthus roseus*. PhD thesis, Leiden University
- Hedhili S, Courdavault V, Giglioli-Guivarc'h N, Gantet P** (2007) Regulation of the terpene moiety biosynthesis of *Catharanthus roseus* terpene indole alkaloids. *Phytochemistry Rev* **6**: 341-351
- Hong SB, Hughes EH, Shanks JV, San KY, Gibson SI** (2003) Role of the non-mevalonate pathway in indole alkaloid production by *Catharanthus roseus* hairy roots. *Biotechnol Progr* **19**: 1105-1108
- Hong SB, Peebles CAM, Shanks JV, San KY, Gibson SI** (2006) Expression of the *Arabidopsis* feedback-insensitive anthranilate synthase holoenzyme and tryptophan decarboxylase genes in *Catharanthus roseus* hairy roots. *J Biotechnol* **122**: 28-38

- Hughes EH, Hong SB, Gibson SI, Shanks JV, San KY** (2004) Metabolic engineering of the indole pathway in *Catharanthus roseus* hairy roots and increased accumulation of tryptamine and serpentine. *Metab Eng* **6**: 268-276
- Höfer R, Dong L, André F, Ginglinger JF, Lugan R, Gavira C, Grec S, Lang G, Memelink J, Van Der Krol S, Bouwmeester H, Werck-Reichhart D** (2013) Geraniol hydroxylase and hydroxygeraniol oxidase activities of the CYP76 family of cytochrome P450 enzymes and potential for engineering the early (seco)iridoid pathway. *Metab Eng.* Available online 7 August 2013
- Iijima Y, Gang DR, Fridman E, Lewinsohn E, Pichersky E** (2004) Characterization of Geraniol Synthase from the Glands of Sweet Basil. *Plant Physiol* **134**: 370-379
- Ikedo H, Esaki N, Nakai S, Hashimoto K, Uesato S, Soda K, Fujita T** (1991) Acyclic Monoterpene Primary Alcohol:NADP+ Oxidoreductase of *Rauwolfia serpentina* Cells: The Key Enzyme in Biosynthesis of Monoterpene Alcohols. *J Biochem* **109**: 341-347
- Irmiler S, Schröder G, St-Pierre B, Crouch NP, Hotze M, Schmidt J, Strack D, Matern U, Schröder J** (2000) Indole alkaloid biosynthesis in *Catharanthus roseus*: new enzyme activities and identification of cytochrome P450CYP72A1 as secologanin synthase. *Plant J* **24**: 797-804
- Ito M, Honda G** (2007) Geraniol synthases from *perilla* and their taxonomical significance. *Phytochemistry* **68**: 446-453
- IUPAC-IUB Joint Commission on Biochemical Nomenclature (JCBN)** (1987) Prenol nomenclature, Recommendations 1986. *Eur J Biochem* **167**:181-184
- Jaggi M, Kumar S, Sinha AK** (2011) Overexpression of an apoplastic peroxidase gene *CrPrx* in transgenic hairy root lines of *Catharanthus roseus* *Appl Microbiol Biotechnol* **90**:1005-1016
- Jensen SR, Kirk O, Nielsen BJ** (1987) application of the Vilsmeier formylation in the synthesis of 11-<sup>13</sup>C labelled iridoids. *Tetrahedron* **4**: 1949 -1954
- Jensen SR, Kirk O, Nielsen BJ** (1989) Biosynthesis of the iridoid glucoside cornin in *Verbena officinalis*. *Phytochemistry* **28**: 97-105
- Katano N, Yamamoto H, Iio R, Inouea K** (2001) 7-Deoxyloganin 7-hydroxylase in *Lonicera japonica* cell cultures. *Phytochemistry* **58**: 53-58

- Lee-Parsons CWT, Royce AJ** (2006) Precursor limitations in methyl jasmonate-induced *Catharanthus roseus* cell cultures. *Plant Cell Rep* **25**: 607-612
- Liu DH, Ren WW, Cui LJ, Zhang LD, Sun XF, Tang KX** (2011) Enhanced accumulation of catharanthine and vindoline in *Catharanthus roseus* hairy roots by overexpression of transcriptional factor ORCA2. *Afr J Biotechnol* **10**:3260-3268
- Luijendijk TJC, Stevens LH, Verpoorte R** (1998) Purification and characterisation of strictosidine beta-D-glucosidase from *Catharanthus roseus* cell suspension cultures. *Plant Physiol Bioch* **36**: 419-425
- McGarvey DJ, Croteau R** (1995) Terpenoid Metabolism. *Plant Cell* **7**: 1015-1026
- McKnight TD, Roessner CA, Devagupta R, Scott AJ, Nessler CL** (1990) Nucleotide sequence of a cDNA encoding the vacuolar protein strictosidine synthase from *Catharanthus roseus*. *Nucleic Acids Res* **18**: 4939
- Mackenzie P, Owens IS, Burchell B, Bock KW, Bairoch A, Belanger A, Gigleux SF, Green M, Hum DW, Iyanagi T, Lancet D, Louisot P, Magdalou J, Roy Chowdhury J, Ritter, JK, Tephly, TR, Schachter H, Tephly T, Tipton, KF, Nebert, DW** (1997) The UDP glycosyltransferase gene superfamily: recommended nomenclature update based on evolutionary divergence. *Pharmacogenetics* **7**: 255-269
- Madyastha KM, Coscia CJ** (1979) Detergent-solubilized NADPH-cytochrome *c* (P-450) reductase from the higher plant, *Catharanthus roseus*. Purification and characterization. *J Biol Chem* **254**: 2419-2427
- Magnotta M, Murata J, Chen J, De Luca V** (2007) Expression of deacetylvindoline-4-O-acetyltransferase in *Catharanthus roseus* hairy roots. *Phytochemistry* **68**: 1922-1931
- Meehan TD, Coscia CJ** (1973) Hydroxylation of geraniol and nerol by a monooxygenase from *Vinca rosea*. *Biochem bioph res com* **53**: 1043
- Meijer AH, Lopes Cardoso MI, Voskuilen JT, de Waal A, Verpoorte R, Hoge JHC** (1993) Isolation and characterization of a cDNA clone from *Catharanthus roseus* encoding NADPH:cytochrome P-450 reductase, an enzyme essential for reactions catalysed by cytochrome P450 mono-oxygenases in plants. *Plant J* **4**: 47-60
- Menke FLH, Champion A, Kijne JW, Memelink J** (1999) A novel jasmonate- and elicitor-responsive element in the periwinkle secondary metabolite biosynthetic gene STR

- interacts with a jasmonate- and elicitor-inducible AP2-domain transcription factor, ORCA2. *EMBO J* **18**: 4455-4463
- Mizukami H, Nordlov H, Lee SL, Scott AI** (1979) Purification and properties of strictosidine synthetase (an enzyme condensing tryptamine and secologanin) from *Catharanthus roseus* cultured cells. *Biochemistry* **18**: 3760-3763
- Morgan JA, Shanks JV** (2000) Determination of metabolic rate-limitations by precursor feeding in *Catharanthus roseus* hairy root cultures. *J Biotechnol* **79**: 137-145
- Murata J, Bienzle D, Brandle JE, Sensen CW, De Luca V** (2006) Expressed sequence tags from Madagascar periwinkle (*Catharanthus roseus*) *FEBS Lett* **580**: 4501-4507
- Murata J, Roepke J, Gordon H, De Luca V** (2008) The leaf epidermome of *Catharanthus roseus* reveals its biochemical specialization. *Plant Cell* **20**: 524-542
- Nagatoshi M, Terasaka K, Nagatsu A, Mizukami H** (2011) Iridoid-specific Glucosyltransferase from *Gardenia jasminoides*. *J. Biol Chem* **286**: 32866-32874
- O'Connor SE, Maresh JJ** (2006) Chemistry and biology of monoterpene indole alkaloid biosynthesis. *Nat Prod Rep* **23**: 532-547
- Oudin A, Courtois M, Rideau M, Clastre M** (2007a) The iridoid pathway in *Catharanthus roseus* alkaloid biosynthesis. *Phytochem Rev* **6**: 259-276
- Oudin A, Mahroug S, Courdavault V, Hervouet N, Zelwer C, Rodriguez-Concepcion M, St-Pierre B, Burlat V** (2007b) Spatial distribution and hormonal regulation of gene products from methyl erythritol phosphate and monoterpene-secoiridoid pathways in *Catharanthus roseus*. *Plant Mol Biol* **65**: 13-30
- Pan Q, Wang Q, Yuan F, Xing S, Zhao J, Choi YH, Verpoorte R, Tian Y, Wang G, Tang K** (2012) Overexpression of ORCA3 and G10H in *Catharanthus roseus* Plants Regulated Alkaloid Biosynthesis and Metabolism Revealed by NMR-Metabolomics. *PLOS ONE* **7**: e43038
- Pauw B, Hilliou FAO, Sandonis Martin V, Chatel G, de Wolf CJV, Champion A, Pré M, van Duijn B, Kijne JW, van der Fits L, Memelink J** (2004) Zinc Finger Proteins Act as Transcriptional Repressors of Alkaloid Biosynthesis Genes in *Catharanthus roseus*. *J Biol Chem* **279**: 52940-52948

- Pasquali G, Goddijn OJM, de Waal A, Verpoorte R, Schilperoort RA, Hoge JHC, Memelink J** (1992) Coordinated regulation of two indole alkaloid biosynthetic genes from *Catharanthus roseus* by auxin and elicitors. *Plant Mol Biol* **18**: 1121-1131
- Peebles CAM, Hughes EH, Shanks JV, San KY** (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. *Metab Eng* **11**: 76-86
- Peebles CAM, Sander GW, Hughes EH, Peacock R, Shanks JV, San KY** (2011) The expression of 1-deoxy-D-xylulose synthase and geraniol-10-hydroxylase or anthranilate synthase increases terpenoidindole alkaloid accumulation in *Catharanthus roseus* hairy roots. *Metab Eng* **13**: 234-240
- Pennings EJM, Groen BW, Duine JA, Verpoorte R** (1989) Tryptophan decarboxylase from *Catharanthus roseus* is a pyridoxoquinoprotein. *FEBS lett* **255**: 97-100
- Pfützner U, Zenk MH** (1989) Homogeneous strictosidine synthase isoenzymes from cell suspension cultures of *Catharanthus roseus*. *Planta Med* 525-530
- Pomahacova B, Dusek J, Duskova J, Yazakid K, Roytrakul S, Verpoorte R** (2009) Improved accumulation of ajmalicine and tetrahydroalstonine in *Catharanthus* cells expressing an ABC transporter. *J Plant Physiol* **166**: 1405–1412
- Rai A, Smita, SS, Singh AK, Shanker K, Nagegowda DA** (2013) Heteromeric and homomeric geranyl diphosphate synthases from *catharanthusroseus* and their role in monoterpene indole alkaloid biosynthesis. *Mol Plant*. Published online March 29, 2013
- Roepke J, Salim V, Wu M, Thamm AM, Murata J, Ploss K, Boland W, De Luca V** (2010) Vincadrag components accumulate exclusively in leaf exudates of Madagascar periwinkle *P Natl Acad Sci USA* **107**:15287–15292
- Rischer H, Oresic M, Seppänen-Laakso T, Katajamaa M, Lammertyn F, Ardiles-Diaz W, Van Montagu MCE, Inze D, Oksman-Caldentey KM, Goossens A** (2006) Gene-to-metabolite networks for terpenoid indole alkaloid biosynthesis in *Catharanthus roseus* cells. *P Natl Acad Sci USA* **103**: 5614–5619
- Scott AI, Lee SL, Wan W** (1977) Indole alkaloid biosynthesis - partial-purification of ajmalicine synthetase from *Catharanthus roseus*. *Biochem Biophys Res Co* **75**: 1004-1009
- Shoji T, Ogawa T, Hashimoto T** (2008) Jasmonate-induced nicotine formation in tobacco is

mediated by tobacco COI1 and JAZ Genes. *Plant Cell Physiol* **49**: 1003-1012

- Shoji T, Kajikawa M, Hashimoto T** (2010) Clustered Transcription Factor Genes Regulate Nicotine Biosynthesis in Tobacco. *Plant Cell* **22**: 3390-3409
- Shoji T, Hashimoto T** (2011) Tobacco MYC2 Regulates Jasmonate-Inducible Nicotine Biosynthesis Genes Directly and By Way of the NIC2-Locus ERF Genes. *Plant Cell Physiol* **52**: 1117-1130
- Siberil Y, Benhamron S, Memelink J, Giglioli-Guivarc'h N, Thiersault M, Boisson B, Doireau Gantet P** (2001) *Catharanthus roseus* G-box binding factors 1 and 2 act as repressors of strictosidine synthase gene expression in cell cultures. *Plant Mol Biol* **45**: 477-488
- St-Pierre B, Vazquez-Flota FA, De Luca V** (1999) Multicellular compartmentation of *Catharanthus roseus* alkaloid biosynthesis predicts intercellular translocation of a pathway intermediate. *Plant Cell* **11**: 887-900
- Suttipanta N, Pattanaik S, Gunjan S, Xie CH, Littleton J, Yuan L** (2007) Promoter analysis of the *Catharanthus roseus* geraniol 10-hydroxylase gene involved in terpenoid indole alkaloid biosynthesis. *BBA-Gene Struct Expr* **1769**: 139-148
- Uesato S, Matsuda S, Inouye H** (1984) Mechanism for iridane skeleton formation from acyclic monoterpenes in the biosynthesis of secologanin and vindoline in *Catharanthus roseus* and *Lonicera morrowii*. *Chem Pharm Bull* **32**: 1671-1674
- Uesato S, Ogawa Y, Inouye H, Saiki K, Zenk MH** (1986a) synthesis of iridodial by cell free extracts from *Rauwolfia serpentina* cell suspension cultures. *Tetrahedron Lett* **27**: 2893-2896
- Uesato S, Kanomi S, Iida A, Inouye H, Zenk MH** (1986b) mechanism for iridane skeleton formation in the biosynthesis of secologanin and indole alkaloids in *Lonicera tatarica*, *Catharanthus roseus* and suspension cultures of *Rauwolfia serpentina*. *Phytochemistry* **25**: 839-842
- Uesato S, Ikeda H, Fujita T, Inouye H, Zenk MH** (1987) Elucidation of iridodial formation mechanism partial purification and characterization of the novel monoterpene cyclase from *Rauwolfia serpentina* cell suspension cultures. *Tetrahedron Lett* **28**: 4431-4434



- van der Fits L, Memelink J** (2000) ORCA3, a jasmonate-responsive transcriptional regulator of plant primary and secondary metabolism. *Science*. **289**: 295-297
- van der Heijden R, Jacobs DI, Snoeijer W, Hallard D, Verpoorte R** (2004) The *Catharanthus* alkaloids: Pharmacognosy and biotechnology. *Curr Med Chem* **11**: 607-628
- Van Moerkercke A, Fabris M, Pollier J, Baart GJE, Rombauts S, Hasnain G, Rischer H, Memelink J, Oksman-Caldentey KM, Goossens A** (2013) CathaCyc, a Metabolic Pathway Database Built from *Catharanthus roseus* RNA-Seq Data. *Plant Cell Physiol* **54**: 673-685
- Veau B, Courtois M, Oudin A, Chenieux JC, Rideau M, Clastre M** (2000) Cloning and expression of cDNAs encoding two enzymes of the MEP pathway in *Catharanthus roseus*. *BBA-Gene Struct Expr* **1517**: 159-163
- Vogt T, Jones P** (2000) Glycosyltransferases in plant natural product synthesis: characterization of a supergene family. *Trends Plant Sci* **5**: 380
- Wang CT, Liu H, Gao XS, Zhang HX** (2010) Overexpression of G10H and ORCA3 in the hairy roots of *Catharanthus roseus* improves catharanthine production. *Plant Cell Rep* **29**: 887-894
- Xiao M, Zhang Y, Chen X, Lee EJ, Barber CJS, Chakrabarty R, Desgagné-Penix I, Haslam TM, Kim YB, Liu E, MacNevin G, Masada-Atsumi S, Reed DW, Stout JM, Zerbe P, Zhang Y, Bohlmann J, Patrick S, Covello P, De Luca V, Page JE, Ro DK, Martin VJJ, Facchini PJ, Sensen CW** (2013) Transcriptome analysis based on next-generation sequencing of non-model plants producing specialized metabolites of biotechnological interest. *J Biotechnol* **166**:122-134
- Yahia A, Kevers C, Gaspar T, Chenieux JC, Rideau M, Creche J** (1998) Cytokinins and ethylene stimulate indole alkaloid accumulation in cell suspension cultures of *Catharanthus roseus* by two distinct mechanisms. *Plant Sci* **133**: 9-15
- Yamada Y, Koyama T, Sato F** (2011) Basic Helix-Loop-Helix transcription factors and regulation of alkaloid biosynthesis. *Plant Signal Behav* **6**: 1627-1630
- Yamamoto H, Katano N, Ooi A, Inoue K** (1999) Transformation of loganin and 7-deoxyloganin into secologanin by *Lonicera japonica* cell suspension cultures. *Phytochemistry* **50**: 417-422

- Yamamoto H, Sha M, Kitamura Y, Yamaguchi M, Katano N, Inoue K** (2002) Iridoid Biosynthesis: 7-Deoxyloganetic Acid I- O- Glucosyltransferase in Cultured *Lonicera japonica* Cells. *Plant Biotechnol* **19**: 295-301
- Yang T, Li J, Wang HX, Zeng Y** (2005) A geraniol-synthase gene from *Cinnamomum tenuipilum*. *Phytochemistry* **66**: 285-293
- Zhang H** (2008) Jasmonate-responsive transcriptional regulation in *Catharanthus roseus*. PhD thesis, Leiden University
- Zhang H, Hedhili H, Montiel G, Zhang Y, Chatel G, Pre M, Gantet P, Memelink J** (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 J* **67**: 61-71
- Zhou ML, Zhu XM, Shao JR, Wu YM, Tang YX** (2010) Transcriptional response of the catharanthine biosynthesis pathway to methyl jasmonate/nitric oxide elicitation in *Catharanthus roseus* hairy root culture. *Appl Microbiol Biotechnol* **88**: 737-750
- Zhou ML, Hou HL, Zhu XM, Shao JR, Wu YM, Tang YX** (2011) Soybean transcription factor GmMYBZ2 represses catharanthine biosynthesis in hairy roots of *Catharanthus roseus*. *Appl Microbiol Biotechnol* **91**: 1095-1105



# Characterization of the plastidial geraniol synthase from Madagascar periwinkle which initiates the monoterpene branch of the alkaloid pathway in internal phloem associated parenchyma

*Phytochemistry* (2013) 85:36-43

Andrew J. Simkin<sup>a,1,†</sup>, Karel Miettinen<sup>b,†</sup>, Patricia Claudel<sup>c</sup>, Vincent Burlat<sup>d,e</sup>, Grégory Guirimand<sup>a</sup>, Vincent Courdavault<sup>a</sup>, Nicolas Papon<sup>a</sup>, Sophie Meyer<sup>c</sup>, Stéphanie Godet<sup>a,2</sup>, Benoit St-Pierre<sup>a</sup>, Nathalie Giglioli-Guivarc'h<sup>a</sup>, Marc J.C. Fischer<sup>c</sup>, Johan Memelink<sup>b</sup>, Marc Clastre<sup>a,\*</sup>

† These authors contributed equally to this work.

<sup>(a)</sup> Université François-Rabelais, EA2106, Biomolécules et Biotechnologies Végétales, 31 avenue Monge, 37200 Tours, France

<sup>(b)</sup> Institute of Biology, Sylvius Laboratory, Sylviusweg 72, PO Box 9505, 2300 RA Leiden, The Netherlands

<sup>(c)</sup> Université de Strasbourg, INRA, Métabolisme secondaire de la vigne, Unité mixte de Recherche Santé Vigne et Qualité des vins, 28 rue de Herrlisheim, 68021 Colmar, France

<sup>(d)</sup> Université de Toulouse, UPS, UMR5546, Laboratoire de Recherche en Sciences Végétales, BP 42617, 31326 Castanet-Tolosan, France

<sup>(e)</sup> CNRS, UMR5546, BP 42617, Castanet-Tolosan, France

<sup>1</sup> Current address: School of Biological Sciences, Wivenhoe Park, University of Essex, Colchester, CO4 3SQ, United Kingdom

<sup>2</sup> Current address : Agrocampus Ouest, Centre d'Angers, Institut National d'Horticulture et de Paysage, 2 rue André Le Nôtre, 49045 Angers cedex 01, France

\* Corresponding author. Tel.: +33 247 36 72 13; fax: +33 247 27 66 60.

E-mail address: [marc.clastre@univ-tours.fr](mailto:marc.clastre@univ-tours.fr)

## Abstract

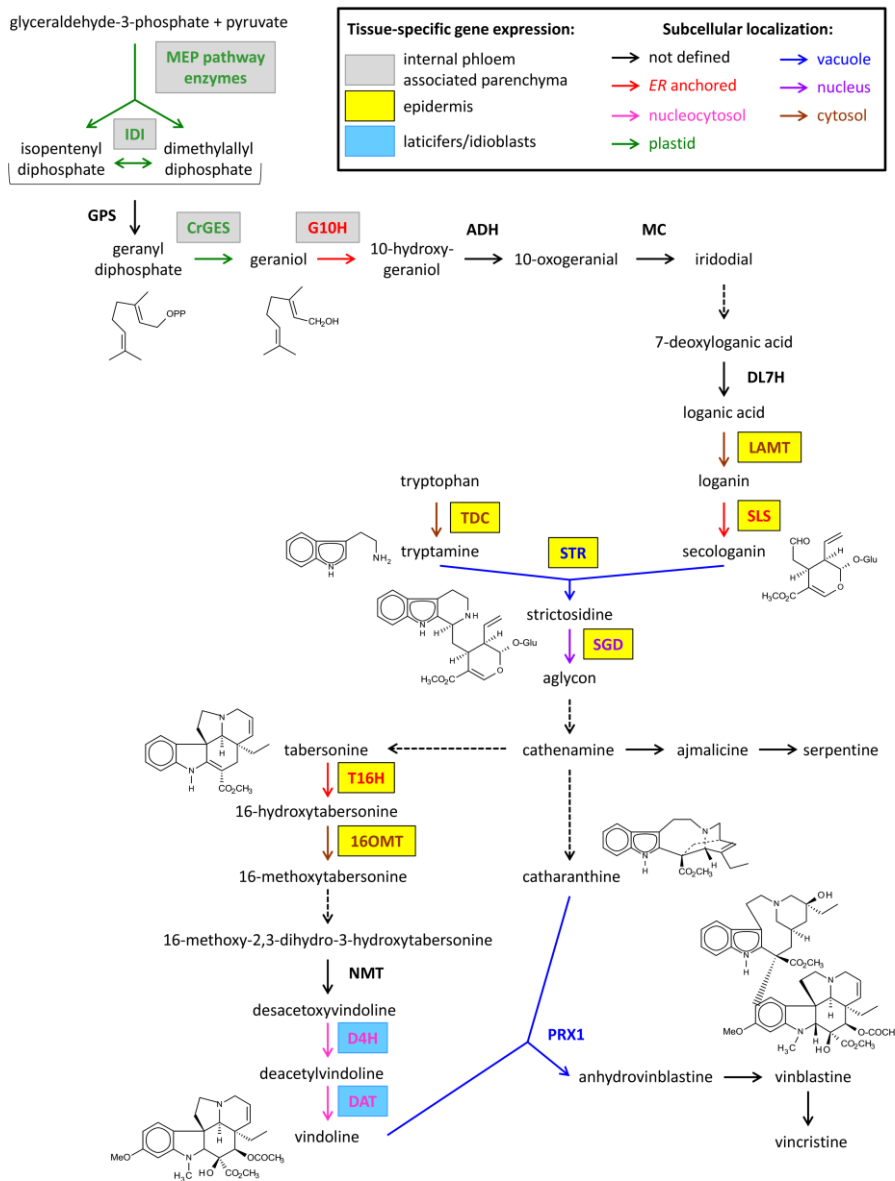
Madagascar periwinkle (*Catharanthus roseus* [L.] G. Don, Apocynaceae) produces monoterpene indole alkaloids (MIAs), secondary metabolites of high interest due to their therapeutic value. A key step in the biosynthesis is the generation of geraniol from geranyl diphosphate (GPP) in the monoterpene branch of the MIA pathway. Here we report on the cloning and functional characterization of *C. roseus* geraniol synthase (CrGES). The full-length CrGES was over-expressed in *Escherichia coli* and the purified recombinant protein catalyzed the conversion of GPP into geraniol with a  $K_m$  value of 58.5  $\mu\text{M}$  for GPP. *In vivo* CrGES activity was evaluated by heterologous expression in a *Saccharomyces cerevisiae* strain mutated in the farnesyl diphosphate synthase gene. Analysis of culture extracts by gas chromatography-mass spectrometry confirmed the excretion of geraniol into the growth medium. Transient transformation of *C. roseus* cells with a Yellow Fluorescent Protein-fusion construct revealed that CrGES is localized in plastid stroma and stromules. In aerial plant organs, RNA *in situ* hybridization showed specific labelling of CrGES transcripts in the internal phloem associated parenchyma as observed for other characterized genes involved in the early steps of MIA biosynthesis. Finally, when cultures of *Catharanthus* cells were treated with the alkaloid-inducing hormone methyl jasmonate, an increase in CrGES transcript levels was observed. This observation coupled with the tissue-specific expression and the subcellular compartmentalization support the idea that CrGES initiates the monoterpene branch of the MIA biosynthetic pathway.

## 1. Introduction

Madagascar periwinkle (*Catharanthus roseus* [L.] G. Don, Apocynaceae) is a pantropical medicinal plant which synthesizes a wide range of complex secondary metabolites known as monoterpene indole alkaloids (MIAs). Several of them are valuable therapeutic compounds, including monomers such as ajmalicine and serpentine used in the treatment of circulatory diseases and anxiety, and heterodimers such as vinblastine and vincristine known as powerful anticancer drugs (van der Heijden et al., 2004). Due to the pharmacological importance of these compounds, the MIA metabolic pathway has been highly studied in the whole plant and in cell suspension culture systems (Zhou et al., 2011). MIAs originate from two convergent pathways. Tryptamine (provided by the indole pathway through decarboxylation of L-tryptophan) and secologanin (provided by the monoterpenoid pathway also known as the iridoid pathway) are condensed into strictosidine, the precursor of all other MIAs (van der Heijden et al., 2004) (Supplementary Fig. S1).

Studies of fluxes in the pathways leading to the formation of MIAs by precursor feeding highlighted that the monoterpenoid branch is limiting for the biosynthesis of alkaloids in cell and tissue cultures of *C. roseus* (Oudin et al., 2007a). A key step in the formation of MIAs is the biosynthesis of the monoterpenoid geraniol from geranyl diphosphate (GPP).

Geraniol feeding of *C. roseus* cell (Lee-Parsons and Royce, 2006) and hairy root cultures (Morgan and Shanks, 2000) resulted in an increase in the formation of the MIAs ajmalicine and tabersonine, respectively, suggesting that the formation of geraniol is a critical step in MIA biosynthesis. To date, periwinkle geraniol synthase (GES) has not been characterized at the molecular level. The present work focuses on the cloning and functional characterisation of *C. roseus* GES, the study of the corresponding gene expression in response to methyl jasmonate, the *in situ* localization of GES mRNA and the subcellular localization of the enzyme.



**Supplementary Fig. S1.** Tissue-specific gene expression and subcellular localization of the enzymes of the monoterpene indole alkaloid pathway of *C. roseus*. Broken arrows represent uncharacterized reactions. The localization of CrGES was determined in the present study. 16OMT: 16-hydroxytabersonine-16-*O*-methyltransferase; ADH: acyclic monoterpene primary alcohol dehydrogenase; CrGES: geraniol synthase; D4H: deacetoxyvindoline-4-hydroxylase; DAT: deacetylvindoline 4-*O*-acetyltransferase; DL7H: 7-deoxyloganin 7-hydroxylase; G10H: geraniol 10-hydroxylase; GPS: geranyl diphosphate synthase; IDI: isopentenyl diphosphate isomerase; LAMT: S-adenosyl-L-methionine:loganic acid methyl transferase; MC: monoterpene cyclase; MEP pathway: methyl erythritol 4-phosphate pathway; NMT: 16-methoxy-2,3-dihydro-3-hydroxytabersonine N-methyltransferase; PRX1: peroxidase; SGD: strictosidine  $\beta$ -glucosidase; SLS: secologanin synthase; STR: strictosidine synthase; T16H: tabersonine 16-hydroxylase; TDC: tryptophan decarboxylase. For references see text.

## 2. Results and discussion

### 2.1. Isolation of the *C. roseus* geraniol synthase (CrGES) full-length cDNA

A partial sequence of 554 bp (Genbank ID: EG558318) displaying similarities with known geraniol synthases was identified in a *Catharanthus roseus* EST database. The full-length coding sequence for CrGES (*C. roseus* geraniol synthase) was recovered by 3' RACE, PhageWalker and GenomeWalker protocols as described in the experimental method section. The CrGES sequence has been deposited at NCBI under Genbank ID: JN882024.

The CrGES open reading frame of 1770 bp encodes a protein of 589 amino acids in length with a calculated mass of 67.7 kDa. The CrGES protein contains the highly conserved aspartate-rich motif DDxxD (positions 343-347) and the less conserved NSE/DTE motif (positions 485-496) with the consensus sequence (L,V)(V,L,A)-(N,D)D(L,I,V)x(S,T)xxxE (Supplementary Fig. S2). Both motifs are found in several terpene synthases and are involved in the fixation of the diphosphate substrate (Christianson, 2006). Amino acid sequence comparison revealed that CrGES showed similarities with terpene synthases. A high degree of similarity was found with previously characterized geraniol synthases (Supplementary Fig. S2). The highest similarity (64% of identity) was found with the geraniol synthase from *Lippia dulcis* (Genbank ID: GU136162; Yang et al., 2011). Furthermore, CrGES possessed 59% of identity with the *Ocimum basilicum* ortholog (Genbank ID: AY362553, Iijima et al., 2004), 36% with the *Cinnamomum tenuipilum* GES (Genbank ID: AJ457070, Yang et al., 2005) and 33% with the GES from *Perilla frutescens* (Genbank ID: DQ234300, Ito et al., 2007) and *P. citriodora* (Genbank ID: DQ088667, Ito et al., 2007).

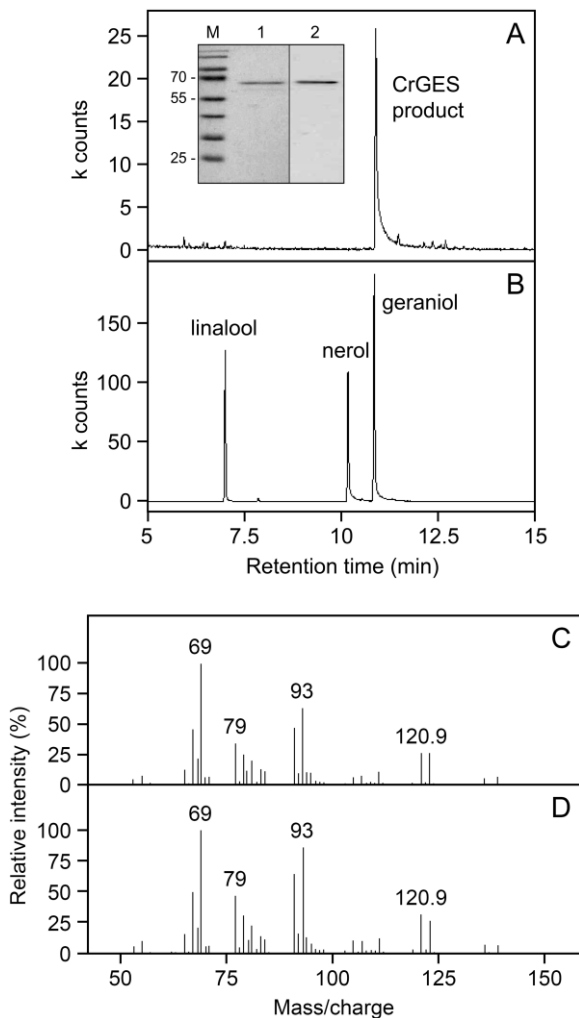




## 2.2. Functional characterization of the purified recombinant CrGES

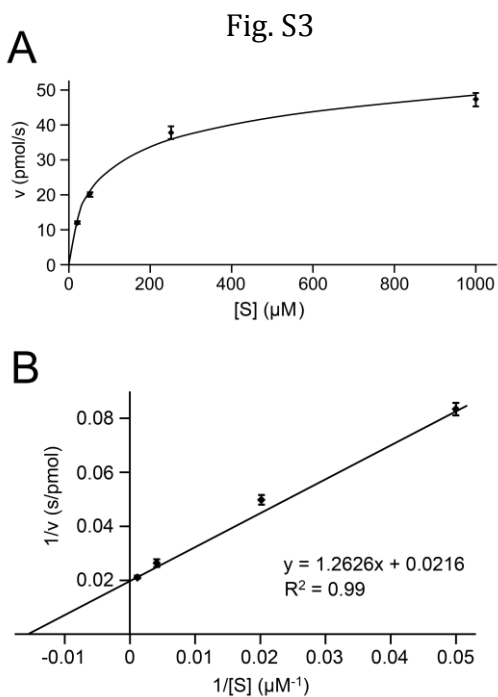
The close homology with previously identified geraniol synthases suggested that CrGES catalyzes the conversion of GPP into geraniol. Functional expression was thus carried out to investigate the catalytic activity. The complete *CrGES* open reading frame including the putative N-terminal plastidial targeting peptide flanked by a C-terminal His tag and an N-terminal Strep tag was expressed in *Escherichia coli* and purified by sequential Ni-NTA and Strep-tactin affinity chromatography. Analysis of the recombinant protein by SDS-PAGE and Coomassie Brilliant Blue staining or western blotting and immunoprobings with anti-His antibodies showed the presence of one major band (Fig. 1A). Analysis of reaction products formed after incubation of the protein with GPP showed the presence of a single peak (Fig. 1A), which was identified as geraniol based on the retention time (Fig. 1B) and on the mass spectrum (Fig. 1C and D). A control reaction with boiled protein did not result in detectable products (data not shown). Under the reaction conditions employed, the protein catalyzed the conversion of GPP to geraniol with a  $K_m$  value of 58.5  $\mu\text{M}$  for GPP (Supplementary Fig. S3), which is close to the apparent  $K_m$  of 55.8  $\mu\text{M}$  reported for GES from the evergreen camphor tree *C. tenuipilum* assayed under similar conditions (Yang et al., 2005).

Fig. 1



**Fig. 1.** Analysis of the CrGES-catalyzed reaction product. GC analysis of reaction products with the substrate geranyl diphosphate and recombinant CrGES (A) and of authentic geraniol, nerol and linalool (B). Mass spectra of authentic geraniol (C) and the peak from A (D). The inset in (A) shows the analysis of recombinant CrGES protein. The protein was separated by 10% SDS-PAGE and either stained with Coomassie Brilliant Blue (lane 1) or visualized after western blotting using anti-His antibodies (lane 2). Sizes of relevant marker (M) bands are indicated in kDa.

**Supplementary Fig. S3.** Michaelis-Menten (A) and Lineweaver-Burk (B) plots of the reaction rate with CrGES and GPP. The data are means  $\pm$  standard errors of three replicates.



### 2.3. Expression of CrGES in *Saccharomyces cerevisiae* FPS mutants

Farnesyl diphosphate synthase (FPS) catalyses the formation of the C<sub>15</sub> product farnesyl diphosphate (FPP) by two sequential reactions: the initial condensation of isopentenyl diphosphate (IPP) and dimethylallyl diphosphate leading to GPP followed by the condensation of GPP with a second molecule of IPP producing FPP. It is generally accepted that FPS produces only FPP and that no GPP is released from the catalytic site of this enzyme. However, yeast mutant strains containing a mutated FPS, with lower FPS-specific activity, also produced GPP available for the synthesis of geraniol and other related monoterpenes. These compounds originate from GPP and result from endogenous yeast enzymes activities and/or chemical instability of some monoterpenols (Fischer et al., 2011). The over-expression of the heterologous sweet basil (*O. basilicum*) GES in an FPS yeast mutant strain resulted in a marked increase in the production of geraniol (Oswald et al., 2007). Optimisation of the FPS mutation (mutation K197G) coupled to GES expression strongly enhanced (10-20 fold) monoterpene production and led to the excretion of up to 5 mg/l geraniol in the culture medium (Fischer et al., 2011).

We used the yeast strain with the FPS-K197G mutation to evaluate CrGES activity *in vivo*. A truncated coding sequence of *CrGES*, devoid of the plastidial targeting peptide, was cloned into the galactose-inducible expression cassette of the pYES2 vector and introduced into the K197G strain. When transformed with the empty vector, the yeast strain produced a low quantity of various monoterpenes where the major component was geraniol (Table 1). Expression of CrGES greatly increased monoterpenoid production (8.4 fold for geraniol) with the major product geraniol reaching a concentration of approximately 3.3 mg/l in the culture medium. Similar observations were made with the GES from sweet basil, which produced only geraniol when assayed as a recombinant protein *in vitro* (Iijima et al., 2004), whereas expression in the K197G yeast strain resulted in an increase of several monoterpenes with geraniol as the major product (Fischer et al., 2011).

	[Linalool] μg/l	[α-terpineol] μg/l	[Citronellol] μg/l	[Neral] μg/l	[Geranial] μg/l	[Nerol] μg/l	[Geraniol] μg/l
pYES2	104 ± 22	6 ± 3	11 ± 1	ND	ND	9 ± 7	394 ± 27
pYES2+Cr GES	257 ± 36	18 ± 3	69 ± 14	31 ± 1	54 ± 3	14 ± 3	3.310 ± 221

**Table 1.** Monoterpene content in the culture medium (μg/l) of a yeast strain with a FPS-K197G mutation transformed with the empty vector pYES2 or with the vector pYES2 containing the *CrGES* coding sequence without the plastidial targeting peptide. Terpenoids were extracted from minimal medium at stationary growth phase. Results are the mean of 5 independent transformants ± standard deviation. ND: Not Detected.

In summary, the results on the heterologous expression of *CrGES* in yeast mutants are in agreement with the product specificity of the recombinant enzyme and demonstrated that, *in vivo*, *CrGES* displayed mainly geraniol synthase activity.

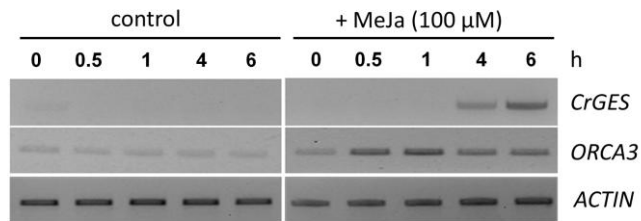
#### 2.4. *CrGES* transcript accumulation

MIA accumulation is induced in *C. roseus* cell cultures growing in a 2,4-D-free medium and can be further increased by the addition of methyl jasmonate (MeJa) (Gantet et al., 1998). This phytohormone stimulates the biosynthesis of alkaloids by inducing the expression of the ESMB genes (encoding enzymes catalyzing Early Steps in Monoterpenoid Biosynthesis) such as those of the methyl erythritol 4-phosphate (MEP) pathway and the geraniol 10-hydroxylase (*G10H*) gene (Collu et al., 2001; Oudin et al., 2007b) (Supplementary Fig. S1). MeJa also induces the expression of the gene encoding the transcription factor *ORCA3* which controls many MIA biosynthetic genes (van der Fits and Memelink, 2000; 2001).

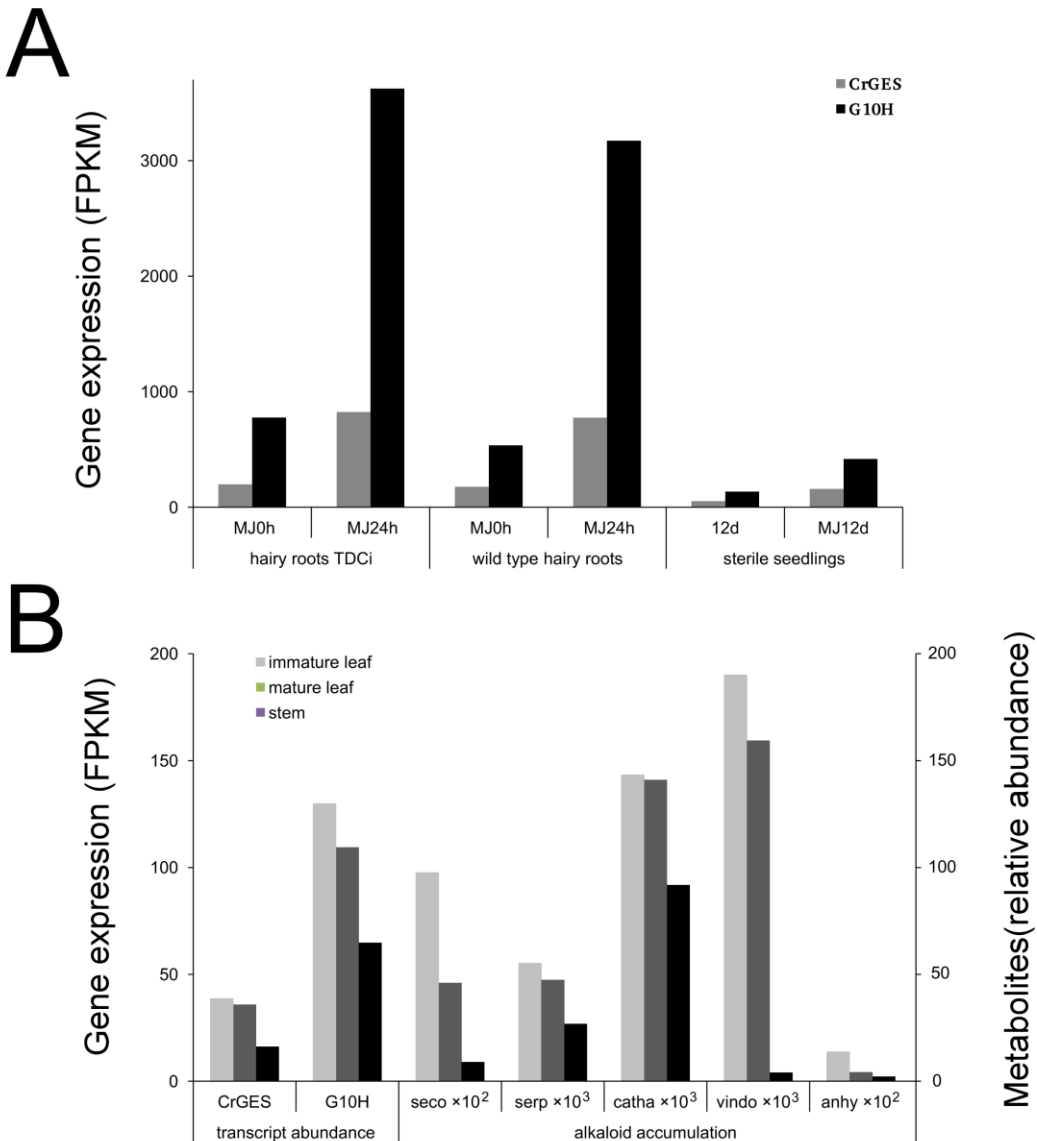
Analysis of transcript levels showed that *CrGES* gene expression was undetectable in untreated cells but strongly induced by MeJa 4 hours post-treatment (Fig. 2). The expression of the *ORCA3* gene, used as a positive control, was strongly increased 30 min post-treatment as previously described (van der Fits and Memelink, 2001). The slower kinetics of gene induction of *CrGES* upon addition of MeJa correlated well with those of *G10H* whose

transcripts also accumulated 4 hours after treatment of periwinkle cell suspensions (Collu et al., 2001).

These results are corroborated by exploiting transcriptomic data from MPGR (<http://medicinalplantgenomics.msu.edu/>) showing that *CrGES* and *G10H* were both strongly induced by MeJa in hairy roots and young seedlings (Supplementary Fig. S4A). When coupled with the periwinkle metabolome from MPGR, the transcriptomic data revealed that *CrGES* and *G10H* transcript abundance had a similar pattern of distribution in aerial organs (immature leaves, matures leaves and stem, respectively) with those of secologanin and some MIA (Supplementary Fig. S4B). Overall, the comparison of the expression values of *CrGES* and *G10H* within the 22 available periwinkle samples from MPGR revealed that both genes had a very similar expression pattern with a very high Pearson Product-Moment Correlation Coefficient (0.977).



**Fig. 2.** Time course of *CrGES* mRNA levels in *C. roseus* cell suspension cultures. The cells were not treated or treated with 100 μM MeJa, harvested at the times (h) indicated above the lanes and then analyzed by RT-PCR. The *ACTIN* gene was used as a constitutive control while the *ORCA3* gene was used as a positive control for the MeJa treatment.



**Supplementary Fig. S4.** Co-expression of *CrGES* and *G10H* and correlation with alkaloid accumulation. (A) Time course of *CrGES* and *G10H* gene expression (FPKM: fragments per kilobase per transcript per million mapped reads) in two types of hairy roots (TDCi and wild type) and in sterile seedlings treated with MeJa. MJ0h: control at time 0 h; MJ24h: 24 h after 250  $\mu$ M MeJa treatment; 12d: 12 days old seedlings without MeJa treatment; MJ12d: seedlings treated with 6  $\mu$ M MeJa for 12 days. (B) Distribution patterns of *CrGES* and *G10H* gene expression (FPKM) and alkaloid relative abundance in aerial organs (immature leaf, mature leaf and stem). Anhy: anhydrovinblastine; catha: catharanthine; seco: secologanin; serp: serpentine; vindo: vindoline. The data were retrieved from the Medicinal Plant Genomics Resource (MPGR, <http://medicinalplantgenomics.msu.edu>).

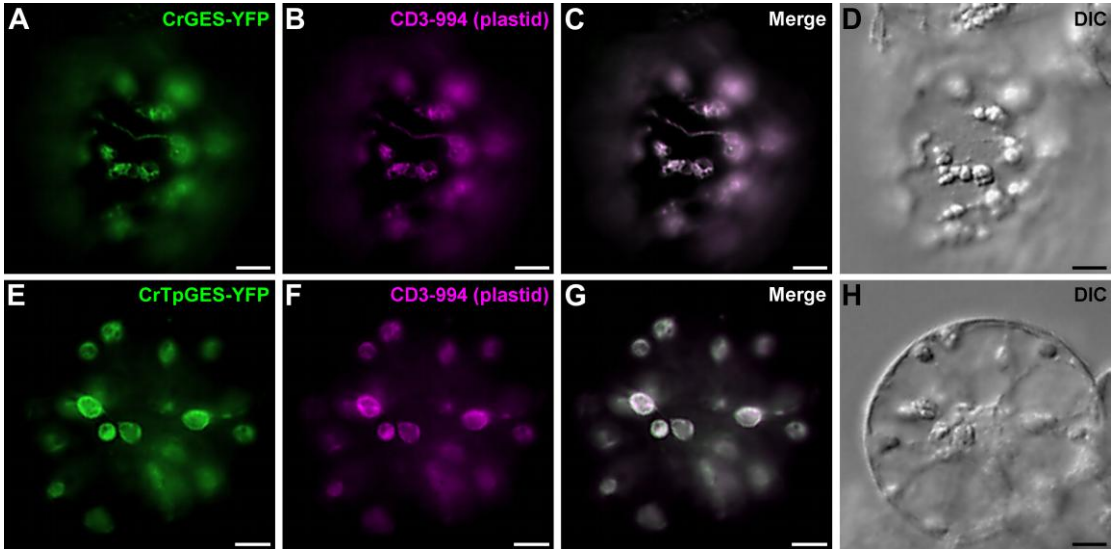
## 2.5. Subcellular localization of the CrGES protein

It is generally accepted that monoterpene synthases that use GPP as a substrate are localized in plastids (Chen et al., 2011). We used ChloroP and PSORT programs to assign confidence values to the presence of an N-terminal plastid transit peptide (Tp) in CrGES. This Tp was predicted by the ChloroP program to be 43 amino-acids in length. To confirm the predicted plastid localization, we determined the subcellular distribution of CrGES through a biolistic-mediated transient expression approach in *C. roseus* cells according to Guirimand et al. (2009). In transiently transformed *C. roseus* cells, the full-length CrGES-YFP fusion protein displayed a pattern of fluorescence which co-localized perfectly with those of the plastidial marker. Fig 3A-C revealed that the full-length CrGES was imported in the plastid stroma but also in stromules (Supplementary Fig. S5), which are thin envelope-bound extensions of the stroma (Natesan et al., 2005). The CrTpGES-YFP fusion protein (corresponding to the first 43 amino acids of CrGES fused to YFP) was also shown to be targeted to the plastid compartment (Fig. 3E-G). This result indicates that the first 43 amino acids are sufficient to direct efficient plastidial import.

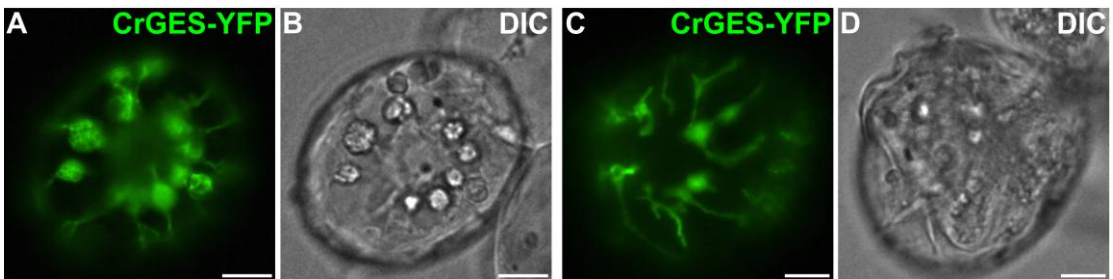
The present work constitutes the first demonstration of the presence of a monoterpene synthase within stromules since previous studies have only reported the localization of this family of enzymes in the plastid stroma (Turner et al., 1999; Nagegowda et al., 2008; Lin et al., 2008). The subcellular localization of CrGES in stroma and stromules is also consistent with our previous work showing that the earlier enzymatic steps leading to the formation of geraniol, including the MEP pathway enzyme hydroxymethylbutenyl 4-diphosphate synthase (Oudin et al., 2007; Guirimand et al., 2009) and a long isoform of isopentenyl diphosphate isomerase (IDI) (Guirimand et al., 2012), have a similar compartmentalization in *C. roseus* (Supplementary Fig. S1). Furthermore, it has been demonstrated that stromules are in close association with the endoplasmic reticulum (ER) suggesting that an exchange of metabolites between the two organelles occurs *in vivo* (Guirimand et al., 2009; Schattat et al., 2011). Such a scenario fits well with the next step of the monoterpene branch consisting of the conversion of geraniol into 10-hydroxygeraniol catalyzed by G10H (Collu et al., 2001). This enzyme is a cytochrome P450 anchored to the ER membrane with its catalytic domain



likely exposed to the cytosol and in close vicinity with stromules (Guirimand et al., 2009). Thus, it could be hypothesized that stromules facilitate the export of geraniol into the cytosol and thereby its conversion into 10-hydroxygeraniol by the ER-anchored G10H.



**Fig. 3.** Subcellular localisation of CrGES. *C. roseus* cells were transiently co-transformed with the plasmid expressing CrGES-YFP (A) or CrTpGES-YFP (E) and the “plastid”-CFP marker CD3-994 (B, F). Co-localization of the two fluorescence signals appeared in white (C, G) when merging the two individual (green/magenta) false colour images. The morphology (D, H) is observed with differential interference contrast (DIC). Bars correspond to 10  $\mu\text{m}$ .



**Supplementary Fig. S5.** CrGES subcellular compartmentation highlighting localization in stromules. *C. roseus* cells were transiently transformed with the plasmid expressing CrGES-YFP (A, C). The morphology (B, D) is observed with differential interference contrast (DIC). Bars correspond to 10  $\mu\text{m}$ .

## 2.6. Cell specific gene expression of *CrGES*

Considerable progress has been made in the past decade in the understanding of the cellular architecture of the MIA biosynthetic pathway revealing its multicellular compartmentalization in aerial organs of *C. roseus* (Supplementary Fig. S1). The late steps, including desacetoxyvindoline-4-hydroxylase and deacetylvindoline-4-O-acetyltransferase, have been localized to laticifers/idioblasts which are specialized alkaloid-accumulating cells (St-Pierre et al., 1999). The intermediate part of the pathway, from the step catalysed by loganic acid *O*-methyltransferase until the step catalyzed by 16-hydroxytabersonine 16-*O*-methyltransferase, occurs in epidermal cells (St-Pierre et al., 1999; Irmeler et al., 2000; Burlat et al., 2004; Murata and De Luca, 2005; Murata et al., 2008; Guirimand et al., 2011a; Guirimand et al., 2011b). The early steps (ESMB, Oudin et al., 2007b), including the MEP pathway, the IDI reaction and the synthesis of 10-hydroxygeraniol by G10H, are present in the internal phloem associated parenchyma (IPAP) cells (Burlat et al., 2004; Oudin et al., 2007; Guirimand et al., 2009; Guirimand et al., 2012).

The cellular distribution of *CrGES* transcripts was investigated by RNA *in situ* hybridization performed on young developing leaves. Using the antisense probe, *CrGES* transcripts were specifically detected in the adaxial part of the leaf vascular region (Fig. 4A, B), corresponding to the IPAP cells as confirmed by the co-localization with *G10H* mRNA (Fig. 4E, F; Burlat et al., 2004). No signal was observed with the *CrGES* sense probe, used as a negative control (Fig. 4C, D). This specific labeling of *CrGES* and *G10H* in IPAP cells was also observed in other aerial organs, such as the cotyledons of young seedlings (Supplementary Fig. S6) as well as carpels and stamens (Supplementary Fig. S7). These results clearly indicate that the two consecutive enzymatic steps involving *CrGES* and *G10H* occur in IPAP cells.

Fig 4

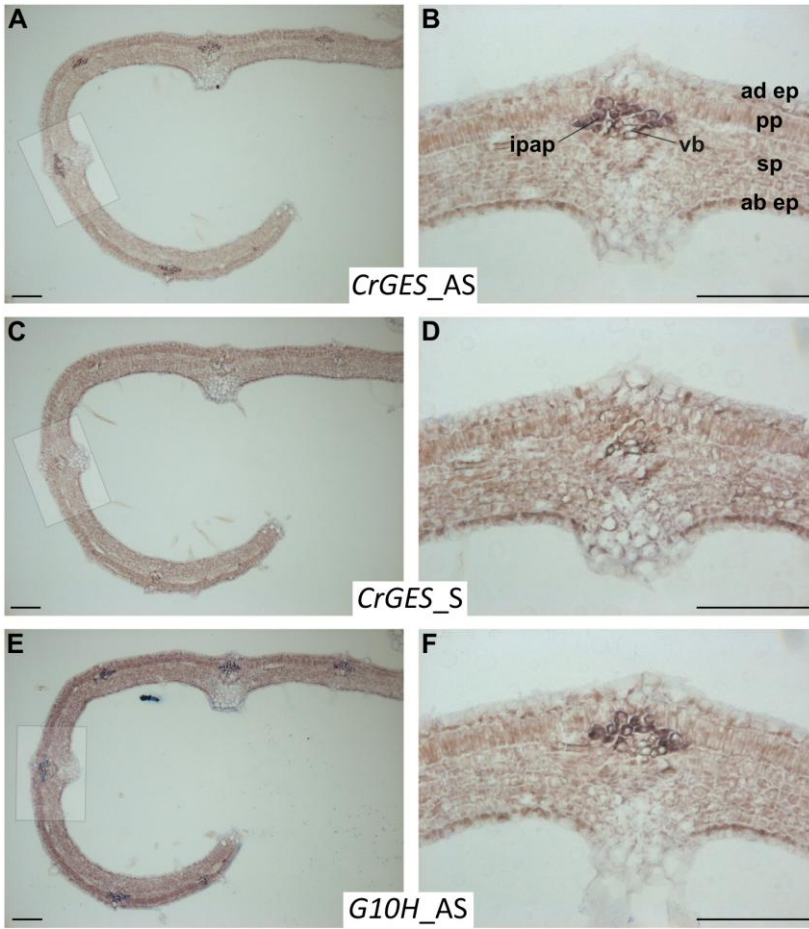
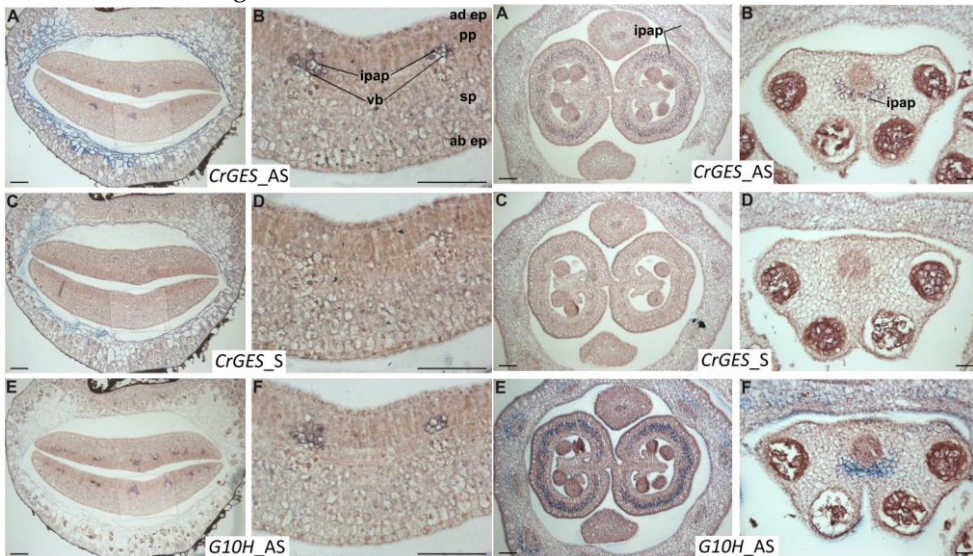


Fig S6

Fig S7



**Fig. 4.** *In situ* localization of *CrGES* mRNA in young leaves of *C. roseus*. Paraffin-embedded serial longitudinal sections of young leaves were hybridized with digoxigenin-labelled riboprobes, which were subsequently localized with antidigoxigenin-alkaline phosphatase conjugates followed by nitro blue tetrazolium chloride (NBT)/5-bromo 4-chloro-3-indolyl phosphate (BCIP) color development. The antisense probe *CrGES\_AS* was used for RNA labeling (A, B) and a control hybridization was performed with the sense probe *CrGES\_S* (C, D). Hybridization of the *G10H* transcripts with the antisense probe *G10H\_AS* was used as a positive control (E, F). ab ep: abaxial epidermis; ad ep: adaxial epidermis; ipap: internal parenchyma associated phloem; pp: palisade parenchyma; sp: spongy parenchyma; vb: vascular bundle. Bars correspond to 100  $\mu$ m.

**Supplementary Fig. S6.** *In situ* localization of *CrGES* mRNA in cotyledons of young seedlings of *C. roseus*. Paraffin-embedded serial longitudinal sections of emerging cotyledons were hybridized with digoxigenin-labelled riboprobes, which were subsequently localized with antidigoxigenin-alkaline phosphatase conjugates followed by nitro blue tetrazolium chloride (NBT)/5-bromo 4-chloro-3-indolyl phosphate (BCIP) color development. The antisense probe *CrGES\_AS* was used for RNA labeling (A, B) and the control hybridization was performed with sense probe *CrGES\_S* (C, D). Hybridization of *G10H* transcripts with the antisense probe *G10H\_AS* was used as a positive control (E, F). ab ep: abaxial epidermis; ad ep: adaxial epidermis; ipap: internal parenchyma associated phloem; pp: palisade parenchyma; sp: spongy parenchyma; vb: vascular bundle. Bars correspond to 100  $\mu$ m.

**Supplementary Fig. S7.** *In situ* localization of *CrGES* mRNA in carpels and stamens of *C. roseus*. Paraffin-embedded serial transversal sections of carpels (A, C, E) and stamens (B, D, F) were hybridized with digoxigenin-labelled riboprobes, which were subsequently localized with antidigoxigenin-alkaline phosphatase conjugates followed by nitro blue tetrazolium chloride (NBT)/5-bromo 4-chloro-3-indolyl phosphate (BCIP) color development. The antisense probe *CrGES\_AS* was used for RNA labeling (A, B) and the control hybridization was performed with sense probe *CrGES\_S* (C, D). Hybridization of *G10H* transcripts with the antisense probe *G10H\_AS* was used as a positive control (E, F). Lines show labeling within the internal parenchyma associated phloem (ipap). Bars correspond to 100  $\mu$ m.

### 3. Conclusions

In this report we describe the cloning and functional characterization of the periwinkle cDNA *CrGES* encoding geraniol synthase. The recombinant enzyme produced in *E. coli* catalysed the *in vitro* conversion of GPP into geraniol only. In the yeast FPS mutant K197G expressing CrGES, geraniol was the main enzymatic product formed. The plastidial localization of CrGES suggests that the availability of its product geraniol to the next enzyme G10H could be facilitated through stromule/ER interactions. Furthermore, the induction of *CrGES* expression by MeJa, a phytohormone which stimulates alkaloid accumulation in *C. roseus* cell suspensions, the positive correlation between *CrGES* and *G10H* expression in different tissues of periwinkle and the *in situ* co-localization of *CrGES* mRNA with the ESMB transcripts in IPAP cells of aerial organs support the view that CrGES initiates the monoterpenoid branch of the MIA pathway in *C. roseus*.

## 4. Experimental

### 4.1. Plant material and growth conditions

Periwinkle (*Catharanthus roseus* [L.] G. Don, Apocynaceae) cell suspensions (line C20D) were maintained on a 7-day-growth cycle in the B5 medium of Gamborg et al. (1968) supplemented with 58 mM sucrose and 4.5  $\mu$ M 2,4-dichlorophenoxyacetic acid (2,4-D). For experimental purposes, 7-day-old cells were subcultured in the same medium depleted of 2,4-D. Methyl jasmonate (MeJa) (Sigma-Aldrich, <http://www.sigmaaldrich.com>) at a final concentration of 100  $\mu$ M, was added at the fourth day of culture. Accumulation of MIAs is induced in *C. roseus* cells growing in a 2,4-D-free medium and can be further increased by the addition of MeJa (Gantet et al., 1998). Mature *C. roseus* plants, grown in a greenhouse, were used for *in situ* hybridization studies.

### 4.2. RNA extraction and cDNA synthesis

Frozen cells were ground to a fine powder in liquid nitrogen. Total RNAs were extracted with RNAsasy Plant mini kit (Qiagen, <http://www.qiagen.com>). The cDNAs were synthesized from 2  $\mu$ g RNA using the oligo-dT AP primer (Supplementary Table S1) according to the protocol in the Superscript II Reverse Transcriptase kit (Invitrogen, <http://www.invitrogen.com>) in a final volume of 20  $\mu$ l.

### 4.3. Cloning of the CrGES cDNA

Using the BLAST algorithm and the NCBI database, we recovered a *C. roseus* EST (Genbank ID: EG558318) of 554 bp encoding an amino acid sequence which showed similarity with an internal fragment of known geraniol synthases. The EST sequence was used to design primers for the isolation of the 5' and 3' ends of the coding sequence of the cDNA.

The 3'-end of the cDNA was recovered by 3' RACE according to the manufacturer's instructions (Invitrogen). Reverse transcription was performed with the primer AP followed by a first PCR with primers GOS4 and AUAP (from the 3' RACE kit) and a second nested PCR with primers GOS6 and AUAP (Supplementary Table S1). The resulting PCR fragment was cloned into the vector pGEM-T Easy (Promega, <http://www.promega.com>) and sequenced.

The 5'-end of the cDNA was recovered by Nested PhageWalker using a *C. roseus* cDNA library (Simkin AJ, unpublished data) constructed with the ZAP Express System (Stratagene, <http://www.stratagene.com>). The first PCR reaction was carried out with the reverse primer PWGES1 and the forward primer corresponding to the phage plasmid T3 primer (Supplementary Table S1). One  $\mu$ l of this PCR reaction was used for a second PCR reaction with primer T3 and the nested primer PWGES2 (Supplementary Table S1). The resulting amplicon was cloned in pGEM-T Easy and sequenced. The remaining 5'-end 55 bp including the start ATG codon was cloned using the GenomeWalker kit (Clontech, <http://www.clontech.com>). PCR amplification of genomic DNA extracted from periwinkle leaf was done with primers GWGES4 and AP1 (from the GenomeWalker kit) followed by nested PCR using primers GWGES5 and AP2 (from the GenomeWalker kit) (Supplementary Table S1). This generated a DNA fragment, which was cloned in pGEM-T Easy and sequenced. Following *in-silico* assembly, the full-length open reading frame was identified.

#### 4.4. CrGES expression in *Escherichia coli* and enzyme purification

The CrGES open reading frame was PCR amplified with the primers EcGES1 and EcGES2 (Supplementary Table S1) using a pACTII cDNA library of *C. roseus* MP183L cells treated with yeast extract as template and cloned into pASK-IBA45plus (IBA Biotagnology, <http://www.iba-go.com>) with BamHI/PstI.

Double Strep/His-tagged CrGES protein was expressed from plasmid pASK-IBA45plus in *Escherichia coli* strain BL21 (DE3) pLysS and purified by sequential Ni-NTA agarose (Qiagen) and Strep-Tactin sepharose (IBA Biotagnology) chromatography. For quality analysis the recombinant protein was run on a 10% (w/v) SDS-PAA gel, transferred to Protran nitrocellulose (Whatman, <http://www.whatman.com>) by semidry electroblotting,

and the western blot was probed with mouse monoclonal anti-His horseradish peroxidase (HRP) conjugate antibodies (5Prime, <http://www.5prime.com>). Antibody binding was detected by incubation in 250  $\mu$ M sodium luminol, 0.1 M Tris-HCl, pH 8.6, 3 mM H<sub>2</sub>O<sub>2</sub>, 67  $\mu$ M p-coumaric acid and exposure to X-ray film.

#### 4.5. Enzymatic assays

Enzymatic assays were performed as 1 ml reactions containing 100 mM HEPES-KOH pH 7, 1 mM MgCl<sub>2</sub>, 100  $\mu$ M MnCl<sub>2</sub>, GPP and 4  $\mu$ g of purified enzyme. For determination of the K<sub>m</sub> value reactions contained 20, 50, 250 or 1000  $\mu$ M GPP. Reaction mixtures were overlaid with 1 ml hexane and incubated at 32 °C for 30 min. For quantitative analysis 5  $\mu$ g of citronellol was added as an internal standard after incubation. The contents of the tubes were thoroughly mixed by vortexing and kept on ice for 10 min. The tubes were centrifuged at 4000 $\times$ g for 10 min, and the supernatant hexane phase was collected. The extraction was repeated with 1 ml hexane. The pooled hexane phase was dehydrated by passing it through a layer of anhydrous Na<sub>2</sub>SO<sub>4</sub> for gas chromatography-mass spectrometry (GC-MS) and GC-flame ionization detector (GC-FID) analysis. Amounts of geraniol formed in enzyme assays were calculated from the resultant GC-FID integral using the relative response factor with respect to the citronellol internal standard. A Lineweaver-Burk plot was constructed to obtain the K<sub>m</sub> value.

#### 4.6. GC-MS and GC-FID analysis of the enzymatic product from CrGES expressed in *E. coli*

For compound identification the CrGES reaction products were subjected to capillary GC-MS, using a Varian Saturn 2000 ion trap mass spectrometer run in electron ionization mode (70 eV). The hexane extract of the reaction mixture was separated on a Varian 3800 gas chromatograph equipped with a DB-5 capillary column (30 m  $\times$  0.25 mm, film thickness of 0.25  $\mu$ m) (Agilent Technologies, <http://www.home.agilent.com>) using nitrogen as carrier gas at a flow rate of 1.2 ml/min. The separation conditions were: injection split ratio 1:20, injection volume 1  $\mu$ l, injector temperature 230°C, initial oven temperature 60°C, then linear gradient to 150°C at a rate of 5°C/min followed by a linear gradient to 240°C at a rate of



20°C/min. The transfer line temperature was 275°C, the ion trap temperature 220°C and the manifold temperature 60°C. The mass scan range was 41-500 u with a scan range time of 1 s.

For quantification of geraniol reaction mixtures were subjected to capillary GC-FID. The hexane extract of the reaction mixture was separated on an Agilent 6890 GC equipped with a DB-5 capillary column using nitrogen as carrier gas at a flow rate of 1.2 ml/min. The separation conditions were: splitless mode, injection volume 4 µl, injector temperature 230°C, initial oven temperature 100°C, then linear gradient to 140°C at a rate of 10°C/min followed by a linear gradient to 240°C at a rate of 35°C/min. The detector temperature was 250°C.

### 4.7. *CrGES* expression in yeast strains

The 1611 bp partial-length cDNA was amplified with primers GES\_F\_pYES2 and GES\_R\_pYES2 (Supplementary Table S1), digested with SacI and BamHI and directionally cloned into the pYES2 expression vector resulting in pYES2-*CrGES*.

Yeast strains were grown aerobically in minimal medium (1.7 g/l Yeast Nitrogen Base (Difco), 5 g/l ammonium sulphate (Merck) supplemented with 1% galactose as carbon source and the required amino acids (His, Leu, Ura) at 50 µg/ml (Euromedex, <http://www.euromedex.com>). The haploid FPPS mutated strain AE9K197G (*Mat<sup>α</sup>; his3; leu2Δ0; ura3; trp1Δ63; YJL167W::kanMX4 [pLB41ERG20K197G]*) (Fischer et al., 2011) was transformed with pYES2 or pYES2-*CrGES* plasmids. Transformed cells were selected for uracil prototrophy on minimal medium supplemented with required amino acids (His, Leu).

### 4.8. Yeast monoterpene analysis

The cells from a stationary phase culture were harvested by centrifugation (5000×g for 5 min). 3-Octanol (4 µg) and Ethylheptanoate (4 µg) were added as internal standard to 20 ml culture medium supernatant. Monoterpenoids were analysed by Stir Bar Sorptive Extraction and liquid desorption followed by gas chromatography-mass spectrometry method (SBSE-LD-GC-MS) (Coelho *et al.*, 2008) adapted to the laboratory conditions, with 1 µl injection volume. The analyses were performed on an Agilent 6890N gas chromatograph equipped with a Gerstel MultiPurpose Sampler MPS 2 (Gerstel, <http://www.gerstel.de>)

coupled to an Agilent 5975B inert MSD (Agilent Technologies). The gas chromatograph was fitted with a DB-Wax capillary column (60 m × 0.32 mm i.d. × 0.50 μm film thickness, Agilent Technologies) and helium was used as carrier gas (1 ml/min constant flow). The GC oven temperature was programmed from 45°C to 82°C at 20°C/min then to 235°C at 2.7°C/min (hold 15 min). The injector was set to 230°C and used in pulsed splitless mode (25 psi for 0.50 min). The transfer line, MS ion source and quadrupole analyzer temperatures were maintained at 270°C, 230°C and 150°C, respectively. Electron ionization mass spectra in the range of 29-450 *m/z* were recorded at 70eV. The mass spectra were obtained in full-scan mode and compared with the Wiley 7 MS library and NIST 05 mass spectral databases. Agilent MSD ChemStation software (G1701DA, Rev D.03.00) was used for instrument control and data processing. Total amounts of monoterpenoids were determined using linear calibration curves with an R<sup>2</sup> value of 0.99 over a concentration range from 0 to 6 mg/l.

#### 4.9. *In situ* hybridization

A *CrGES* cDNA obtained by PCR amplification with primers *GES\_Fwd* and *GES\_Rev* (Supplementary Table S1) was cloned in pGEM-T Easy and used for the synthesis of sense (*SpeI* linearization / T7 RNA polymerase) and antisense (*SacII* linearization / SP6 RNA polymerase) digoxigenin-labelled RNA riboprobes as previously described (Mahroug et al., 2006). For *CrG10H*, a previously described plasmid was used for riboprobe transcription (Burlat et al., 2004). Paraffin-embedded serial longitudinal sections of young developing leaves and of emerging cotyledons and serial transversal sections of carpels and stamens were hybridized with digoxigenin-labelled riboprobes and localized with anti-digoxigenin-alkaline phosphatase conjugates according to Mahroug et al. (2006).

#### 4.10. Generation of constructs for subcellular localisation studies

The full-length *CrGES* cDNA was amplified from *C. roseus* RT-generated cDNA using primers *GES\_Fyfp* and *GES\_Ryfp* (Supplementary Table S1). The full-length *CrGES* cDNA was digested with *NheI* and cloned into the *SpeI* site of vector pSCA-cassette YFPi resulting in the plasmid pSCA-*CrGES-YFP*.

A truncated *CrGES* cDNA representing the first 43 amino acids of the transit peptide was amplified using primers *GES\_Fyfp* and reverse primer *GES\_Ryfp\_TP1* (Supplementary Table S1). This amplicon was cloned into the *SpeI* site of vector pSCA-cassette YFPi resulting in plasmid pSCA-*CrTpGES-YFP*.

#### 4.11. Transient transformation of *C. roseus* cells and epifluorescence microscopy

Transient transformation of *C. roseus* cells by particle bombardment and YFP imaging were performed following the procedures described by Guirimand et al. (2009, 2010). Briefly, *C. roseus* cells were plated onto solid culture medium and bombarded with DNA-coated gold particles (1  $\mu\text{m}$ ) using a 1,100 psi rupture disk at a stopping-screen-to-target distance of 6 cm, using the Bio-Rad PDS1000/He system. Cells were cultivated for 15h to 48h and the protein subcellular localization was determined using an Olympus BX-51 epifluorescence microscope equipped with an Olympus DP-71 digital camera. The “plastid”-CFP (CD3-994) marker (Nelson et al., 2007) obtained from the ABRC (<http://www.arabidopsis.org>) was used for co-transformation studies with the *CrGES-YFP* and *CrTpGES-YFP* constructs.

#### 4.12. Semi-quantitative RT-PCR

The gene expression levels of *CrGES*, *ORCA3* and *ACTIN* were analysed semi-quantitatively using RT-PCR. The first strand cDNA was synthesized by reverse transcriptase as described above and was used as template for PCR amplification with the primer pairs indicated in Supplementary Table S1. The *ACTIN* gene was chosen as a constitutive external control. The *ORCA3* gene, whose expression is induced by MeJa, was chosen as a positive control for MeJa treatments. PCR reactions were performed with GoTaq polymerase according to the manufacturer’s instructions (Promega). Thermocycling conditions were 94°C for 5 min followed by 27 cycles of 94°C for 30 s, 56°C for 30 s and 72°C for 45 s and a final 5 min extension at 72°C. The amplified products were resolved on a 1.5% (w/v) agarose gel and visualized by ethidium bromide staining.

#### 4.13. Sequence analysis

Database searches for similar protein sequences were performed using *NCBI's BLAST network service* (<http://blast.ncbi.nlm.nih.gov/Blast.cgi>). Protein sequence alignment was performed using ClustalW from the Mac Vector program (MacVector, <http://www.macvector.com>). The theoretical molecular weight was calculated using the Compute pI/Mw tool ([http://expasy.org/tools/pi\\_tool.html](http://expasy.org/tools/pi_tool.html)). The chloroplast targeting peptide was predicted using ChloroP (<http://www.cbs.dtu.dk/services/ChloroP/>) and PSORT (<http://wolfpsort.org/>) programs.

#### 4.14. Transcriptomic and metabolomic data

The *C. roseus* transcriptome and metabolome are available on the Medicinal Plant Genomics Resource web site (MPGR, <http://medicinalplantgenomics.msu.edu>). Gene expression levels are provided in the form of FPKM values (Fragments per kilobase per transcript per million mapped reads) for 22 periwinkle samples. The Microsoft Excel PEARSON function calculating the Pearson Product-Moment Correlation Coefficient (PCC) for two sets of expression values (*CrGES* and *G10H*) within the 22 samples was used. The metabolome database provides information about the relative distribution of metabolites (including MIAs) in 19 samples. The relative abundances of individual metabolites were measured on the basis of their molecular mass.

## Primer used for reverse transcription

AP 5'-GGCCACGCGTCGACTAGTACTTTTTTTTTTTTTTTTTTTTTT-3'

## Primers used for RT-PCR

GESf 5'-AGAAAATGGAGTTAGAACAGCAGG-3'  
 GESr 5'-TACACAACCTGGGAAGCTCTGC-3'  
 ORCA3f 5'-CCGAGTTCGAAAGCTGTCAA-3'  
 ORCA3r 5'-TAGAAGGCTCCGCAGGGAAAC-3'  
 ACTINF 5'-TGGTGTGATGGTGGGAATGG-3'  
 ACTINr 5'-TTCCCGTTCGCTGAGGTTGT-3'

## Primers used for 3'RACE

GOS4 5'-TGAATCTCTCATCAAGTTCTTGCG-3'  
 GOS6 5'-ATGCAGAAAAACCATTGGAGGC-3'  
 AUAP 5'-GGCCACGCGTCGACTAGTAC-3'

## Primers used for nested phage walker

PWGES1 5'-ACTGATCTTATGCCCATTTGTGTCGAAGC-3'  
 PWGES2 5'-AGTGAGGAGATCTTCATGAGTTTGGCTGC-3'  
 T3 5'-AATTAACCCTCACTAAAGGG-3'

## Primers used for genome walker

GWGES4 5'-TCAATGGAGTTGCCAAAGGCAGAGACATGG-3'  
 GWGES5 5'-AGTCGATGAAGTTTTAGGCCTTTCTAGCC-3'  
 AP1 5'-GTAATACGACTCACTATAGGGC-3'  
 AP2 5'-ACTATAGGGCACGCGTGGT-3'

Primers used for CrGES expression in *Escherichia coli*

EcGES1 5'-CCGGTTGGATCCAATGGCAGCCACAATTAGTAACC-3'  
 EcGES2 5'-AACCGGCTGCAGTAAAAACAAGGTGTAAAAAACAAAGC-3'

Primers used for CrGES expression in *Saccharomyces cerevisiae*

GES\_F\_pYES2 5'-ATAGGGAATATTGAGCICATGTCTCTGCCTTTGGCAACTCC-3'  
 GES\_R\_pYES2 5'-ATAGGGAATATTGGATCCTTAAAAACAAGGTGTAAAAAACAAAGC-3'

## Primers used for subcellular localization

GES\_Fyfp 5'-CTGAGAGCTAGCATGGCAGCCACAATTAGTAACC-3  
 GES\_Ryfp 5'-CTGAGAGCTAGCAAACAAGGTGTAAAAAACAAAGCTTTTAC-3'  
 GES\_Ryfp\_TP1 5'-CTGAGAGCTAGCTGATGGCATAGACATGCAAATAGTCG-3

Primers used for *in situ* localization

GES\_Fwd 5'-CTGAGAGGATCCTCTCTGCCTTTGGCAACTCC-3'  
 GES\_Rev 5'-CTGAGAGAGCTCTTAAAAACAAGGTGTAAAAAACAAAGC-3'

**Supplementary Table S1.** Primers used in this study. Restriction sites incorporated for downstream cloning of the amplification fragments are underlined.

## Acknowledgements

The research leading to these results has received funding from the European Union Seventh Framework Programme FP7/2007-2013 under grant agreement number 222716 – SMARTCELL. We thank “Le STUDIUM” (Agency for Research and Hosting Foreign associated Researchers in the Centre region, France) for the financial support of Andrew J. Simkin.

## References

- Burlat, V., Oudin, A., Courtois, M., Rideau, M., St Pierre, B., 2004. Co-expression of three MEP pathway genes and geraniol-10-hydroxylase in internal phloem parenchyma of *Catharanthus roseus* implicates multicellular translocation of intermediates during the biosynthesis of monoterpene indole alkaloids and isoprenoid-derived primary metabolites. *Plant J.* 38, 131-141.
- Chen, F., Tholl, D., Bohlmann, J., Pichersky, E., 2011. The family of terpene synthases in plants: a mid-size family of genes for specialized metabolism that is highly diversified throughout the kingdom. *Plant J.* 66, 212-229.
- Christianson, D.W., 2006. Structural biology and chemistry of the terpenoid cyclases. *Chem. Rev.* 106, 3412-3442.
- Coelho, E., Perestrelo, R., Neng, N.R., Câmara, J.S., Coimbra, M.A., Nogueira, J.M.F., Rocha, S.M., 2008. Optimisation of stir bar sorptive extraction and liquid desorption combined with large volume injection-gas chromatography-quadrupole mass spectrometry for the determination of volatile compounds in wines. *Anal. Chim. Acta* 624, 79-89.
- Collu, G., Unver, N., Peltenburg-Looman, A.M.G., van der Heijden, R., Verpoorte, R., Memelink, J., 2001. Geraniol 10-hydroxylase, a cytochrome P450 enzyme involved in terpenoid indole alkaloid biosynthesis. *FEBS Lett.* 508, 215-220.
- Fischer, M.J.C., Meyer, S., Claudel, P., Bergdoll, M., Karst, F., 2011. Metabolic engineering of monoterpene synthesis in yeast. *Biotechnol. Bioeng.* 108, 1883-1892.

- Gamborg, O.L., Miller, R.A., Ojima, K., 1968. Nutrient requirements of suspension cultures of soybean root cells. *Exp. Cell Res.* 50, 151-158.
- Gantet, P., Imbault, N., Thiersault, M., Doireau, P., 1998. Necessity of a functional octadecanoic pathway for indole alkaloid synthesis by *Catharanthus roseus* cell suspension cultured in an auxin-starved medium. *Plant Cell Physiol.* 39, 220-225.
- Guirimand, G., Burlat, V., Oudin, A., Lanoue, A., St-Pierre, B., Courdavault, V., 2009. Optimization of the transient transformation of *Catharanthus roseus* cells by particle bombardment and its application to the subcellular localization of hydroxymethylbutenyl 4-diphosphate synthase and geraniol 10-hydroxylase. *Plant Cell Rep.* 28, 1215-1234.
- Guirimand, G., Courdavault, V., Lanoue, A., Mahroug, S., Guihur, A., Blanc, N., Giglioli-Guivarc'h, N., St-Pierre, B., Burlat, V., 2010. Strictosidine activation in Apocynaceae: towards a "nuclear time bomb"? *BMC Plant Biol.* 10, 182.
- Guirimand, G., Guihur, A., Ginis, O., Poutrain, P., Héricourt, F., Oudin, A., Lanoue, A., St-Pierre, B., Burlat, V., Courdavault, V., 2011a. The subcellular organization of strictosidine biosynthesis in *Catharanthus roseus* epidermis highlights several tonoplast translocations of intermediate metabolites. *FEBS J.* 278, 749-763.
- Guirimand, G., Guihur, A., Phillips, M.A., Oudin, A., Glévarec, G., Melin, C., Papon, N., Clastre, M., St-Pierre, B., Rodríguez-Concepción, M., Burlat, V., Courdavault, V., 2012. A single gene encodes isopentenyl diphosphate isomerase isoforms targeted to plastids, mitochondria and peroxisomes in *Catharanthus roseus*. *Plant Mol. Biol.* 79, 443-459.
- Guirimand, G., Guihur, A., Poutrain, P., Héricourt, F., Mahroug, S., St-Pierre, B., Burlat, V., Courdavault, V., 2011b. Spatial organization of the vindoline biosynthetic pathway in *Catharanthus roseus*. *J. Plant Physiol.* 168, 549-557.
- Iijima, Y., Gang, D.R., Fridman, E., Lewinsohn, E., Pichersky, E., 2004. Characterization of geraniol synthase from the peltate glands of sweet basil. *Plant Physiol.* 134, 370-379.
- Irmeler, S., Schroder, G., St-Pierre, B., Crouch, N.P., Hotze, M., Schmidt, J., Strack, D., Matern, U., Schroder, J., 2000. Indole alkaloid biosynthesis in *Catharanthus roseus*: new enzyme activities and identification of cytochrome P450 CYP72A1 as secologanin synthase. *Plant J.* 24, 797-804.
- Ito, M., Honda, G., 2007. Geraniol synthases from perilla and their taxonomical significance.

Phytochemistry 68, 446-453.

- Lee-Parsons, C.W., Royce, A.J., 2006. Precursor limitation in methyl jasmonate-induced *Catharanthus roseus* cell cultures. *Plant Cell Rep.* 25, 607-612.
- Lin, C., Shen, B., Xu, Z., Kollner, T.G., Degenhardt, J., Dooner, H.K., 2008. Characterisation of the monoterpene synthase gene *tps26*, the ortholog of a gene induced by insect herbivory in maize. *Plant Physiol.* 146, 940-951.
- Mahroug, S., Courdavault, V., Thiersault, M., St-Pierre, B., Burlat, V., 2006. Epidermis is a pivotal site of at least four secondary metabolic pathways in *Catharanthus roseus* aerial organs. *Planta* 223, 1191-1200.
- Morgan, J.A., Shanks, J.V., 2000. Determination of metabolic rate-limitations by precursor feeding in *Catharanthus roseus* hairy root cultures. *J. Biotechnol.* 79, 137-145.
- Murata, J., De Luca, V., 2005. Localization of tabersonine 16-hydroxylase and 16-OH tabersonine-16-O-methyltransferase to leaf epidermal cells defines them as a major site of precursor biosynthesis in the vindoline pathway in *Catharanthus roseus*. *Plant J.* 44, 581-594.
- Murata, J., Roepke, J., Gordon, H., De Luca, V., 2008. The leaf epidermome of *Catharanthus roseus* reveals its biochemical specialization. *Plant Cell* 20, 524-542.
- Nagegowda, D.A., Gutensohn, M., Wilkerson, C.G., Dudareva, N., 2008. Two nearly identical terpene synthases catalyze the formation of nerolidol and linalool in snapdragon flowers. *Plant J.* 55, 224-239.
- Nelson, B.K., Cai, X., Nebenführ, A., 2007. A multicolored set of *in vivo* organelle markers for co-localization studies in *Arabidopsis* and other plants. *Plant J.* 51, 1126-1136.
- Natesan, S.K., Sullivan, J.A., Gray, J.C., 2005. Stromules: a characteristic cell-specific feature of plastid morphology. *J. Exp. Bot.* 56, 787-797.
- Oswald, M., Fischer, M., Dirninger, N., Karst, F., 2007. Monoterpenoid biosynthesis in *Saccharomyces cerevisiae*. *FEMS Yeast Res.* 7, 413-421.
- Oudin, A., Courtois, M., Rideau, M., Clastre, M., 2007a. The iridoid pathway in *Catharanthus roseus* alkaloid biosynthesis. *Phytochem. Rev.* 6, 259-276.
- Oudin, A., Mahroug, S., Courdavault, V., Hervouet, N., Zelwer, C., Rodriguez-Concepcion, M., St-Pierre, B., Burlat, V., 2007b. Spatial distribution and hormonal regulation of gene



- products from methyl erythritol phosphate and monoterpene-secoiridoid pathways in *Catharanthus roseus*. *Plant Mol. Biol.* 65, 13-30.
- Schattat, M., Barton, K., Baudisch, B., Klösigen, R.B., Mathur, J., 2011. Plastid stromule branching coincides with contiguous endoplasmic reticulum dynamics. *Plant Physiol.* 155, 1667-1677.
- St-Pierre, B., Vazquez-Flota, F.A., de Luca, V., 1999. Multicellular compartmentation of *Catharanthus roseus* alkaloid biosynthesis predicts intercellular translocation of a pathway intermediate. *Plant Cell* 11, 887-900.
- Turner, G., Gershenzon, J., Nielson, E.E., Froehlich, J.E., Croteau, R., 1999. Limonene synthase, the enzyme responsible for monoterpene biosynthesis in peppermint, is localized to leucoplasts of oil gland secretory cells. *Plant Physiol.* 120, 879-886.
- van der Fits, L., Memelink, J. 2000. ORCA3, a jasmonate-responsive transcriptional regulator of plant primary and secondary metabolism. *Science* 289, 295-297.
- van der Fits, L., Memelink, J., 2001. The jasmonate-inducible AP2/ERF-domain transcription factor ORCA3 activates gene expression via interaction with a jasmonate-responsive promoter element. *Plant J.* 25, 43-53.
- van der Heijden, R., Jacobs, D.I., Snoeijer, W., Hallard, D., Verpoorte, R., 2004. The *Catharanthus* Alkaloids: Pharmacognosy et Biotechnology. *Curr. Med. Chem.* 11, 1241-1253.
- Yang, T., Li, J., Wang, H.X., Zeng, Y., 2005. A geraniol-synthase gene from *Cinnamomum tenuipilum*. *Phytochemistry* 66, 285-293.
- Yang, T., Stoopen, G., Yalpani, N., Vervoort, J., de Vos, R., Voster, A., Verstappen, F.W., Bouwmeester, H.J., Jongsma, M.A., 2011. Metabolic engineering of geranic acid in maize to achieve fungal resistance is compromised by novel glycosylation patterns. *Metab. Eng.* 13, 414-425.
- Zhou, M.L., Hou, H.L., Zhu, X.M., Shao, J.R., Wu, Y.M., Tang, Y.X., 2011. Molecular regulation of terpenoid indole alkaloids pathway in the medicinal plant, *Catharanthus roseus*. *J. Med. Plants Res.* 5, 663-676.

# Characterization of two geraniol synthases from *Valeriana officinalis* and *Lippia dulcis*: similar activity but difference in subcellular localization

Lemeng Dong<sup>a</sup>, **Karel Miettinen**<sup>b</sup>, Miriam Goedbloed<sup>a</sup>, Francel W.A. Verstappen<sup>a</sup>, Alessandra Voster<sup>c</sup>, Maarten A. Jongsma<sup>c</sup>, Johan Memelink<sup>b</sup>, Sander van der Krol<sup>a</sup>, Harro J. Bouwmeester<sup>a</sup>

<sup>a</sup> Laboratory of Plant Physiology, Wageningen UR, P.O. Box 658, 6700 AR Wageningen, The Netherlands

<sup>b</sup> Institute of Biology, Sylvius Laboratory, Leiden university, P.O. Box 9505, 2300 RA Leiden, The Netherlands

<sup>c</sup> Plant Research International, Wageningen UR, P.O. Box 619, 6700 AP Wageningen, The Netherlands

Corresponding author: Harro J. Bouwmeester

Tel: +31317489859

Email: [harro.bouwmeester@wur.nl](mailto:harro.bouwmeester@wur.nl)

## Abstract

Two geraniol synthases (GES), from *Valeriana officinalis* (VoGES) and *Lippia dulcis* (LdGES), were isolated and were shown to have geraniol biosynthetic activity with Km value of 32  $\mu$ M and 51  $\mu$ M for GPP, respectively, upon expression in *E. coli*. The *in planta* enzymatic activity and sub-cellular localization of VoGES and LdGES were characterized in stable transformed tobacco and using transient expression in *Nicotiana benthamiana*. Transgenic tobacco expressing VoGES or LdGES accumulate geraniol, oxidized geraniol compounds like geranial, geranic acid and hexose conjugates of these compounds to similar levels. Geraniol emission of leaves was lower than that of flowers, which could be related to higher levels of competing geraniol-conjugating activities in leaves. GFP-fusions of the two GES proteins show that VoGES resides (as expected) predominantly in the plastids, while LdGES import into to the plastid is clearly impaired compared to that of VoGES, resulting in both cytosolic and plastidic localization. Geraniol production by VoGES and LdGES in *N. benthamiana* was nonetheless very similar. Expression of a truncated version of *VoGES* or *LdGES* (cytosolic targeting) resulted in the accumulation of 30% less geraniol glycosides than with the plastid targeted VoGES and LdGES, suggesting that the substrate geranyl diphosphate is readily available, both in the plastids as well as in the cytosol. The potential role of GES in the engineering of the TIA pathway in heterologous hosts is discussed.

## 1. Introduction

Plants are estimated to produce more than 500,000 secondary metabolites of various classes (isoprenoids, phenylpropanoids, alkaloids) (Hadacek, 2002). Of these, the isoprenoids represent the largest family based on their diverse structural features which relate to numerous biological activities. Isoprenoids have been shown to affect many physiological processes such as respiration, signal transduction, cell division, membrane architecture, photosynthesis, and growth. In addition, isoprenoids have ecological significance as they play an important roles in the exchange of signals between plants and between plants and microorganisms or in defense against pathogens and herbivores. Also the applications of

isoprenoids in foods, cosmetics and pharmaceutical drugs make specific terpenoids interesting commercial targets.

Although isoprenoids are extraordinarily diverse, they all originate from the condensation of the universal five-carbon precursors, isopentenyl diphosphate (IPP) and dimethyl allyl diphosphate (DMAPP). In higher plants, two independent pathways, located in separate intracellular compartments, are involved in the biosynthesis of IPP and DMAPP. In the cytosol, IPP is derived from the classic mevalonic acid (MVA) pathway that starts from acetyl-CoA (Porter et al., 1981), whereas in plastids, IPP is formed from pyruvate and glyceraldehyde 3-phosphate via the methylerythritol phosphate (MEP or non-mevalonate) pathway (Eisenreich et al., 2001; Lichtenthaler, 1999). Cytosolic IPP and DMAPP are converted to farnesyl diphosphate (FPP, C15), which serves as a precursor of sesquiterpene and triterpene biosynthesis in the cytosol. In contrast, the plastidial pool of IPP/DMAPP is converted to geranyl diphosphate (GPP, C10) and geranylgeranyl diphosphate (GGPP, C20) which serve as precursors for monoterpenes, and diterpenes and tetraterpenes, respectively, in the plastid (Lange et al., 2001; McConkey et al., 2000; Tholl and Lee, 2011; Turner et al., 1999).

Geraniol is an acyclic monoterpene alcohol that is synthesized in one step from GPP. Geraniol is a component of essential oils present in many fragrant plant species (Antonelli et al., 1997; Bakkali et al., 2008; Bayrak and Akgül, 1994; Sangwan et al., 2001; Yang et al., 2005). It has a rose-like odor and is commonly used in perfumes (Chen, 2006; Rastogi et al., 2003) and aromatic fragrance in wine (Herrero et al., 2008; Pedersen et al., 2003). Geraniol also has pharmaceutical properties, as it can inhibit the growth of human colon cancer cells (Carnesecchi et al., 2001) and interfere with membrane functions in *Candida albicans* and *Saccharomyces cerevisiae* (Bard et al., 1988). In some plant species geraniol is the precursor for terpenoid indole alkaloid (TIA) biosynthesis. For instance, in *Catharanthus roseus* the anticancer agents vinblastine and vincristine are synthesized from geraniol (monoterpene iridoid branch) and tryptophan (indole branch) in the TIA pathway.

Multiple approaches have been tested to increase TIA production. For example, overexpress gene 1-deoxy-D-xylulose synthase and geraniol-10-hydroxylase gene were shown to increase the flux towards vinblastine and vincristine in *C. roseus* hairy root (Peebles et al., 2011). Attempts to boost transcription of TIA biosynthetic genes in the hairy roots or

suspension cells were only partially successful (Liu et al., 2011; Memelink and Gantet, 2007; Montiel et al., 2007). For example, ORCA3 is a jasmonate responsive transcription factor that promote transcription of TIA biosynthesis genes (Vom Endt et al., 2007). However, when ORCA3 is overexpressed, also repressor activity is activated, which in the long term actually caused a decrease in several TIA metabolites in *C. roseus* (Peebles et al., 2009). Expression of the TIA pathway biosynthesis genes in a heterologous host may provide a way to overcome such feedback regulation problems.

The objective of the present study was the efficient production of the monoterpene geraniol as the first step in a larger program to rebuild the complete monoterpene iridoid branch of the TIA biosynthesis pathway in a heterologous host. To achieve this, a geraniol synthase (GES) was cloned from *Valeriana officinalis* L. (Valerianaceae) (VoGES) and compared to the previously isolated *LdGES* from *Lippia dulcis* (Yang et al., 2011). Both proteins showed similar geraniol synthase activity *in vitro* and *in planta*. VoGES was subsequently used in a number of transient and stable metabolic engineering approaches to explore the possibility to reconstitute the monoterpene branch of TIA biosynthesis in tobacco.

## 2. Materials and methods

### 2.1. Cloning and sequence analysis of geraniol synthase gene

For the cloning of the geraniol synthase gene, *Valeriana officinalis* L. (VoGES) total RNA was isolated from *V. officinalis* leaves using SV Total RNA Isolation System (Promega). Based on conserved domains of known geraniol synthases, the degenerate primers (forward primer 5'-GAYGARAAYGGIAARTTYAARGA-3' and reverse primer 5'-CCRTAIGCRTRCAAIGTRTCRTC -3') were designed to amplify partial cDNA fragment by reverse transcription PCR (RT-PCR). Full length sequences of the cDNAs were obtained by rapid amplification of cDNA ends (RACE).

Putative VoGES sequence was blasted against the GenBank ENTREZ database (NCBI Blast 2.2.23) (Altschul et al., 1997) and GES sequences were aligned using CLUSTALX 1.83 (Thompson et al., 1997) using standard settings. Prediction of the subcellular localization was from the targeting prediction programs PREDOTAR version 1.03

(<http://urgi.versailles.inra.fr/predotar/>) (Small et al., 2004) and TARGETP 1.1 Server (<http://www.cbs.dtu.dk/services/TargetP/>) (Emanuelsson et al., 2000).

## 2.2. Heterologous expression of VoGES and LdGES protein in *Escherichia coli*

For the *in vitro* functional analysis of the putative geraniol synthase from *Valeriana officinalis* and comparison with *LdGES*, the truncated cDNAs  $\Delta$ NVoGES (bp 178-1785) and  $\Delta$ NLdGES (bp 139-1755) were subcloned into the multiple cloning site of the expression vector pRSET A (Invitrogen) to yield constructs pRSET- $\Delta$ NVoGES and pRSET- $\Delta$ NLdGES. Primer sequences for PCR amplification and restriction sites for each primer are listed in Table S1. After full re-sequencing to check integrity, constructs were transformed into *E. coli* BL21 (DE3) (Invitrogen) and expression was induced by isopropyl  $\beta$ -D thiogalactopyranoside (IPTG) in transformed *E. coli* BL21 (DE3) cell cultures. The His-tagged proteins were isolated by passing through Ni-NTA Spin columns according to the manufacturers recommendations (Qiagen). For quality analysis, the recombinant protein was confirmed with 12.5 % (w/v) SDS-PAGE gel electrophoresis followed by Western blotting using mouse monoclonal anti-His horseradish peroxidase (HPR) conjugate antibodies (5Prime, <http://www.5prime.com>). Antibody binding was detected by incubation in 250  $\mu$ M sodium luminol, 0.1 M Tris-HCl (pH 8.6), 3 mM H<sub>2</sub>O<sub>2</sub>, 67  $\mu$ M *p*-coumaric acid and exposure to X-ray film.

**Table S1** Primer sequences for PCR amplification and restriction site for *E. coli* expression and GFP fusion constructs

An enzyme assay was carried out for functional characterization, using geranyl diphosphate (GPP) and farnesyl diphosphate (FPP) as substrates. Enzymatic assays were done in 0.5 ml reaction buffer containing 50 mM Tris-HCl, 1 mM MgCl<sub>2</sub>, 0.1 mM MnCl<sub>2</sub>, and 10, 20, 50 or 100  $\mu$ M GPP (or 62.5  $\mu$ M FPP) and 0.5  $\mu$ g (VoGES) and 2  $\mu$ g (LdGES) of purified enzyme. The reaction mixture was incubated at 32 °C for 5 min. For quantitative analysis citronellol was added to a concentration of 50  $\mu$ M as an internal standard into the reaction tube after incubation. The reaction was stopped by adding 1 volume of hexane, mixing thoroughly by vortexing and keeping on ice for 10 min. The tubes were centrifuged at 4000 g

for 10 min, and the supernatant hexane phase was collected. The extraction was repeated with hexane (0.5 ml). Then the hexane phase was collected and dehydrated and then subjected to capillary gas chromatography-flame ionizing detector (GC-FID) and gas chromatography-mass spectrometry (GC-MS, supplementary method). For the latter, the hexane extract was separated on a Agilent GC 6890 series equipped with a DB-5 capillary column (30 m × 0.25 mm, film thickness of 0.25 µm) (J&W Scientific) using nitrogen as carrier gas at a flow rate of 1.2 ml min<sup>-1</sup>. The separation conditions were: split mode 1: 5, injection volume 5 µl, injector temperature 230 °C, initial oven temperature 100 °C, then linear gradient to 140 °C at a rate of 10 °C min<sup>-1</sup> followed by a linear gradient to 240 °C at a rate of 35 °C min<sup>-1</sup>. Amounts of geraniol formed in enzyme assays were calculated from the resultant GC/FID integral using the relative response factor with respect to the citronellol internal standard. Lineweaver-Burk plots of VoGES and LdGES activity were used to obtain the *K<sub>m</sub>* values for GPP.

### 2.3. VoGES and LdGES subcellular localization studies

For analysis of the subcellular targeting, the coding sequences of *EGFP* was fused to the N-terminus or C-terminus of full length *VoGES* and *LdGES*. In addition a truncated version of *LdGES* lacking the first 46 AA ( $\Delta$ N*LdGES*: bp 139-1755) and a truncated version of *VoGES* lacking the first 56 AA ( $\Delta$ N*VoGES*: bp 178 -1785) was made using standard cloning techniques and the C-terminal coding sequence of these genes was fused in frame to that of GFP. The *VoGES-GFP*, *LdGES-GFP*, *GFP-VoGES*, *GFP-LdGES*,  $\Delta$ N*VoGES-GFP* and  $\Delta$ N*LdGES-GFP* were cloned into impact vector pIV2A 2.1 ([www.pri.wur.nl/UK/products/ImpactVector/](http://www.pri.wur.nl/UK/products/ImpactVector/)) under control of the CaMV 35S-promoter. In addition, the truncated versions of *VoGES* and *LdGES* were provided with a heterologous plastid import signal by cloning into impact vector pIV2A 2.4 which carries an artificial plastid targeting signal ([www.pri.wur.nl/UK/products/ImpactVector/](http://www.pri.wur.nl/UK/products/ImpactVector/)) (Wong et al., 1992). The fusion constructs were sequenced to check integrity, before transferring the fusion cassettes to the binary vector pBIN+ (van Engelen et al., 1995) using LR recombination (Gateway technology) (Karimi et al., 2002). Primer sequences for PCR amplification and restriction sites for each primer are listed in Table S1. Finally, the binary expression

constructs were transformed into *Agrobacterium tumefaciens* strain AGL0 (Lazo et al., 1991) which was used for transient expression in *Nicotiana benthamiana* as described below. Expression and localization were analyzed at different days post-agroinfiltration in small leaf samples (~0.5 cm<sup>2</sup> leaf material from at least three independent agro-infiltrated plants) by confocal laser scanning.

Microscopy using an Axiovert 200 M with a Zeiss LSM 5 PASCAL laser scanning microscope (Carl Zeiss) and a 20 × (N.A. 0.5) Plan NeoFluar (Zeiss) or a 63 × (N.A. 1.4 oil) Plan Apochromat (Zeiss) objective. Samples were excited with 1 % of a 488 nm laser (emission from a 30 mW argon tube) for *EGFP* and chlorophyll excitation and green *EGFP* fluorescence and red chlorophyll fluorescence were collected using two emission filters. Band pass was set to 505 to 530 nm for *EGFP* fluorescence detection and long pass was set to > 560 nm for chlorophyll autofluorescence detection. Images were exported from .lsm files to .tif files using Zeiss LSM image browser version 3.5 without further processing.

#### 2.4. Tobacco stable plant transformation

The *VoGES* and *LdGES* full length cDNA sequences were subcloned into the binary expression vector pBIN+ using standard cloning techniques. In this vector *VoGES* and *LdGES* are under the control of the constitutive CaMV-d35S promoter and a nopaline synthase terminator. Both the 35S::*LdGES* and 35S::*VoGES* pBIN+ constructs were introduced into *A. tumefaciens* AGL0 and *in vitro* grown wild-type *N. tabacum* 'Samsun NN' was transformed using the leaf disc method as described (Horsch et al., 1985). Transformed shoots were selected on medium with 100 mg/l kanamycin. Primary transformed shoots were rooted on non-selective medium, checked by PCR for presence of the expression construct and positive T0 shoots were transferred to soil and grown until seed set. T1 plants were grown until seed set and from the T2 population five homozygous lines 35S::*LdGES* and three homozygous lines 35S::*VoGES* with single insert were selected.

#### 2.5. Transient expression in *Nicotiana benthamiana*

*A. tumefaciens* infiltration (agro-infiltration) was performed according to the description of van Herpen *et al.* (van Herpen *et al.*, 2010). Briefly, *A. tumefaciens* was grown at 28 °C for 24 hours in LB media with kanamycin (50 mg/L) and rifampicillin (34 mg/L). Cells



were harvested by centrifugation for 20 min at 4000 g and 20 °C and then resuspended in infiltration buffer containing 10 mM MES (2-(N-morpholino) ethanesulfonic acid, Duchefa Biochemie), 10 mM MgCl<sub>2</sub> and 100 μM acetosyringone (4'-hydroxy-3', 5'-dimethoxyacetophenone, Sigma) to a final OD<sub>600</sub> of ~0.5, followed by incubation at room temperature under gentle shaking at 50 rpm for 150 min. In all experiments, *A. tumefaciens* harboring TBSV *p19* was included to maximize protein production by suppression of gene silencing (Voinnet et al., 2003). *A. tumefaciens* harbouring constructs with *GFP-GES* or *GES* were infiltrated into leaves of five-week-old *N. benthamiana* plants by pressing a 1 mL syringe without metal needle against the abaxial side of the leaf and slowly injecting the bacterium suspension into the leaf. *N. benthamiana* plants were grown from seeds on soil in the greenhouse with a minimum of 16 hour light. Day temperatures were approximately 28 °C, night temperatures 25 °C. After agro-infiltration the plants remained under the same greenhouse conditions until further analysis.

## 2.6. Volatile GC-MS analysis

For analysis of the production of volatiles by 35S::*LdGES* or 35S::*VoGES* tobacco seedling, T2 seeds were sterilized by chlorine bleach and placed on ½ MS (Duchefa, Netherlands) medium. Petri dishes with ~50 seeds were incubated at 4°C in the dark to ensure synchronous germination. Subsequently, petri dishes were transferred to growth cabinets with 16 L/ 8 D, 28 °C day, 25 °C night. After 2 weeks of growing, petri dishes were placed in 0.5 L glass jars, and volatiles were trapped for 5 hours in the light using Tenax TA (20/35 mesh, Alltech, Breda, the Netherlands).

Headspace analysis of stable transformed *VoGES* tobacco was done using detached leaves and flowers. Before they were enclosed in 1 L glass jars with a Teflon-lined lid equipped with in- and outlet, detached leaves and flowers from tobacco plants at different development stages were placed into small glass bottles with water in them while different parts of flowers were placed on wet filter paper. A vacuum pump was used to draw air through the glass jar at approximately 100 ml.min<sup>-1</sup>, with the incoming air being purified through a glass cartridge (140 × 4 mm) containing 150 mg Tenax TA. At the outlet the volatiles emitted by the samples were trapped on a similar Tenax cartridge. After 5 h of

trapping, the trapped volatiles were analyzed by Thermodesorption GC-MS using a thermal desorber (Unity, Markes International Limited) and a Trace GC Ultra (Thermo Electron Corporation) coupled with DSQ mass spectrometer (Thermo Electron Corporation). The tubes were first purged to remove water vapor and oxygen for 2 min at room temperature with helium flow of 50 ml min<sup>-1</sup>. Then trapped volatiles were desorbed from the Tenax in the thermal desorber at 250 °C for 5 minutes. Volatiles were collected in an electrically cooled sorbent trap (Unity; Markes, Llantrisant) at 10 °C and injected into the analytical column (ZB-5MSI, 30 m × 0.25 mm ID, 1.0 µm-film thickness, Zebron, Phenomenex). The temperature program of the gas chromatograph started at 40 °C (3 min hold) and rose to 280 °C at 12 °C min<sup>-1</sup>, with a hold at final temperature for 2 min. The mass spectrometer was set to scan from 45 to 300 m/z with a scan time of 5.4 scans s<sup>-1</sup>. The helium flow was constant at 1.0 ml min<sup>-1</sup>. Ionization potential was set at 70 eV. For quantification, a geraniol calibration curve was made with a series of standard solutions. The GC-MS results were analyzed using Xcalibur software (Thermo, Waltham). After headspace trapping for 5 h in the light, samples were frozen in liquid nitrogen and stored at -80 °C for further analysis.

### 2.7. GC-MS analysis of extracts

Volatile compounds accumulated in the plant material were extracted with dichloromethane (DCM) and measured by GC-MS as described above. Aliquots of 500 mg of frozen, powdered material in pre-cooled glass tubes were extracted twice with 3 ml DCM. Extracts were shortly vortexed, sonicated for 10 min, centrifuged for 10 min at 1200 g and filtered through a small glass column containing anhydrous Na<sub>2</sub>SO<sub>4</sub>. The eluent was concentrated extracts under a flow of nitrogen and 1 µl concentrated extracts were injected into the GC-MS column. The initial oven temperature was 45 °C for 1 min, and was increased to 300 °C at a rate of 10 °C min<sup>-1</sup> and held for 5 min at 300 °C. For quantification of geraniol and oxidised geraniol products, standards of geraniol, geranial and geranic acid were injected at different concentrations to establish calibration curves. Each sample was spiked with 2 µg/µl *cis*-nerolidol as internal standard.

### 2.8. Analysis of geraniol-derived conjugates by LC- QTOF-MS and LTQ-Orbitrap-MS<sup>n</sup>

Analysis of non-volatile compounds in transgenic tobacco extracts was done by liquid chromatography coupled to quadrupole time-of-flight mass spectrometry (LC-QTOF-MS) (De Vos et al., 2007). Aliquots of 200 mg of frozen, powdered material were extracted with 0.6 ml 99.9% MeOH / 0.133% formic acid in 1.5 ml Eppendorf vial. After short vortex and 15 min sonication, the extracts were centrifuged and filtered through 0.45  $\mu\text{m}$  filters (SRP4, Sartorius, Germany) and 5  $\mu\text{l}$  of the filtered extract was analysed using a Waters Alliance 2795 HPLC connected to a QTOF Ultima V4.00.00 mass spectrometer (Waters, MS technologies, UK). Measurements were in negative ionization mode and leucine enkephalin ( $[\text{M} - \text{H}]^- = 554.2620$ ) was used as a lock mass for online mass calibration.

Acquisition of LC-MS data was performed under MassLynx 4.0 (Waters). MassLynx was used for visualization and manual processing of LC-PDA-MS data. Mass data were processed using metAlign version 1.0 ([www.metalignment.nl](http://www.metalignment.nl)). Baseline and noise calculations were performed from scan number 70 to 2,480. The maximum amplitude was set to 25,000 and peaks below three times the local noise were discarded. Multiple mass signals derived from the same compound were grouped using MSClust software ([biotools.wurnet.nl](http://biotools.wurnet.nl)) by Multivariate Mass Spectra Reconstruction (MMSR) (Tikunov et al., 2005). The selected mass intensities were normalized using log<sub>2</sub> transformation and standardized using range scaling, in which each value in a certain row, corresponding to the internal standard leucine enkephalin, was divided by the intensity range observed for this row throughout all samples analysed. Each row was then mean centred. Finally, the normalized and log-transformed data matrix was used for Principal Components Analysis implemented in GeneMath XT version 2.1.

Significance of differences in intensity of each aligned mass signal between samples was assessed using student t-test (level of significance set at 0.05). Masses showing significant difference were manually checked in MassLynx. Putative identification of metabolites was by determining the best fit elemental composition using C, H and O with MassLynx software. For multiple possible molecular formulas (tolerance <5 ppm) the best matches were

searched in the Dictionary of Natural Products and SciFinder databases for possible structures.

For further identification, selected compounds were targeted for fragmentation by an Accela HPLC tower connected to a LTQ Orbitrap hybrid mass spectrometer (Thermo Fisher Scientific). The instrument settings of HPLC and LTQ Orbitrap were used as described earlier (van der Hooff et al., 2011). The LTQ was programmed to use a window of 10 D to isolate the mass of interest in MS1. The data-dependent fragmentation was set as follows: MS2 fragmentation of most intense ion in MS1; MS3 fragmentation of the 5 most intense fragment ions in MS2; MS4 fragmentation of the 5 most intense fragment ions in each MS3.

### 2.9. Glycosidase treatment

To quantify how much of the geraniol related products were modified by glycosylation, extracts were treated with glycosidase to release geraniol or geraniol related compounds. For this purpose, 200 mg infiltrated leaf material was ground in liquid nitrogen and extracted with 1 ml citrate phosphate buffer, pH 5.4. The extracts were prepared by brief vortexing and sonication for 15 min. 0.2 ml of Viscozyme L (Sigma) was added and overlaid with 1ml of pentane to trap released volatiles and the samples were again vortexed. To trap released volatiles, 1 ml pentane was added on top of the extract. The mixture was incubated overnight at 37 °C, and subsequently extracted twice with 2 ml of pentane. Extracts were dehydrated using anhydrous Na<sub>2</sub>SO<sub>4</sub> and concentrated to approximately 100 µl. An internal standard, *cis*-nerolidol was used to quantify the products. Samples were analyzed using GC-MS as described under 2.7.

## Supplemental methods

### Western blotting

Western blotting was used to check the integrity of the LdGES-GFP fusion proteins after transient expression in *N. benthamiana*. 10mg liquid nitrogen grounded agro-infiltrated *N. benthamiana* leaf was suspended in sample buffer (50 mM Tris pH 6.8, 2% SDS, 0.1% BFB, 10% glycerol, 4% β-me). The suspension together with sample was heated 10 min at 94 °C and

centrifuged 1 min at 13000 rpm. 15  $\mu$ l supernatant was first run in a 12% SDS-PAGE gel, subsequently was blotted on to a PDVF membrane (Biorad) for 1 hr at 100 V. The membrane was blocked overnight at 4°C in TBS-T (Tris buffered saline with 0.05% Tween 20) containing 3% non-fat dried milk. Next the blot was incubated with 1: 20000 rabbit anti-GFP (ab290, Abcam) for 1hr at room temperature, and washed three times in TBS-T. Then the blot was incubated with donkey anti rabbit-HRP 1: 50000 dilution (Thermo scientific) in TBS-T, followed by washing four times in TBS-T. 3 ml Supersignal® West Dura Extended Duration Substrate (Thermo scientific) was used.

### *GC-MS analysis of the enzymatic product from VoGES and LdGES expressed in E. coli*

Products were analyzed by GC-MS using a gas chromatograph (7809A, Agilent Technologies, USA) equipped with a 30 m  $\times$  0.25 mm, 0.25 mm film thickness column (ZB-5, Phenomenex) and a Triple-Axis detector (model 5975C, Hewlett Packard, Agilent Technologies). The injection port (splitless mode), interface and MS source temperatures were 250 °C, 290 °C and 180 °C respectively. The injection volume was 2  $\mu$ l. The oven was programmed at an initial temperature of 45 °C for 1 min, with a ramp of 10 °C min<sup>-1</sup> to 280 °C, and final time of 5 min. Scanning was performed from 45-300 atomic mass unit (amu). The helium inlet pressure was checked by electronic pressure control to achieve a constant column flow of 1.0 ml min<sup>-1</sup>. Ionization potential was set at 70 eV. Products were identified by comparing mass spectra to the National Institute of Standards and Technology (NIST) mass spectra library and by calculating the Kovats Index (Kovats, 1965) based on the retention time relative to alkane standards. An empty vector (original pRSET A vector without any insert) was used as negative control.

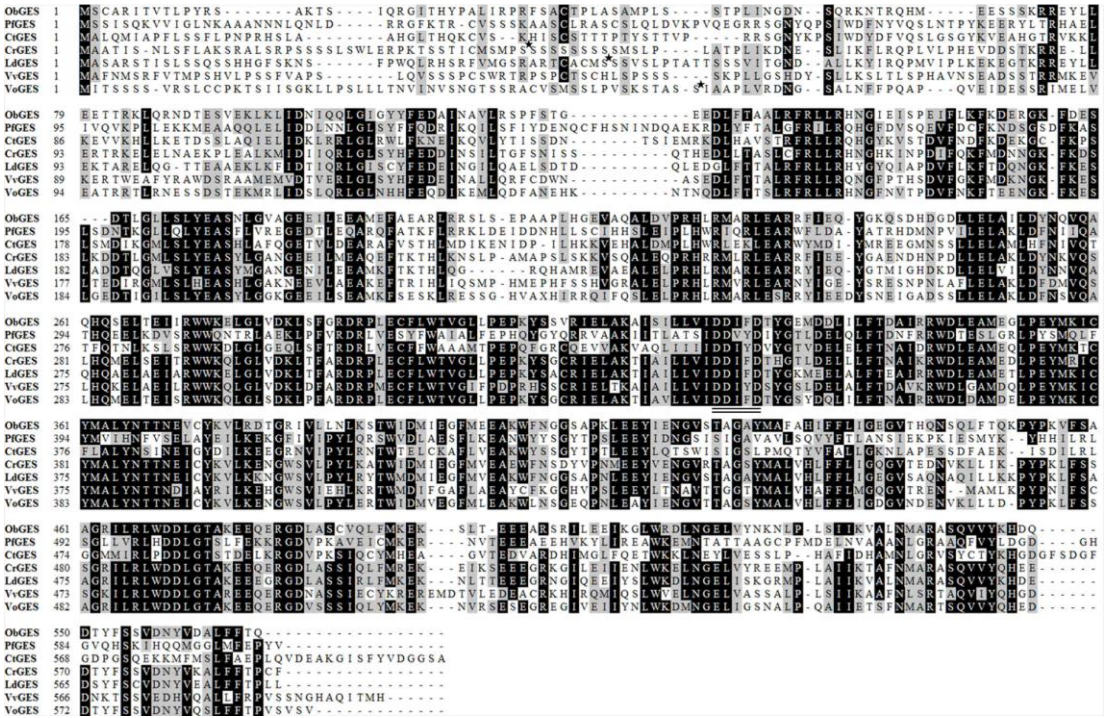
### 3. Results

#### 3.1. Functional characterization of geraniol synthase from *Valeriana officinalis* and *Lippia dulcis* *in vitro*

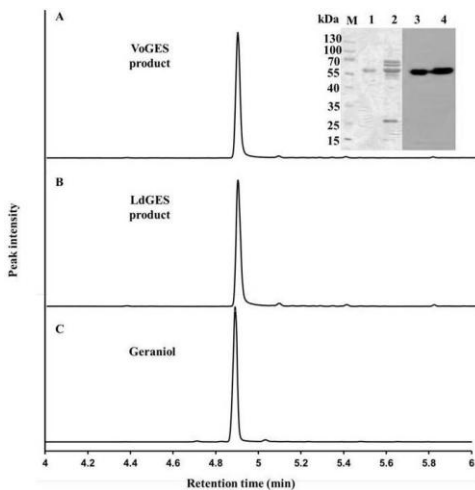
The GES gene from *Valeriana officinalis* (*VoGES*) encodes a protein which shares 37-67 % identity with GES proteins from other plant species (Table 1, Figure 1). Previous characterized GES gene from *Lippia dulcis* (*LdGES*) shares 63% identity with *VoGES*. The alignment shows that the Aspartate-rich DDxxD-motif for metal-dependent ionization of the prenyl diphosphate substrate (Bohlmann et al., 1998; Tarshis et al., 1996; Wendt and Schulz, 1998) is present in all GES sequences (Figure 1). Both the *VoGES* and *LdGES* protein with N-terminal truncation were produced in *E. coli* (see methods) to compare the *in vitro* activity. Analysis of both recombinant protein by SDS-PAGE and Coomassie Brilliant Blue staining and immunoprobng with anti-His antibodies showed the presence of one major band (Figure 2A). In the presence of GPP, both enzymes catalyzed the formation of the acyclic monoterpene alcohol geraniol, which was identified based on comparison with the retention time of authentic geraniol (Figure 2) and the mass spectrum comparison with the NIST library (Figure S2). Truncated *VoGES* and *LdGES* catalyzed the conversion of GPP to geraniol with a  $K_m$  value of 32  $\mu\text{M}$  and 51  $\mu\text{M}$ , respectively (Figure S3). No product was detected when *VoGES* or *LdGES* were supplied with FPP (data not shown).

Gene	Identities (%)							Full Length	Predicted Localization
	<i>VoGES</i>	<i>LdGES</i>	<i>CrGES</i>	<i>ObGES</i>	<i>CtGES</i>	<i>PfGES</i>	<i>VvGES</i>		
<i>VoGES</i>	100	63	67	59	42	37	57	594aa	Chloroplast
<i>LdGES</i>	63	100	66	68	39	36	54	584aa	Chloroplast
<i>CrGES</i>	67	66	100	63	42	36	57	589aa	Chloroplast
<i>ObGES</i>	59	68	63	100	41	35	50	567aa	Mitochondrion
<i>CtGES</i>	42	39	42	41	100	45	42	603aa	Chloroplast
<i>PfGES</i>	37	36	36	35	45	100	37	603aa	Chloroplast
<i>VvGES</i>	57	54	57	50	42	37	100	595aa	Chloroplast

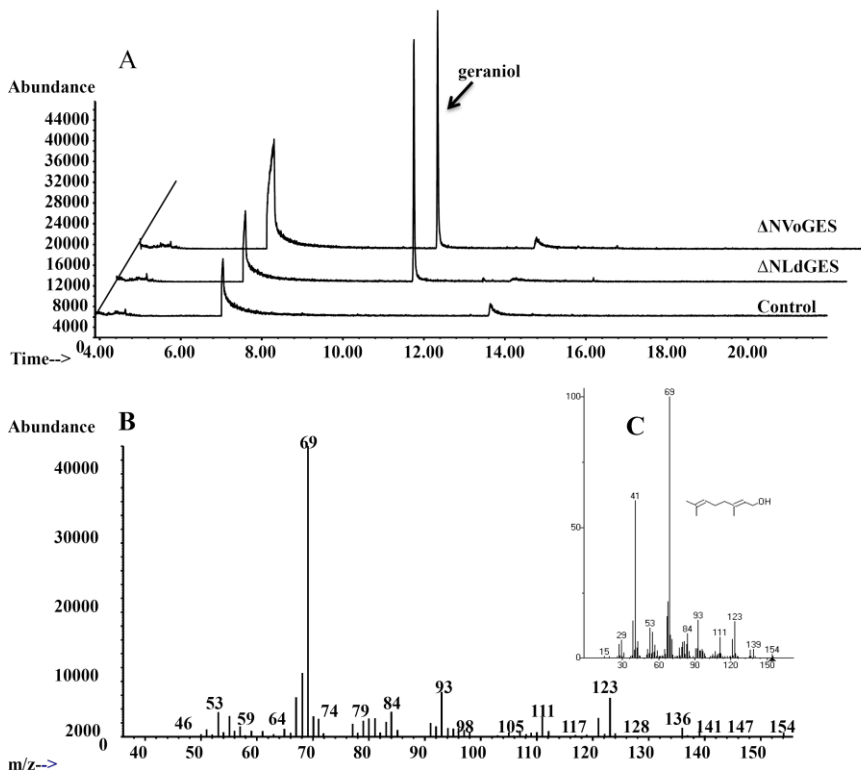
**Table 1** Comparison of Geraniol synthase proteins from different plant species and their predicted location by Target P(<http://www.cbs.dtu.dk/services/TargetP/>).



**Figure 1** Alignment of deduced amino acid sequences of ObGES (*Ocimum basilicum* geraniol synthase, AY362553), PfGES (*Perilla frutescens* geraniol synthase, DQ234300), CtGES (*Cinnamomum tenuipilum* geraniol synthase, AJ457070), CrGES (*Catharanthus roseus* geraniol synthase, AFD64744), LdGES (*Lippia dulcis* geraniol synthase, GU136162), VvGES (*Vitis vinifera* geraniol synthase, ADR74218), and VoGES (*Valeriana officinalis* geraniol synthase) using CLUSTALX 1.83 program. The metal ion-binding motif DDXD is double underlined. The black star indicates the putative cleavage site of the targeting signal.

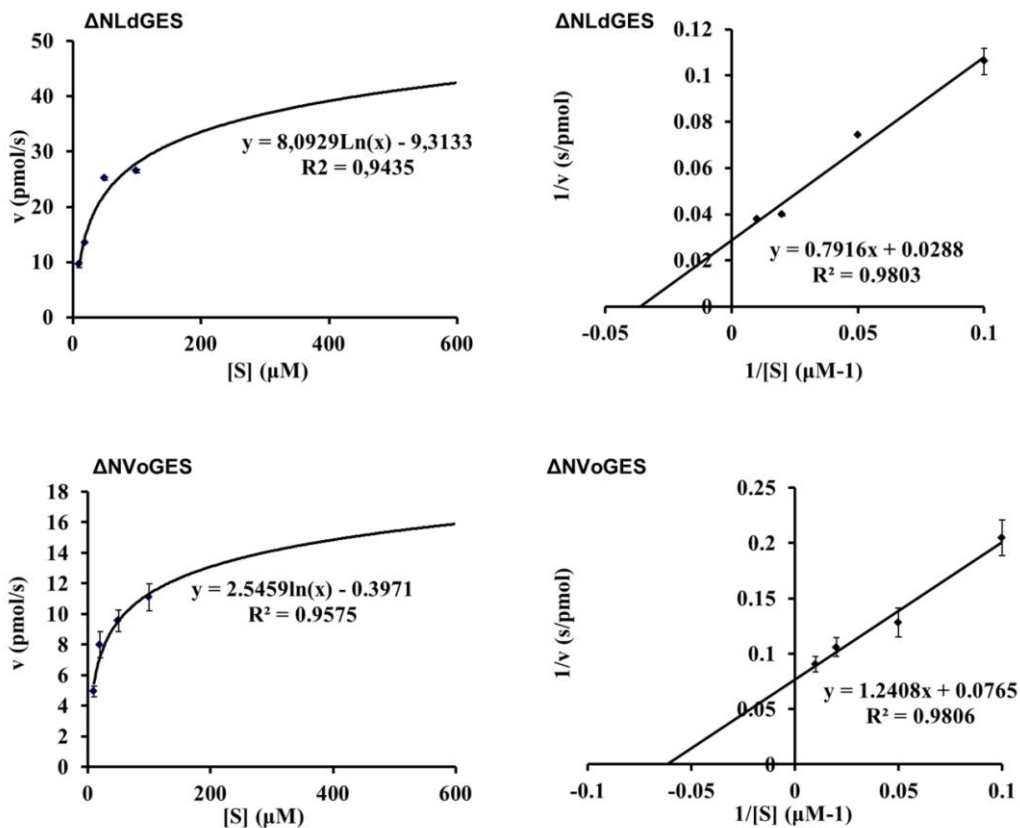


**Figure 2** GC-FID analysis of volatiles produced *in vitro* from GPP by ΔNVoGES(A) and ΔNLdGES(B) and of authentic geraniol (C). The insert in (A) shows the analysis of recombinant ΔNVoGES and ΔNLdGES protein. The protein was separated by SDS-PAGE and either stained with Coomassie Brilliant Blue (lane 1, ΔNVoGES; lane 2, ΔNLdGES) or visualized with Western blotting using anti-His antibodies (lane 3, ΔNVoGES; lane 4, ΔNLdGES).



**Figure S2** GC-MS profile of volatiles produced *in vitro* from GPP by N-terminal truncated VoGES and LdGES. A. GC chromatogram, B. Mass spectrum of the peak at 11.16 min, C. Mass spectrum of geraniol from the NIST library.



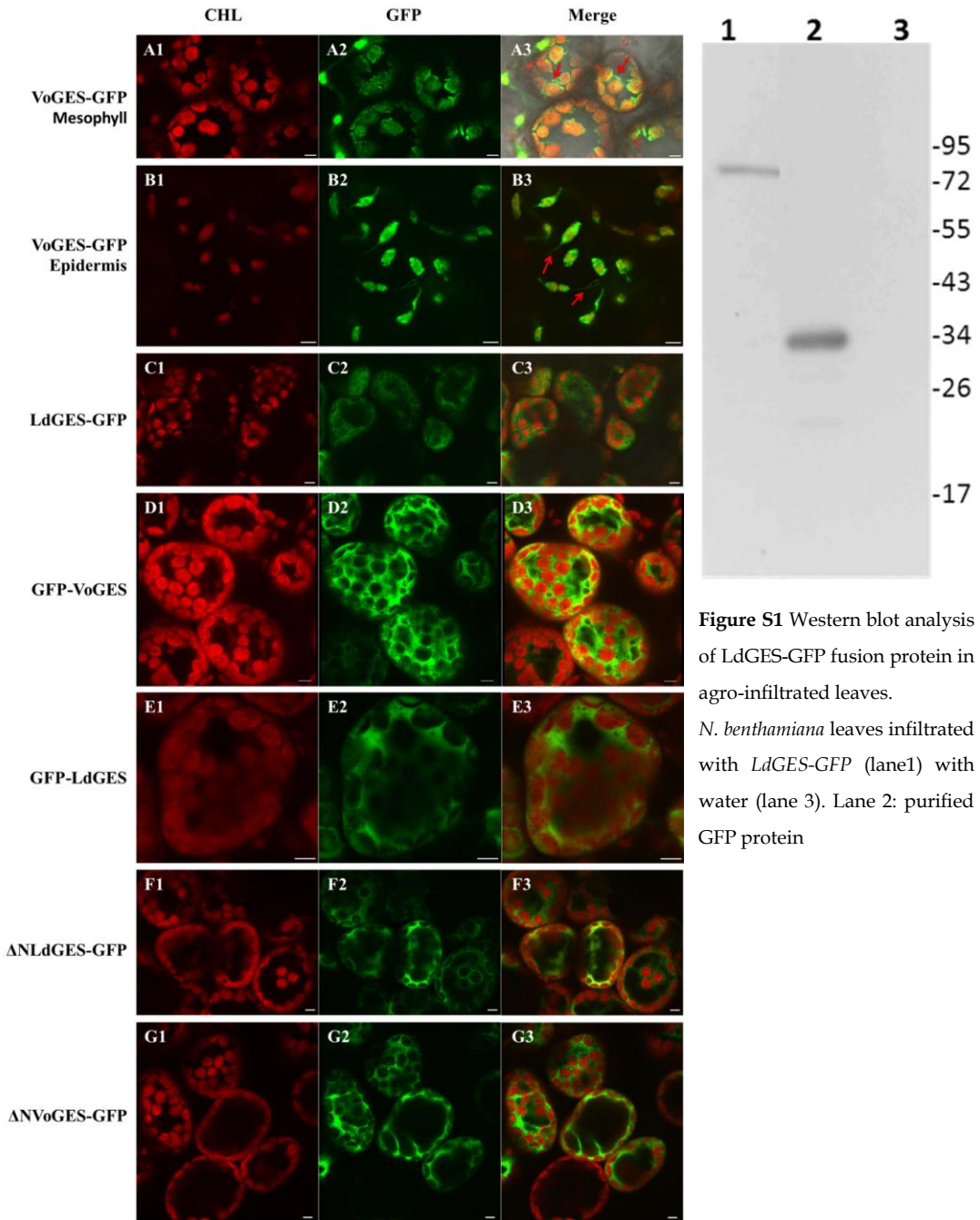


**Figure S3** Michaelis-Menten (left) and Lineweaver-Burk (right) plots of the geraniol production from GPP as catalyzed by truncated VoGES and LdGES. The data are means  $\pm$  standard errors of three replicates.

### 3.2. Different subcellular localization of VoGES and LdGES

Monoterpene synthases are generally believed to be localized in the plastids, and also their precursor GPP is supposed to be produced in the plastids. Target P (Emanuelsson et al., 2000) predicts plastid targeting for all GES proteins except ObGES, which is predicted to be targeted to the mitochondria (Table 1). Transient expression of the GES-GFP fusion proteins in *N. benthamiana* leaves showed that for VoGES-GFP the fluorescence signal was located in the plastids (Figure 3A). The fluorescence of VoGES-GFP was also detected in the stromules that emanate from the plastids, which was most clearly observed in epidermal cells (Figure 3B). However, for LdGES-GFP, the GFP fluorescence was observed in the cytosol and around the plastids (Figure 3C). No stromule labelling was observed with LdGES-GFP in

epidermal cells, confirming impaired import of LdGES into plastids and suggesting preferred localization in the cytosol. The cytosolic signal of GFP fluorescence can also be explained by instability of the LdGES-GFP fusion protein: if the GFP moiety is cleaved from the fusion protein in the cells, this would result in cytosolic localization. We therefore checked the integrity of the fusion proteins after transient expression in *N. benthamiana* on Western blots (Figure S1). The western blot shows that the signal detected by anti-GFP is of the predicted size of the LdGES-GFP fusion protein (91 kD), indicating that the fluorescence signal is derived from the intact fusion protein, which for LdGES-GFP is located in mostly in the cytosol. For the N-terminal fusion proteins, both GFP-VoGES and GFP-LdGES (full length and truncated one) located (as expected) to the cytosol (Figure 3D and E, result only show the GFP with full length). The fluorescence signal for LdGES-GFP and GFP-LdGES fusion proteins was consistently lower than that of VoGES-GFP, suggesting a lower stability for the LdGES protein.



**Figure S1** Western blot analysis of LdGES-GFP fusion protein in agro-infiltrated leaves. *N. benthamiana* leaves infiltrated with *LdGES-GFP* (lane1) with water (lane 3). Lane 2: purified GFP protein

**Figure 3** Visualization of subcellular compartments of *N. benthamiana* cells using transient transformation with VoGES and LdGES fusion with epi-fluorescence GFP (e-GFP). A. VoGES-GFP (mesophyll cell), B. VoGES-GFP (epidermal cell), C. LdGES-GFP, D. GFP-VoGES, E. GFP-LdGES. F.  $\Delta$ NLdGES-GFP, G.  $\Delta$ NVoGES-GFP. CHL, chlorophyll autofluorescence (red), GFP, GFP fluorescence (green), Merge, merged green and red images. Arrows indicate stromules. Bar represents 5  $\mu$ m.

### 3.3. The different subcellular localization of VoGES and LdGES does not affect transient *in planta* activity

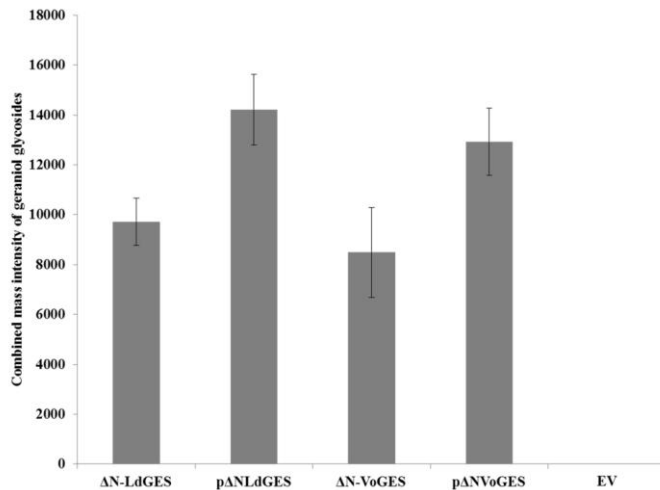
GES is a monoterpene synthase which uses the substrate GPP, which is mostly synthesized in the plastidic MEP pathway. Because of the different subcellular location of VoGES and LdGES (Figure 3), it was of interest to compare their *in planta* activity. For this purpose, the native VoGES, LdGES and the VoGES-GFP and LdGES-GFP fusion proteins were expressed transiently in *N. benthamiana* leaves. Ten days after agro-infiltration the plastidic (VoGES-GFP) and cytosolic (LdGES-GFP) targeting of the fusion protein was again confirmed by confocal fluorescence microscopy. No free geraniol was detected in the headspace of leaves expressing VoGES or LdGES and their GFP fusions when assayed at ten days post-agroinfiltration (not shown). However, previous experiments with expression of LdGES in maize had shown that the LdGES product geraniol may be converted by endogenous enzymes to geranial, geranic acid and other geraniol related products, which subsequently are sequestered as glycosylated compounds (Yang et al., 2011). Therefore, *N. benthamiana* leaves expressing the different constructs were extracted with citrate phosphate buffer and either mock treated or treated with Viscozyme, to release any glycosylated geraniol compounds. The released geraniol, geranial and geranic acid were trapped in the pentane overlay. Indeed, without glycosidase treatment, the levels of free geraniol and geraniol derivatives were below the level of detection by GC-MS in the citrate phosphate leaf extracts. In contrast, with Viscozyme treatment of the leaf samples, geraniol and the geraniol related compounds geranial, geranic acid, nerol, and neral were detected by GC-MS (Table 2). No geraniol or geraniol related products were detected after Viscozyme treatment of *N. benthamiana* leaves infiltrated with empty vector (Table 2). There was no significant difference in geraniol and the different geraniol derived products from free GES or the GES-GFP fusion proteins, indicating that the C-terminal fusion of GFP does not impair the GES activity. Moreover, there was not much difference in GES related product accumulation for the full length VoGES and LdGES (Table 2). The presence of the products geranial, geranic acid, nerol, and neral indicates that geraniol was further converted to these products, likely by endogenous *N. benthamiana* enzymes.

	Geraniol		Geranial		Geranic acid		Nerol		Neral	
	Vis-	Vis+	Vis-	Vis+	Vis-	Vis+	Vis-	Vis+	Vis-	Vis+
<i>N. benthamiana</i> VoGES	nd	47.82 ± 5.14	nd	1.05 ± 0.11	nd	6.10 ± 0.27	nd	0.95 ± 0.95	nd	0.40 ± 0.06
<i>N. benthamiana</i> LdGES	nd	45.10 ± 10.24	nd	0.58 ± 0.29	nd	5.56 ± 1.15	nd	nd	nd	0.16 ± 0.08
<i>N. benthamiana</i> VoGES-GFP	nd	46.23 ± 3.85	nd	0.95 ± 0.18	nd	7.94 ± 1.25	nd	0.48 ± 0.08	nd	0.36 ± 0.06
<i>N. benthamiana</i> LdGES-GFP	nd	72.07 ± 17.55	nd	0.53 ± 0.38	nd	31.95 ± 10.51	nd	1.17 ± 0.48	nd	0.32 ± 0.16
<i>N. benthamiana</i> empty vector	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd
<i>N. tobacum</i> VoGES leaf	4.25 ± 1.16	32.04 ± 9.53	nd	0.35 ± 0.05	nd	3.78 ± 0.22	nd	nd	nd	nd
<i>N. tobacum</i> VoGES flower	1.27 ± 0.21	6.39 ± 0.79	nd	0.27 ± 0.04	nd	5.79 ± 1.39	nd	nd	nd	nd
<i>N. tobacum</i> wild type	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd

Results are the mean of 3 replicates ± standard deviation (µg/g fw). nd: not detected. Vis+ : with Viscozyme treatment; Vis-: without Viscozyme treatment.

**Table 2** GC-MS quantification of geraniol and its derivatives in *N. benthamiana* leaves, *N. tobacum* leaves and flowers extracts

To determine if geraniol production is possible from a cytosolic localized GES we tested an N-terminal truncated version of both *VoGES* and *LdGES* ( $\Delta NVoGES$ -GFP and  $\Delta NLdGES$ -GFP, respectively), which lack the plastid import signal. As expected, both these constructs showed cytosolic localization when expressed in *N. benthamiana* (Figure 3F and 3G). In addition, we made expression constructs of the  $\Delta NLdGES$  and  $\Delta NVoGES$  with an N-terminal fusion to the same *A. thaliana* ribulose-1,5-bisphosphate carboxylase small subunit transit peptide ( $p\Delta NLdGES$  and  $p\Delta NVoGES$ , respectively) (Wong et al., 1992). The transient expression of  $\Delta NLdGES$  and  $\Delta NVoGES$  in *N. benthamiana* leaves resulted in similar levels of geraniol glycosides, but levels were about 30% lower than that produced by truncated genes fused to the heterologous plastid targeted signal ( $p\Delta NLdGES$  and  $p\Delta NVoGES$ ) of the native *VoGES* and *LdGES* (Figure 4).



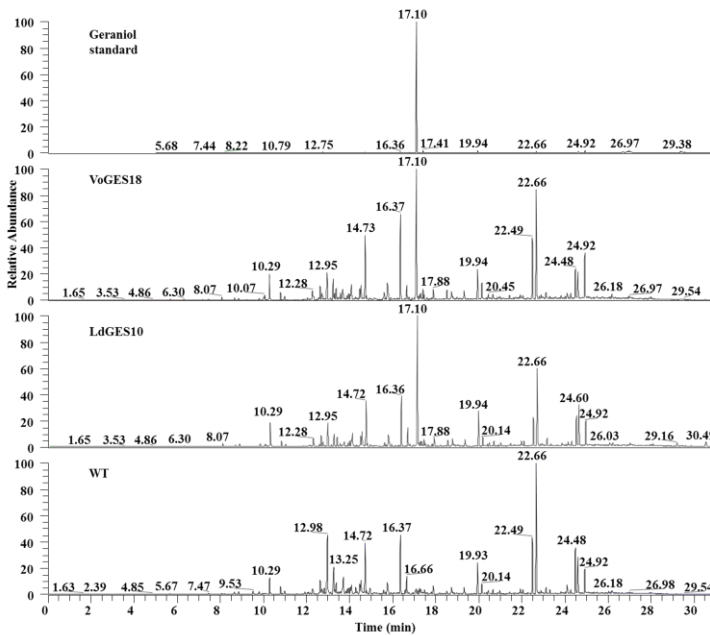
**Figure 4** Relative quantitative analysis of geraniol glycosides in agro-infiltrated *N. benthamiana* leaves with truncate ( $\Delta N$ ), artificial plastidic target(p) LdGES and VoGES and empty vector control (EV) by LC-MS.

Bars represent the total mass intensity of all geraniol-derived glycosides (malonyl-hexosyl-geraniol, acetyl-hexosyl-hydroxy-dihydro-geranic acid, acetyl-dihexosyl-geraniol, hexosyl-carboxygeranic acid, hexosyl-hydroxygeraniol, pentosyl-hexosyl-geraniol and pentosyl-hexosyl-hydroxy-tetrahydro-geranic acid).

3.4. *Similar activity of VoGES and LdGES in stably transformed tobacco plants*

Headspace analysis of stable transformants of *Nicotiana tabacum* L. 'Samsun NN' showed no geraniol in the headspace of wild type plantlets, but plantlets transformed with either *VoGES* or *LdGES* expression constructs did emit geraniol into the headspace (Figure 5A). Geraniol production differed between lines, but the production of geraniol was similar in the highest producer of *VoGES* and highest producer of *LdGES* transformants (Figure 5B).

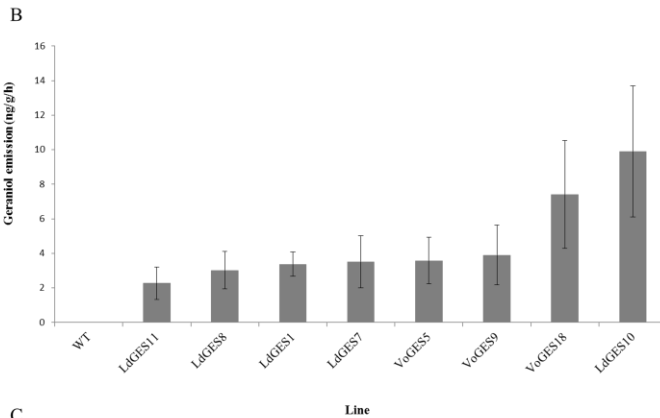
To analyse the production of non-volatile geraniol-derived products produced *in planta* in *VoGES* or *LdGES* transformed plants, the semipolar, non-volatile metabolites were analysed by LC-QTOF-MS. PCA of reconstituted mass peaks (see materials and methods) shows clustering of the replicate groups and separation between WT and transgenic lines in the first component (49.7% of the variation; Figure 5C). There was no separation between *VoGES* and *LdGES*, confirming that both GES proteins have very similar *in planta* activity, despite their different subcellular localization. Because of strong similarity in chemical profiles between the *VoGES* and *LdGES* lines, subsequently only *VoGES*18 was analyzed in more detail.



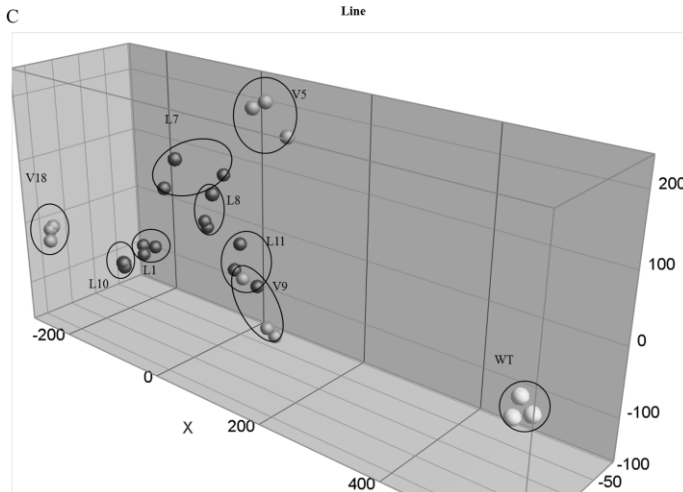
**Figure 5** GC-MS headspace analysis of volatiles and principal component analysis of non-volatiles on transgenic and wild type tobacco plantlets

A. GC chromatogram showing selected ion 69 of a geraniol standard, transgenic line VoGES18, LdGES10 and WT. Geraniol eluted at retention time 17.10 min.

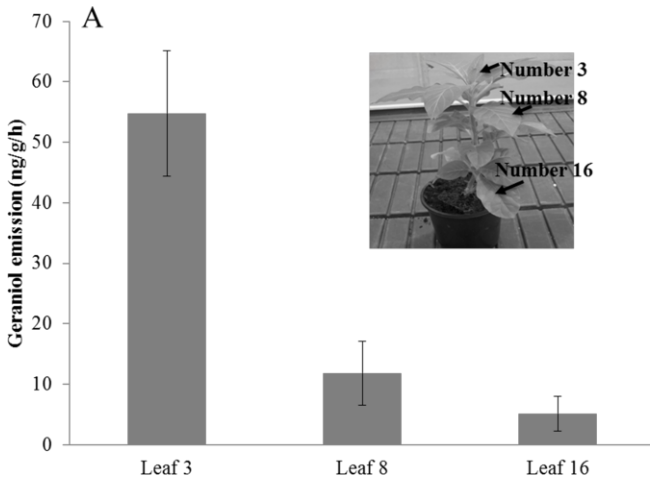
B. Geraniol concentration in the headspace of tobacco plantlets (n= 3 groups ~50 plants)



C. Principal component analysis of the LC-MS data on tobacco plantlets. PC1, PC2 and PC3 describe 49.7%, 21.5% and 15.3% of the total metabolic variation, respectively. V5, V9 and V18 indicate independent transgenic lines of VoGES; L1, L7, L8, L10 and L11 indicate independent transgenic lines of LdGES (three biological replicates of each line).





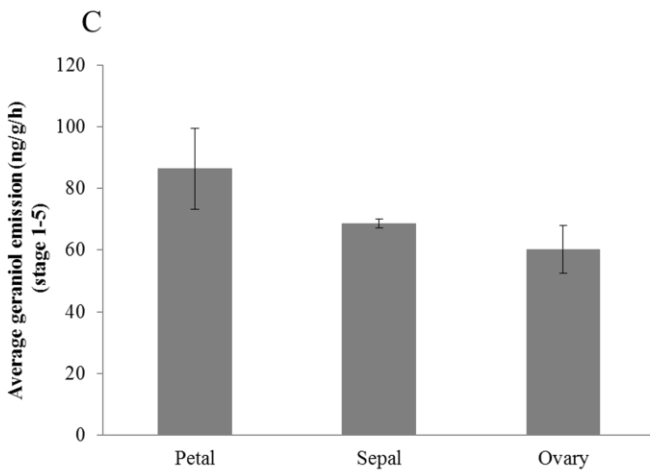
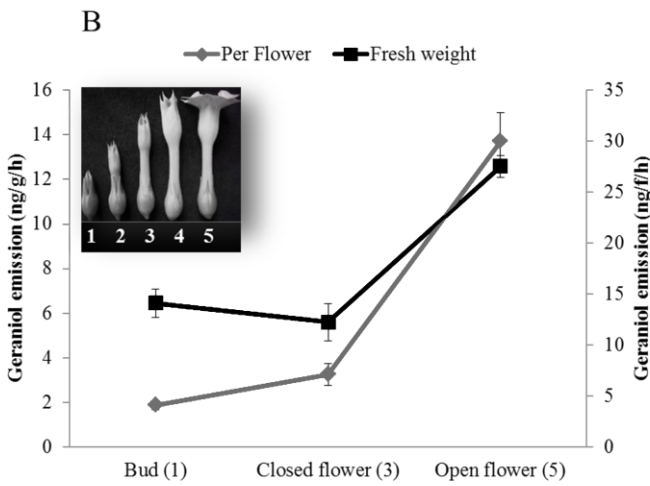


**Figure 6** Headspace analysis of different tissues of transgenic VoGES tobacco plants

A. Geraniol concentration of different developmental stage leaves. The numbers 3, 8 and 16 represent young (8 cm), middle (20 cm) and old (20 cm) age leaves, respectively.

B. Geraniol emission of different developmental stage flowers.

C. Geraniol emission of different parts of flowers.



### 3.5. Different geraniol related compounds in flowers and leaves of transformed plants expressing GES

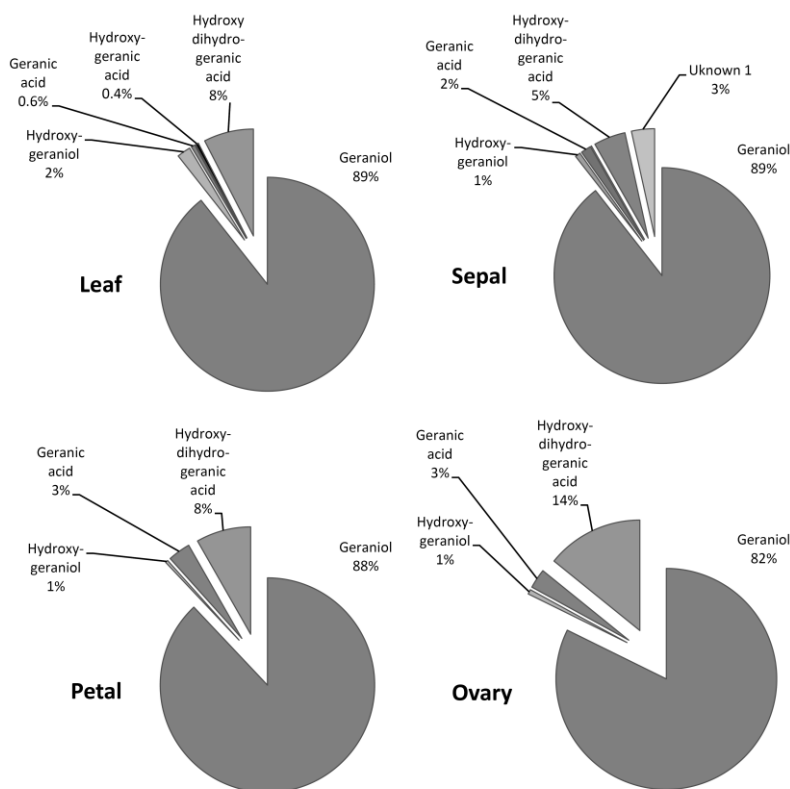
To determine free geraniol emission capacity of different parts of the plant, the headspace of leaves and flowers of different developmental stages was analyzed. Leaves were harvested from the plant at the onset of flowering. The headspace volatiles were collected for 5 hours during the day. Analysis by GC-MS showed that the youngest leaves emit most geraniol and emission decreased in older leaves, while there was no significant difference between middle and old leaves (Figure 6A). Analysis of flowers in different stages of development showed that emission was highest in freshly opened flowers (Figure 6B). To understand which part of the flower contributes most to the geraniol emission, different parts of the tobacco flowers were isolated and separately analysed for headspace production. Figure 6C shows that petal, sepal and ovary contribute almost equally to the geraniol in the total flower headspace. However, we noted that after dissection of the flower the production of geraniol increased almost 10-fold compared with the intact flower, possibly as a result of a wound induced increase in the flux through the MEP pathway (Bonaventure and Baldwin, 2010; Cordoba et al., 2009). Alternatively, it is possible that wounding results in induction of glycosidases that could liberate glycosidically bound geraniol (Edwards and Wratten, 1983; Hoagland, 1990).

To determine the level of geraniol-derived glycosides in different tissues, leaves and flowers were treated with Viscozyme. Analysis of the released volatiles by GC-MS showed that without glycosidase treatment only geraniol is detected in the tissue extracts. Viscozyme treatment strongly increased geraniol production and now also geranial, and geranic acid were detected (Table 2). Combined, the level of glycosylated geraniol products in leaves (32.04  $\mu\text{g/g}$  fw) were about 5-fold higher than in flowers (6.39  $\mu\text{g/g}$  fw).

For analysis of the conjugated GES related products in the transformants the LC-MS chromatograms of WT and VoGES18 plants were compared and 19 compounds were identified with significantly higher signal in the transgenic line (Table 3). For identification of these putative geraniol-related products, the masses were targeted for fragmentation by LTQ Orbitrap MSn. According to the MSn spectra from 18 of the differential masses, the compounds are likely hexose, pentose and malonyl or acetyl conjugates of geraniol, hydroxy-

Retention time	Observed mass	Putative identity	Calculated mass	$\Delta$ Mass (ppm)	Elemental formula	MS(n) fragment	Tissue
24.04	345.1563	Hexosyl-hydroxy-geranic acid	345.1555	2.35	$C_{16}H_{25}O_8$	131, 143, 149, 161, 167, 179, 183, 206, 219, 261, 282, 299, 307, 327, 331	L
30.95	435.1862	Acetyl-hexosyl-hydroxy-dihydro-geranic acid (formic acid adduct)	435.1866	-0.92	$C_{19}H_{31}O_{11}$	157, 185, 205, 231, 249, 267, 291, 309, 315, 355, 375, 389, 417	L, S, P, O
32.98	435.1869	Acetyl-hexosyl-hydroxy-dihydro-geranic acid (formic acid adduct)	435.1866	0.69	$C_{19}H_{31}O_{11}$	131, 157, 185, 205, 231, 249, 267, 291, 309, 359, 373, 391, 417	L, S, P, O
40.20	447.2242	Pentosyl-hexosyl-geraniol	447.2235	1.52	$C_{21}H_{35}O_{10}$	131, 143, 149, 161, 179, 191, 233, 251, 293, 311, 315, 375, 401, 430	L, S, P, O
39.88	447.2243	Pentosyl-hexosyl-geraniol	447.2235	1.87	$C_{21}H_{35}O_{10}$	131, 143, 149, 161, 179, 191, 221, 239, 251, 275, 281, 293, 311, 315	L, S, P, O
39.17	447.2245	Pentosyl-hexosyl-geraniol	447.2235	2.28	$C_{21}H_{35}O_{10}$	131, 143, 149, 161, 179, 191, 217, 221, 233, 251, 281, 293, 311, 347, 420	L, S, P, O
38.45	461.2036	Pentosyl-hexosyl-geranic acid	461.2023	2.82	$C_{21}H_{33}O_{11}$	149, 191, 221, 239, 279, 293, 311, 339, 373, 389	L, S, P, O
42.41	461.2393	Deoxyhexosyl-hexosyl-geraniol	461.2392	0.21	$C_{22}H_{37}O_{10}$	161, 179, 213, 267, 295, 307, 323, 373, 385, 397, 415, 447	L, S, P, O
23.54	463.2178	Pentosyl-hexosyl-hydroxy-geraniol	463.2185	-1.40	$C_{21}H_{35}O_{11}$	143, 149, 161, 191, 221, 251, 275, 293, 311, 331, 349, 387, 425	L, S
36.50	477.2335	Dihexosyl-geraniol	477.2331	0.75	$C_{22}H_{37}O_{11}$	161, 179, 221, 251, 281, 307, 323, 341, 361, 421, 449	L, S, P, O
25.37	611.2547	Dipentosyl-hexosyl-hydroxy-dihydro-geranic acid	611.2551	-0.65	$C_{26}H_{43}O_{16}$	149, 161, 191, 221, 251, 281, 293, 317, 347, 395, 429, 441, 461, 479	L, S, P
24.95	611.2560	Dipentosyl-hexosyl-hydroxy-dihydro-geranic acid	611.2551	1.47	$C_{26}H_{43}O_{16}$	149, 185, 221, 251, 275, 293, 311, 333, 360, 377, 435, 449, 461, 479	L, S, P
44.84	489.2329	Acetyl-pentosyl-hexosyl-geraniol	489.2336	-1.43	$C_{23}H_{37}O_{11}$	131, 149, 161, 191, 203, 233, 315, 327, 335, 367, 447, 463	L, S, P, O
47.39	503.2486	Acetyl-deoxyhexose-hexose-geraniol	503.2492	-1.19	$C_{24}H_{39}O_{11}$	143, 161, 203, 247, 307, 329, 443, 461, 477, 233, 293, 352, 372, 401, 425, 463, 493, 511, 567	L, S, P, O
37.06	579.2661	Dipentosyl-hexosyl-geraniol	579.2653	1.40	$C_{26}H_{43}O_{14}$	191, 242, 267, 299, 315, 355, 387, 413, 447, 455, 475, 499, 537, 571	L, S, P, O
34.30	609.2729	Pentosyl-dihexosyl-geraniol	609.2753	-3.94	$C_{27}H_{45}O_{15}$	161, 173, 399, 445, 449, 467, 495, 541, 583, 609, 663, 699	S
34.08	709.3562	Unknown 1	709.3588	-3.67	$C_{40}H_{53}O_{11}$	131, 149, 161, 221, 329, 357, 401, 450, 501, 597, 643, 717	L, S, P
47.95	803.3712	Malonyl-hexosyl-geraniol (dimer)	803.3701	1.31	$C_{38}H_{50}O_{18}$		
48.72	831.3299	Malonyl-hexosyl-geranic acid (dimer)	831.3287	1.44	$C_{38}H_{49}O_{20}$	161, 167, 191, 265, 307, 349, 371, 415	S, P, O

geraniol, geranic acid, hydroxy-geranic acid and hydroxy-dihydro-geranic acid. One mass (709.3562), which is specific for sepal tissue could not be identified as a geraniol related product (Table 3). A malonylated hexosyl geranic acid conjugate was present in flower tissues but not in leaves, whereas a hexosyl hydroxyl-geranic acid conjugate was specific for leaves. Combined, the results suggest that the glycosylating activities for geraniol and geraniol derivatives are tissue-dependent. According to the mass intensity signal of the different glycoside conjugates in leaves, 89% were directly derived from geraniol, while 11% were apparently derived from geraniol after further modification to hydroxy-geraniol, geranic acid, hydroxy-geranic acid or hydroxy-dihydro-geranic acid (Figure 7). In flowers, the proportion of geraniol and further modified products was slightly different from that in leaves in sepal, petal and ovary (Figure 7), likely as a result of different endogenous enzyme activities in different tissues.



**Figure 7** Components of geraniol related aglycons based on mass intensity of putative glycosides in VoGES tobacco leaves and flowers.

## 4. Discussion

### 4.1. Localisation of VoGES in the plastids and LdGES mostly in the cytosol has no effect on *in planta* activity

In this study, we show that both VoGES and LdGES only produce the product geraniol *in vitro* as specific product from GPP (Figure 2). Both VoGES and LdGES contain a putative plastid signal peptide (Table 1) and monoterpene synthases are generally believed to be active in the plastids (Bouvier et al., 2000; Dudareva et al., 2005; Turner et al., 1999). For instance, CrGES from *C. roseus* has been shown to be localized in the plastids (Simkin et al., 2012). Our localization experiments showed that VoGES directs the GFP protein to the plastids, as expected, and no cytosolic signal was observed from transient expression of VoGES-GFP in agroinfiltrated leaves of *N. benthamiana*, indicating that import into the plastids was not saturated. The plastidic localization of VoGES-GFP was also evident from GFP signal in the stromules extending from the plastids (Figure 3A-B). In contrast, the signal of LdGES-GFP was mostly in the cytosol. Although some fluorescence in or around the plastids was observed, no LdGES-GFP signal was localized to the stromules of the plastids (Figure 2C). Control experiments show that the cytosolic signal from LdGES-GFP does not result from cleavage of GFP from the fusion protein (Figure S1). The GFP localization studies thus suggest a predominantly cytosolic localization of LdGES. Because product analysis (Table 2 and Figure 5) shows that native VoGES and LdGES have similar activity *in planta*, this suggests that the substrate pool of GPP is not limiting for native LdGES in the cytosol. However, we also note that activity of the truncated VoGES and LdGES proteins (both localized to the cytosol) was only 30% of plastid localized VoGES and LdGES, suggesting that (although for these cytosolic GES proteins the GPP substrate is available) GPP is not as abundant available as in plastids. Indeed, it recently was shown that in tomato fruits cytosolic monoterpene biosynthesis can be supported by plastid-generated geranyl diphosphate substrate (Gutensohn et al., 2013). The higher activity of native (mostly) cytosolic LdGES and truncated  $\Delta$ NLdGES could be related to the partial plastidic localization of native LdGES.

#### 4.2. Multiple oxidized and glycosylated forms of geraniol produced by endogenous enzymes

In our experiments approximately 11% of the produced geraniol was converted to the oxidized products geranial, hydroxy-geraniol, geranic acid, hydroxy-geranic acid and hydroxy-dihydro-geranic acid, which were mostly present as conjugates (Figure 7, Table 3). These conversions of geraniol are likely catalyzed by endogenous tobacco enzymes. For examples, geranic acid is likely to be synthesized from geranial by an NAD<sup>+</sup> dependent dehydrogenase (Boyer and Petersen, 1991; Campos-García, 2010; Davidovich-Rikanati et al., 2007; Lüddecke et al., 2012). Geranic acid was only present in glycosylated forms in the transgenic plants and not as the free form (Figure 8). Because low levels of geranial were detected only after and not before de-glycosylation, presumably this product is formed from oxidation of geraniol during the de-glycosylation reaction.

#### 4.3. Biotransformation of geraniol in different hosts

In this study, VoGES and LdGES were expressed in both a microbial, *E. coli*, and plant host (*N. benthamiana* and *N. tabacum*). With *E. coli* protein, we could only detect geraniol as product, while geraniol, geranial and geranic acid were observed in transiently expressing *N. benthamiana* as well as stable transformed *N. tabacum*. In addition, in *N. benthamiana* small amounts of nerol and neral were detected (Table 2). Previously it was shown that LdGES expressed in maize also results in geranyl acetate production in leaves (Yang et al., 2011), while this product was absent in *N. benthamiana* or tobacco plants expressing the same LdGES (this paper). This shows that different heterologous GES hosts metabolize the produced geraniol differently, and that especially plants tend to produce a wider spectrum of geraniol derivatives. Similar observation were made with expression of *Ocimum basilicum* geraniol synthase (ObGES) in yeast, *E. coli*, grape and tobacco (Fischer et al., 2012) and tomato (Davidovich-Rikanati et al., 2007). As expected, in all hosts the main product of ObGES was geraniol, but minor product (linalool, citronellol, nerol,  $\alpha$ -terpineol...) formation differed between the heterologous hosts.

#### 4.4. *Different glycosylation patterns of geraniol products in different tissues and stages of development*

Geraniol and geraniol derivatives accumulated in glycosylated form, some of which were further modified by endogenous malonyltransferases or acetyltransferases. The glycosides were mono-glycoside, di-glycosides and tri-glycosides. The hexoside, ubiquitous in nature, which can potentially be attached to geraniol is glucopyranoside, whereas the pentosides could be arabinofuranoside, arabinopyranoside, apiofuranoside, xylopyranoside or rhamnopyranoside (Genovés et al., 2005; Maicas and Mateo, 2005; Mateo and Jiménez, 2000). Presumably, the glycosylated geraniol products are sequestered in the vacuole (Wink, 2010; Winterhalter and Skouroumounis, 1997). The profile of glycosylated geraniol products differed between the agro-infiltrated *N. benthamiana* leaves and leaves of the stable transformed tobacco plants (Table 2). The glycosylation profile also differed between tissues and developmental stages within the same plant (Table 3). For instance, in the mature leaves of stable transformed tobacco plants expressing GES, the acetylated and malonylated glycosides of geraniol are ~5-fold higher than in seedlings (data not shown), suggesting that acetylation and malonylation are induced later during development. In total, the bioconversion by endogenous enzyme activities distributed the product of VoGES, geraniol, over at least 19 different compounds. Clearly, for effective reconstruction of a multiple step terpene biosynthesis pathway, such diversion of products needs to be avoided in order to raise the desired product yield. The use of different promoters, driving expression in tissues more specialized for the production and/or storage of terpenoids, could be one way to achieve this.

#### 4.5. *No pleiotropic effects of GES overexpression*

Introduction of GES into tobacco could potentially result in competition for the substrate GPP with other monoterpene synthases or result in depletion of the substrates isopentenyl diphosphate and dimethylallyl diphosphate which are important for the formation of other terpenes, carotenoids and gibberellins, and this could have a dramatic effect on the plant

phenotype (Aharoni et al., 2003; Davidovich-Rikanati et al., 2007). However, no leaf or flower phenotype was observed in our stable transformed tobacco plants expressing VoGES or LdGES, despite high transgenic product levels suggesting that sufficient GPP is available in both leaves and flowers. Also, in the LC-MS and GC-MS analysis of transgenic plants or after transient expression of GES genes in *N. benthamiana*, only new products were detected but no obvious reduction in endogenous products, which suggests there is indeed no competition for substrate.

In the selected transgenic lines no pleiotropic effect of GES overexpression was observed. We cannot exclude that in the regeneration of stable transformed tobacco plants with 35S-VoGES or 35S-LdGES a selection occurred against very high GES expression, resulting in either toxic levels of geraniol or creating a drain of GPP required for other processes. Boosting of the GES production in the stable transformed tobacco plants (e.g. by combining with GPPS overexpression) could result in toxic effects of geraniol, but these effects may be eliminated by simultaneous co-expression with P450 enzymes which convert geraniol to less reactive intermediates/end products and/or by using tissue-specific promoters to drive geraniol production in tissues more specialized for terpene production and storage (for example trichomes).

#### 4.6. Higher fluxes through geraniol sequestration than through geraniol emission

We note that emission of geraniol in stable transformed tobacco plants changed during development. For instance, in young seedlings emission was ~8 ng/g h, while in mature plants emission by leaves declined from young to old leaves from 55 to 5 ng/g h (Figure 5 and Figure 6). In flowers of stable transformants emission of geraniol increased until flowers were fully open from 12 to 28 ng/g h (Figure 6). Based on the average headspace emission, the loss of geraniol from leaves would be ~17 µg/g month and from flowers ~13 µg/g month, while 37 µg/g fw of geraniol (derived) products (determined by de-glycosylation) are stored in 1 month old tobacco leaves (Table 2). *N. benthamiana* leaves at four days post-agroinfiltration show an average emission rate of 0.13 ng/g h (not shown), while no geraniol was emitted into the headspace ten days post-agro-infiltration (data not shown). This might suggest that after ten days not only has the T-DNA with the expression construct degraded,



but that also the GES protein is no longer active in the cells, indicating a relatively high turnover of terpene synthase protein.

Assuming that (1) effective expression in transient assays is from day 4-7 and that (2) geraniol production is at least 0.13 ng/g h in *N. benthamiana* leaves and that (3) flux through the pathway is 24 hr/day (which may not be the case if precursors are preferably produced in the light (Aharoni et al., 2003), the cumulative production of emitted geraniol over 4 days of transient expression would be  $\sim 0.012 \mu\text{g/g fw}$ . However, at ten days post agroinfiltration, the combined level of accumulated geraniol (derived) products (determined by deglycosylation) was around  $56 \mu\text{g/g fw}$  (Table 2).

Assuming that geraniol derived products in *N. benthamiana* accumulate during four days of continuous activity, this suggests a geraniol production of  $\sim 0.58 \mu\text{g/g h fw}$ . Total level of geraniol derived products in the stable transformants is  $\sim 37 \mu\text{g/g fw}$  and that is produced in 1 month, the geraniol production rate of stable transformants in tobacco is only  $0.05 \mu\text{g/g h fw}$ . The high production in the transient assay is most likely due to an extremely high gene dosage number of the infiltrated GES expression construct compared to the low copy number in stable transformants. At present it is not clear why the high GES activity in the transient expression assay is not matched by a higher emission of geraniol, but one explanation may be that the conjugation capacity of infiltrated *N. benthamiana* leaves is highly efficient.

#### 4.7. Towards a monoterpene iridoid pathway in heterologous plants

Expression of GES in tobacco is the first step in a larger effort to introduce the terpene indole alkaloid (TIA) biosynthesis into tobacco plants, consisting of a monoterpene iridoid branch and an indole branch (Mahroug et al., 2007; Simkin et al., 2012). The second step of this pathway is the conversion of geraniol to 10-hydroxy-geraniol by geraniol 10-hydroxylase (G10H), but (as described above) this conversion may already take place by endogenous tobacco enzymes (Table 3). We are currently testing whether co-expression of VoGES with a G10H can effectively compete for the endogenous unwanted oxidation and glycosylation of the primary product geraniol. In the context of multiple-step monoterpene pathway reconstruction in heterologous plant hosts, elimination of the conversion by endogenous

enzyme activities may become an important target for efficient channeling towards the desired product(s).

## Acknowledgements

This research was funded by the European Community's Framework VII Program FP7/2007-2013 to the SMARTCELL project KBBE-2007-3-1-01. Lemeng Dong was supported by a PhD fellowship from China Scholarship Council. We thank Bert Schipper and Ric de Vos for their help with LC-MS analysis.

## References

- Aharoni, A., Giri, A. P., Deuerlein, S., Griepink, F., de Kogel, W. J., Verstappen, F. W. A., Verhoeven, H. A., Jongsma, M. A., Schwab, W., Bouwmeester, H. J., 2003. Terpenoid metabolism in wild-type and transgenic *Arabidopsis* plants. *Plant Cell* 15, 2866-2884.
- Aharoni, A., Giri, A. P., Verstappen, F. W. A., Berteaux, C. M., Sevenier, R., Sun, Z., Jongsma, M. A., Schwab, W., Bouwmeester, H. J., 2004. Gain and loss of fruit flavor compounds produced by wild and cultivated strawberry species. *Plant Cell* 16, 3110-3131.
- Altschul, S., Madden, T., Schaffer, A., Zhang, J., Zhang, Z., Miller, W., Lipman, D., 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* 25, 3389-3402.
- Antonelli, A., Fabbri, C., Giorgioni, M. E., Bazzocchi, R., 1997. Characterization of 24 old garden roses from their volatile compositions. *J. Agric. Food Chem.* 45, 4435-4439.
- Bakkali, F., Averbeck, S., Averbeck, D., Idaomar, M., 2008. Biological effects of essential oils-a review. *Food Chem. Toxicol.* 46, 446-475.
- Bard, M., Albrecht, M. R., Gupta, N., Guynn, C. J., Stillwell, W., 1988. Geraniol interferes with membrane functions in strains of *Candida* and *Saccharomyces*. *Lipids* 23, 534-538.
- Bayrak, A., Akgül, A., 1994. Volatile oil composition of Turkish rose (*Rosa damascena*). *J. Sci. Food Agric.* 64, 441-448.

- Bohlmann, J., Meyer-Gauen, G., Croteau, R., 1998. Plant terpenoid synthases: molecular biology and phylogenetic analysis. *P. Natl. Acad. Sci. USA* 95, 4126-4133.
- Bonaventure, G., Baldwin, I. T., 2010. Transduction of wound and herbivory signals in plastids. *Commun. Integr. Biol.* 3, 313-317.
- Bouvier, F., Suire, C., d'Harlingue, A., Backhaus, R. A., Camara, B., 2000. Molecular cloning of geranyl diphosphate synthase and compartmentation of monoterpene synthesis in plant cells. *Plant J.* 24, 241-252.
- Boyer, C. S., Petersen, D. R., 1991. The metabolism of 3, 7-dimethyl-2, 6-octadienal (citral) in rat hepatic mitochondrial and cytosolic fractions. Interactions with aldehyde and alcohol dehydrogenases. *Drug Metab. Dispos.* 19, 81-86.
- Campos-García, J., 2010. Metabolism of acyclic terpenes by *Pseudomonas*. *Pseudomonas* 235-253.
- Carnesecchi, S., Schneider, Y., Ceraline, J., Duranton, B., Gosse, F., Seiler, N., Raul, F., 2001. Geraniol, a component of plant essential oils, inhibits growth and polyamine biosynthesis in human colon cancer cells. *J. Pharmacol. Exp. Ther.* 298, 197-200.
- Chen, Y., 2006. Quantification of perfume compounds in shampoo using solid phase microextraction. *Flavour Fragr. J.* 21, 822-832.
- Cordoba, E., Salmi, M., León, P., 2009. Unravelling the regulatory mechanisms that modulate the MEP pathway in higher plants. *J. Exp. Bot.* 60, 2933-2943.
- Davidovich-Rikanati, R., Sitrit, Y., Tadmor, Y., Iijima, Y., Bilenko, N., Bar, E., Carmona, B., Fallik, E., Dudai, N., Simon, J. E., Pichersky, E., Lewinsohn, E., 2007. Enrichment of tomato flavor by diversion of the early plastidial terpenoid pathway. *Nat. Biotechnol.* 25, 899-901.
- Davidovich-Rikanati, R., Lewinsohn, E., Bar, E., Iijima, Y., Pichersky, E., Sitrit, Y., 2008. Overexpression of the lemon basil  $\alpha$ -zingiberene synthase gene increases both mono- and sesquiterpene contents in tomato fruit. *Plant J.* 56, 228-238.
- De Vos, R. C. H., Moco, S., Lommen, A., Keurentjes, J. J. B., Bino, R. J., Hall, R. D., 2007. Untargeted large-scale plant metabolomics using liquid chromatography coupled to mass spectrometry. *Nat. Protoc.* 2, 778-791.

- Dudareva, N., Andersson, S., Orlova, I., Gatto, N., Reichelt, M., Rhodes, D., Boland, W., Gershenzon, J., 2005. The nonmevalonate pathway supports both monoterpene and sesquiterpene formation in snapdragon flowers. *P. Natl. Acad. Sci. USA* 102, 933-938.
- Edwards, P., Wratten, S., 1983. Wound induced defences in plants and their consequences for patterns of insect grazing. *Oecologia* 59, 88-93.
- Eisenreich, W., Rohdich, F., Bacher, A., 2001. Deoxyxylulose phosphate pathway to terpenoids. *Trends Plant Sci.* 6, 78-84.
- Emanuelsson, O., Nielsen, H., Brunak, S., von Heijne, G., 2000. Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. *J. Mol. Biol.* 300, 1005-1016.
- Fischer, M. J., Meyer, S., Claudel, P., Perrin, M., Ginglinger, J. F., Gertz, C., Masson, J. E., Werck-Reinhardt, D., Hugueney, P., Karst, F., 2012. Specificity of *Ocimum basilicum* geraniol synthase modified by its expression in different heterologous systems. *J. Biotechnol.* 163, 24-29.
- Genovés, S., Gil, J. V., Vallés, S., Casas, J. A., Manzanares, P., 2005. Assessment of the aromatic potential of palomino fino grape must using glycosidases. *Am. J. Enol. Vitic.* 56, 188-191.
- Gutensohn, M., Orlova, I., Nguyen, T. T., Davidovich-Rikanati, R., Ferruzzi, M. G., Sitrit, Y., Lewinsohn, E., Pichersky, E., Dudareva, N., 2013. Cytosolic monoterpene biosynthesis is supported by plastid-generated geranyl diphosphate substrate in transgenic tomato fruits. *Plant J.* doi: 10.1111/tpj.12212.
- Hadacek, F., 2002. Secondary metabolites as plant traits: current assessment and future perspectives. *Crit. Rev. Plant Sci.* 21, 273-322.
- Herrero, Ó., Ramón, D., Orejas, M., 2008. Engineering the *Saccharomyces cerevisiae* isoprenoid pathway for *de novo* production of aromatic monoterpenes in wine. *Metab. Eng.* 10, 78-86.
- Hoagland, R. E., 1990. Biochemical responses of plants to pathogens, in: Hoagland, R. E. (Eds.), *Microbes and Microbial Products as Herbicides*. Am. Chem. Soc. Symp. Ser. No. 439. ACS Books, Washington, DC, pp. 87-113.
- Horsch, R., Fry, J., Hoffmann, N., Eichholtz, D., Rogers, S., Fraley, R., 1985. A simple and general method for transferring genes into plants. *Science* 227, 1229-1231.

- Karimi, M., Inzé, D., Depicker, A., 2002. GATEWAY™ vectors for *Agrobacterium*-mediated plant transformation. *Trends Plant Sci.* 7, 193-195.
- Lange, B. M., Ketchum, R. E. B., Croteau, R. B., 2001. Isoprenoid biosynthesis. Metabolite profiling of peppermint oil gland secretory cells and application to herbicide target analysis. *Plant Physiol.* 127, 305-314.
- Lazo, G. R., Stein, P. A., Ludwig, R. A., 1991. A DNA transformation-competent *Arabidopsis* genomic library in *Agrobacterium*. *Nat. Biotechnol.* 9, 963-967.
- Lichtenthaler, H. K., 1999. The 1-deoxy-D-xylulose-5-phosphate pathway of isoprenoid biosynthesis in plants. *Annu. Rev. Plant Biol.* 50, 47-65.
- Liu, D. H., Ren, W. W., Cui, L. J., Zhang, L. D., Sun, X. F., Tang, K. X., 2011. Enhanced accumulation of catharanthine and vindoline in *Catharanthus roseus* hairy roots by overexpression of transcriptional factor ORCA2. *Afr. J. Biotechnol.* 10, 3260-3268.
- Lüddecke, F., Wülfing, A., Timke, M., Germer, F., Weber, J., Dikfidan, A., Rahnfeld, T., Linder, D., Meyerdierks, A., Harder, J., 2012. Geraniol and geranial dehydrogenases induced in anaerobic monoterpene degradation by *Castellaniella defragrans*. *Appl. Environ. Microbiol.* 78, 2128-2136.
- Mahroug, S., Burlat, V., St-Pierre, B., 2007. Cellular and sub-cellular organisation of the monoterpene indole alkaloid pathway in *Catharanthus roseus*. *Phytochem. Rev.* 6, 363-381.
- Maicas, S., Mateo, J. J., 2005. Hydrolysis of terpenyl glycosides in grape juice and other fruit juices: a review. *Appl. Microbiol. Biotechnol.* 67, 322-335.
- Mateo, J., Jiménez, M., 2000. Monoterpenes in grape juice and wines. *J. Chromatogr. A.* 881, 557-567.
- McConkey, M. E., Gershenzon, J., Croteau, R. B., 2000. Developmental regulation of monoterpene biosynthesis in the glandular trichomes of peppermint. *Plant Physiol.* 122, 215-224.
- Memelink, J., Gantet, P., 2007. Transcription factors involved in terpenoid indole alkaloid biosynthesis in *Catharanthus roseus*. *Phytochem. Rev.* 6, 353-362.
- Montiel, G., Breton, C., Thiersault, M., Burlat, V., Jay-Allemand, C., Gantet, P., 2007. Transcription factor Agamous-like 12 from *Arabidopsis* promotes tissue-like

- organization and alkaloid biosynthesis in *Catharanthus roseus* suspension cells. *Metab. Eng.* 9, 125-132.
- Pedersen, D. S., Capone, D. L., Skouroumounis, G. K., Pollnitz, A. P., Sefton, M. A., 2003. Quantitative analysis of geraniol, nerol, linalool, and  $\alpha$ -terpineol in wine. *Anal. Bioanal. Chem.* 375, 517-522.
- Peebles, C. A., Hughes, E. H., Shanks, J. V., San, K. Y., 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. *Metab. Eng.* 11, 76-86.
- Peebles, C. A., Sander, G. W., Hughes, E. H., Peacock, R., Shanks, J. V., San, K. Y., 2011. The expression of 1-deoxy-d-xylulose synthase and geraniol-10-hydroxylase or anthranilate synthase increases terpenoid indole alkaloid accumulation in *Catharanthus roseus* hairy roots. *Metab. Eng.* 13, 234-240.
- Porter, J. W., Spurgeon, S. L., 1981. Biosynthesis of isoprenoid compounds, in: Porter, J. W., Spurgeon, S. L. (Eds). John Wiley and Sons Inc., New York, pp. 1-46.
- Rastogi, S. C., Menne, T., Johansen, J. D., 2003. The composition of fine fragrances is changing. *Contact Dermatitis.* 48, 130-132.
- Sangwan, N., Farooqi, A., Shabih, F., Sangwan, R., 2001. Regulation of essential oil production in plants. *Plant Growth Regul.* 34, 3-21.
- Schmidt, A., Wächtler, B., Temp, U., Krekling, T., Séguin, A., Gershenzon, J., 2010. A bifunctional geranyl and geranylgeranyl diphosphate synthase is involved in terpene oleoresin formation in *Picea abies*. *Plant Physiol.* 152, 639-655.
- Simkin, A. J., Miettinen, K., Claudel, P., Burlat, V., Guirimand, G., Courdavault, V., Papon, N., Meyer, S., Godet, S., St-Pierre, B., 2012. Characterization of the plastidial geraniol synthase from *Madagascar periwinkle* which initiates the monoterpene branch of the alkaloid pathway in internal phloem associated parenchyma. *Phytochem.* 85, 36-43.
- Small, I., Peeters, N., Legeai, F., Lurin, C., 2004. Predotar: A tool for rapidly screening proteomes for N-terminal targeting sequences. *Proteomics.* 4, 1581-1590.
- Tarshis, L., Proteau, P. J., Kellogg, B. A., Sacchettini, J. C., Poulter, C. D., 1996. Regulation of product chain length by isoprenyl diphosphate synthases. *P. Natl. Acad. Sci. USA* 93, 15018.

- Tholl, D., Lee, S., 2011. Elucidating the metabolism of plant terpene volatiles: alternative tools for engineering plant defenses, in: Gang, D.R. (Eds.), *The Biological Activity of Phytochemicals*. Springer Inc., Washington, pp. 159-178.
- Thompson, J., Gibson, T., Plewniak, F., Jeanmougin, F., Higgins, D., 1997. The CLUSTAL\_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucl. Acids Res.* 25, 4876-4882.
- Tikunov, Y., Lommen, A., De Vos, C., Verhoeven, H. A., Bino, R. J., Hall, R. D., Bovy, A. G., 2005. A novel approach for nontargeted data analysis for metabolomics. Large-scale profiling of tomato fruit volatiles. *Plant Physiol.* 139, 1125-1137.
- Turner, G., Gershenzon, J., Nielson, E. E., Froehlich, J. E., Croteau, R., 1999. Limonene synthase, the enzyme responsible for monoterpene biosynthesis in peppermint, is localized to leucoplasts of oil gland secretory cells. *Plant Physiol.* 120, 879-886.
- van der Hooft, J. J. J., Vervoort, J., Bino, R. J., de Vos, R. C. H., 2011. Spectral trees as a robust annotation tool in LC-MS based metabolomics. *Metabolomics* 8, 691-703.
- van Engelen, F. A., Molthoff, J. W., Conner, A. J., Nap, J. P., Pereira, A., Stiekema, W. J., 1995. pBINPLUS: an improved plant transformation vector based on pBIN19. *Transgenic Res.* 4, 288-290.
- van Herpen, T. W. J. M., Cankar, K., Nogueira, M., Bosch, D., Bouwmeester, H. J., Beekwilder, J., 2010. *Nicotiana benthamiana* as a production platform for *Artemisinin* precursors. *PLoS One* 5, e14222.
- Voinnet, O., Rivas, S., Mestre, P., Baulcombe, D., 2003. An enhanced transient expression system in plants based on suppression of gene silencing by the p19 protein of tomato bushy stunt virus. *Plant J.* 33, 949-956.
- Vom Endt, D., Silva, M. S., Kijne, J. W., Pasquali, G., Memelink, J., 2007. Identification of a bipartite jasmonate-responsive promoter element in the *Catharanthus roseus* ORCA3 transcription factor gene that interacts specifically with AT-Hook DNA-binding proteins. *Plant Physiol.* 144, 1680-1689.
- Wendt, K. U., Schulz, G. E., 1998. Isoprenoid biosynthesis: manifold chemistry catalyzed by similar enzymes. *Structure* 6, 127-133.

- Wink, M., 2010. Introduction: biochemistry, physiology and ecological functions of secondary metabolites, in: Wink, M. (Eds.), Annual plant reviews: Biochemistry of plant secondary metabolism, second ed. John Wiley & Sons, Inc., Oxford, UK, pp. 1-19.
- Winterhalter, P., Skouroumounis, G., 1997. Glycoconjugated aroma compounds: occurrence, role and biotechnological transformation. Adv. Biochem. Eng. Biotechnol. 55, 73-105.
- Wong, E. Y., Hironaka, C. M., Fischhoff, D. A., 1992. *Arabidopsis thaliana* small subunit leader and transit peptide enhance the expression of *Bacillus thuringiensis* proteins in transgenic plants. Plant Mol. Biol. 20, 81-93.
- Yang, T., Li, J., Wang, H. X., Zeng, Y., 2005. A geraniol-synthase gene from *Cinnamomum tenuipilum*. Phytochem. 66, 285-293.
- Yang, T., Stoopen, G., Yalpani, N., Vervoort, J., De Vos, R., Voster, A., Verstappen, F. W. A., Bouwmeester, H. J., Jongsma, M. A., 2011. Metabolic engineering of geranic acid in maize to achieve fungal resistance is compromised by novel glycosylation patterns. Metab. Eng. 13, 414-425.





# Discovery and reconstitution of the secoiridoid pathway from *Catharanthus roseus*

**Karel Miettinen**<sup>1†</sup>, Lemeng Dong<sup>2†</sup>, Nicolas Navrot<sup>3†</sup>, Thomas Schneider<sup>4‡</sup>, Vincent Burlat<sup>5</sup>, Jacob Pollier<sup>6</sup>, Lotte Woittiez<sup>4§</sup>, Sander van der Krol<sup>2</sup>, Raphaël Lugin<sup>3</sup>, Tina Ilc<sup>3</sup>, Robert Verpoorte<sup>1</sup>, Kirsi-Marja Oksman-Caldentey<sup>7</sup>, Enrico Martinoia<sup>4</sup>, Harro Bouwmeester<sup>2</sup>, Alain Goossens<sup>6</sup>, Johan Memelink<sup>1\*</sup> and Danièle Werck-Reichhart<sup>3\*</sup>

<sup>1</sup>Institute of Biology Leiden, Sylvius Laboratory, Sylviusweg 72, Leiden University, PO Box 9505, 2300 RA Leiden, The Netherlands.

<sup>2</sup>Laboratory of Plant Physiology, Wageningen University, Droevendaalsesteeg 1, 6708 PB Wageningen, the Netherlands.

<sup>3</sup>Institut de Biologie Moléculaire des Plantes, Unité Propre de Recherche 2357 du Centre National de la Recherche Scientifique, Université de Strasbourg, 28 rue Goethe, 67000 Strasbourg, France.

<sup>4</sup>Institute of Plant Biology, University Zurich, Zollikerstrasse 10, CH-8008 Zurich, Switzerland.

<sup>5</sup>CNRS; UMR 5546, Université de Toulouse ; UPS ; UMR 5546, Laboratoire de Recherche en Sciences Végétales ; BP 42617 Auzeville, F-31326 Castanet-Tolosan, France.

<sup>6</sup>Department of Plant Systems Biology, VIB and Department of Plant Biotechnology and Bioinformatics, Ghent University, Technologiepark 927, B-9052 Gent, Belgium.

<sup>7</sup>Bio and Process Technology, VTT Technical Research Centre of Finland, P.O.Box 1000, FI-02044 VTT (Espoo), Finland.

**† These authors contributed equally to this work.**

<sup>‡</sup>Present address: Philip Morris Products S.A., Quai Jeanrenaud 5, 2000 Neuchatel.

<sup>§</sup>Present address: Plant Production Systems Group, Wageningen University, P.O. Box 430, 6700 AK Wageningen, The Netherlands.

\*Corresponding author. E.mail: j.memelink@biology.leidenuniv.nl (J.M.); werck@unistra.fr (D.W.)

## Abstract

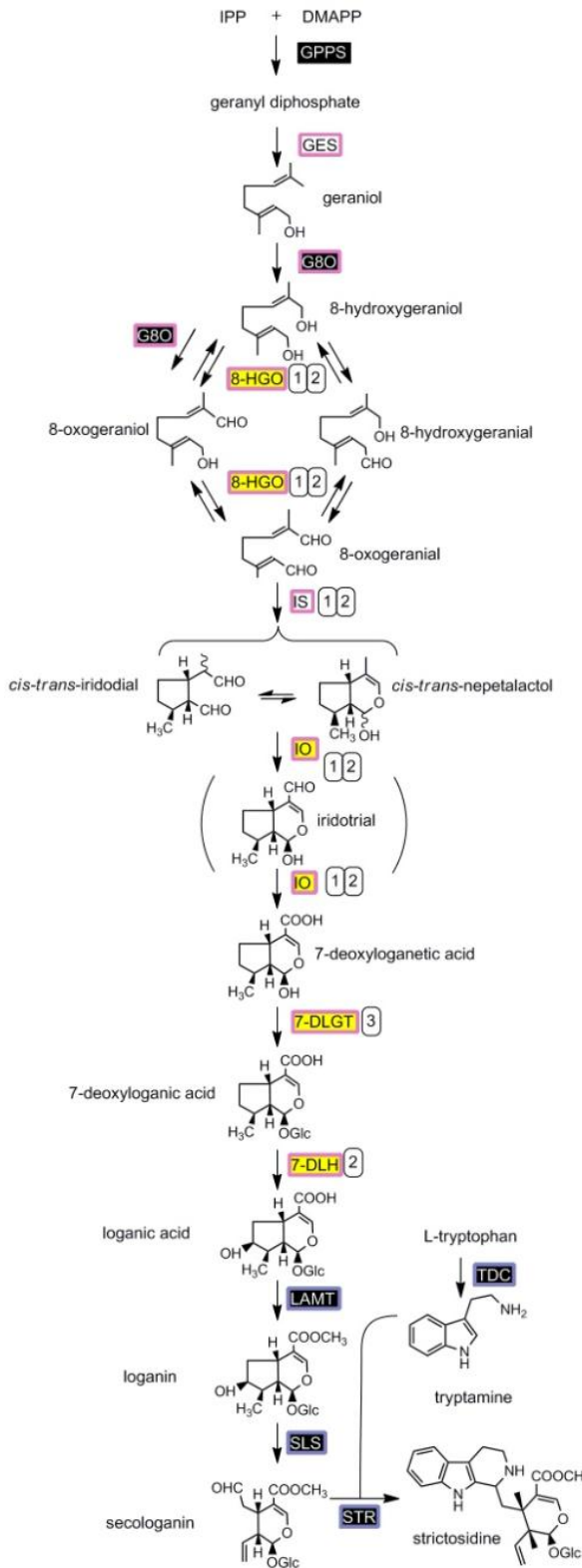
The (seco)iridoids and their derivatives, the monoterpenoid indole alkaloids (MIAs), form two large families of plant-derived bioactive compounds with a wide spectrum of high-value pharmacological and insect-repellent activities. Vinblastine and vincristine, MIAs used as anti-cancer drugs, are produced by *Catharanthus roseus* in extremely low levels, leading to high market prices and poor availability. Their biotechnological production is hampered by the fragmentary knowledge of their biosynthesis. Here we report the discovery of the last four missing steps of the (seco)iridoid biosynthesis pathway. Expression of the eight genes encoding this pathway together with two genes boosting precursor formation and two downstream alkaloid biosynthesis genes in an alternative plant host allowed the heterologous production of the complex MIA strictosidine. This confirms the functionality of all enzymes of the pathway and highlights their utility for synthetic biology programs towards a sustainable biotechnological production of valuable (seco)iridoids and alkaloids with pharmaceutical and agricultural applications.

## Introduction

Monoterpenoid indole alkaloids (MIAs) are a large group of plant-derived natural products with a range of pharmacological properties. Examples of MIAs are camptothecin used to treat cancer and quinine, the antimalarial drug of choice till the mid of the last century. Madagascar periwinkle, *Catharanthus roseus*, the best-characterized MIA-producing plant species, is the source of the valuable MIAs vincristine and vinblastine, which are used directly or as derivatives for the treatment of several cancer types. Because of the extremely low concentrations (0.0002 % fresh weight), production of vincristine and vinblastine is expensive (3000 USD/g) and availability of the drug is sensitive to environmental and political instability in the production countries. Therefore, biotechnology-based production of MIAs in microorganisms or alternative plant hosts has been proposed as a sustainable substitute but progress has been hampered by the lack of knowledge of the enzymes responsible for MIA biosynthesis, particularly in the secologanin pathway (Fig. 1).

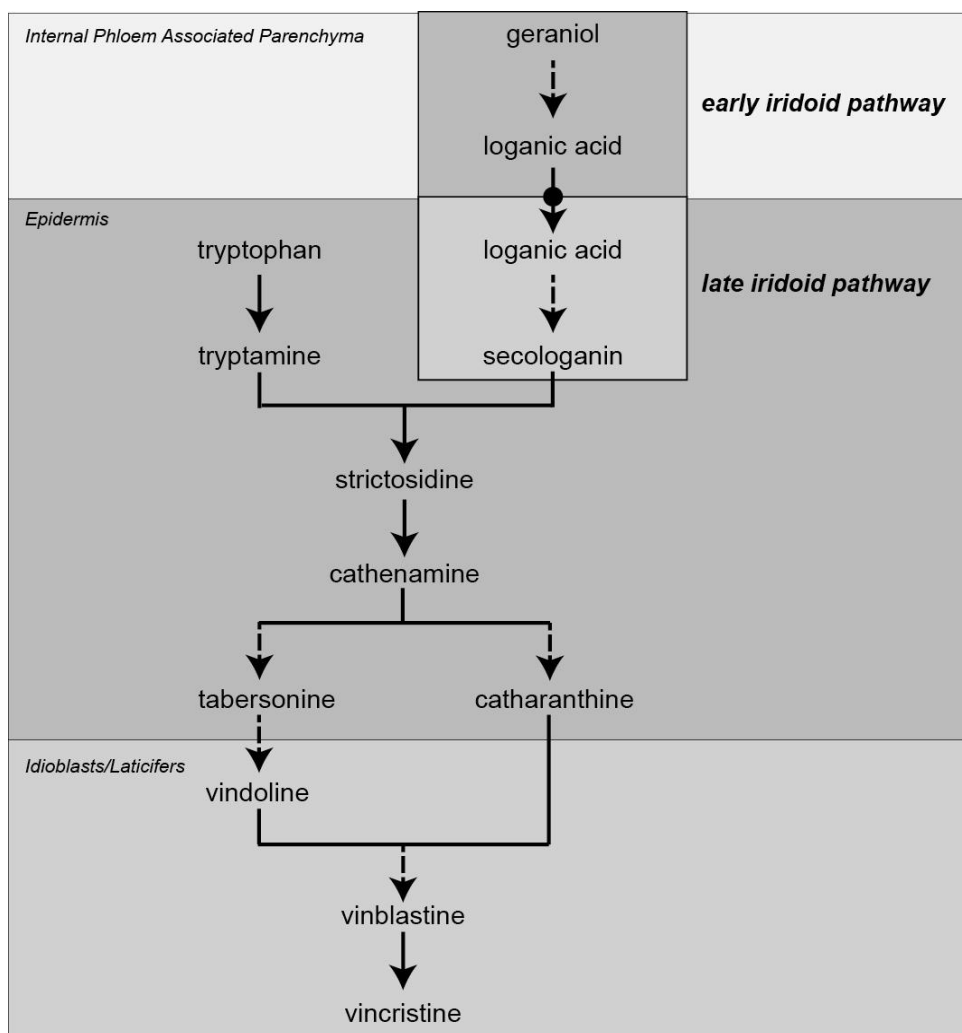
Secologanin is the monoterpenoid (also called iridoid or secoiridoid) branch end point and is coupled to tryptamine by strictosidine synthase (STR) to form strictosidine, the universal MIA precursor. The secologanin pathway has broad importance as many plant species accumulate iridoids and secoiridoids (including secologanin) as end products without incorporating them into complex alkaloids. Many (seco)iridoids are bioactive themselves, with among others anticancer, anti-microbial and anti-inflammatory activities<sup>1-4</sup>. Iridoids are also pheromones in some insect species, and as such can be employed for pest management in agriculture and the control of insect disease vectors<sup>5,6</sup>.

The (seco)iridoid pathway is still largely unresolved. It starts with geraniol and is thought to comprise approximately 10 enzymes catalyzing successive oxidation, reduction, glycosylation and methylation reactions (Fig. 1). Although the pathway has been investigated for decades<sup>7,8</sup>, only the first step (geraniol 10-hydroxylase/8-oxidase, G8O)<sup>9</sup> and the two last steps (loganic acid *O*-methyltransferase, LAMT; and secologanin synthase, SLS)<sup>10,11</sup> are well established. Only recently, two additional enzymes were identified, iridoid synthase (IS) responsible for the reductive cyclisation step<sup>12</sup> and geraniol synthase (GES)<sup>13</sup>. A complicating factor for gene discovery as well as biotechnological production is that the MIA pathway in *C. roseus* is organized in a complex manner, with the enzymes localized in different cell types and subcellular compartments<sup>14,15</sup> (Fig. S1).



**Fig. 1.** The secologanin-strictosidine pathway.

Genes indicated in boxes were published before (black background) or during (white background) the present study, or are reported here (yellow background). Frames indicate mRNA localization in the leaf internal phloem-associated parenchyma (IPAP) (pink) or epidermis (blue). Numbers indicate predicted enzyme classes in the initial gene discovery strategy. 1: oxidoreductase, 2: cytochrome P450, 3: UDP-glycosyltransferase (UGT). GPPS, geranyldiphosphate synthase; GES, geraniol synthase; G8O, geraniol 8-oxidase; 8-HGO, 8-hydroxygeraniol oxidoreductase; IS, iridoid synthase; IO, iridoid oxidase; 7-DLGT, 7-deoxyloganic acid glucosyltransferase; 7-DLH, 7-deoxyloganic acid hydroxylase; LAMT, loganic acid *O*-methyltransferase; SLS, secologanin synthase; STR, strictosidine synthase; TDC, tryptophan decarboxylase.



**Fig. S1. Overview of the MIA pathway and the cell-specific localization of the branches leading to vinblastine and vincristine.** Solid arrows represent single enzymatic steps, dashed arrows multiple enzymatic steps. The arrow bearing a black circle represents the transport of loganic acid from one type of cells to the other.

Here we report the characterization of the last missing steps of the *C. roseus* secoiridoid pathway. We used an integrated transcriptomics and proteomics approach for gene discovery, followed by biochemical characterization of the isolated candidates. Furthermore, we reconstituted the entire MIA pathway up to strictosidine in the plant host *Nicotiana benthamiana*, by heterologous expression of the newly identified genes in combination with the previously known biosynthesis genes. This work provides essential tools that will allow development of synthetic biology platforms for the production of bioactive iridoids, secoiridoids and complex MIAs with a wide range of agricultural and pharmaceutical applications, including the treatment of cancer.

### Results and Discussion

#### *Gene discovery*

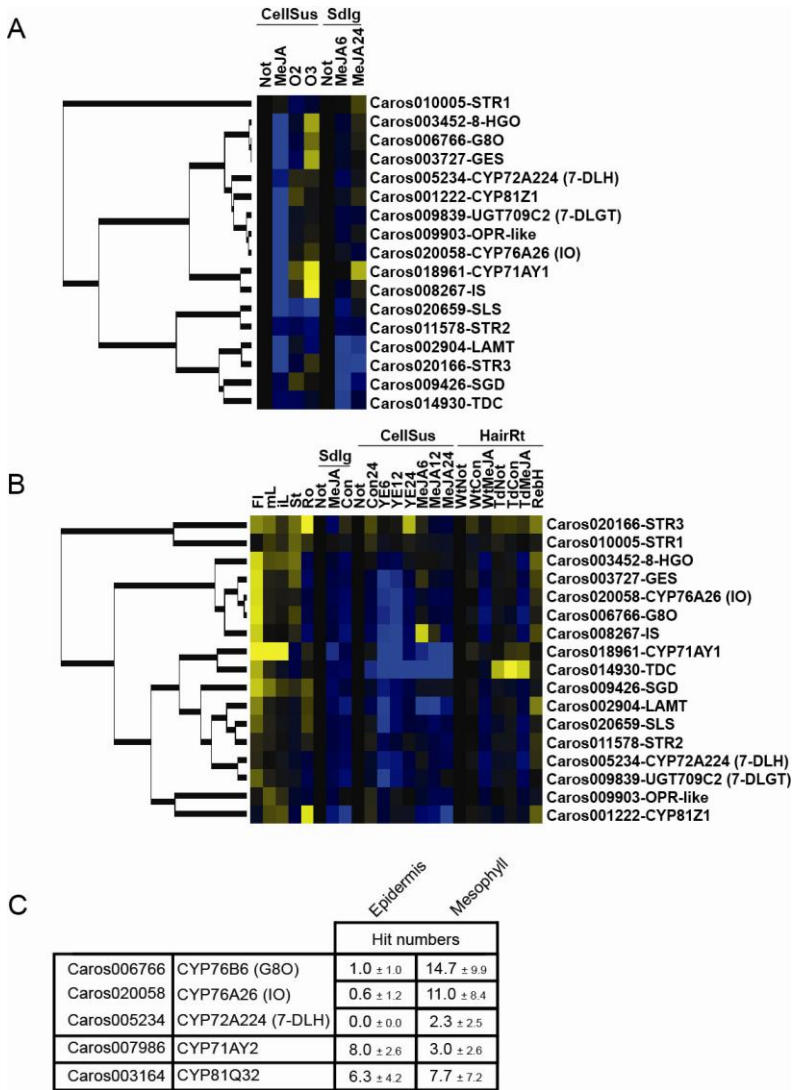
We previously reported the assembly of CathaCyc, a *C. roseus* metabolic pathway database based on Illumina HiSeq2000 RNA sequencing data<sup>16</sup>. The dataset was derived from *C. roseus* suspension cells and shoots treated with the plant hormone methyl jasmonate (MeJA)<sup>16</sup>. Here, we complemented this dataset with RNA-Seq data from cell suspensions overexpressing either *ORCA2* or *ORCA3*, transcription factors that regulate the expression of *LAMT*, *SLS* and several other genes in the MIA biosynthesis pathway<sup>17</sup>, but not *GES* and *G8O* (Fig. 2A). These four and all other known MIA genes are induced by MeJA both in cell suspension cultures and whole *C. roseus* plants, although *GES/G8O* and *LAMT/SLS* show different induction characteristics (Fig. 2B). Furthermore, *GES* and *G8O* are expressed in the internal phloem-associated parenchyma (IPAP) cells, whereas *LAMT* and *SLS* are expressed in the leaf epidermis<sup>10, 11, 13, 18, 19</sup>. The differential induction and *in situ* expression data suggested that the first part of the pathway (up to 7-deoxyloganic acid) corresponds to one transcriptional regulon, whereas all subsequent steps up to the synthesis of secologanin would comprise a second regulon.

Based on the hypothetical pathway and predicted enzyme activities catalyzing the hitherto missing steps (Fig. 1), we screened our dataset for genes encoding NAD(P)-binding Rossmann fold domain-type oxidoreductases, cytochrome P450 monooxygenases (P450s) and

UDP-glycosyltransferases (UGTs) that display co-expression with *GES/G8O*, and for P450s that show co-expression with *LAMT/SLS*. Three genes encoding putative oxidoreductases showed a high degree of co-expression with *GES/G8O* (Fig. 2A). The first (accession number Caros008267) was annotated as a progesterone reductase, the second (Caros003452) as an aldehyde dehydrogenase and the third (Caros009903) as a 12-oxophytodienoate reductase. We also identified four P450s (Caros020058, Caros001222, Caros018961, and Caros005234) and one UGT (Caros009839) that showed close co-expression with *GES/G8O* (Fig. 2A). The first oxidoreductase was found to encode the recently-described iridoid synthase<sup>12</sup>, thus confirming the validity of our screening strategy. The others were selected for further functional analysis. We also verified the co-expression of these candidate genes in the publicly-available dataset from the Medicinal Plant Genomics Resource consortium (<http://medicinalplantgenomics.msu.edu>) which has been integrated into the ORCAE website (<http://bioinformatics.psb.ugent.be/orcae><sup>16</sup>). This analysis strongly supported our selection of candidate genes (Fig. 2B).

The tissue localization of the enzymes in the leaf was used as a second criterion to pinpoint the most promising candidate genes. This was investigated using proteomics on epidermal and mesophyll protoplasts isolated from *C. roseus* leaves. The proteomics analysis resulted in the identification of 2200 proteins. Three P450s and one UGT (Caros005234, Caros006766, Caros020058 and Caros020739) were enriched in the mesophyll fraction, whereas one P450 (Caros007986), one UGT (Caros004449) and nine oxidoreductases (Caros022489, Caros002459, Caros017236, Caros002170, Caros012730, Caros006689, Caros007544, Caros021570 and Caros003491) were enriched in the epidermal fraction. One P450 (Caros003164) was present in both tissues (Fig. 2C and Table S)





**Fig. 2. Gene discovery strategy.** (A-B) Complete-linkage hierarchical clustering of early MIA pathway gene expression in *C. roseus* based on our data (A) or the Medicinal Plant Genomics Resource consortium (<http://medicinalplantgenomics.msu.edu>). (B) Colors indicate transcriptional activation (blue) or repression (yellow) relative to untreated samples. Tissues: Fl, flower; mL, mature leaves; iL, immature leaves; St, stem; Ro, root; Sdlg, seedling. Suspension cells (CellSus): Wt, wild-type; O2, ORCA2; O3, ORCA3. Hairy roots (HairRt): Wt, wild-type; Td, TDCi; RebH, RebH\_F. Treatments: Not, no treatment; MeJA, methyl jasmonate (6, 12 or 24 h); Con, mock; YE, yeast extract. (C) Candidate P450 proteins in epidermis and mesophyll from proteomics analysis. GES, geraniol synthase; G8O, geraniol 8-oxidase; 8-HGO, 8-hydroxygeraniol oxidoreductase; IS, iridoid synthase; IO, iridoid oxidase; 7-DLGT, 7-deoxy-loganetic acid glucosyltransferase; 7-DLH, 7-deoxyloganic acid hydroxylase; LAMT, loganic acid *O*-methyl-transferase; SGD, strictosidine- $\beta$ -D-glucosidase; SLS, secologanin synthase; STR, strictosidine synthase (1-3: three related genes); TDC, tryptophan decarboxylase

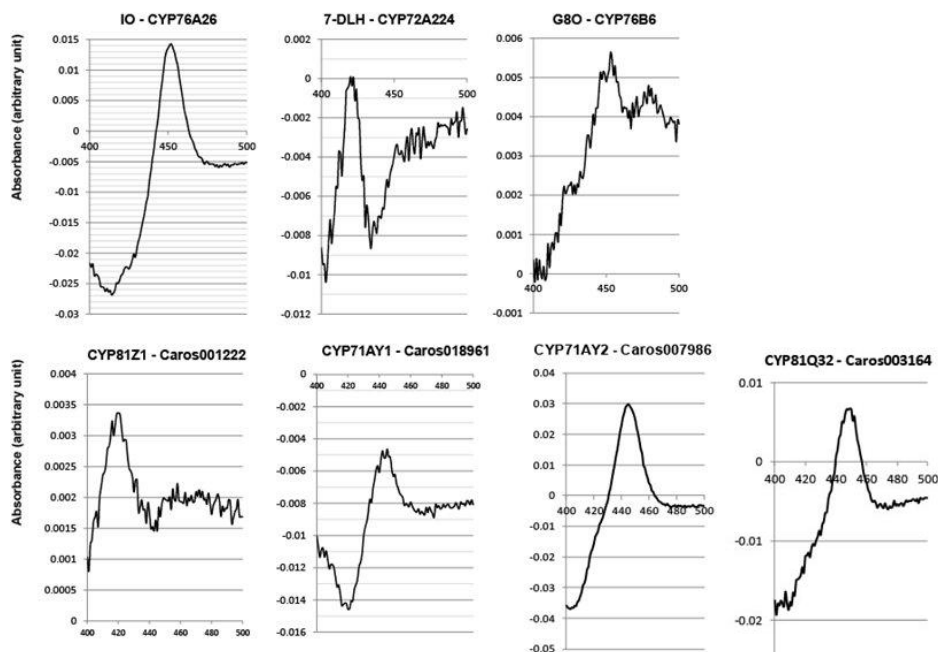
Enzyme Class	Enzyme name	Accession number	proteomics hits			Enrichment	
			Epidermis Mean $\pm$ SD	Mesophyll Mean $\pm$ SD	Statistics Fisher's exact test Significance level (%)		p-value
Cyt P450	CYP72A224	Caros005234	0.0 $\pm$ 0.0	2.3 $\pm$ 2.5	95	(0.013)	Mesophyll
Cyt P450	CYP76B6	Caros006766	1.0 $\pm$ 1.0	14.7 $\pm$ 9.9	95	(0.000000025)	Mesophyll
Cyt P450	CYP81Q32	Caros003164	6.3 $\pm$ 4.2	7.7 $\pm$ 7.2	0	(0.52)	n.s.
Cyt P450	CYP76A26	Caros020058	0.67 $\pm$ 1.2	11.0 $\pm$ 8.4	95	(0.00000018)	Mesophyll
Cyt P450	CYP71AY2	Caros007986	8.0 $\pm$ 2.6	3.0 $\pm$ 2.6	95	(0.0018)	Epidermis
oxidoreductase (alcohol dehydrogenase)	CrADH2	Caros002459	5.3 $\pm$ 4.2	0.7 $\pm$ 1.2	95	(0.00021)	Epidermis
oxidoreductase (alcohol dehydrogenase)	CrADH3	Caros022489	15.0 $\pm$ 10.8	3.3 $\pm$ 3.2	95	(0.000000056)	Epidermis
oxidoreductase (alcohol dehydrogenase)	CrADH5	Caros021570	11.7 $\pm$ 7.8	4.3 $\pm$ 3.2	95	(0.00015)	Epidermis
oxidoreductase (alcohol dehydrogenase)	CrADH8	Caros012730	27.3 $\pm$ 17.1	20.0 $\pm$ 16.1	95	(0.0036)	Epidermis
oxidoreductase (alcohol dehydrogenase)	CrADH9	Caros017236	16.0 $\pm$ 4.6	3.7 $\pm$ 6.4	95	(0.000000029)	Epidermis
oxidoreductase (alcohol dehydrogenase)	CrADH11	Caros007544	36.0 $\pm$ 6.2	18.7 $\pm$ 15.1	95	(0.00000029)	Epidermis
oxidoreductase (alcohol dehydrogenase)	CrADH14	Caros006689	48.0 $\pm$ 13.7	11.0 $\pm$ 9.5	95	(0.000000000000000000000049)	Epidermis
oxidoreductase (alcohol dehydrogenase)	CrADH15	Caros002170	14.0 $\pm$ 5.0	3.0 $\pm$ 5.2	95	(0.00000011)	Epidermis
oxidoreductase	CrBBE1	Caros003491	2.0 $\pm$ 2.6	0.0 $\pm$ 0.0	95	(0.0097)	Epidermis
UGT	CrUGT2	Caros004449	16.3 $\pm$ 9.8	5.3 $\pm$ 4.5	95	(0.00000015)	Epidermis
UGT	CrUGT3	Caros020739	0.0 $\pm$ 0.0	10.3 $\pm$ 14.6	95	(0.000000046)	Mesophyll

**Table S1. Candidate enzymes in mesophyll and epidermal cells.** A proteomic approach with microsomes isolated from enriched epidermal or mesophyll protoplast fractions was carried out as described in Methods. Shown are candidate enzymes found among 2200 identified proteins and their distribution between mesophyll and epidermis. n.s.: not significant.

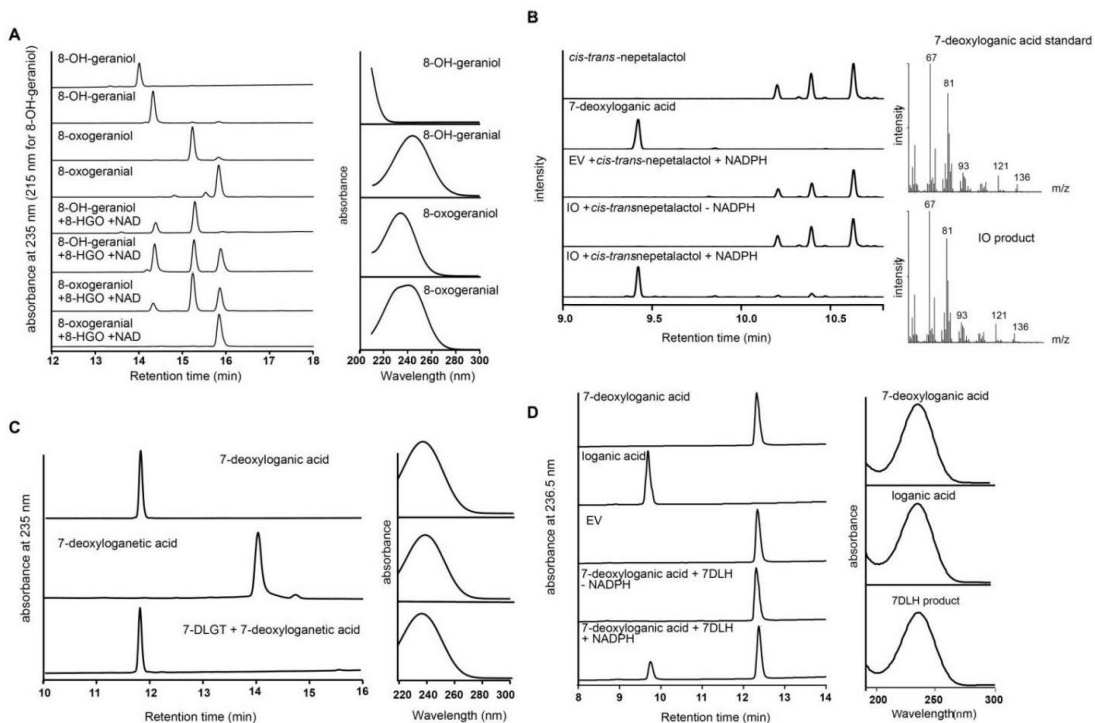
### *Candidate enzyme activities*

The three oxidoreductases (Caros008267, Caros003452 and Caros009903) were produced in *Escherichia coli* and purified for *in vitro* enzyme assays. Confirming a previous report<sup>12</sup>, the putative progesterone reductase (Caros008267) was shown to possess iridoid synthase (IS) activity in the presence of 8-oxogeraniol (data not shown), yielding a mixture of *cis-trans*-iridodial and *cis-trans*-nepetalactol. Caros003452 was active with the substrates 8-OH-geraniol, 8-OH-geraniol and 8-oxogeraniol in the presence of NAD<sup>+</sup>, yielding mixtures of the three compounds and 8-oxogeraniol in varying relative amounts depending on the combination and the incubation time (Fig. 3A). The enzyme was therefore coined 8-hydroxygeraniol oxidoreductase (8-HGO). With the co-factor NAD<sup>+</sup> it did not convert 8-oxogeraniol (Fig. 3A), and it was not active with any of the substrates listed above in the presence of NADP<sup>+</sup>/NADPH. Given the complex kinetics, with four interconvertible compounds and eight possible reactions, the reaction constants could not be determined. G8O was recently shown to also produce 8-oxo-geraniol from geraniol<sup>20</sup>, thus the two enzymes G8O and 8-HGO appear to catalyze partially overlapping (and in the case of G8O, monodirectional) oxidation reactions that result in the production of 8-oxogeraniol from geraniol (Fig. 1). Six candidate P450 genes (*CYP76A26*, *CYP81Z1*, *CYP81Q32*, *CYP72A224* and *CYP71AY1*, *CYP71AY2*) were transferred to a yeast expression vector and co-expressed in *Saccharomyces cerevisiae* together with the P450 reductase *ATR1* from *Arabidopsis thaliana*<sup>21</sup> (Fig. S2). Functional screening was carried out as described<sup>20</sup> with geraniol, 8-hydroxygeraniol, 8-hydroxygeraniol, 8-oxogeraniol, 8-oxogeraniol, iridodial, iridotrial, 7-deoxyloganic acid and 7-deoxyloganic acid as potential substrates. *CYP76A26* converted both iridodial and iridotrial into 7-deoxyloganic acid. The *cis*-iridodial and *trans*-iridodial freely interconverted with *cis-trans*-nepetalactol<sup>12</sup>, and although *CYP76A26* seemed to use the bicyclic iridodial nepetalactol as the preferred substrate, the monocyclic *cis*- and *trans*-iridodials were also utilized, possibly after spontaneous conversion into nepetalactol (Fig. 3B). The interconversion and sequential metabolism of nepetalactol and the iridodials in aqueous solution and the limited availability of pure iridodial prevented evaluation of the

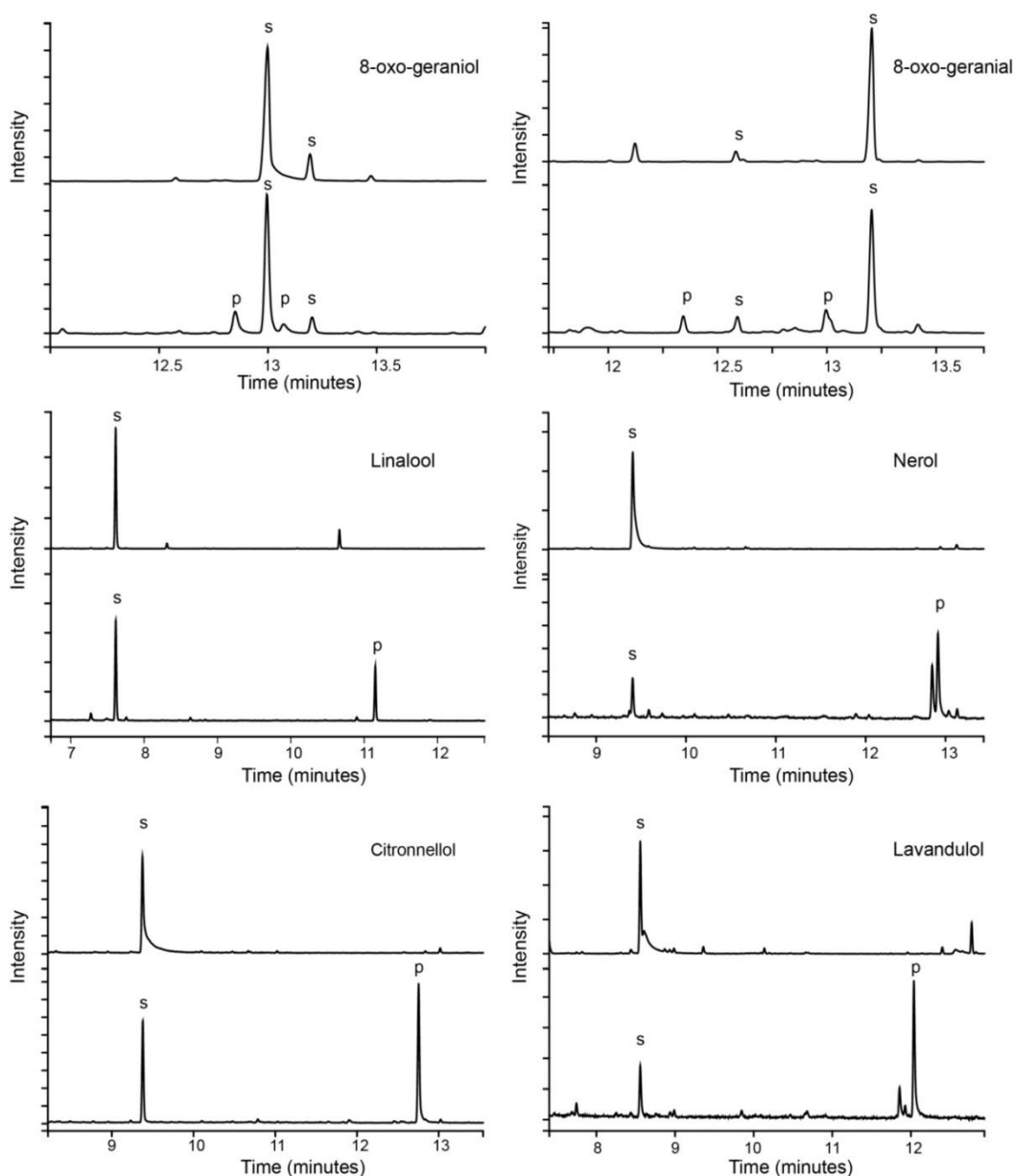
catalytic parameters with these substrate(s). Although we never detected



**Fig. S2.** Evaluation of the expression of P450 candidates in transformed yeast microsomes. Differential absorbance of the CO-saturated-reduced versus reduced microsomes was recorded between 400 and 500 nm and cytochrome P450 concentration was determined according to<sup>45</sup>. x-axis: wavelength in nm. Iridoid oxidase (IO, CYP76A26, Caros003676), 7-deoxyloganic acid hydroxylase (7-DLH, CYP72A224, Caros005234), Geraniol-8-oxidase (G8O, CYP76B6, Caros006766), CYP81Z1 (Caros001222), CYP71AY1 (Caros018961), CYP71AY2 (Caros007986), CYP81Q32 (Caros003164).



**Fig. 3. Functional characterization of recombinant pathway enzymes.** The enzymatic activities of 8-hydroxygeraniol oxidoreductase (8-HGO) and 7-deoxyloganic acid glucosyltransferase (7-DLGT) were tested using affinity-purified proteins expressed in *E. coli*. Cytochrome P450s were expressed in yeast and assayed as microsomal preparations. (A) 8-HGO; (B) iridoid oxidase (IO); (C) 7-DLGT; (D), 7-deoxyloganic acid hydroxylase (7-DLH). Left: GC-FID and HPLC-DAD profiles show the specific formation of products by the enzymes, compared to negative controls and authentic standards. Right: Product identity was confirmed by comparison of MS or UV-spectra with authentic standards



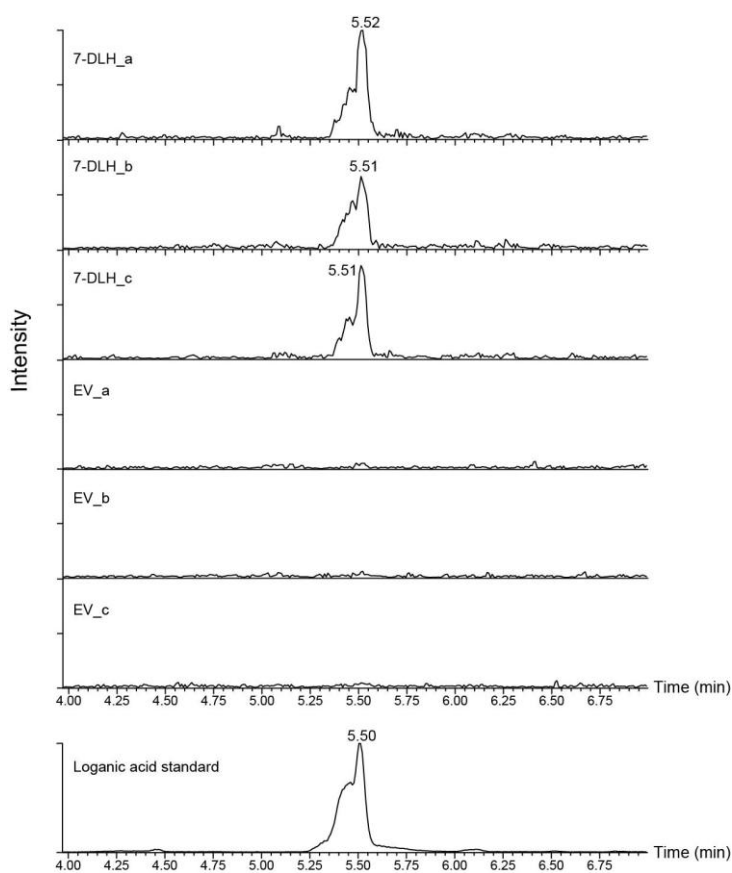
**Fig. S3. Activity of iridoid oxidase (IO; CYP76A26) with different substrates.** Substrates (100  $\mu\text{M}$ ) were incubated for 20 min at 27  $^{\circ}\text{C}$  with 23  $\mu\text{M}$  of CYP76A26 in absence (top) or in presence (bottom) of NADPH. Samples were then extracted with ethyl acetate and analyzed on GC-FID. s: substrate peak(s); p: peak(s) of product(s).

iridotrial as an intermediate in the iridodial to 7-deoxyloganetic acid conversion, iridotrial was converted into 7-deoxyloganetic acid with a  $K_{mapp}$  of 45  $\mu\text{M}$  and  $k_{cat}$  of 57  $\text{s}^{-1}$ . Because the related G8O (CYP76B6) was previously shown to oxidize several (other) monoterpene

alcohols in addition to geraniol<sup>9</sup>, these compounds were also tested as potential substrates for CYP76A26. The enzyme converted 8-oxogeraniol into an unidentified product (Fig. S3), albeit with a low efficiency, and catalyzed the hydroxylation of linalool, nerol, citronellol and lavandulol, but not geraniol (Fig. S3). CYP76A26 thus catalyzes the direct conversion of iridodial into 7-deoxyloganetic acid in the secologanin pathway, and was consequently named iridoid oxidase (IO) (Fig. 1).

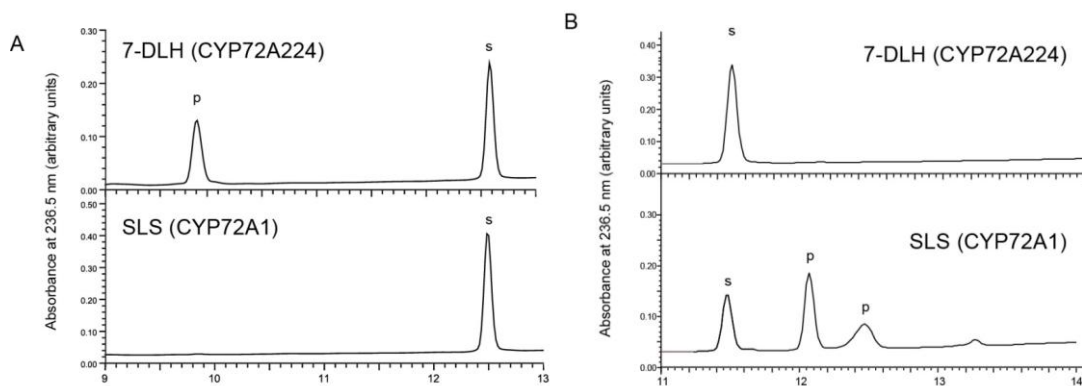
The UGT (Caros009839 or UGT709C2), produced in *E. coli*, catalyzed the glucosylation of 7-deoxyloganetic acid to form 7-deoxyloganic acid using UDP-glucose as the sugar donor (Fig. 3C). The enzyme had a  $K_{mapp}$  of 9.8  $\mu\text{M}$  and a  $k_{cat}$  of 1.25  $\text{s}^{-1}$ . The enzyme was inactive with loganetic acid, loganetin, iridodial, iridotrial, 8-OH-geraniol, jasmonic acid, gibberellic acid, indole acetic acid, salicylic acid, abscisic acid, zeatin and luteolin. It thus behaved as a selective 7-deoxyloganetic acid glucosyltransferase (7-DLGT) (Fig. 1).

In despite of its poor expression in yeast, CYP72A224 catalyzed the conversion of 7-deoxyloganic acid into loganic acid in yeast microsomes with a  $K_{mapp}$  of 400  $\mu\text{M}$  and a  $V_{max}$  of 0.5  $\text{pmol}\cdot\text{min}^{-1}\cdot\text{mg}^{-1}$  microsomal protein (Fig. 3D). The activity was confirmed in a *N. benthamiana* leaf-disc assay (Fig. S4). The aglycone derivative of 7-deoxyloganic acid (7-deoxyloganetic acid) was not a substrate for CYP72A224, confirming that glycosylation precedes hydroxylation of the cyclopentane ring. CYP72A224 was thus named 7-deoxyloganic acid hydroxylase (7-DLH). It belongs to the same P450 subfamily as secologanin synthase (SLS; CYP72A1), which catalyzes the conversion of loganin into secologanin<sup>10</sup>. Both of these P450s use glycosylated substrates and have similar regiospecificities, suggesting that SLS evolved from 7-DLH thus extending the iridoid pathway to the secoiridoids. This evolutionary process apparently resulted in a complete change in activity because SLS shows no 7-DLH activity and vice versa (Fig. S5).



**Fig. S4.** Evaluation of 7-deoxyloganic acid conversion into loganic acid by CYP72A224 in *N. benthamiana* leaf-disc assay. Discs from leaves agro-infiltrated with a binary vector containing the 7-deoxyloganic acid hydroxylase (7-DLH; CYP72A224) sequence were excised 5 days post-infiltration and incubated for 3 hours on buffer containing 7-deoxyloganic acid. A leaf methanol extract was analyzed on UPLC-MS. Multiple reaction monitoring in positive mode with the transition 215.1>108.9 is shown. EV: extracts of discs agro-transfected with the empty vector incubated with 7-deoxyloganic acid.





**Fig. S5. Comparison of the activities of 7-deoxyloganic acid hydroxylase and secologanin synthase expressed in yeast with 7-deoxyloganic acid and loganin.** Microsomes from yeast expressing CYP72A224 (7-DLH) or CYP72A1 (SLS) were tested for activity with 7-deoxyloganic acid (A) and with loganin (B) and analyzed by HPLC as described earlier. No 7-DLH activity was found for SLS. No activity with loganin was observed for 7-DLH, whereas SLS forms the two peaks of products (p) corresponding to reference secologanin (RT: 12.065 and 12.468 minutes). Experimental conditions were the same as used for testing 7-DLH activity (see main text). s: substrate; p: product.

### *Pathway reconstitution in Nicotiana benthamiana*

The identification of the four enzymes described above allowed us to propose a complete secologanin pathway (Fig. 1), which we tested by stepwise combinatorial transient expression of the corresponding genes in *N. benthamiana* (Fig. 4A). To boost substrate availability for the pathway a geranyldiphosphate synthase from *Piceaabies* (*PaGPPS*)<sup>22</sup> and a geraniol synthase from *Valerianaofficinalis* (*VoGES*)<sup>23</sup> were used because the *C. roseus* orthologs were not available at the onset of these studies. The transient expression of *PaGPPS*+*VoGES* in *N. benthamiana* resulted in the formation of geraniol, but also a series of additional oxidized and glycosylated derivatives produced by endogenous *N. benthamiana* enzymes as previously reported<sup>23</sup>. The stepwise addition of *G8O* and *IS* resulted in the generation of new compounds, including new derivatives of the anticipated pathway intermediates (combinations 1, 2 and 4 in Fig. 4A, B; Fig. S6). In contrast, the addition of *8-HGO* to the combination *PaGPPS*+*VoGES*+*G8O* only modified the existing product profile but did not generate any new compounds (combination 3 in Fig. 4A, B; Fig. S6). When *IO* was co-expressed with *PaGPPS*+*VoGES*+*G8O*+*8-HGO*+*IS* (combination 5), the metabolic profile did not change, and 7-deoxyloganic acid and its derivatives were not detected (Fig. 4A, B).

However, reconstitution of the pathway up to 7-DLGT(combination 6) was successful and resulted in the production of 7-deoxyloganic acid and putative acetylated 7-deoxyloganic acid (Fig. 4A, C). Without *IO*, these products were not detected (Fig. S7), indicating that *IO* is functional in *N. benthamiana* and forms an essential part of the biosynthesis pathway. These findings also illustrate the importance of full pathway coverage for functional analysis of individual enzymes.

Subsequently, the entire postulated secologanin pathway (*PaGPPSto SLS*) was introduced by agroinfiltration, but in first instance this only yielded products up to 7-deoxyloganic acid. Therefore, we artificially increased the input halfway into the pathway by infiltrating the intermediates iridodial, iridotrial or 7-deoxyloganic acid in combination with the pathway genes. In all cases this resulted in the production of secologanin, indicating that the second half of the pathway is also functional (Fig. 4D). Finally we tested whether the biosynthesis pathway up to secologanin can be functionally combined with the reconstituted tryptamine branch of the MIA pathway. When the secologanin biosynthesis pathway genes were co-infiltrated with the tryptophan decarboxylase (*TDC*) and strictosidine synthase (*STR*) genes, and the flux through the pathway was boosted by infiltration of the intermediates iridodial, iridotrial or 7-deoxyloganic acid, strictosidine was indeed produced (combination 10 in Fig. 4D).

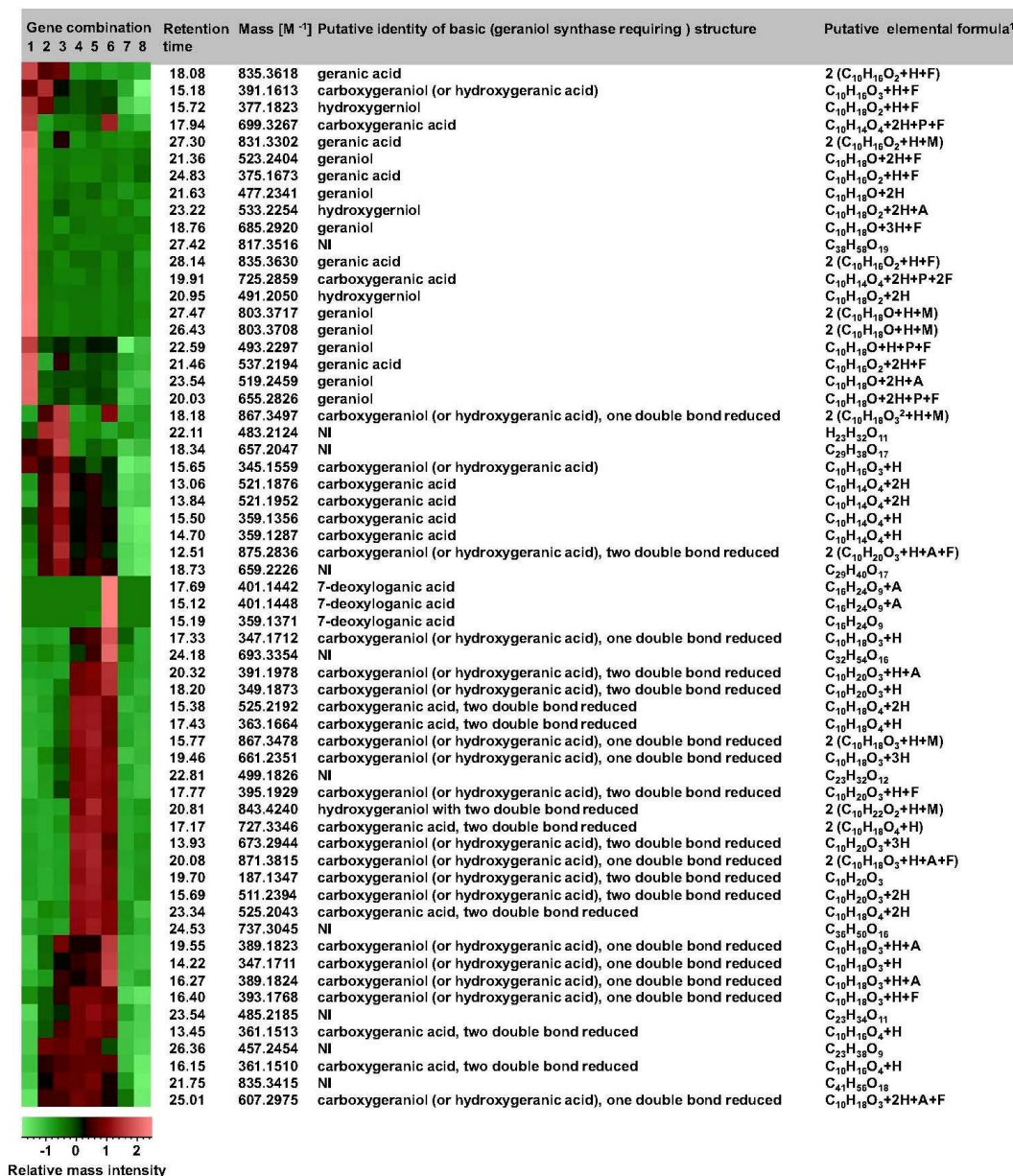
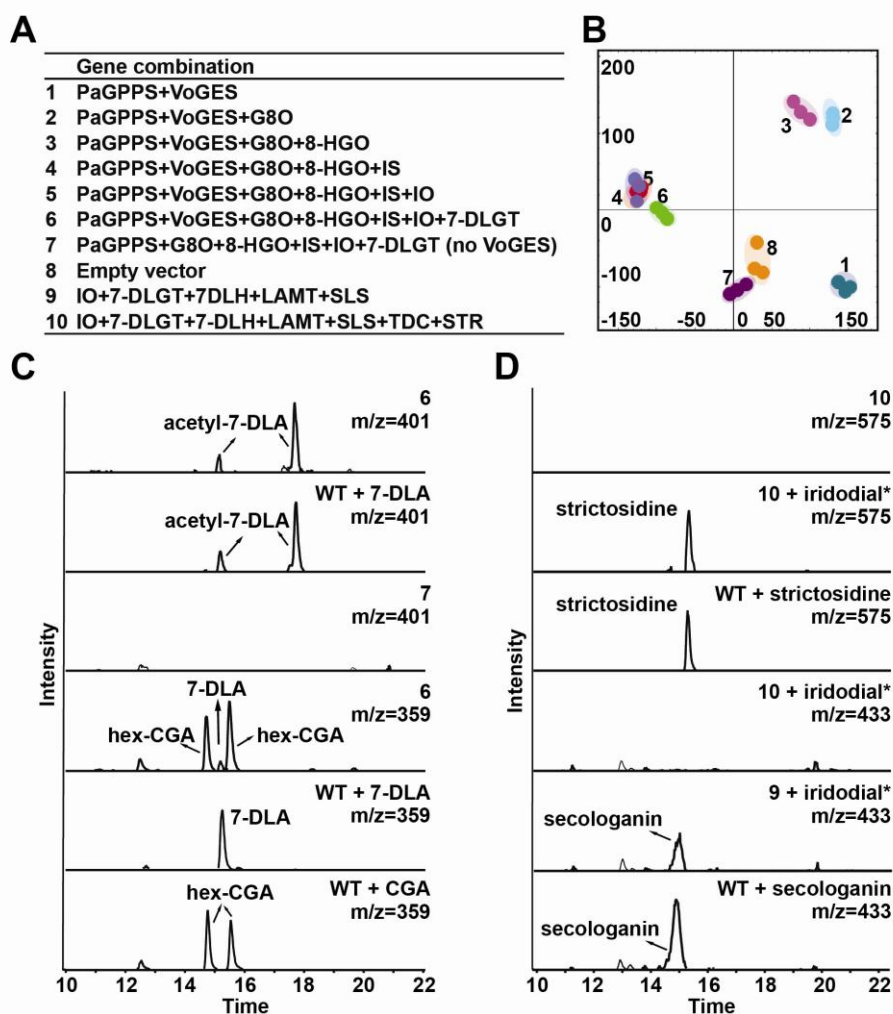
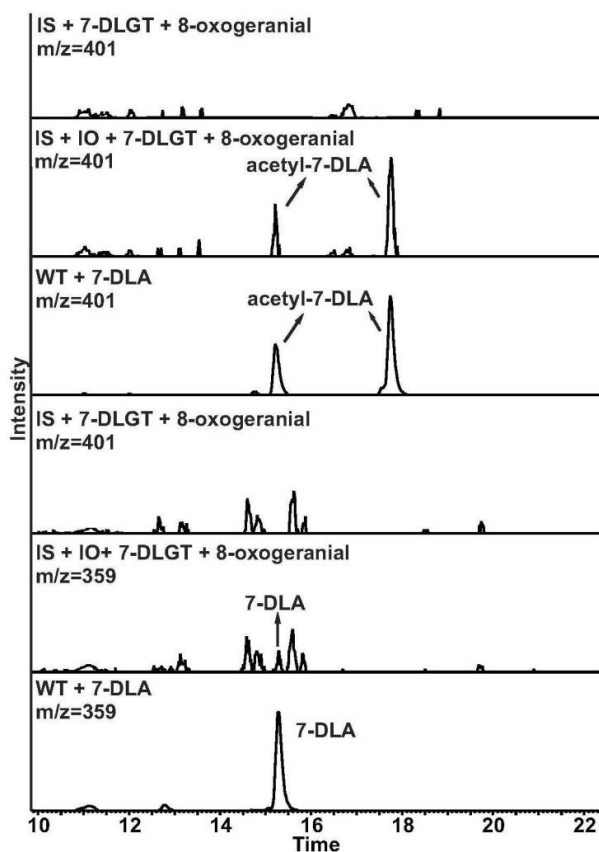


Fig. S6. Masses and quantitative changes of compounds detected upon step-wise reconstitution of the secologanin pathway in *N. benthamiana*. Gene combinations are as listed in Fig. 4A. Heatmap shows relative mass intensity changes.

<sup>1</sup> F, formic acid adduct; A, acetyl group adduct; M, malonyl group adduct; number 2, dimer; H, hexose, P, pentose. NI: not identified.



**Fig. 4. Reconstitution of the strictosidine pathway in *N. benthamiana*.** (A). Gene combinations infiltrated in leaves in triplicate. (B). Principal component analysis. PC1 and PC2 describe 36.2% and 31.1% of the total mass variation, respectively. (C). LC-MS analysis showing selected masses 401 and 359 representing (acetylated) 7-deoxyloganic acid (7-DLA) from infiltrations with 8-carboxygeranic acid (CGA), 7-DLA or gene combinations 6 or 7 (negative control). (D). LC-MS analysis showing selected masses 433 (formic acid adduct of secologanin) and 575 (formic acid adduct of strictosidine) from infiltrations with secologanin or strictosidine, or with gene combinations 9 or 10, with or without iridodial. \*Identical profiles with iridodial or 7-DLH. Hex = hexosyl.



**Fig. S7. Iridoid oxidase (IO) is an essential component of the pathway.** LC-MS analysis on selected mass 359 (7-deoxyloganic acid ; 7-DLA) and 401 (acetylated 7-DLA) of *N. benthamiana* leaves infiltrated with 7-DLA or 8-oxogeranial co- agro-infiltrated with gene combination *IS+IO+7-DLGT* or *IS+7-DLGT*.

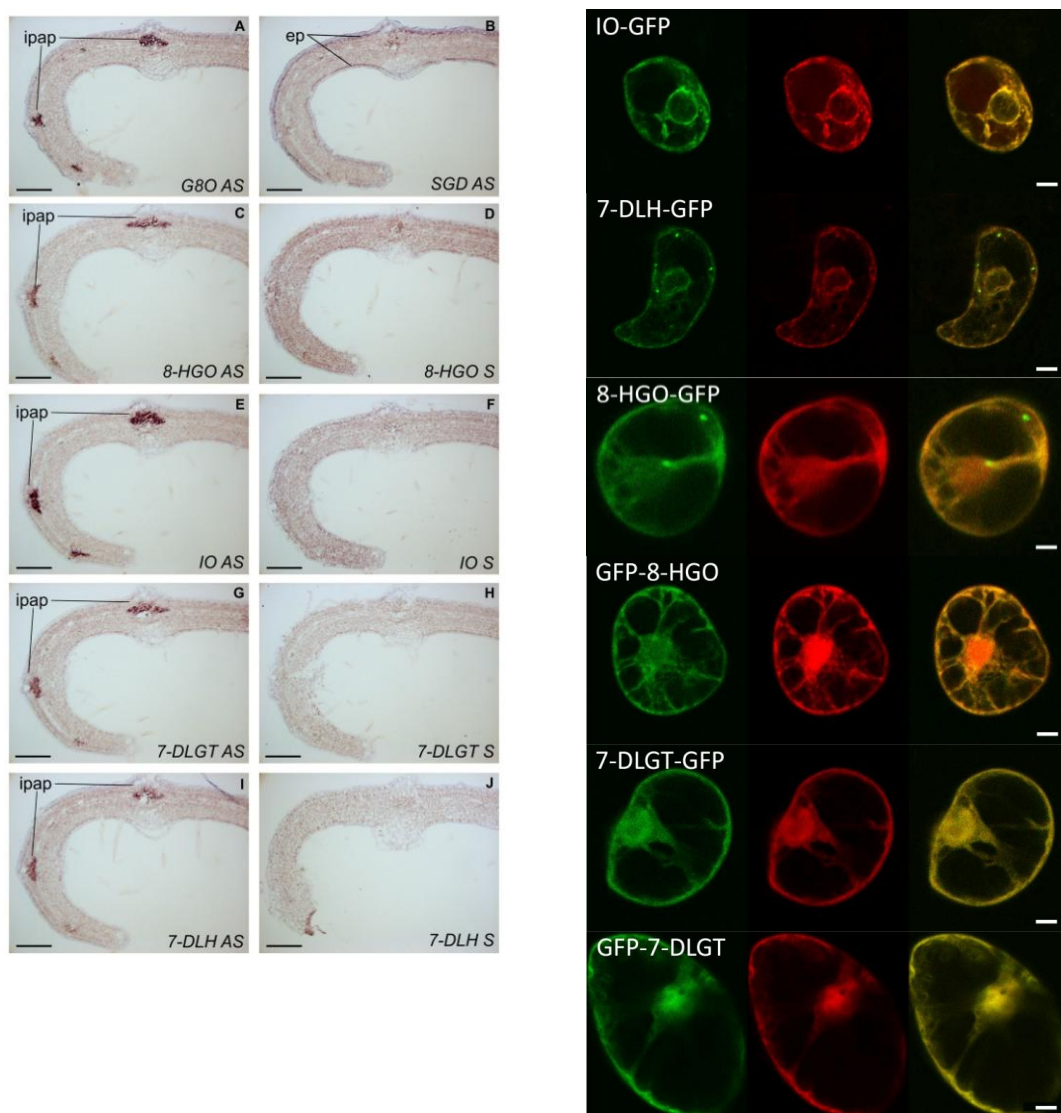
Pathway reconstitution experiments thus validated the enzyme sequence leading to the formation of strictosidine (Fig. 1). The low flux through the pathway probably reflects the substantial conversion of intermediate products to other oxidized, acetylated, malonylated and glycosylated derivatives by endogenous *N. benthamiana* enzymes, which can be addressed by further engineering but may also open new avenues for combinatorial biochemistry programs to produce novel bioactive (seco)iridoids (Fig. S6).

## Pathway localization

Efficient pathway engineering also requires precise knowledge of the cellular and subcellular organization of different pathway components. IO, 7-DLH, 8-HGO and 7-DLGT were therefore expressed as green fluorescent protein (GFP) fusions in *C. roseus* cells together with mCherry markers for nucleus, cytosol, endoplasmic reticulum (ER), plastids, mitochondria or peroxisomes. This revealed that IO and 7-DLH are ER-associated as predicted, whereas 8-HGO and 7-DLGT are soluble proteins found in the cytosol and the nucleus (Fig. S8). Previous *in situ* transcript analysis<sup>24</sup> has suggested that a pathway intermediate is translocated from IPAP cells (the site of the early biosynthetic steps) to the epidermis, where the final steps of the pathway occur. *In situ* hybridization showed that 8-HGO, IO, 7-DLGT and 7-DLH are expressed in the same tissue as G8O (Fig. 5). The separate clustering of these genes in Fig. 2A and the joint transcript localization as shown in Fig. 5 suggest that these genes constitute a transcriptional regulon for the production of loganic acid in IPAP cells. The epidermis-specific positive control *SGD* confirms the localization of the next transcriptional regulon, consisting of *LAMT*, *SLS*, *STR* and *TDC*, in the epidermis. The existence of this regulon is supported by the highly similar co-expression profiles of these genes as shown in Fig. 2A. The expression of 7-DLH in IPAP cells indicates that loganic acid is the mobile intermediate transferred to the epidermis, and hence that glycosylation by 7-DLGT is not sufficient for mobility but that further hydroxylation by 7-DLH is also required. This tissue-specific expression may increase the flux through the pathway by alleviating feedback inhibition by intermediates and products, and/or it may segregate the iridoid and secoiridoid/MIA pathways, allowing them to fulfill tissue-specific functions.

In conclusion, we have identified the four last missing enzymes in the secologanin pathway. In combination with the iridoid genes previously identified, the genes encoding these four enzymes are sufficient to engineer secologanin production and, together with *TDC* and *STR*, biosynthesis of the complex alkaloid strictosidine in a heterologous plant system. Notably, although different segments of the strictosidine pathway are localized in different cell types in *C. roseus*, our results show that the entire pathway can be successfully reconstituted in a single *N. benthamiana* organ. This paves the way for the biotechnological production of

valuable iridoids and iridoid-derived compounds, such as the MIAs vincristine and vinblastine, making these important anti-cancer drugs available to more people and at a lower price.



**Fig. 5.** Expression of the transcriptional regulon required for loganic acid biosynthesis in the internal phloem-associated leaf parenchyma (IPAP) cells. *In situ* hybridization on serial longitudinal sections of young developing leaves was carried out with antisense (AS) probes and sense (S) probes as controls. Geraniol 8-oxidase (*G8O*) and strictosidine  $\beta$ -D-glucosidase (*SGD*) AS probes were used as IPAP and epidermis markers, respectively. Sense probe controls gave no signals. 8-HGO, 8-hydroxygeraniol oxidoreductase; IO, iridoid oxidase; 7-DLGT, 7-deoxyloganetic acid glucosyltransferase; 7-DLH, 7-deoxyloganic acid hydroxylase; ipap, internal phloem associated parenchyma; ep, epidermis. Scale bar = 100  $\mu$ m.

**Fig. S8.** Subcellular localization of (seco)iridoid pathway enzymes. *C. roseus* cells were transiently co-transformed with plasmids expressing pathway enzymes fused to GFP (green; left) and mCherry organelle markers (red; middle). Co-localization of the fluorescent signals appears in the merged pictures in yellow (right). Iridoid oxidase (IO) and 7-deoxyloganic acid hydroxylase (7-DLH) were co-transformed with ER marker and 8-hydroxygeraniol oxidoreductase (8-HGO) and 7-deoxyloganetic acid-O-glucosyltransferase (7-DLGT) with cytosol/nucleus marker. Scale bars = 10  $\mu$ m.



**References and Notes**

1. B. Dinda, S. Debnath, Y. Harigaya, Naturally occurring iridoids. A review, part 1. *Chem. Pharm. Bull. (Tokyo)* **55**, 159-222 (2007).
2. B. Dinda, S. Debnath, Y. Harigaya, Naturally occurring secoiridoids and bioactivity of naturally occurring iridoids and secoiridoids. A review, part 2. *Chem. Pharm. Bull. (Tokyo)* **55**, 689-728 (2007).
3. R. Tundis, M. R. Loizzo, F. Menichini, G. A. Statti, F. Menichini, Biological and pharmacological activities of iridoids: Recent developments. *Mini Rev. Med. Chem.* **8**, 399-420 (2008).
4. A. Viljoen, N. Mncwangi, I. Vermaak, Anti-Inflammatory Iridoids of Botanical Origin. *Curr. Med. Chem.* **19**, 2104-2127 (2012).
5. M. A. Birkett, J. A. Pickett, Aphid sex pheromones: from discovery to commercial production. *Phytochemistry* **62**, 651-656 (2003).
6. M. A. Birkett, A. Hassanali, S. Høglund, J. Pettersson, J. A. Pickett, Repellent activity of catmint, *Nepeta cataria*, and iridoid nepetalactone isomers against Afro-tropical mosquitoes, ixodid ticks and red poultry mites. *Phytochemistry* **72**, 109-114 (2011).
7. A. Oudin *et al.*, Spatial distribution and hormonal regulation of gene products from methyl erythritol phosphate and monoterpene-secoiridoid pathways in *Catharanthus roseus*. *Plant Mol. Biol.* **65**, 13-30 (2007).
8. V. M. Loyola-Vargas, R. M. Galaz-Ávalos, R. Kú-Cauich, *Catharanthus* biosynthetic enzymes: the road ahead. *Phytochemistry Rev.* **6**, 307-339 (2007).
9. G. Collu *et al.*, Geraniol 10-hydroxylase, a cytochrome P450 enzyme involved in terpenoid indole alkaloid biosynthesis. *FEBS Lett.* **508**, 215-220 (2001).
10. S. Irmeler *et al.*, Indole alkaloid biosynthesis in *Catharanthus roseus*: new enzyme activities and identification of cytochrome P450 CYP72A1 as secologanin synthase. *Plant J.* **24**, 797-804 (2000).
11. J. Murata, J. Roepke, H. Gordon, V. De Luca, The leaf epidermome of *Catharanthus roseus* reveals its biochemical specialization. *Plant Cell* **20**, 524-542 (2008).

12. F. Geu-Flores *et al.*, An alternative route to cyclic terpenes by reductive cyclization in iridoid biosynthesis. *Nature* **492**, 138-142 (2012).
13. A. J. Simkin *et al.*, Characterization of the plastidial geraniol synthase from Madagascar periwinkle which initiates the monoterpene branch of the alkaloid pathway in internal phloem associated parenchyma. *Phytochemistry* **85**, 36-43 (2013).
14. S. Mahroug, V. Burlat, B. Saint-Pierre, Cellular and sub-cellular organisation of the monoterpene indole alkaloid pathway in *Catharanthus roseus*. *Phytochemistry Rev.* **6**, 363-381 (2007).
15. P. Verma, A. K. Mathur, A. Srivastava, A. Mathur, Emerging trends in research on spatial and temporal organization of terpene indole alkaloid pathway in *Catharanthus roseus*: a literature update. *Protoplasma* **249**, 255-268 (2012).
16. A. Van Moerkercke *et al.*, CathaCyc, a Metabolic Pathway Database Built from *Catharanthus roseus* RNA-Seq Data. *Plant Cell Physiol.* **54**, 673-685 (2013).
17. L. van der Fits, J. Memelink, ORCA3, a jasmonate-responsive transcriptional regulator of plant primary and secondary metabolism. *Science* **289**, 295-297 (2000).
18. G. Guirimand *et al.*, The subcellular organization of strictosidine biosynthesis in *Catharanthus roseus* epidermis highlights several trans-tonoplast translocations of intermediate metabolites. *FEBS J.* **278**, 749-763 (2011).
19. V. Burlat, A. Oudin, M. Courtois, M. Rideau, B. St-Pierre, Co-expression of three MEP pathway genes and geraniol 10-hydroxylase in internal phloem parenchyma of *Catharanthus roseus* implicates multicellular translocation of intermediates during the biosynthesis of monoterpene indole alkaloids and isoprenoid-derived primary metabolites. *Plant J.* **38**, 131-141 (2004).
20. R. Höfer *et al.*, Geraniol hydroxylase and hydroxygeraniol oxidase activities of the CYP76 family of cytochrome P450 enzymes and potential for engineering the early (seco)iridoid pathway. *Metab. Eng.* in press.
21. D. Pompon, B. Louerat, A. Bronine, P. Urban, Yeast expression of animal and plant P450s in optimized redox environments. *Cytochrome P450, Pt B* **272**, 51-64 (1996).
22. A. Schmidt *et al.*, A bifunctional geranyl and geranylgeranyl diphosphate synthase is involved in terpene oleoresin formation in *Picea abies*. *Plant Physiol.* **152**, 639-655 (2010).

23. L. Dong *et al.*, Characterization of two geraniol synthases from *Valeriana officinalis* and *Lippia dulcis*: similar activity but difference in subcellular localization. submitted.
24. G. Guirimand *et al.*, Spatial organization of the vindoline biosynthetic pathway in *Catharanthus roseus*. *J. Plant Physiol.***168**, 549-557 (2011).
25. S. R. Jensen, O. Kirk, B. J. Nielsen, Application of the Vilsmeier Formylation in the Synthesis of 11-C-13 Labeled Iridoids. *Tetrahedron***43**, 1949-1954 (1987).
26. S. R. Jensen, O. Kirk, B. J. Nielsen, Biosynthesis of the Iridoid Glucoside Cornin in *Verbena-Officinalis*. *Phytochemistry***28**, 97-105 (1989).
27. L. van der Fits, J. Memelink, Comparison of the activities of CaMV 35S and FMV 34S promoter derivatives in *Catharanthus roseus* cells transiently and stably transformed by particle bombardment. *Plant Mol. Biol.***33**, 943-946 (1997).
28. J. Zuo, Q. W. Niu, N. H. Chua, Technical advance: An estrogen receptor-based transactivator XVE mediates highly inducible gene expression in transgenic plants. *Plant J.***24**, 265-273 (2000).
29. M. B. Eisen, P. T. Spellman, P. O. Brown, D. Botstein, Cluster analysis and display of genome-wide expression patterns. *P. Natl. Acad. Sci. USA***95**, 14863-14868 (1998).
30. T. Tohge *et al.*, Toward the Storage Metabolome: Profiling the Barley Vacuole. *Plant Physiol.***157**, 1469-1482 (2011).
31. U. K. Laemmli, Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature***227**, 680-685 (1970).
32. A. Shevchenko, M. Wilm, O. Vorm, M. Mann, Mass spectrometric sequencing of proteins silver-stained polyacrylamide gels. *Anal. Chem.***68**, 850-858 (1996).
33. T. Schneider *et al.*, A proteomics approach to investigate the process of Zn hyperaccumulation in *Noccaea caerulea* (J & C. Presl) F.K. Meyer. *Plant J.***73**, 131-142 (2013).
34. A. Keller, A. I. Nesvizhskii, E. Kolker, R. Aebersold, Empirical statistical model to estimate the accuracy of peptide identifications made by MS/MS and database search. *Anal. Chem.***74**, 5383-5392 (2002).
35. A. I. Nesvizhskii, A. Keller, E. Kolker, R. Aebersold, A statistical model for identifying proteins by tandem mass spectrometry. *Anal. Chem.***75**, 4646-4658 (2003).

36. F. L. H. Menke, A. Champion, J. W. Kijne, J. Memelink, A novel jasmonate- and elicitor-responsive element in the periwinkle secondary metabolite biosynthetic gene *Str* interacts with a jasmonate- and elicitor-inducible AP2-domain transcription factor, ORCA2. *EMBO J.***18**, 4455-4463 (1999).
37. R. Töpfer, V. Matzeit, B. Gronenborn, J. Schell, H. H. Steinbiss, A set of plant expression vectors for transcriptional and translational fusions. *Nucleic Acids Res.***15**, 5890 (1987).
38. W. Chiu *et al.*, Engineered GFP as a vital reporter in plants. *Curr. Biol.***6**, 325-330 (1996).
39. Y. Niwa, T. Hirano, K. Yoshimoto, M. Shimizu, H. Kobayashi, Non-invasive quantitative detection and applications of non-toxic, S65T-type green fluorescent protein in living plants. *Plant J.***18**, 455-463 (1999).
40. B. K. Nelson, X. Cai, A. Nebenfuhr, A multicolored set of in vivo organelle markers for co-localization studies in Arabidopsis and other plants. *Plant J.***51**, 1126-1136 (2007).
41. H. K. Kim, Y. H. Choi, R. Verpoorte, NMR-based metabolomic analysis of plants. *Nat. Protoc.***5**, 536-549 (2010).
42. G. Guirimand *et al.*, Strictosidine activation in Apocynaceae: towards a "nuclear time bomb"? *BMC Plant Biol.***10**, (2010).
43. S. Mahroug, V. Courdavault, M. Thiersault, B. St-Pierre, V. Burlat, Epidermis is a pivotal site of at least four secondary metabolic pathways in *Catharanthus roseus* aerial organs. *Planta***223**, 1191-1200 (2006).
44. H. H. Nour-Eldin, B. G. Hansen, M. H. Norholm, J. K. Jensen, B. A. Halkier, Advancing uracil-excision based cloning towards an ideal technique for cloning PCR fragments. *Nucleic Acids Res.***34**, e122 (2006).
45. T. Omura, R. Sato, The carbon monoxide-binding pigment of liver microsomes. I. Evidence for its hemoprotein nature. *J. Biol. Chem.***239**, 2370-2378 (1964).
46. O. Voinnet, S. Rivas, P. Mestre, D. Baulcombe, An enhanced transient expression system in plants based on suppression of gene silencing by the p19 protein of tomato bushy stunt virus. *Plant J.***33**, 949-956 (2003).

**Acknowledgments:** We are grateful to Richard Twyman for critical reading of the manuscript and to Søren Rosendal Jensen for advice on iridoid substrate synthesis. David Nelson is acknowledged for naming P450 enzymes. The research leading to these results has received funding from the European Union Seventh Framework Programme FP7/2007-2013 under grant agreement number 222716 – SMARTCELL. TI was supported by the Marie Curie ITN action P4Fifty. The data reported in this paper are presented in the main text and supplementary materials. Transcriptomic data are available from NCBI Short Read Archive under the accession number SRP026417. Sequences of genes mentioned in the text have been deposited in GenBank/EMBL/DDBJ/SRA with accession numbers KF302066 (IO), KF302067 (7DLH), KF302068 (7DLGT), KF302069 (8-HGO), KF302070 (CYP81Q32), KF302071 (CYP81C13), KF302072 (CYP71AY2), KF302073 (CYP71BT1), KF302074 (CYP72A225), KF302075 (CYP76T24), KF302076 (CYP71D2), KF302077 (ADH9), KF302078 (ADH13), KF302079 (ADH14), KF302080 (UGT9) and KF309243 (CYP71AY1).

## Materials and methods

**Chemicals.** The substrates 8-carboxygeranial, 8-carboxygeranic acid, 8-oxogeranic acid and 8-hydroxygeranic acid were synthesized on order by Synthelor (Vandoeuvre-Lès-Nancy, France), whereas 8-OH-geraniol, 8-oxogeraniol, 8-OH-geranial and 8-oxogeranial were synthesized by Chiralix B.V. (Nijmegen, Netherlands). Iridodial-glucoside, iridotrial-glucoside and 7-deoxyloganic acid were synthesized from aucubin extracted from *Aucuba japonica* leaves by Chiralix B.V. as described<sup>25,26</sup>. The aglucone iridoid pathway intermediates were produced by incubation with almond  $\beta$ -glucosidase (Sigma Aldrich) in 50 mM acetate buffer (pH 5). Loganetic acid and loganetin were produced by the deglycosylation of loganic acid and loganin (Extrasynthese). Aglycones were extracted with diethyl ether, evaporated under N<sub>2</sub> and quantified by <sup>1</sup>H-NMR.

## **Transcriptomic analysis.**

Transgenic derivatives of *C. roseus* cell line MP183L (overexpressing the ORCA transcription factors) were generated by particle bombardment<sup>27</sup> with derivatives of the pER8 plasmid<sup>28</sup> carrying either the ORCA2 or ORCA3 ORFs. Selected transgenic lines were treated for 24 h with 10  $\mu$ M estradiol, and RNA was isolated as described<sup>16</sup>. Illumina HiSeq2000 RNA sequencing, assembly, annotation and mapping of the RNA-Seq reads onto the reference transcriptome was carried out as described<sup>16</sup>. Complete linkage hierarchical cluster analysis was achieved using the CLUSTER and TREEVIEW software<sup>29</sup> and the  $\log_{10}$  transformed values of the normalized FPKM values were used as input for CLUSTER.

## **Plant material.**

*Catharanthus roseus* var. Little Bright Eyes seeds were sown in sterilized soil and covered with transparent plastic until germination. The soil was kept humid. The plants were fertilized weekly with 0.2% liquid Wuxal (29 g/l nitrate-N; 46 g/l ammonium-N; 25 g/l carbamide-N; 100 g/l P<sub>2</sub>O<sub>5</sub> total phosphate; 75 g/l K<sub>2</sub>O; 124 mg/l B; 50 mg/l Cu; 248 mg/l Fe; Cu, Fe, Mn and Zn as EDTA-chelate). The plants were re-potted twice during further growth.

## **Isolation of epidermal protoplasts.**

Young, light-green *C. roseus* leaves (length 4–7 cm) from side branches (without buds or flowers) of 8 to 11-week-old plants were harvested for protoplast isolation. The mid-vein was removed and the leaves were cut into 1–2 mm strips. Protoplasts were isolated as described<sup>30</sup>. To collect epidermal protoplasts, a layer of 1 ml betaine solution (0.5 M betaine, 1 mM CaCl<sub>2</sub>, 10 mM MES, pH 5.6 with KOH) was added on top and the tubes containing the protoplast mix. After centrifugation for 7 min at 1500 rpm and 4°C, epidermal protoplasts were collected from the upper interphase. The suspension was mixed with 4 ml protoplast solution and 25% Percoll (pH 6), and overlaid with 1.5 ml of betaine solution for a second purification step. The tubes were centrifuged at 700 rpm for 30 min and the epidermal

protoplasts were again collected from the upper interphase. Protoplasts were pelleted in the betaine solution.

### **Isolation of mesophyll protoplasts.**

Mesophyll protoplasts were isolated as described for the epidermal protoplasts although with the MCP solution was replaced with TEX solution (3.2 g/l Gamborg's B5 medium, 3.1 mM NH<sub>4</sub>NO<sub>3</sub>, 5.1 mM CaCl<sub>2</sub>, 2.6 mM MES, 0.4 M sucrose, pH 5.6-5.8 with 0.5 M KOH), and the betaine solution replaced with mannitol/W5 (1 mM D-glucose, 30 mM NaCl, 25 mM CaCl<sub>2</sub>, 1 mM KCl, 0.3 mM MES, 320 mM  $\alpha$ -mannitol, pH 5.6-5.8 with KOH). For the first gradient, 5–10% Percoll (pH 6) was added instead of 3–5 %.

### **Collection of protoplast microsomes.**

The protoplast pellet was resuspended in 2–4 ml buffer A (20 mM HEPES pH 7.2, 1 tablet/10 ml Roche Protease Inhibitor, 1 mM PMSF). The mixture was pressed 10–20 times through a syringe with needle and centrifuged for 10 min at 3000 rpm and 4°C. The samples were ultracentrifuged for 1 h at 30,000 rpm. The membrane pellets were dissolved in 1 ml washing solution (0.3 M NaCl, 20 mM HEPES buffer, pH 7.2 (with KOH), 1 tablet/10 ml Roche Protease Inhibitor, 1 mM PMSF) by stirring with a small brush, and then vortexed for 30 s. The microsomes were frozen at –80°C for at least 1 h, thawed and then centrifuged for 1 h at 30,000 rpm and 4°C. The pellets were dissolved in 50–100  $\mu$ l buffer A for further analysis. The protein concentration was determined using the Bio-Rad DC protein assay according to the manufacturer's instructions.

### **Mass spectrometry.**

Approximately 50  $\mu$ g of protein was separated by one-dimensional gel electrophoresis<sup>31</sup> in 10% polyacrylamide gels to reduce sample complexity. The gels were stained with Coomassie Brilliant Blue, the lanes were cut into ten slices, the proteins were reduced with tris(carboxyethyl)phosphine hydrochloride (TCEP) and sulhydryl groups were blocked with iodoacetamide. In-gel digestion with sequencing-grade modified trypsin (Promega, reference

V5111) was carried out overnight at 37°C<sup>32</sup>. The resulting peptides were recovered by adding 40 mM Tris-HCl buffer (pH 8.0) and 50% acetonitrile containing 1% formic acid. The peptide mixtures were desalted by solid-phase extraction on C<sub>18</sub> reversed-phase columns and analyzed on an LTQ Orbitrap mass spectrometer (Thermo Fischer Scientific, Bremen, Germany) coupled to an Eksigent Nano HPLC system (Eksigent Technologies, Dublin, CA, USA) as previously described<sup>33</sup>.

### **Database searching and protein identification.**

Protein databases were searched using Mascot v2.3. Raw data were searched against a composite database consisting of all entries in the NCBI Viridiplantae database (released in November 2010), all publicly-available *C. roseus* expressed sequence tags (downloaded from NCBI in November 2010) and the reference transcriptome released on CathaCyc and ORCAE<sup>16</sup> (database contained forward and reverse protein entries, total number of protein entries 1,166,013). The parameters for precursor and fragment ion mass tolerance were set to 5 ppm and 0.8 Da, respectively. One missed trypsin cleavage was allowed. Carboxamidomethylation of cysteine was specified as fixed modification, and oxidation of methionine and pyroglutamate formation from glutamine were selected as variable modifications.

### **Data processing and protein quantitation.**

Scaffold v3.0 (Proteome Software, Portland, OR, USA) was used to validate and quantify MS/MS-based peptide and protein identifications. Peptide identifications were accepted if they were established at greater than 95% probability as specified by the Peptide Prophet algorithm<sup>34</sup>. Protein identifications were accepted if they were established at greater than 90% probability and at least one peptide was uniquely assigned to a corresponding protein in a minimum of two of our samples. Protein probability was assigned by the Protein Prophet algorithm<sup>35</sup>. Proteins that were identified with the same set of peptides and could not be differentiated by MS/MS analysis were grouped to protein clusters to satisfy the principles of parsimony. Peptide and protein false-discovery rates (FDR) were determined by the Scaffold



software. A peptide FDR of 0.01% and a protein FDR of 0.2% were computed. The Scaffold software was also used to determine protein abundance in the mesophyll and epidermal fractions based on the number of spectra assigned to each protein. An F-test was applied to assess significant differences in protein abundances.

**Gene isolation.** Open reading frames (ORFs) were amplified by PCR from a pACT2 cDNA library of a *C. roseus* cell culture elicited with yeast extract<sup>36</sup> using the primers listed in Extended Data Table 2. For expression in plants, the ORFs were transferred to vector pRT101<sup>37</sup> to bring them under the control of the Cauliflower mosaic virus 35S promoter, and the entire expression cassettes were then transferred to the binary vector pCAMBIA1300 (Cambia). For expression in *E. coli*, the ORFs were transferred to vector pASK-IBA45plus (IBA) and/or pET16-H (Novagen pET-16b derivative). Probes for *in situ* hybridization were prepared from the same ORFs cloned in pBluescript II SK+ (Stratagene). For localization analysis the ORFs were transferred to vector pTH2<sup>38, 39</sup> and/or pTH2BN (a derivative of pTH2). The marker for nucleocytosolic localization (pRT101-mCherry) was prepared by amplifying the mCherry ORF from plasmid ER-rk<sup>40</sup> (The *Arabidopsis* Information Resources, TAIR, clone CD3-959).

### **Isolation of His-tagged recombinant proteins.**

Recombinant proteins carrying a His<sub>6</sub> tag were expressed using plasmid pASK-IBA45plus and/or pET16-H in *E. coli* strain BL21 (DE3) pLysS and purified by Ni-NTA agarose chromatography (Qiagen). For quality analysis, the recombinant proteins were separated by 12.5% (w/v) SDS-PAGE, transferred to Protran nitrocellulose membranes (Whatman) by semidry electroblotting, and western blots were probed with mouse monoclonal anti-His horseradish peroxidase (HRP) conjugate antibodies (5Prime). Antibody binding was detected by incubation in 250  $\mu$ M sodium luminol, 0.1 M Tris-HCl (pH 8.6), 3 mM H<sub>2</sub>O<sub>2</sub> and 67  $\mu$ M p-coumaric acid, followed by exposure to X-ray film.

**Enzymatic assays of UGT and oxidoreductases.** UGT activity was detected in 0.1 ml reaction buffer containing 50 mM potassium phosphate (pH 7.5), 2 mM UDP-glucose, 5–1000  $\mu$ M 7-deoxyloganic acid or 2 mM of other tested compounds and 50–1000 ng of purified enzyme. Reactions were incubated at 32°C for 15 min and stopped by adding 1 volume of methanol,

mixed by vortexing and kept on ice for 10 min. The tubes were centrifuged at 4000 x g for 10 min, and the supernatants were passed through 0.22- $\mu$ m nylon filters.

Oxidoreductase activity was detected in 1 ml reaction buffer containing 50 mM bis-tris propane (pH 9 for oxidation and pH 7.5 for reduction), 2–1000  $\mu$ M 8-OH-geraniol, 8-oxogeraniol, 8-OH-geranial, 8-oxogeranial or other tested compounds, 2–2000  $\mu$ M NAD<sup>+</sup> or NADH and 50–1000 ng of purified enzyme. Reactions were incubated for 15 min at 32°C, stopped by adding 0.2 volumes of 1 M sodium citrate (pH 3) and centrifuged and filtered as above. Quantitative assays were carried out by measuring NADH production at 340 nm in a Nanodrop2000c (Thermoscientific)

### **Chromatographic analysis.**

Qualitative and quantitative analysis of 7-DLGT and 8-HGO products was carried out using an Agilent series 1200 HPLC with a diode array detector (DAD) and a Polymer Laboratories PL-ELS 2100 ICE evaporative light scattering detector (ELSD) and a Phenomenex Luna 5 micron 150 x 4.6 mm C<sub>18</sub> column. The injection volumes were 10 or 100  $\mu$ l. The binary solvent system consisted of acetonitrile and 0.1% trifluoroacetic acid in water. The elution program was: 5 min isocratic 10% acetonitrile and then 25 min gradient until 95% acetonitrile. Peak areas were calculated using Agilent ChemStation.

### **NMR.**

Structures of enzyme products were analyzed by nuclear magnetic resonance (NMR) spectroscopy in 750  $\mu$ l of acetone-d<sub>6</sub> or methanol-d<sub>4</sub> in a 5 mm NMR glass tube<sup>41</sup>.

### **Subcellular localization studies.**

*C. roseus* MP183L cell suspension cultures were maintained by weekly 10-fold dilution in 50 ml Linsmaier&Skoog (LS) medium containing 88 mM sucrose, 2.7  $\mu$ M 1-NAA and 0.23  $\mu$ M kinetin (LS-13). The cells were grown at 25°C with a 18/6h light dark cycle. The cells were bombarded<sup>27</sup> using the plasmids pTH2-IO, pTH2-7-DLH, pTH2-7-DLGT, pTH2BN-7-DLGT, pTH2-8-HGO and pTH2BN-8-HGO. The first two were combined with equal amounts of the

ER marker ER-rk, and the others with the nucleocytosolic marker mCherry. Bombarded cells were placed on Petri dishes with LS-13 medium and viewed after 24 h using a Zeiss Observer laser scanning microscope.

### **In situ hybridization.**

pBluescript plasmid derivatives containing the cDNAs for 8-HGO, IO, 7-DLGT and 7-DL were used for the synthesis of antisense and sense digoxigenin-labeled riboprobes as previously described<sup>19</sup>. G8O antisense probes<sup>19</sup> and SGD antisense probes<sup>42</sup> were used as internal phloem-associated parenchyma and epidermis markers, respectively. Paraffin-embedded serial longitudinal sections of young developing leaves were hybridized with digoxigenin-labeled riboprobes and localized with antidigoxigenin-alkaline phosphatase conjugated antibodies<sup>43</sup>.

### **P450 expression in yeast and corresponding enzyme assays.**

P450 coding sequences were amplified from pRT100 source vectors using specific primers to introduce USER™ sites at the 5' and 3' ends of each sequence. The genes were subsequently transferred to the plasmid pYeDP60 using the USER™ cloning technique (New England Biolabs, Ipswich, UK) as previously described<sup>44</sup>. The resulting recombinant plasmids were introduced into *S. cerevisiae* strain WAT11, cultivated at 28°C and P450 expression was induced as described<sup>21</sup>. Cells were harvested by centrifugation and manually broken with 0.45-mm glass beads in 50 mM Tris-HCl buffer (pH 7.5) containing 1 mM EDTA and 600 mM sorbitol. The homogenate was centrifuged for 10 min at 10,000 × g and the supernatant was centrifuged for 1 h at 30,000 × g. The pellet, comprising microsomal membranes, was resuspended in 50 mM Tris-HCl (pH 7.4), 1 mM EDTA and 30% (v/v) glycerol with a Potter-Elvehjem homogenizer and stored at -20°C. All procedures for microsomal preparation were carried out at 0–4°C. P450 expression was evaluated as described<sup>45</sup> and enzymatic activities were determined in a standard 0.1-ml assay comprising for IO 2.3 nmol cytochrome P450, 0.6 mM NADPH and substrate in 20 mM sodium phosphate (pH 7.4). The reaction was initiated by the addition of NADPH and was stopped after 5 min by the

addition of 10  $\mu$ l 1M HCl. Iridoids were extracted in 1 ml ethyl acetate, and the organic phase was concentrated to 200  $\mu$ l before GC-FID analysis. For 7-DLH, 10  $\mu$ l of yeast microsomes expressing 7-DLH (130  $\mu$ g of microsomal protein) were incubated for 20 min at 27°C, in 0.1 ml of 20 mM Na-phosphate (pH 7.4) containing 0.6 mM NADPH and substrate. The reaction was initiated by the addition of NADPH and was stopped after 20 min on ice. After addition of 50  $\mu$ l of 50 % acetic acid, tubes were vortexed and centrifuged. The supernatant was run on reverse-phase HPLC (Alliance 2695 Waters system, NOVA-PAK C18 4.6 x 250 mm column) with photo-diode array detection at 236.5 nm. Peak areas of the product(s) were used to calculate the catalytic parameters of each enzyme.

**Leaf disc assays.** Five-week-old *N. benthamiana* leaves were infiltrated with *A. tumefaciens* transformed with vector pC1300 containing the relevant candidate genes, plus the helper plasmid p19. Five days post-infiltration, leaves were used in a leaf disc assays as previously described<sup>20</sup>.

### Pathway reconstruction in *N. benthamiana*.

The pathway genes were transiently expressed in the leaves of five-week-old *N. benthamiana* plants by agroinfiltration as previously described<sup>23</sup>. Briefly, bacteria carrying the relevant expression constructs (*PaGPPS*, *VoGES*, *G8O*, *8-HGO*, *IS*, *IO*, *7-DLGT*, *7-DLH*, *LAMT*, *SLS*, *TDC*, *STR*, empty vector or TBSV p19<sup>46</sup>) were grown individually at 28°C for 24 h. Cells were harvested by centrifugation and then resuspended in infiltration buffer containing 10 mM MES (DuchefaBiochemie), 10 mM MgCl<sub>2</sub> and 100  $\mu$ M acetosyringone (4'-hydroxy-3',5'-dimethoxyacetophenone, Sigma) to a final OD<sub>600</sub> of ~0.5. For all gene combinations that compared subsequent steps in the pathway, the amounts of cell suspension for each expression construct were kept constant by adding the corresponding amount of *A. tumefaciens* carrying an empty vector. In several experiments, pathway intermediates were injected into the same leaves three days after agroinfiltration. Compounds used for infiltration were diluted to a final concentration of 400  $\mu$ M in methanol/water (1:19), with the same ratio of methanol/water alone as a negative control and 400  $\mu$ M 7-deoxyloganic acid, 8-carboxygeranic acid, secologanin or strictosidine as positive controls. Leaves were harvested

for metabolite analysis 5 days after agroinfiltration. Frozen, powdered *N. benthamiana* leaves (200 mg aliquots) were extracted in 0.6 ml 99.867% methanol, 0.133% formic acid and 5  $\mu$ l of the extract was analyzed using a Waters Alliance 2795 HPLC connected to a QTOF Ultima V4.00.00 mass spectrometer (Waters, MS technologies, UK). Measurements were taken in negative ionization mode. LC-MS data were acquired using MassLynx 4.0 (Waters) and processed using MetAlign version 1.0. The normalized and log-transformed data matrix was used for principal component analysis implemented in GeneMath XT v 2.1.

construct	restriction site	primer F	primer R
pRT101 IO	XhoI, KpnI	GGCCITTCGAGATGGCGACCACACTTCG	AAGGCCGGTACCTTAGATGAACCTCTCTTCTTAGGGATG
pRT101 7-DLH	XhoI, KpnI	GGCCITTCGAGATGGAATGAACTTCAACTCAAAITTA	GGCCITTCGAGATGAGTGTGCGAAGATCAAAATGAG
pRT101 7-DLGT	KpnI, SpeI	CCGGTACCAATGGGTCTCAAGAAACAATTTG	CCCTAGTTCAAAATAACAGTGAATTTATGTAATCAAC
pRT101 8-HGO	KpnI, XbaI	CCGGTACCAATGACCAAGACCAATTCCTCC	GGTCTAGTATGAACTTGAATAACAACCTTTGACACAATC
pRT101 CYP81Z1	XhoI, SacI	GGTCTCGAGATGGAGGTTCCITTTTCTAC	ACCGAGCTCTATGGGTTATTTTCCATGG
pRT101 CYP71AY1	XhoI, EcoRI	GGCCITTCGAGATGGATCAGCTGATGAACTTCTC	GGCCITTCGAAATTCCTAGTTCCTTCAACTACAGTTGAGATG
pRT101 IS	XhoI, KpnI	CCGGTTCGAGAAATGATGGTGGTGGGAAGAGG	CCITGGGGTACCCTAAGGAATAAACCTATAATCCCTC
pRT101 SLS	XhoI, KpnI	GTTCTCGAGATGGATGGATGGATGATACCAAT	CCGGTACCCTAGCTCTCAAGCTTCTTGTA
pASKIBA45plus 7-DLGT	XhoI, KpnI	CCGGTACCAATGGTTCCTCAAGAAACAATTTG	GGCTCGAGCAAAATAACAGTGAATTTATGTAATCAAC
pET16H 8-HGO	XhoI, KpnI	CCCTCGAGATGACCAAGACCAATTCCTCC	GGGTACCCTTAGAACTTGAACAACCTTTGACACAATC
pASKIBA45plus IS	XhoI, Sall	CCGGTTCGAGATGATGGTGGTGGGAAGAGG	AACCGGGTGCACAAAGGAATAAACCTATAATCCCTC
pET16H IS	XhoI, KpnI	CCGGTTCGAGATGATGGTGGTGGGAAGAGG	AACCGGGTACCCTAAGGAATAAACCTATAATCCCTC
pTH2 IO	Sall	CCGTCGACAAAATGGCGACCACATCTTCG	CCGTCGACGATGAACCTCTCTCTTAGGGATG
pTH2 7-DLH	Sall	CCGTCGACAAAATGGAAITGAACTTCAAAATCAATTA	CCGTCGACGAGTGTGCGAAGATCAAAATGAG
pTH2 7-DLGT	Sall	CCGTCGACAAAATGGGTCTCAAGAAACAATTTG	CCGTCGACAAATAACAGTGAATTTATGTAATCAAC
pTH2BN 7-DLGT	XhoI, SpeI	GGCTCGAGATGGGTTCCTCAAGAAACAATTTG	GGACTAGTTCAAAATAACAGTGAATTTATGTAATCAAC
pTH2 8-HGO	Sall	CCGGCCGTCGACAAAATGACCAAGACCAATTCCTCC	CCGGCCGTCGACGAACTTGAATAACAACCTTTGACACAATC
pTH2BN 8-HGO	XhoI, SpeI	CCGGGGCTCGAGATGACCAAGACCAATTCCTCC	CCGGGGACTAGTTAGAACTTGAATAACAACCTTTGACACAATC
pRT101 mCherry	XhoI, KpnI	GGTTCCTCGAGAAAATGGTGGCAAGGGCGAGGAGGA	GGCCITTCGATACCTTACITGTTACAGCTGCTCCATGCCGCCCGG
pYeDP60 CYP81Z1	USER	GGCTTAAUATGGAGGTTTCTTTTCTACAC	GGTTTAAUATATGGGTTAATTTCCATGGG
pYeDP60 CYP71AY1	USER	GGCTTAAUATGGATCAGCTGATGAACTTCTCTC	GGTTTAAUATAGTTCCTTCAACTACAGTTGAGATG
pYeDP60 IO/CYP76A26	USER	GGCTTAAUATGGCGACCACACTTCTCTC	GGTTTAAUATAGATATGAACTCTCTCTT
pYeDP60 7-DLH/CYP72A224	USER	GGCTTAAUATGGAAITGAACTTCAAAATCA	GGTTTAAUATAGAGTTTGGCAGAAATCAA
pYeDP60 G80/CYP76B6	USER	GGCTTAAUATGGATTACCTTACCATAAATTAAC	GGTTTAAUATAAAGGGTCTTGGTACAGC

Table S2. Sequences of primers and cloning methods used for plasmid constructions. Primers are displayed from 5' to 3' end.

**Author notes**

J.M., A.G., K.MO-C., R.V., E.M., H.B., N.N. and D.W. conceived the project and general strategy; A.G. and J.P. performed the expression analysis and gene selection. K.M. isolated the candidate genes and prepared most vectors, expressed and functionally characterized 8-HGO and 7DLGT, performed subcellular localization experiments and performed parts of the synthesis of iridoid substrates; L.D. did and analysed the pathway reconstruction in *N. benthamiana*; N.N. assisted by T.I. expressed and functionally characterized the candidate P450 genes; T.S. assisted by LW did the proteomics analysis; V.B. carried out the *in situ* hybridizations; N.N., RL. contributed to metabolite analysis; D.W., J.M., A.G., S.vdK. and H.B. wrote the manuscript; D.W. acted as coordinator.

**Summary**



*Catharanthus roseus* (Madagascar periwinkle) is the best studied medicinal plant. It produces the important class of secondary metabolites the monoterpene indole alkaloids (MIA) and their precursors the (seco)iridoids, bioactive compounds with a wide spectrum of high-value pharmacological and insect-repellent activities. From the over 130 known *C. roseus* MIAs many have pharmaceutical applications such as the anti-hypertensive drugs serpentine and ajmalicine and the potent antitumor agents bisindole alkaloids vincristine and vinblastine that are widely used to treat several types of cancer such as Hodgkins disease, Kaposi's sarcoma, breast cancer, bladder cancer and testicular cancer. Vinblastine and vincristine are produced by *C. roseus* in extremely low levels, leading to high market prices and poor availability. Because of their complex structures total synthesis is unfeasible (van der Heijden et al., 2004). After decades of research only parts of the MIA biosynthetic pathway are known. Complete knowledge of the biosynthesis is essential for biotechnological production of MIA's and (seco)iridoids.

The aim of this work was elucidation and further description of the iridoid pathway from *C. roseus* from a metabolic engineering perspective. A candidate based approach was taken to find enzymes catalyzing both hypothetical and completely unknown intermediate reactions in iridoid biosynthesis and to find suitable enzymes for biotechnological applications. The candidates were picked based on amino acid and nucleotide sequence homology with known enzymes and the screening was refined by gathering additional information such as expression pattern in different conditions and established tissue localization of proteins and transcripts. Based on the evidence at the onset of these studies the pathway was missing 4-6 enzymes with completely new functionalities. In addition the gene coding for enzyme geraniol synthase (GES) was known from *Ocimum basilicum* (Iijima et al., 2004), *Cinnamomum tenuipilum* (Yang et al., 2005), and *Perilla citriodora* (Ito and Honda, 2007) but *C.roseus* GES (*CrGES*) was not identified. Several oxidoreductases of different classes, cytochrome P450's, glucosyltransferases and a terpene synthase were chosen as candidates according to our pathway model (Fig. 1). These candidates were expressed in *Escherichia coli* and *Saccharomyces cerevisiae* and used for candidate screening with *in vitro* biochemical assays. The pathway was completely reconstituted by transiently expressing the final four candidates and known iridoid biosynthesis genes in *Nicotiana benthamiana* and new metabolites were analyzed by

LC-MS, *CrGES* was validated by *in vitro* biochemical assays of an enzyme produced in *E. coli* and it was further characterized by expression in a farnesyl biosynthesis mutant *S. cerevisiae* strain and by *in situ* hybridization and by transient expression of a GFP-fusion in a *C. roseus* cell culture. Geraniol synthases from *Valeriana officinalis* (*VoGES*) and *Lippia dulcis* (*LdGES*) were characterized *in vitro* and *in planta* to assess biotechnological aspects of monoterpene biosynthesis.

**Chapter 1** serves as a general overview of MIA and iridoid biosynthesis and present the research strategy for this work.

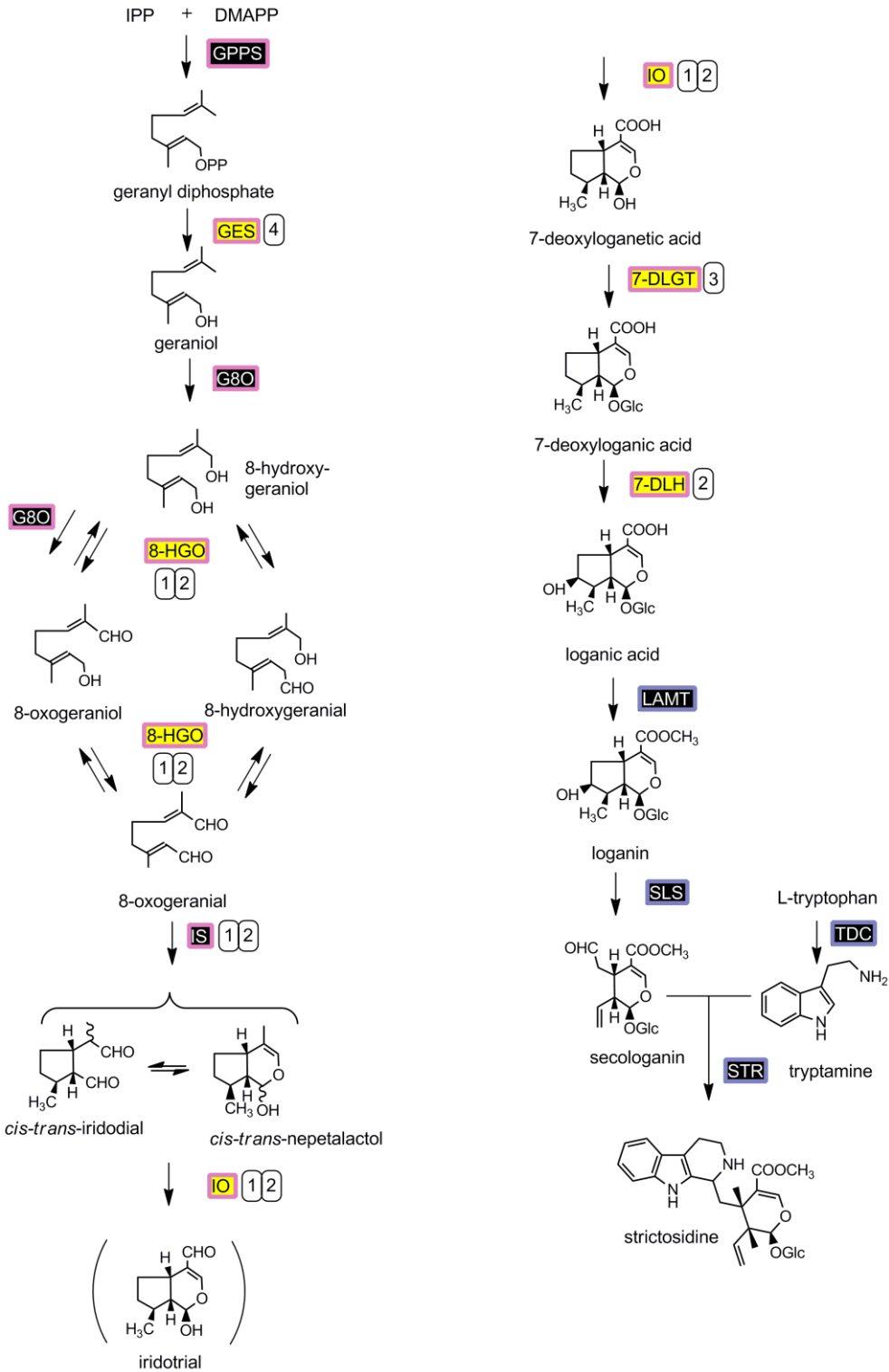
**Chapter 2** presents the cloning and functional characterization of the *C. roseus* cDNA for *CrGES* encoding geraniol synthase (*CrGES*). The full-length *CrGES* was over-expressed in *E. coli* and the purified recombinant protein catalyzed the *in vitro* conversion of GPP into geraniol only with a  $K_m$  value of 58.5  $\mu\text{M}$  for GPP. *In vivo* *CrGES* activity was evaluated by heterologous expression in *S. cerevisiae* strain K197G mutated in the farnesyl diphosphate synthase gene. Analysis of culture extracts by gas chromatography-mass spectrometry confirmed the excretion of geraniol into the growth medium. Transient transformation of *C. roseus* cells with a Yellow Fluorescent Protein-fusion construct revealed that *CrGES* is localized in the plastid stroma and the stromules. This could suggest that the availability of its product geraniol to the next enzyme G10H could be facilitated through stromule/ER interactions. In aerial plant organs, RNA *in situ* hybridization showed specific labelling of *CrGES* transcripts in the internal phloem associated parenchyma as observed for other characterized genes involved in the early steps of MIA biosynthesis. Finally, when cultures of *Catharanthus* cells were treated with the alkaloid-inducing hormone methyl jasmonate, an increase in *CrGES* transcript levels was observed. This observation coupled with the tissue-specific expression and the subcellular compartmentalization support the idea that *CrGES* initiates the monoterpene branch of the MIA biosynthetic pathway.

**Chapter 3** discloses the recombinant expression and characterization of two geraniol synthases (*GES*). *GES*s from *Valeriana officinalis* (*VoGES*) and *Lippia dulcis* (*LdGES*), were isolated from *E. coli* and were shown to turn geranyl diphosphate (GPP) into geraniol as a specific product *in vitro* with  $K_m$  values of 32  $\mu\text{M}$  and 51  $\mu\text{M}$  for GPP, respectively. The *in*

*planta* enzymatic activity and sub-cellular localization of VoGES and LdGES were characterized in stably transformed tobacco plants and using transient expression in *N. benthamiana*. Transgenic tobacco expressing VoGES or LdGES accumulate geraniol, oxidized geraniol compounds like geranial, geranic acid and hexose conjugates of these compounds to similar levels. Geraniol emission of leaves was lower than that of flowers, which could be related to higher levels of competing geraniol-conjugating activities in leaves. GFP-fusions of the two GES proteins show that VoGES resides (as expected) predominantly in the plastids, while LdGES import into the plastid is clearly impaired compared to that of VoGES, resulting in both cytosolic and plastidic localization. Geraniol production by VoGES and LdGES in *N. benthamiana* was nonetheless very similar. Expression of a truncated version of VoGES or LdGES (cytosolic targeting) resulted in the accumulation of 30% less geraniol glycosides than with the plastid targeted VoGES and LdGES, suggesting that the substrate geranyl diphosphate is readily available, both in the plastids as well as in the cytosol. The potential use of GES in the engineering of the MIA pathway in heterologous hosts is discussed.

**Chapter 4** reports the discovery of the last five missing steps of the (seco)iridoid biosynthesis pathway from *C. roseus* (Fig. 1) using an integrated approach combining gene expression and coexpression analysis, tissue-specific proteomics and biochemical assays. Five new biosynthesis genes were cloned and the corresponding enzymes characterized. One of the enzymes, the alternative terpene cyclase iridoid synthase (IS) was recently published while this work was in progress (Geu-Flores et al., 2012). The novel enzyme 8-hydroxygeraniol oxidoreductase (8-HGO) catalyzes the oxidation of 8-hydroxygeraniol both at position 1 and 8 to ultimately yield 8-oxogeranial. Another novel enzyme, iridoid oxidase (IO, CYP76A26) was found to turn *cis-trans*-nepetalactol (also known as iridodial hemiacetal form) and *cis-trans*-iridodials into 7-deoxyloganetic acid. Also iridotrial was utilized by the enzyme resulting in the same product. 7-deoxyloganetic acid was in turn used by 7-deoxyloganetic acid glucosyltransferase (7DLGT, UGT709C2) to produce 7-deoxyloganic acid. This compound was hydroxylated by a fourth enzyme, 7-deoxyloganic acid hydroxylase (7DLH, CYP72A224), into loganic acid. All of the enzymes were found to be expressed in vascular tissue of *C. roseus* aerial organs, notably the internal phloem associated parenchyma (IPAP) cells, by in situ hybridization. The novel biosynthesis genes showed a very similar expression

pattern in whole plants and in tissues cultures in multiple conditions. The subcellular localization of GFP fusions of the CYP450 enzymes (IO, 7DLH) was found to be in the ER as expected. 8-HGO and 7DLGT N- and C-terminal fusion proteins were found localize in the cytosol. Expression of the eight genes encoding this pathway together with two genes boosting precursor formation and two downstream alkaloid biosynthesis genes in an alternative plant host allowed the heterologous production of the complex MIA strictosidine.



**Fig.1 The secoiridoid pathway in *C. roseus*:** Reactions for which the corresponding *C.roseus* gene has been described in literature have a black background, reactions for which the corresponding gene has been newly described in this thesis have a yellow background. IPP: isopentenyl diphosphate, DMAPP: dimethylallyl diphosphate, GPPS: geranyl diphosphate synthase, GES: geraniol synthase, G8O: geraniol-8-oxidase, 8-HGO: 8-hydroxygeraniol oxidoreductase, IS: iridoid synthase, IO: iridoid oxidase, 7DLGT: 7-deoxyloganetic acid glucosyltransferase, 7DLH: 7-deoxyloganic acid hydroxylase, LAMT: loganic acid methyltransferase, SLS secologanin synthase, TDC: tryptophan decarboxylase, STR: strictosidine synthase. Frames indicate mRNA localization in the leaf internal phloem-associated parenchyma (IPAP) (pink) or epidermis (blue). Numbers indicate predicted enzyme classes in the initial gene discovery strategy. 1: oxidoreductase, 2: cytochrome P450, 3:UDP-glucose-glucosyltransferase, 4: terpene synthase.

This confirmed the functionality of all enzymes of the pathway and highlights their utility for synthetic biology programs towards sustainable biotechnological production of valuable (seco)iridoids and alkaloids with pharmaceutical and agricultural applications. Notably, although different segments of the strictosidine pathway are localized in different cell types in *C. roseus*, our results show that the entire pathway can be successfully reconstituted in a single *N. benthamiana* organ. This paves the way for the biotechnological production of valuable iridoids and iridoid-derived compounds, such as the MIAs vincristine and vinblastine, making these important anti-cancer drugs available to more people and at a lower price.

As a conclusion this thesis presents the complete set of iridoid biosynthesis genes and enzymes. The tissue specific expression of the genes confirms the model proposing that the early part of iridoid synthesis occurs in the IPAP cells associated with vascular tissue and the late part in the epidermis. The nearly very similar expression patterns of all the early pathway genes reinforce this model and suggest regulation by common transcription factors. The elucidation of the pathway validates the approach taken in this work of coexpression analysis and tissuespecific proteomics for finding specific biosynthesis genes. The reconstitution of the whole pathway in *N. benthamiana* acts as proof of concept for expressing whole biosynthetic pathways in plants. Each enzyme was found to fulfill its hypothesized function and to be required to complete the pathway. The metabolite analysis from stably transformed tobacco with the GESs and *N. benthamiana* transiently transformed with GESs or the whole pathway and the first MIA biosynthesis genes can help to identify and solve bottlenecks in future metabolic engineering efforts. As formation of the monoterpenoid

moiety of MIAs (MEP pathway and iridoid pathway) is considered to be the rate limiting step in MIA biosynthesis (Morgan and Shanks, 1999), the availability of pathway genes and additional information on their characteristics can bring about new more elaborate metabolic engineering schemes for producing high amounts MIAs and iridoids in heterologous hosts.

## References

- Geu-Flores F, Sherden NH, Courdavault V, Burlat V, Glenn WS, Wu C, Nims E, Cui Y and O'Connor SE** (2012) An alternative route to cyclic terpenes by reductive cyclization in iridoid biosynthesis. *Nature* **492**: 138-142
- Iijima Y, Gang DR, Fridman E, Lewinsohn E, Pichersky E** (2004) Characterization of geraniol synthase from the glands of sweet basil. *Plant Physiol* **134**: 370-379
- Ito M, Honda G** (2007) Geraniol synthases from *perilla* and their taxonomical significance. *Phytochemistry* **68**: 446-453
- Morgan JA, Shanks JV** (2000) Determination of metabolic rate-limitations by precursor feeding in *Catharanthus roseus* hairy root cultures. *J Biotechnol* **79**: 137-145
- Yang T, Li J, Wang HX, Zeng Y** (2005) A geraniol-synthase gene from *Cinnamomum tenuipilum*. *Phytochemistry* **66**: 285-293
- van der Heijden R, Jacobs DI, Snoeijer W, Hallard D, Verpoorte R** (2004) The *Catharanthus* alkaloids: Pharmacognosy and biotechnology. *Curr Med Chem* **11**: 607-628

Samenvatting



*Catharanthus roseus* (roze maagdenpalm) is de best bestudeerde medicinale plant. Het produceert de monoterpenoïde indoolalkaloïden (MIAs) en hun voorlopers de (seco)iridoïden. Deze secundaire metabolieten bezitten een breed spectrum aan farmacologische en insectenwerende activiteiten. Van de meer dan 130 bekende MIAs uit *C. roseus* hebben velen farmaceutische toepassingen, zoals serpentine en ajmalicine tegen hoge bloeddruk en de bisindoolalkaloïden vincristine en vinblastine die op grote schaal worden gebruikt om verschillende soorten kanker te behandelen, zoals leukemie, de ziekte van Hodgkin, Kaposi-saroom, borstkanker, blaaskanker en testiculaire kanker. Vinblastine en vincristine worden door *C. roseus* geproduceerd in extreem kleine hoeveelheden, wat leidt tot hoge marktprijzen en slechte beschikbaarheid. Vanwege hun complexe structuren is chemische totaalsynthese onhaalbaar (van der Heijden et al., 2004). Ondanks tientallen jaren van onderzoek zijn slechts delen van de MIA biosyntheseroute bekend. Volledige kennis van de biosyntheseroute is essentieel voor biotechnologische productie van MIAs en (seco)iridoïden.

Het doel van dit werk was opheldering en nadere omschrijving van de iridoïdroute in *C. roseus*, en met name identificatie van de enzymen die zowel de veronderstelde als de onbekende intermediaire reacties katalyseren. Kandidaatgenen voor deze enzymen werden gekozen op basis van een hypothetisch model van de biosyntheseroute waarin bepaalde enzymactiviteiten verondersteld werden, en vervolgens werden DNA sequenties coderend voor deze veronderstelde enzymen gekozen uit het *C. roseus* transcriptoom op basis van overeenkomsten in de aminozuurvolgorde met bekende vergelijkbare enzymen van andere planten. De selectie werd verfijnd door het verzamelen van toekomstige informatie zoals genexpressiepatronen onder verschillende omstandigheden en de lokalisatie van gentranscripten in verschillende plantenweefsels. De wetenschappelijke stand van zaken bij het begin van deze studie gaf aan dat er 4-6 enzymen met compleet nieuwe functies ontbraken in de iridoïdroute. Het gen dat codeert voor het eerste enzym in de route, geraniol synthase (GES), was bekend van basilicum (*Ocimum basilicum*; Iijima et al., 2004), *Cinnamomum tenuipilum* (Yang et al., 2005) en *Perilla citriodora* (Ito en Honda, 2007), maar niet

van *C. roseus*. Meerdere oxidoreductasen van verschillende klassen, cytochroom P450s, glucosyltransferases en terpeensynthases werden als kandidaten gekozen volgens het model voor de biosynthetische route (Afbeelding 1). Deze kandidaten werden tot expressie gebracht in de darmbacterie *Escherichia coli* of in de bakkersgist *Saccharomyces cerevisiae* en enzymactiviteiten werden vervolgens getest met *in vitro* biochemische assays. De route werd volledig opgelost door tijdelijke expressie van de vier nieuwe kandidaten samen met de bekende iridoïdbiosynthese genen in *Nicotiana benthamiana* en analyse van de nieuwe metabolieten met vloeistofchromatografie-massaspectrometrie (LC-MS). De identiteit van het GES enzym uit *C. roseus* werd bevestigd door expressie in *E. coli* en in een *S. cerevisiae* stam met een mutatie in farnesylbiosynthese gevolgd door biochemische analyse van de producten, door meting van de weefsel-specifieke genexpressie in *C. roseus* via *in situ* hybridisatie en door tijdelijke expressie van een fusie-eiwit met Yellow Fluorescent Protein (YFP) in *C. roseus* cellen. Homologe geraniol synthases van valeriaan (*Valeriana officinalis*; VoGES) en Azteeks zoetkruid (*Lippia dulcis*; LdGES) werden gekarakteriseerd via *in vitro* assays en tijdelijke *in planta* expressie.

**Hoofdstuk 1** geeft een algemeen overzicht van de MIA en de iridoïd biosynthese en beschrijft de onderzoeksstrategie voor dit werk.

**Hoofdstuk 2** beschijft de klonering en functionele karakterisering van geraniol synthase uit *C. roseus* (CrGES). CrGES werd tot expressie gebracht in *E. coli*. Het gezuiverde recombinanteiwit katalyseerde de *in vitro* omzetting van geranyl difosfaat (GPP) in geraniol met een  $K_M$  waarde van 58.5  $\mu\text{M}$  voor GPP. *In vivo* CrGES activiteit werd geëvalueerd door heterologe expressie in de *S. cerevisiae* stam K197G die gemuteerd is in het farnesyl difosfaatsynthase gen. Analyse van cultuurextracten door gaschromatografie-MS (GC-MS) bevestigde de uitscheiding van geraniol in het groeimedium. De tijdelijke expressie in *C. roseus* cellen van een fusie-eiwit met Yellow Fluorescent Protein (YFP) wees uit dat CrGES gelokaliseerd is in het stroma en de stromules van de plastiden. Dit vormt een aanwijzing dat de beschikbaarheid van het product geraniol voor het volgende enzym G10H wordt vergemakkelijkt door interacties tussen de stromules en het endoplasmatisch reticulum (ER) waar G10H zich bevindt. In bovengrondse plantenorganen toonde RNA in

situ hybridisatie aan dat CrGES transcripten voorkomen in de parenchymcellen die zich dicht bij de interne floëmvaten bevinden, net als de transcripten van andere bekende genen betrokken bij de vroege stappen in de MIA biosynthese. Behandeling van *C. roseus* cellen met de signaalstof methyljasmonaat leidde tot een toename van de hoeveelheid CrGES transcripten, net als eerder werd gerapporteerd voor de transcripten van andere genen betrokken bij de vroege stappen van de MIA biosynthese. Gezamenlijk ondersteunen deze waarnemingen het idee dat CrGES aan het begin staat van de monoterpenoïde tak van de MIA biosynthese.

**Hoofdstuk 3** beschrijft de karakterisering van twee geraniol synthases van *Valeriana officinalis* (VoGES) en *Lippia dulcis* (LdGES). De enzymen werden geproduceerd in *E. coli* en waren in staat om GPP om te zetten in geraniol met  $K_M$  waarden van respectievelijk 32  $\mu\text{M}$  en 51  $\mu\text{M}$  voor GPP. De *in planta* enzymatische activiteit en de subcellulaire lokalisatie van VoGES en LdGES werden vastgesteld in respectievelijk stabiel getransformeerde tabaks- en tijdelijk getransformeerde *N. benthamiana* planten. Transgene tabaksplanten met VoGES of LdGES accumuleerden geraniol, geoxideerde geraniolverbindingen zoals geranial en geranisch zuur en hexose-conjugaten van deze verbindingen. Geraniol-emissie van bladeren was lager dan die van bloemen, wat correleerde met hogere niveaus van concurrerende geraniol-conjugerende activiteit in bladeren. Analyse van de lokalisatie van GFP-fusies van de twee GES eiwitten toonde aan dat VoGES zoals verwacht overwegend in plastiden voorkomt, terwijl LdGES ook in het cytosol voorkwam. Ondanks deze verschillen in lokalisatie was geraniolproductie door *N. benthamiana* planten die VoGES of LdGES tot expressie brachten vrijwel hetzelfde. Expressie van verkorte versies van VoGES of LdGES met lokalisatie in het cytosol resulteerde in de accumulatie van 30% minder geraniolglycosiden dan met de VoGES en LdGES varianten die in de plastiden voorkomen, wat suggereert dat het substraat geranyldifosfaat zowel beschikbaar is in de plastiden als in het cytosol. De potentiële toepassing van GES om de MIA biosyntheseroute in heterologe gastheren tot expressie te brengen wordt bediscussieerd.

**Hoofdstuk 4** beschrijft de ontdekking van de laatste vijf ontbrekende stappen van de (seco)iridoïdbiosyntheseroute van *C. roseus* (Afbeelding 1) met behulp van een geïntegreerde

benadering, inclusief analyse van genexpressie en co-expressie, weefsel-specifieke proteomics en biochemische assays. Vijf nieuwe biosynthesegenen werden gekloneerd en de bijbehorende enzymen gekarakteriseerd. Eén van de enzymen, een alternatief terpeencyclase dat iridoid synthase (IS) is genoemd, werd recent gerapporteerd in het tijdschrift Nature (Geu-Flores et al., 2012). Het nieuwe enzym 8-hydroxygeraniol oxidoreductase (8-HGO) katalyseert de oxidatie van 8-hydroxygeraniol zowel op positie 1 als 8 tot het uiteindelijk eindproduct 8-oxogeraniol. Een ander nieuw enzym, iridoid oxidase (IO, CYP76A26), bleek *cis-trans*-nepetalactol (ook wel bekend als de hemiacetaalvorm van iridodial) en *cis-trans*-iridodial om te zetten in 7-deoxyloganetisch zuur. Ook iridotrial werd omgezet door het enzym in hetzelfde product. 7-deoxyloganetisch zuur werd op zijn beurt omgezet door 7-deoxyloganetisch zuur glucosyltransferase (7DLGT, UGT709C2) tot 7-deoxyloganaat. Deze verbinding werd gehydroxyleerd door een vierde enzym, 7-deoxyloganaat hydroxylase (7DLH, CYP72A224) tot loganaat. Transcripten voor alle enzymen werden gedetecteerd in parenchymcellen rond het floëemvaatweefsel in bovengrondse *C. roseus* organen door *in situ* hybridisatie. Onder verschillende omstandigheden vertoonden de nieuwe biosynthesegenen hetzelfde expressiepatroon in planten en celculturen. De subcellulaire lokalisatie van GFP-fusies van de CYP450 enzymen (IO, 7DLH) was zoals verwacht in het ER. GFP fusie-eiwitten met 8-HGO en 7DLGT bevonden zich in het cytosol. Expressie van de acht genen die coderen voor de secoiridoïdroute samen met twee genen die precursorvorming stimuleren en met twee genen voor biosynthese van alkaloiden in een heterologe waardplant leidde tot de productie van het complexe MIA strictosidine. Dit bevestigde de werking van alle enzymen uit de route en geeft een indicatie voor hun toepasbaarheid in synthetische biologieprogramma's voor duurzame biotechnologische productie van waardevolle (seco)iridoïden en alkaloiden met farmaceutische en agrarische toepassingen. Hoewel verschillende delen van de strictosidine biosyntheseroute gelokaliseerd zijn in verschillende celtypen in *C. roseus*, blijkt uit onze resultaten dat de hele route met succes tot expressie kan worden gebracht in een enkel *N. benthamiana* orgaan. Dit effent de weg voor de biotechnologische productie van waardevolle iridoïden en iridoïd-afgeleide verbindingen, zoals de MIAs vincristine en vinblastine, hetgeen deze belangrijke anti-kanker medicijnen beschikbaar zou maken voor meer mensen en tegen een lagere prijs.

**Afbeelding 1 (pagina 148).** De secoiridoïdroute in *C. roseus*. Reacties waarvoor het overeenkomstige *C. roseus* gen in de literatuur is beschreven hebben een zwarte achtergrond, reacties waarvoor het overeenkomstige gen is beschreven in dit proefschrift hebben een gele achtergrond. IPP: isopentenyl-difosfaat, DMAPP: dimethylallyl-difosfaat, GPPS: geranyl-difosfaat synthase, GES: geraniol synthase, G8O: geraniol-8-oxidase, 8-HGO: 8-hydroxygeraniol oxidoreductase, IS: iridoïd synthase, IO: iridoïd oxidase, 7DLGT: 7-deoxyloganetisch zuur glucosyltransferase, 7DLH: 7-deoxyloganaat hydroxylase, LAMT: loganaat methyltransferase, SLS: secologanine synthase, TDC: tryptofaan decarboxylase, STR: strictosidine synthase. De kleur van het kader geeft de mRNA lokalisatie aan in de parenchymcellen rond de floëemvaatcellen (IPAP cellen) (roze) of de epidermiscellen (blauw) in het blad. Nummers geven voorspelde enzymklassen aan in de oorspronkelijke onderzoeksstrategie. 1: oxidoreductase, 2: cytochroom P450, 3: UDP-glucose-glucosyltransferase, 4: terpeen synthase

Kortom, dit proefschrift beschrijft de complete set van iridoïdbiosynthesegenen en enzymen. De weefsel-specifieke expressie van de genen bevestigt het model dat voorstelt dat het begin van iridoïdsynthese plaatsvindt in de parenchymcellen rond het floëemvaatweefsel en het laatste gedeelte in de epidermiscellen. De identieke expressiepatronen van de biosynthesegenen suggereert regulatie door gemeenschappelijke transcriptiefactoren. De opheldering van de biosyntheseroute bevestigt de juistheid van de onderzoeksstrategie voor het vinden van de specifieke biosynthesegenen die gebaseerd was op co-expressie analyse en weefsel-specifieke proteomics. De reconstructie van de hele route in *N. benthamiana* vormt een proof of concept voor het tot expressie brengen van hele biosynthesewegen in planten. Elk enzym bleek de hypothetische functie inderdaad te bezitten en was nodig voor de complete biosyntheseroute. Toekomstige analyse van metabolieten in tabaksplanten of *N. benthamiana* planten getransformeerd met delen van of met de gehele route kan helpen om knelpunten in de biosyntheseroute te identificeren en op te lossen. De biosynthese van het monoterpenoïde deel van de MIAs via de MEP route en de iridoïdroute wordt beschouwd als de snelheidsbepalende stap in de MIA biosynthese (Morgan en Shanks, 1999). De beschikbaarheid van alle biosynthesegenen en de aanvullende informatie over de eigenschappen van de bijbehorende enzymen vormen de basis voor geavanceerde “metabolic engineering” strategieën voor de productie van MIAs en iridoïden in heterologe gastheren.

---

**Referenties**

- Geu-Flores F, Sherden NH, Courdavault V, Burlat V, Glenn WS, Wu C, Nims E, Cui Y and O'Connor SE** (2012) An alternative route to cyclic terpenes by reductive cyclization in iridoid biosynthesis. *Nature* **492**: 138-142
- Iijima Y, Gang DR, Fridman E, Lewinsohn E, Pichersky E** (2004) Characterization of geraniol synthase from the glands of sweet basil. *Plant Physiol* **134**: 370-379
- Ito M, Honda G** (2007) Geraniol synthases from *perilla* and their taxonomical significance. *Phytochemistry* **68**: 446-453
- Morgan JA, Shanks JV** (2000) Determination of metabolic rate-limitations by precursor feeding in *Catharanthus roseus* hairy root cultures. *J Biotechnol* **79**: 137-145
- Yang T, Li J, Wang HX, Zeng Y** (2005) A geraniol-synthase gene from *Cinnamomum tenuipilum*. *Phytochemistry* **66**: 285-293
- van der Heijden R, Jacobs DI, Snoeijer W, Hallard D, Verpoorte R** (2004) The *Catharanthus* alkaloids: Pharmacognosy and biotechnology. *Curr Med Chem* **11**: 607-628

## Acknowledgements

The Smart-Cell project has been special with respect to the particularly high amount of collaboration and material exchange. Without it, it would not have been possible to obtain the results achieved.

First I would like to thank all Smart-Cell partners for intense day-to-day collaboration and inspiration.

I would also like to thank collaborators from outside the Smart-Cell consortium: Marc Clastre and Andrew Simkin from Université François-Rabelais (France) for the *C. roseus* geraniol synthase collaboration, Vincent Burlat from Université de Toulouse (France) for *in situ* hybridization in the *C. roseus* geraniol synthase and iridoid pathway collaborations, Søren Rosendal Jensen from Denmark's Technical University (Denmark) for advice concerning iridoid synthesis and Wolfgang Kreis, Jenny Munkert and Jan Pietersen at the University of Erlangen-Nuremberg (Germany) for the progesterone reductase collaboration.

I would like to thank Young Hae Choi and Justin Fishedick from the Natural Product Lab at Leiden University for help with chemistry and Ward de Winter from the Institute of Biology in Leiden for the tremendous work involved in maintaining our cell cultures.

Finally I would also like to acknowledge present and past members of the Plant Cell Physiology group of the IBL for support and help in everyday lab work and in general the people of the IBL for support and company.

## Curriculum Vitae

Karel Miettinen was born on the 16<sup>th</sup> of June 1982 in Lappeenranta, Finland. He attended high school at the Lappeenrannan Lyseon Lukio. He got acquainted with biology during 2001-2002 in the program "Basics of Genetics" at the Open University of the University of Helsinki. In 2003 he was accepted in the biotechnology BSc/MSc program HEBIOT at the Faculty of Agriculture and Forestry at the University of Helsinki. He conducted his bachelors project dealing with purification of plant produced heterologous microbial epitopes from tobacco and his masters project about purification and description of a  $\beta$ -glucosidase enzyme from *Gerbera hybrida* at the Department of Applied Biology. He obtained his MSc degree in Biotechnology in 2010. In 2009 he started his PhD studies at the department of Plant Cell Physiology at Leiden University with funding from the EU seventh framework project SmartCell under the supervision of Prof. Dr. Johan Memelink.