Gene Discovery and *In Vivo* Characterization of the Iridoid Biosynthesis Pathway in *Nepeta*

Lira Palmer

This thesis is submitted in fulfilment of the requirements of the degree of Doctor of Philosophy at the University of East Anglia

Department of Biochemistry and Metabolism

John Innes Centre

Department of Natural Product Biosynthesis

Max Planck Institute for Chemical Ecology

Norwich & Jena

December 2021

Abstract

This thesis presents the characterization of the *Nepeta* iridoid biosynthetic pathway via gene discovery and the development of a functional genomics tool for *in planta* genetic characterization. The Lamiaceae plant family, colloquially known as the mint family, is well known for its chemical diversity and economical importance, especially amongst members of the Nepetoideae sub-family. Most members of this sub-family are well known for their diverse terpene-based natural products; however, one genus, *Nepeta*, is unique amongst the Nepetoideae for its ability to produce nepetalactone, an iridoid-scaffold compound known for its psychoactive effect on cats and potential use as a bio-based pest control in agriculture due to its influence on various insect species. Chemical profiling on *Nepeta spp*. and within varieties of a single species have revealed the production of different nepetalactone stereoisomers varies widely across plants. Previous work has identified *Nepeta spp*. biosynthetic enzymes that can synthesize different stereoisomers of nepetalactones *in vitro*. Work in this thesis also presents the gene discovery and biochemical characterization of the early steps of this pathway (Chapter 2). The role these biosynthetic genes play *in planta* leading to the synthesis of different stereochemical ratios of nepetalactone has not been addressed. This thesis presents the development of a virus-induced gene silencing (VIGS) tool for *N. cataria* (Chapter 3) to explore the *in vivo* function of the putative biosynthetic genes (Chapter 4). Simultaneously targeting a visual marker gene, magnesium chelatase subunit H (*ChlH*), and the genes involved in the production of the various nepetalactone stereoisomers, allows the precise selection of the tissue under the knockdown phenotype of VIGS and characterisation of this pathway *in vivo*. Furthermore, VIGS provides the possibility to untangle the mechanisms behind isomer regulation and gene expression in nepetalactone production, as well as to understand the effect of this pathway on other physiological processes.

Acknowledgements

My PhD became an unexpected journey which changed who I am. I would like to thank the John Innes Centre for granting me the opportunity to take upon this journey, and Sarah O'Connor for your mentorship. I am grateful for your enthusiasm, patience, and advice throughout. Your love for science, chemistry and plants is infectious and great motivation when motivation is hard to find. Thank you for challenging me to become a scientist. I would also like to thank Anne Osbourn for your continual support from Norwich. I always enjoyed our meetings and the advice you had to give. Both of you are great role models and I am honoured to have worked with you.

I would also like to thank the entirety of the SOC department for each members' time and support. I would especially like to thank Benjamin Lichman, Carlos Rodriguez Lopez, Omar Kamileen, Prashant Sonawane, Nestor Hernandez, Kotaro Yamamoto, Maite Colinas and Maritta Kunert for their hard work and great advice on this project. Thank you for letting me constantly pick your brains and your words of encouragement. Additionally, I would like to thank Ling Chuang and Marlen Siegmund for your assistance on this and other projects, and especially for your great company in the lab. I would like to additionally thank Lorenzo Caputi and Delia Serna for their advice, great humour and understanding during especially difficult and much less difficult times of my PhD.

I am grateful to my fellow PhD students Marianna Boccia, Matilde Florean, Dagny Grzech, Omar Kamileen, and Chloe Langley for your camaraderie, your patient ears to listen to my rants, and your shoulders for me to cry on. I especially would like to thank Chloe and Dagny for the long drive to Frankfurt, and their subsequent care and support. I have learned a lot from all of you.

Although I left the JIC early on, I formed special bonds with many people. I am especially grateful to Marietta Avramova, Isabel Diez-Santos and Raf Zdrzalek, AKA "the B Squat", for their friendship and the genuine support and affection we all share. I thank my rotation PhD cohort Neftaly Cruz Ramirez, Marco D'Ario, Sam Deans and Sebastian Samwald for embarking on this journey together. I am proud to have grown alongside you. I would also like to thank the wonderful groups at the JIC, including the student voice, a small book club, the Spanish and Mexican chats and the Rec Centre volunteers. At the Max Planck, I am

grateful to the friends I found and would like to especially thank Amy Eacock, Cee Hayworth, Paola Rubiano Buitrago and Na Ra Shin for embracing and teaching this budding knitter.

To all my friends outside of the JIC and MPI, thank you for believing in me, offering your support and for creating wonderful memories despite the difficult times. A big special thanks to Inas Wafiya and Hana Subhan for your deep understanding of the challenges I have faced, personally and academically. Thank you to Mei Shen, Shaothing Teoh, Sze Ping, Tomas Lastrilla and Tue Nguyen for pandemic weekend hangouts across the world.

I would like to thank my family who always cheered me on and have never hesitated to tell me how proud you are of me. I miss those who have departed, and those who are still here, yet far. I especially would like to thank my grandfather Stan for keeping up to date on my research and for providing the best example of what life at 90 can be. I am sorry you missed the final stretch of this PhD. I would also like to especially thank my aunts Sandy and April, and my uncle John for continually sending me warm, funny, and encouraging messages throughout this PhD.

To Tom de Hoop, I would like to express how lucky I feel to have a partner who is the most patient, silly, and encouraging person I know. You make me laugh every day, even on the hardest days. Thank you for pushing me to work hard when I had to and for steering me away from the lab when I needed a break. I cannot ever thank you enough for all my tears and rants you have endured, nor for the many adventures you took me on in these 4 years, but I hope I can keep sparking in you the same laughter, sense of adventure and love that you have sparked in me. Thank you for being my best friend.

I would like to thank my mom, Denise, for being a source of inspiration my entire life. I know I share your energy, kindness, ambition, and loud laugh. I am eternally grateful to my dad, Randy, for being the best parent anyone could hope for. I am here today pursuing my dreams, thanks to your love, hard work, and honesty. You will always inspire me to chase my own path and stand up for myself and my values. Thank you for opening the world up to me.

Finally, this thesis is dedicated to my brother, Tristan. I am truly lucky to have been your favourite person. It is an injustice that my biggest fan cannot be here to see this achievement and be proud of your little sister. Your energy, generosity, utter lack of fear, and unabashed honest love will forever push me forward and are deeply missed every day. You are irreplaceable in all our lives.

Table of Contents

List of Figures

List of Tables

Abbreviations

Acetyl-CoA	Acetyl-coenzyme A
ACT	Actin
AU	Arbitrary units
bHLH	basic Helix-Loop-Helix
BIS	bHLH iridoid synthase
BLAST	Basic local alignment search tool
bp	Base-pairs
CaLCuV	Cabbage leaf curl virus
cDNA	complementary DNA
CDS	Coding sequence
ChlH	Magneium chelatase subunit H
COI1	Coronatine insensitive 1
CP	Coat protein
CPR	Cytochrome P450 reductase
Ct	cycle threshold
Del-ros	Delila and Rosea1 tomato lines
DMAPP	Dimethyl allyl diphosphate
DNA	deoxyribunucleic acid
dsRNA	double stranded RNA
DTT	Dithiothreitol
DXS	deoxyxylulose 5-phosphate synthase
EDTA	Ethylenediaminetetraacetic acid
ER	Endoplasmic reticulum
EtAc	Ethyl Acetate
EV	Empty vector, ie. Empty pTRV2
FLAG	Peptide sequence DYKDDDDK
FPKM	Fragments per kilo base per million mapped reads
FPP	Farnesyl pyrophosphate
G3P	Glyceraldehyde-3-phosphate
G8H	Geraniol-8-hydroxylase
GC-MS	Gas chromatography-mass spectrometry
GES	Geraniol synthase

GGPP	Geranylgeranyl pyrophosphate
GOI	Gene of interest
GPP	Geranyl pyrophosphate
GST	Glutathione-S-Transferase
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
Hex	Hexane
HGO	Hydroxygeraniol oxidase
IO	Iridoid oxidase
IPP	Isopentenyl diphosphate
IPTG	Isopropyl β-D-1-thiogalactopyranoside
ISY	Iridoid synthase
LAMT	Loganic acid O-methyltransferase
LC-MS	Liquid chromatography-mass spectrometry
LDS	Lithium dodecyl sulfate
LiAc	Lithium acetate
MCS	Multiple cloning site
MEP	2-*C*-methyl-D-erythritol 4-phosphate/1-deoxy-D-xylulose 5-phosphate
MIA	Monoterpene indole alkaloid
MLPL	Major latex protein like
MOPS	3-(N-morpholino)propanesulfonic acid
MP	Movement protein
MVA	Mevalonate
MVAP	Mevalonate phosphate
NAD^+/H	Nicotinamide adenine dinucleotide
$NADP^+$/H	Nicotinamide adenine dinucleotide phosphate
NCBI	National center for biotechnology information
NEPS	Nepetalactol-related short chain dehydrogenases
NF-Y	Fuclear factor Y
Ni-NTA	Nickel-charged affinity resin
NMT	16-methoxy-2,3-dihydro-3-hydroxytabersonine N-methyltransferase
P5βR	Progesterone 5 beta reductase
PCR	Polymerase chain reaction
PDS	Phytoene desaturase
ppm	Parts per million

PRISE	P5βR-like enzyme
PTGS	Post transcriptional gene silencing
qPCR	quantitative PCR
QTOF	Quadrupole time of flight
RISC	RNA-induced silencing complex
RNA	Rinbnucleic acid
RNAi	RNA interference
RpRP	RNA-dependent RNA polymerases
Rz	Self-cleaving Ribozyme
SDR	Short chain reductase
SDS-PAGE	sodium dodecyl sulphate–polyacrylamide gel electrophoresis
siRNA	small interfering RNA
SLS	Secologanin synthase
SPE	Solid phase extraction
SSIV	Superscript IV
STR	Strictosidine synthase
TAIR	The arabidopsis information resource
tBLASTn	BLAST using translated nucleotides
TCP	Teosinte Branched 1, Cycloidea, and Proliferating4 Cell Factors
TIC	Total-ion chromatogram
TRV	Tobacco rattle virus
UBI	Ubiquitin
UGT	Glycosyl transferase
UTR	Untranslated region
VIGS	Virus-induced gene silencing
WT	Wild-type

Chapter 1

1.1 Introduction

1.1.1 Plants as chemical factories

Plants produce a vast array of chemicals essential for their survival. These chemical compounds are typically classified into primary metabolites and secondary metabolites based on their assigned function. The primary or centralized metabolism is classed as the reactions involved in creating and regulating essential chemicals for growth and basic survival, such as creating the chemical building blocks for cellular structures, proteins or energy storage. [1], [2]. Primary metabolism is often conserved across and beyond the plant kingdom. Variation is often maladaptive, as disruption of these essential metabolites can be extremely costly for an organism's survival. Secondary or specialized metabolism branches off from various primary metabolic pathways and produces compounds used to interact with the external world; for example, defensive chemicals against herbivores or fragrances and colours for signalling to pollinators and symbiotic partners. Variation in the secondary metabolism is often advantageous, and it is in the secondary metabolism that the variability of plant's chemistry can be found [1]–[5]. Humanity has exploited plant secondary metabolites throughout history, in art, textiles, religion and medicine [6].

1.1.2 Plant secondary metabolites

Plant secondary metabolites are broadly categorized into polyketides, phenylpropanoids, terpenes, alkaloids and peptides based on their chemical structures and biosynthetic origins [7], [8]. Plant polyketides and phenylpropanoids contain aromatic ring structures often derived from aromatic amino acid biosynthetic pathways such as the shikimate pathway [7], [8]; terpenes are assembled via the 5-carbon units isopentenyl diphosphate (IPP) and dimethyl allyl pyrophosphate (DMAPP) (fig. 1.1) [7]–[9]; alkaloids are a broad class of nitrogen-containing compounds that arise from different primary and secondary metabolic sources; and peptide secondary metabolites often arise from the ribosome [7], [8].

Amongst the terpenes (fig. 1.1), sub-classes based on the number of 5-carbon units used in their assembly are applied, commonly monoterpenes (C10), sesquiterpenes (C15), diterpenes (C20) and triterpenes (C30), etc. [9]. IPP and DMAPP are produced by the 2-C-methyl-D-erythritol-4-phosphate (MEP), and the mevalonate (MVA) pathways, operating in

the plastid and cytoplasm respectively. IPP and DMAPP are then condensed by trans-prenyltransferases, producing pyrophosphate chains of various lengths that are then used as substrates by a wide family of enzymes, the terpene synthases, that remove the phosphate group to produce the various types of terpene scaffolds [8]–[10] (fig. 1.1). A sub-type of monoterpenes (C10) are iridoid compounds [11], [12]. Iridoids are cyclopentane pyran monoterpenes that serve as lead compounds in the formation of more complex metabolites and can be found throughout the plant kingdom [11], [12].

The various sub-types of terpenes, as well as the broader range of secondary metabolites, can often interconnect in their biosynthesis, creating complex biochemicals emerging from a variety of secondary metabolic pathways [8]. Carbohydrates and saccharides often decorate these secondary metabolites, providing new functions [13]. The range of possibilities within secondary metabolism creates a wide array of evolutionary opportunities for diversification and specializations [1]–[5].

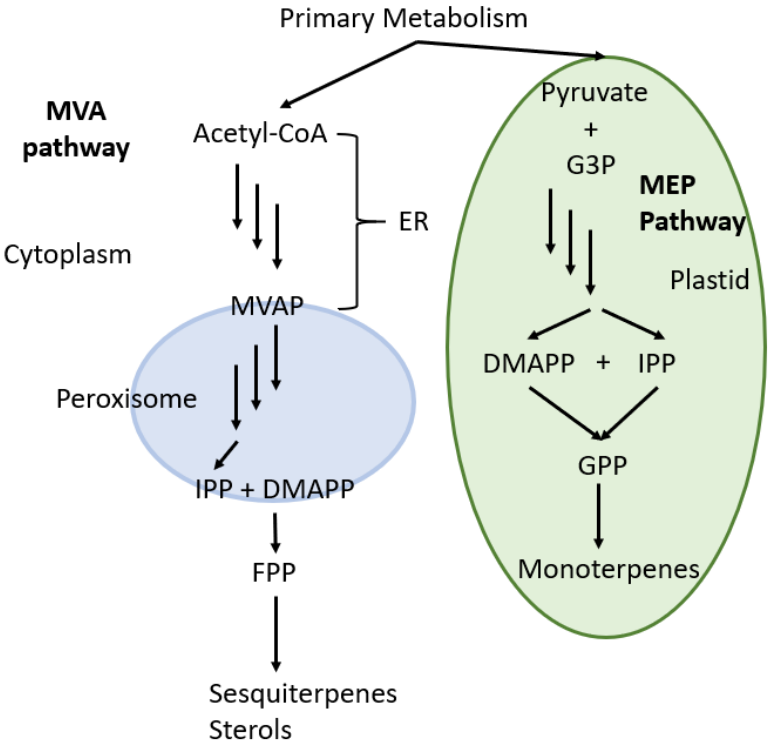


Figure 1.1: Terpene biosynthesis

Terpene biosynthetic pathways and their subcellular compartmentalization in plants: MVA and MEP pathway. The MVA pathway takes place mostly in the cytoplasm, with endoplasmic reticulum (ER) enzymes building carbon chains with acetyl-coenzyme A (acetyl-CoA) to form

mevalonate (MVA) phosphate (MVAP), which is transferred to the peroxisome and further phosphorylated and converted into isopentenyl pyrophosphate (IPP), which can exit the peroxisome into the cytoplasm. IPP and dimethyl allyl pyrophosphate (DMAPP) are five carbon building blocks for terpenes. In the MVA pathway, IPP and DMAPP condense to form the prenyl pyrophosphate intermediate farnesyl pyrophosphate (FPP), which serves as the basis for sesquiterpenes and sterols, or, via the formation of geranylgeranyl pyrophosphate (GGPP), diterpenes. The MEP pathway largely occurs in the plastid, starting with the condensation of pyruvate and glyceraldehyde-3-phosphate (G3P) and, via the intermediate 2-C-methyl-D-erythritol-4-phosphate (MEP), produce DMAPP and IPP. These are condensed into geranyl pyrophosphate (GPP) by GPP synthase. GPP is modified into monoterpenes, which can then become iridoids [14], [15].

1.1.3 The Lamiaceae plant family: specialized terpene producers

The Lamiaceae plant family, colloquially known as the mint family, is the sixth largest family within the Angiosperms [11], [16], [17]. This plant family consists of about 7000 species arranged in 253 genera [11], [16], [17]. This family includes many species of economic and cultural value. Many are culinary herbs, including rosemary (*Rosmarinus spp.),* basil (*Ocimum spp.*) and mint (*Mentha spp.*), others are valued for their fragrance and are often used in the cosmetic industry; for example, the lavender (*Lavandula spp.*) oil industry is expected to be worth 54 million USD by 2025 [18]. Other Lamiaceae are valued as ornamentals for their fragrance, appearance, and anti-pest activities in home gardens [19]–[21].

A study of the chemotaxonomy of the Lamiaceae family revealed the main types of secondary metabolites found in the Lamiaceae and their distribution throughout the family [11]. Of the 48 surveyed species across the family, the specialized metabolites reported were largely monoterpenes, iridoids and sesquiterpenes [11] (fig. 1.2). Monoterpenes and sesquiterpenes could be found in most clades of the family; however, iridoid distribution was mostly absent in one of the main sub-clades of the family, the Nepetoideae. An exception within the Nepetoideae sub-clade that does produce iridoids in abundance is the *Nepeta* genus. Nepetalactone is the main iridoid that is found in the *Nepeta* genus [11], [22]–[24], and is well-known for its psychoactive effect on house cats (*Felis catus*) and other members of the Felidae family [24], and a burgeoning interest into its effect on insects such as mosquitoes and aphids [25], [26].

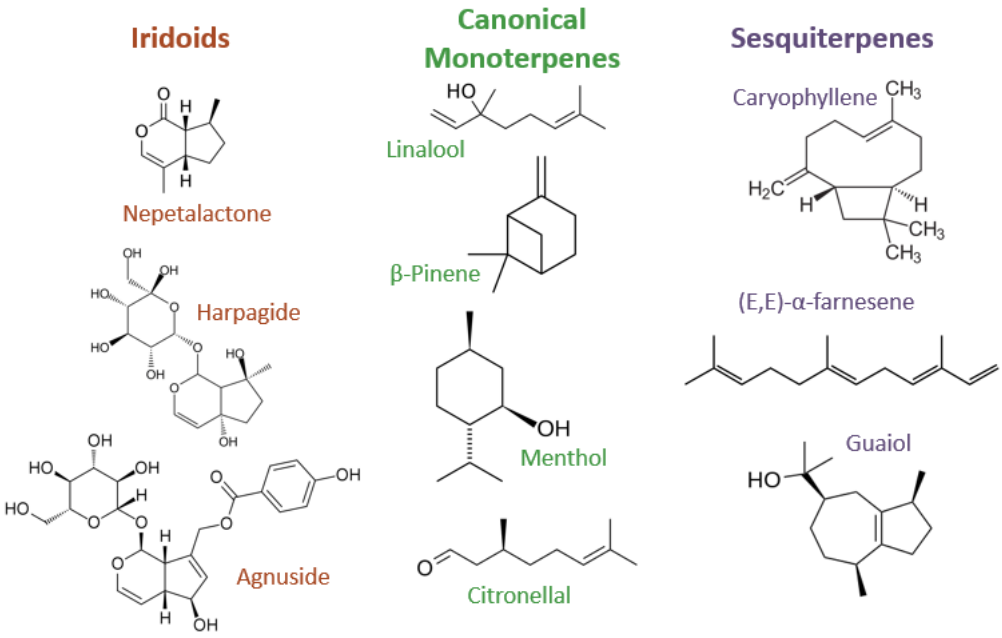


Figure 1.2: Example Lamiaceae terpenoids

Structures from Wikimedia commons.

1.1.4 Iridoid Ecology

Recent research has been published on the effect of iridoids on mosquitoes and the potential ecological role behind the characteristic effect of catnip on cats. In Melo *et al.* [25], the authors found that nepetalactone was an effective repellent against several kinds of insects, including mosquitos, drosophila, and ants. Further *in vitro* research on Drosophila S2 cultures revealed that nepetalactone triggered the TRPA1 receptor homologues of *Drosophila spp*. and *Aedes aegypti*. In behavioural studies, *Ae. aegypti* mosquitoes with an inactive TRPA1 were not repelled by a human arm rubbed with catnip, while the control mosquitoes were repelled.

Recently, Uenoyama *et al.* [27] reported the increased protection against mosquitoes cats have upon rubbing themselves on *Nepeta cataria* (catnip) or *Actinidia polygama* (silver vine), a common response from cats upon encountering these nepetalactol containing plants. They found that nepetalactol activated the μ-opioid system which promoted face rubbing on the plant. Cats were also exposed to mosquitos, and the researchers found that the cats who had carried out the rubbing behaviour had significantly less mosquito bites than the control cats.

As in the research from Uenoyama *et al.* [27] it has been found that other plants containing nepetalactol have a similar effect on cats. These include *Lonicera tartarica* and *Valeriana officinalis*. These plants all contain nepetalactone and other compounds such as actinidine which are thought to have a similar effect on cats as catnip. In a study from Bol *et al.* [28], the authors compared the effect of these nepetalactone containing plants on cat behaviour. They found that cats respond strongest to catnip, and curiously, from the cats that did not respond to catnip, a significant fraction of them responded with the characteristic "catnip" behaviour to the alternative plants presented.

Iridoids in plants are often deployed as anti-herbivory compounds, often in the form of iridoid-glucosides, against invertebrates [26]. Upon ingestion, the glucose moieties are cleaved off resulting in iridoid aglycones which behave as an alkylating agent, capable of denaturing amino acids, proteins, and nucleic acids [26]. Generalist insect herbivores therefore tend to avoid eating iridoid-glycoside containing plants. However, specialist insects have evolved the ability to sequester iridoid aglycones for use as anti-predation agents against predatory insects.

Curiously, iridoids are not only produced in the plant kingdom, but can also be endogenously produced in insects, and serve specific roles in the insect-producers life cycles. Several species of aphids, utilize female-produced iridoid bicyclic monoterpenoid nepetalactol and nepetalactone as sex-hormones, key to the species life-cycle and reproduction [29]. This information has been used in bio-based agricultural pest management to create male-aphid traps or lure aphid predators using these iridoids as chemical cues [29].

Research into iridoid ecology has revealed the complex roles volatile and glycoside iridoids play as plant-produced anti-herbivory compounds, as well as their role in insect defence and reproduction. Furthermore, the recent research into the ecological roles these compounds can play between insects, mammals and plants are introducing novel and complex ecological networks.

1.1.5 Iridoid biosynthesis

Although iridoids and their derivatives can be found across the plant kingdom [30]–[35], key enzymes for their biosynthesis have been reported from Catharanthus roseus, and key enzymes such as iridoid synthase and geraniol synthase have been found to conserved in species such as *Ocimum spp.* and *Antirrhinum majus* [12], [36], [37] (fig. 1.3). As explored in

section 1.1.2 and figure 1.1, the early building blocks for monoterpenes come from two pathways producing DMAPP and IPP, which are spatially separated in the cell [35], [38]. In the plastid, the MEP pathway produces IPP and DMAPP from pyruvate and glyceraldehyde-phosphate and are synthesized into the pre-terpenoid scaffolds GGPP or GPP for pre-terpenoid biosynthesis. Terpenoid synthases are responsible for using these molecules as substrates to produce various types of terpenes [8]–[10]. A sub-specialized terpene synthase, geraniol synthase (GES), synthesizes geraniol from GPP [12], [39]. Geraniol becomes the entry point for iridoid biosynthesis, which can then be further modified into the iridoid scaffold.

The following reaction step is carried out by geraniol-8-hydroxylase (G8H), a class II cytochrome P450 enzyme [12], [40]. This is a super family of hemeproteins which usually carry out oxidation reactions, requiring NADPH electron transfer to the heme cofactor through the NADPH-cytochrome P450 reductase (CPR) [38]. In *C. roseus*, three CPR homologues, under different regulation patterns, have been identified, with two of them playing an active role in iridoid biosynthesis [40]. Specifically, G8H hydroxylates the carbon C8 of geraniol to produce 8-hydroxygeraniol. For iridoids to be cyclized, 8-hydroxygeraniol oxidase (HGO) oxidizes both terminal hydroxyl groups and iridoid synthase (ISY) carries out the cyclization reaction using 8-oxogeranial to produce nepetalactol, which is further oxidized into iridotrial by iridoid oxidase (IO)/7-deoxyloganetic synthase, another cytochrome P450 enzyme [12], [41]–[43] (fig. 1.3-A). Modifications unto the main iridoid scaffold, such as the addition of glycosides, is carried out by a glucosyl transferase (UGT) [12], [43]. Iridoid biosynthesis is further expanded upon in Chapter 2 of this thesis.

1.1.6 Iridoids in *Nepeta*

Initial work into the Lamiaceae chemotaxonomy reported *Nepeta* as the only iridoid producing Nepetoideae and related the gene expression levels of transcripts similar to genes previously reported to be involved in iridoid biosynthesis, as described in section 1.1.5, with the presence/absence of iridoids [11] (fig. 1.3-B). The authors also reported a stereochemical difference in the iridoids produced in *Nepeta* compared to those produced by non-Nepeteoideae Lamiaceae plant species [11] (section 2.1.1, fig. 2.3). Furthermore, *Nepeta spp.* produced various isomers of nepetalactone, depending on the stereochemistry of the 4α and 7α carbons. Research that identified the *Nepeta* ISY revealed that *Nepeta* ISY does not cyclize 8-oxogeranial into the iridoid scaffold, instead it produces an enol intermediate,

8-oxocitronellyl enol [23] (fig. 1.3). This discovery sparked the search for an alternative enzyme or enzymes that would be responsible for cyclizing this enol intermediate to make the resulting nepetalactone isomers.

Isolating the proteome of *Nepeta* glandular trichomes revealed trichome enriched proteins in the iridoid pathway [44]. From this proteome, the enzymes responsible for producing an array of nepetalactone isomers in *Nepeta* were identified and characterized as a family of short-chain reductase (SDR) enzymes named nepetalactol-related SDR (NEPS) [44]. These enzymes carry out stereoselective cyclisation of ISY's enol intermediate into nepetalactol, and further oxidization into nepetalactone (fig. 1.3-B). The crystal structure of NEPS3 bound to nicotine adenine dinucleotide (NAD^+) showed a similar active site to known SDRs, although NAD^+ is not turned over, yet retained as a cofactor. *Nepeta* ISY and NEPS reveal an example of convergent evolution and neofunctionalization [23], [44].

While the chemotaxonomical work previously reported [11] offered tentative candidates for the early enzymes in the *Nepeta* iridoid biosynthesis pathway, in depth characterization into their enzymatic activities and relation to the later step enzymes had not been explored. Chapter 2 of this thesis describes the identification and characterization of these missing early steps.

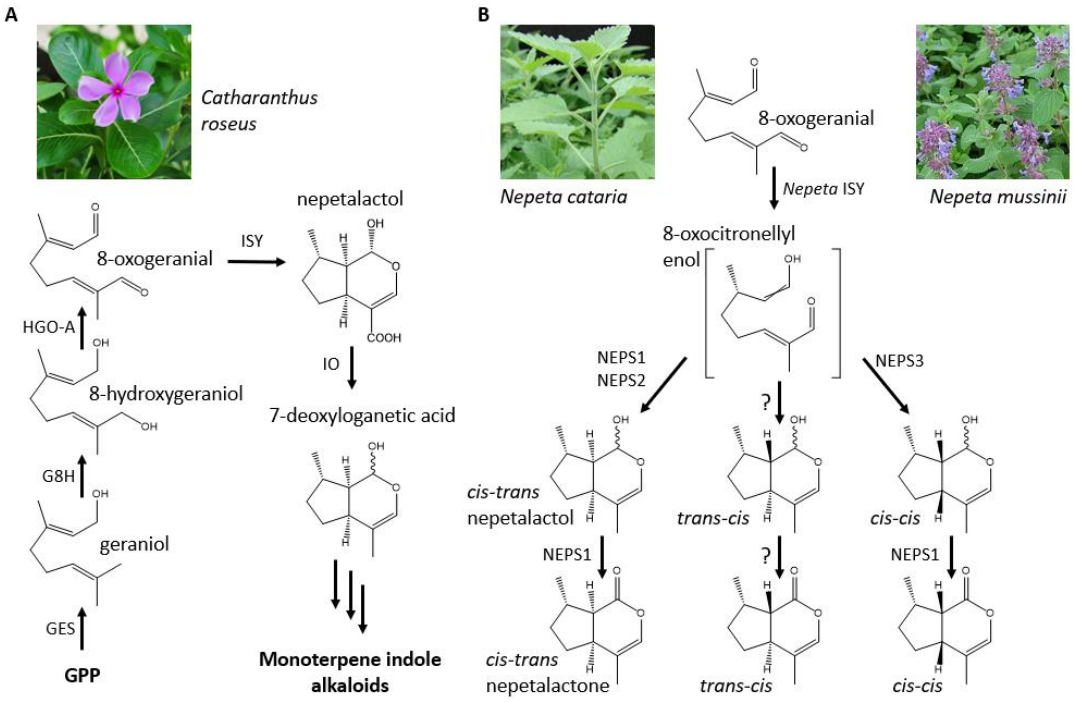


Figure 1.3: Iridoid biosynthesis

Schematic of iridoid biosynthesis as is known from A) *Catharanthus roseus* [12], [40], and what was known of iridoid biosynthesis in B) *Nepeta spp.* as of the start of this thesis [11], [23], [44]. Images from Wikimedia commons; structures were drawn in ChemDraw.

1.1.7 VIGS – Virus-induced gene silencing

Virus induced gene silencing (VIGS) is a transient and versatile post transcriptional genetic knock down technique that uses a plant's innate RNA mediated gene silencing defence system [45], [46]. Along with the viral sequences, inclusion of short homologous sequences to endogenous host genes lead to the formation of short interfering RNA (siRNA) transcripts that guide the RNA induced silencing complex (RISC) machinery to mRNA from the endogenous gene, which is then degraded, preventing translation, a process known as post-transcriptional gene silencing (PTGS) [44] (fig 1.4-A). It has proven to be a versatile tool in the study of specialized pathways, having been adapted to study the secondary metabolism of plants such as *Solanum lycopersicum* (tomato) [47]–[49] and *Catharanthus roseus* (Madagascar periwinkle) [50]–[53]. Amongst the Nepetoideae, VIGS has been previously adapted for use in *Ocimum basilicum* (sweet basil) [54] (fig. 1.4-B).

The success of VIGS in a new plant system often depends on the vector system used. A commonly used viral vector is the TRV vector system, adapted from the plus-strand RNA Tobacco Rattle virus with a bipartite genome [55]. This type of virus can be introduced into *Agrobacterium tumefaciens* which is then used to infect the plant, allowing the viral sequences to trigger the host's post-transcriptional silencing machinery. The TRV vector system is comprised of two vectors which work in tandem: pTRV1, which contains viral replication machinery, and movement protein sequences and pTRV2, which contains the sequences for virion formation and a multiple cloning site to introduce the host sequence to be targeted for silencing. The elements of these vectors are under the 35S promoter from the cauliflower mosaic virus for constitutive expression and are well suited for systematic transient silencing. This pTRV2 vector system has been adapted to a wide range of hosts [49], [50], [54], [56], [57], making it a good candidate for adapting to a new system. Information on VIGS mechanisms is further expanded upon in Chapter 3, section 3.1.

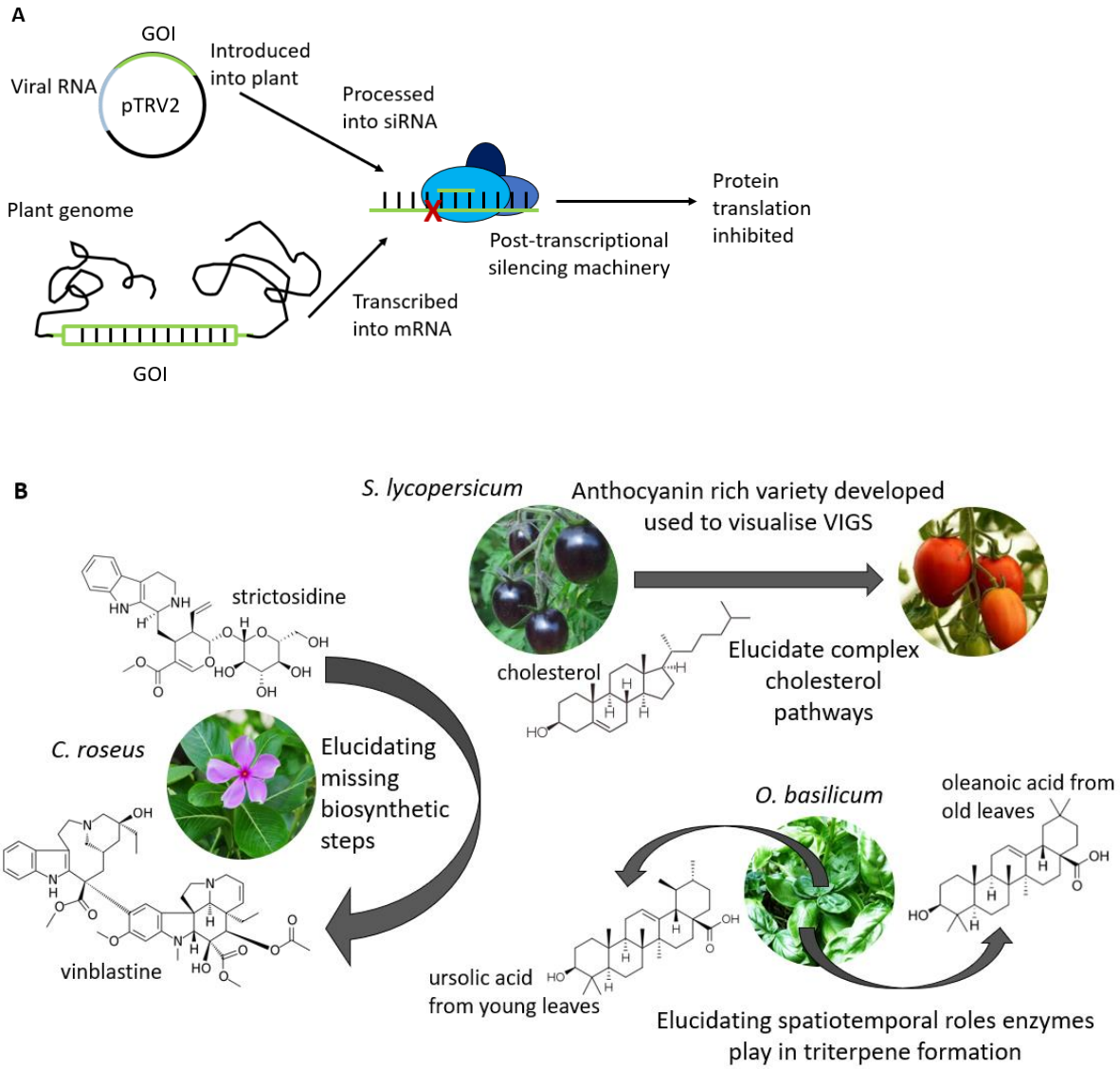


Figure 1.4: VIGS mechanism and examples

In A) a simplified schematic of the VIGS mechanism is shown. A more detailed mechanism schematic can be found in figure 3.1. The pTRV2 vector containing viral RNA (blue) and a fragment of a gene of interest (GOI) (green) is introduced into the plant and processed into the plant to produce small interfering RNA (siRNA). These are recognized by the post-transcriptional silencing machinery, which identifies the plant's genomic mRNA and prevents translation. B) shows example species (*C. roseus*, *S. lycopersicum* & *O. basilicum*) in which VIGS has been used to study secondary metabolic pathways. Images and structures from Wikimedia commons.

1.1.8 Transforming Lamiaceae

As explored in section 1.1.3, Lamiaceae are a large plant family of much economic and scientific interest. Research into the secondary metabolism of Lamiaceae has led to the development of various transformation methods to induce the over-expression of molecules of interest or for functional genomic characterization.

Increasing the production of desired metabolites is often achieved via hairy-root transformation (using *Agrobacterium rhizogenes*) or transformation via *Agrobacterium tumefaciens* co-cultured with explant calluses to introduce over-expression vectors of key metabolic genes. The use of *A. rhizogenes* as a transformation method is often used to create super producing metabolic systems [58]. This system infects the plant with *A. rhizogenes* which induces the over-production of roots (referred to as "hairy-roots") and introduces transfer DNA into the plant host. This is a useful system for harvesting target molecules that may be endogenous in the plant [58]. For example, in *Nepeta cataria*, this system has been used to harvest rosmarinic acid from the newly produced roots [59]. Besides hairy root transformation, glandular trichomes are of special interest in the overproduction of secondary metabolites [60]. Transformation methods targeting glandular trichomes in mints (*Mentha x piperita* and *Mentha spicata*) and Lavender (*Lavandula spp.*), either co-cultured explant calluses with *A. tumefaciens* to introduce over-expression vectors for certain metabolic genes, or introduced RNAi targeting inhibitory transcription factors, or a combination of both [60].

Other transformation methods in Lamiaceae have focused on knocking out genes to study their function. VIGS was applied to sweet basil (*O. baslicum*) to characterize the *in vivo* function of two CYP716A proteins involved in triterpene biosynthesis [54]. CRISPR/Cas9 was also adapted to sweet basil [61] to characterize the effect targeted mutagenesis of the *O. basilicum* homologue of Downy Mildew Resistant 1 gene, which conveys pathogen resistance upon knocking-out via the accumulation of homoserine. In sweet basil, lines with loss-of-function mutations had a dwarfed phenotype. CRISPR/Cas9 has also been used to characterize the *in vivo* function of key genes in the phenolic pathway of *Salvia miltiorrhiza* [62], [63].

1.1.9 *In vivo* characterization of iridoid pathways

Many studies into iridoid metabolism in plants use *in vitro* and *in silico* characterization, for example by testing enzymatic activities of heterologous expressed proteins. However, the iridoid metabolism has also been studied *in vivo*. One of the main systems in which it has been studied is in the monoterpene indole alkaloid (MIA) producer *C. roseus* (fig. 1.5). This plant, as well as other members of the Gentianaceae family, produces a wide variety of MIAs, which use iridoids as an initial scaffold for their biosynthesis [12]. MIAs are important medicinal compounds, used in cancer-chemotherapy treatments, and therefore has been a metabolic pathway of great interest (fig. 1.5).

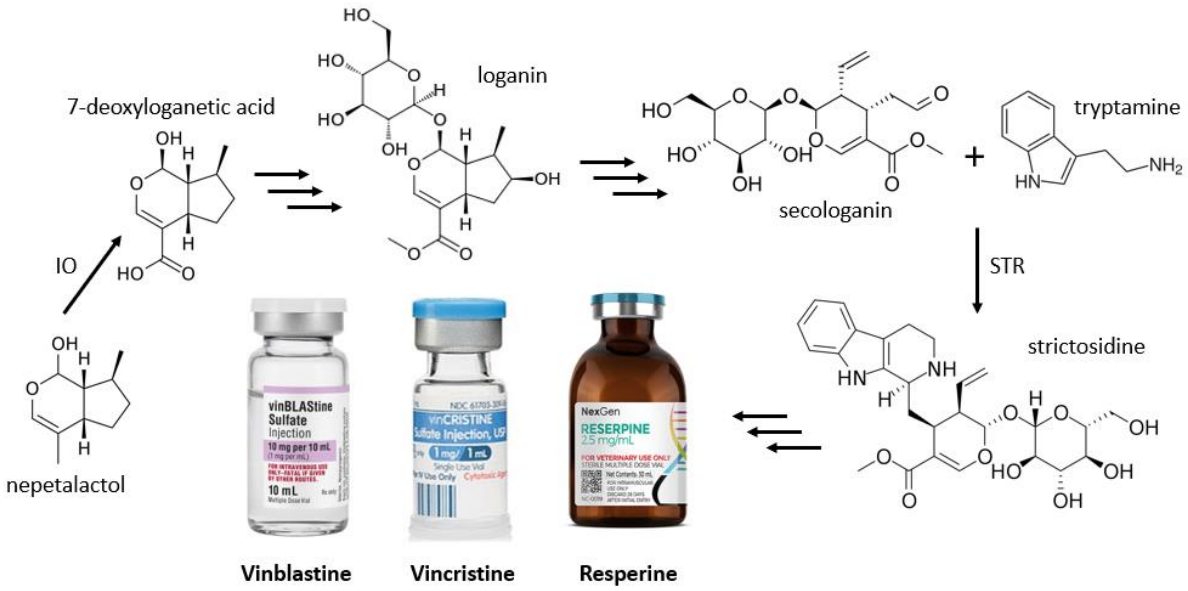


Figure 1.5: MIA biosynthesis from nepetalactol with example MIA medications

Iridoid oxidase (IO) oxidizes nepetalactol into 7-deoxyloganetic acid, which is glycosylated and further oxidized into loganin. Loganin is further modified into secologanin, which condenses with tryptamine in a picter-spengler condensation reaction carried out by strictosidine synthase (STR) with tryptamine into strictosidine. Strictosidine serves as the entry-point alkaloid for further MIA diversity. Vinblastine, vincristine and resperine are example MIA with medical applications. Images from Wikimedia commons.

As explored above, VIGS was developed in *C. roseus* to study the MIA metabolism. While plenty of work has been done to study the downstream alkaloid part of the pathway [51]–[53], VIGS in *C. roseus* has also been implemented to study the early iridoid steps of the pathway, including the discovery and *in vivo* characterization of CrISY [32], the cytochrome P450 enzyme 7-deoxyloganetic acid synthase [42] and the identification and characterization of three UGTs responsible for the modification of several iridoid intermediate steps in MIA biosynthesis [43].

1.2 Aims and Scopes of this Thesis

Identify the Missing *Nepeta* Iridoid Pathway Genes

While work was done to identify the ISY and cyclase enzymes in the iridoid pathway in *Nepeta* [11], [23], [64], the early enzymes leading up to 8-oxogeranial had not been discovered yet. Using homology-based gene discovery, I identified candidates for the missing early iridoid enzymes in *N. cataria* and *Nepeta mussinii*, and the non-iridoid Nepetoideae *Hyssopus officinalis*. I identified GES, G8H, HGO candidates in these plants, analysed their gene expression and carried out *in vitro* characterization of the enzymatic activity of the *Nepeta* candidates, confirming their proposed biochemical activity. This part of the thesis was published in Lichman *et al.* [65].

Develop a Transformation Technique for *Nepeta spp.*

Once the key enzymes for *Nepeta* iridoid biosynthesis had been fully characterized *in vitro* and *in silico*, I wanted to confirm the pathway *in vivo*. I developed a novel and reliable VIGS system in *N. cataria*. VIGS was an ideal functional genomics tool as it has a rapid turnover, it has been adapted to various non-model organisms and the tool-kit was readily available from previous work in the Prof. O'Connor's laboratory. I developed a tandem gene silencing system within the VIGS method to reliably study metabolic pathway genes in *Nepeta* cuttings. Work from this part of the thesis was published in Palmer *et al.* [66].

Characterize the Iridoid Metabolic Genes *in Planta*

After developing this transformation technique, I set out to characterize GES, ISY, the NEPS enzymes, a major latex protein like (MLPL) enzyme reported in Lichman *et al.* [65], and a regulatory candidate (RLK) found in the *Nepeta* gene cluster in *N. cataria.* I verified the *in vitro* activity of GES, ISY and MLP, and revealed the complexity of NEPS enzymatic activity on regulating the isomer production of the pathway. I explored the advantages and disadvantages of VIGS as a functional genomics tool for *N. cataria* secondary metabolism characterization.

Chapter 2: Early Steps of the *Nepeta* Iridoid Biosynthetic Pathway

2.1 Introduction

2.1.1 Metabolic profiles of Lamiaceae

As introduced in section 1.1.3, the Lamiaceae family is a prolific producer of chemical diversity. Amongst the range of metabolites produced in this plant family are iridoids, monoterpenes and sesquiterpenes (fig. 2.1) [11]. This large plant family is divided into several sub-families, each harbouring a wide range of specialised metabolites. The most economically important sub-family is the Nepetoideae, which contains plants commonly used for flavours and fragrances such as lavender (*Lavandula spp*.), mint (*Mentha spp*.) and Sages (*Salvia spp.*) [11], [16]. However, while members of this family are prolific mono- and sesquiterpene producers, notably they do not produce iridoid based compounds except for the *Nepeta* genus. Data produced by Doudareva and co-workers [11], [65] within the Mint Genome Project compared the metabolic profiles of various tissues from *Nepeta cataria*, *N. mussinii* and a non-iridoid Nepetoideae plant *Hyssopous officinalis* (fig. 2.2). These data indicate that *Nepeta spp.* overwhelmingly specializes in nepetalactone production in leaves and floral tissues, while *H. officinalis* produces a variety of monoterpene-based metabolites across its tissues, mostly 3-pinanone (fig. 2.2).

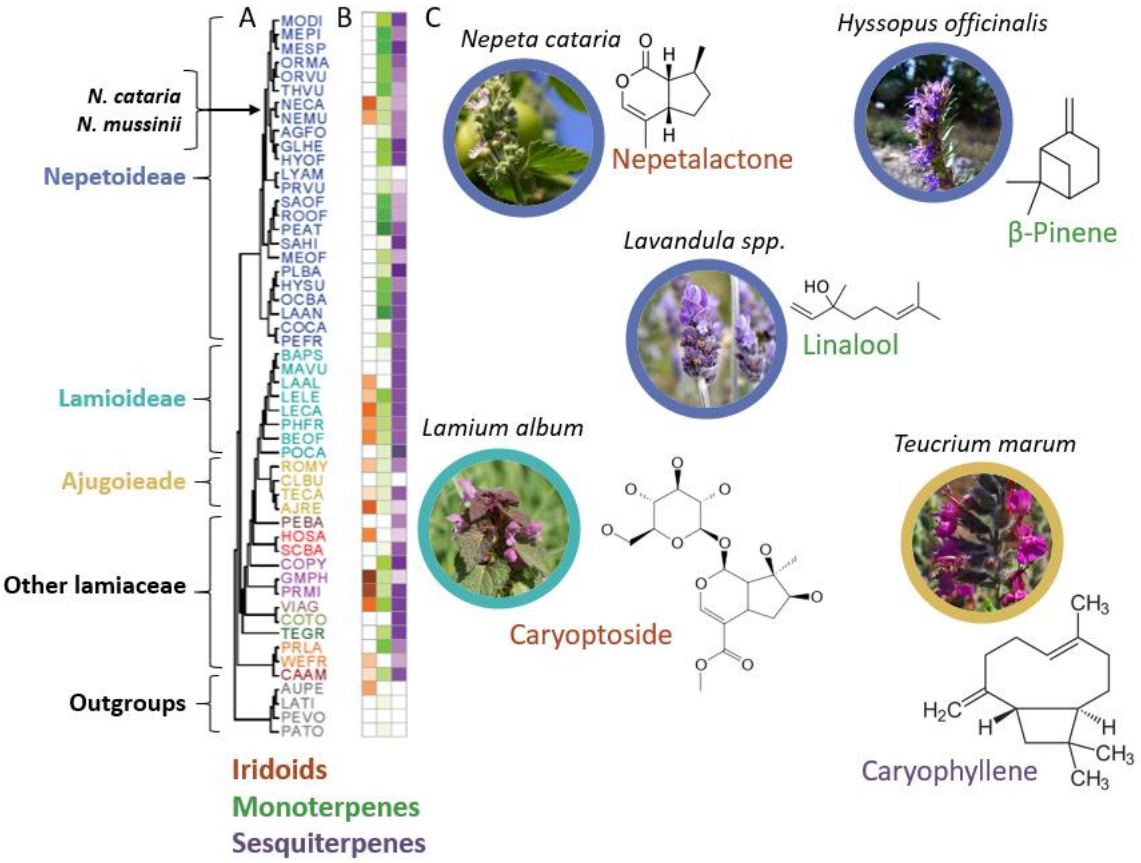


Figure 2.1: Lamiaceae chemotaxonomy

Mint family abbreviated chemotaxonomy adapted from [11]. A) is an abbreviate Lamiaceae phylogeny as proposed by [16], B) is a heatmap of the diversity of iridoid, monoterpene or sesquiterpene compounds found in those species. C) presents some species and structural examples. NECA is *Nepeta cataria* and NEMU is *Nepeta mussinii*. Key for the rest of the species can be found in [11]. The colour of the circular frame on the pictures correspond to their subfamily. Font colour on chemical name correspond to their class of chemical. Plant images and chemical structures from Wikimedia commons. Caryoptoside structure from SpectraBase.

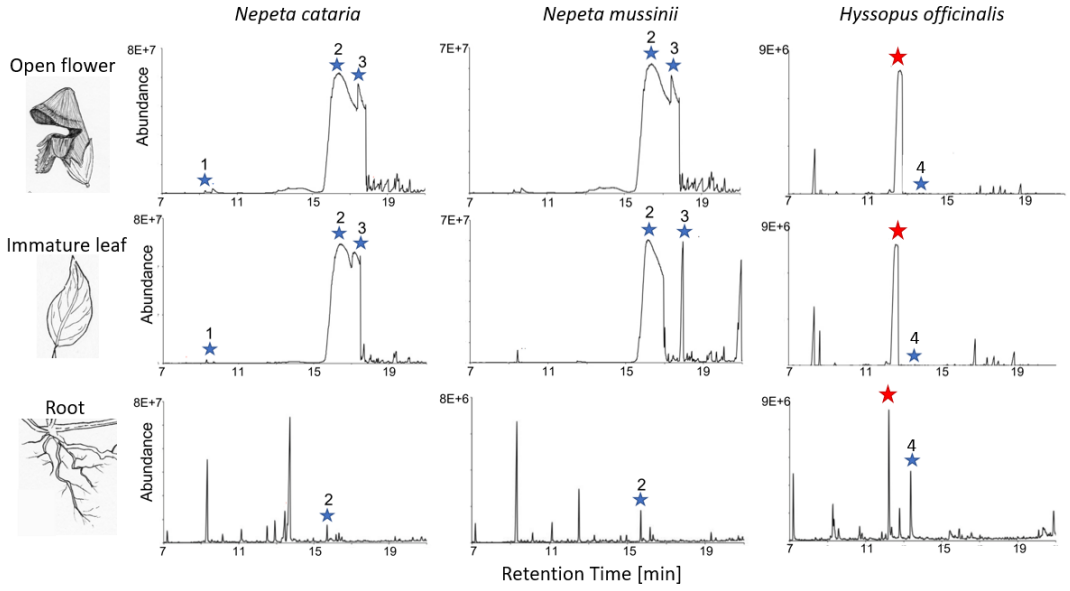


Figure 2.2: *Nepeta* and *Hyssopus* metabolite tissue distribution

Mass spectrometer (MS) traces from open flowers, immature leaves, and roots of *N. cataria, N. mussinii* and *H. officinalis*. Figure adapted from [65]. Blue stars denote metabolites derived from geraniol. 1 is β-citronellol, 2 is nepetalactone, 3 is isonepetalactone and 4 is citral. Red star indicates 3-pinanone, a non-geraniol derived monoterpene. GC/MS data generated by the Doudareva laboratory for the mint genome project [65].

Lamiaceae plants outside of the Nepetedoidea sub-family are able to make iridoid based compounds. Interestingly, iridoid based compounds produced in *Nepeta* have substantial chemical differences compared to those produced in non-Nepetoideae [11], [12], [23], [44], [65]. The majority of iridoids found in *Nepeta* are nepetalactones, which are aglycones, whereas iridoid glycosides are the major iridoid products found outside this subfamily (such as caryoptoside from *Lamium album*, fig. 2.1). The aglycone nepetalactone has an oxidized lactone rather than a glucosyl-oxygen bond. Notably, the nepetalactones are volatile whereas the iridoid glycosides are not. Moreover, nepetalactones present a different stereo-chemistry at carbon 7: (7*S*) compared to (7*R*) found in the rest of the Lamiaceae (fig. 2.3-A). This difference in stereochemistry alludes to an alternative pathway in *Nepeta* compared to non-Nepetoideae species [11]. Furthermore, while iridoids in the rest of the Lamiaceae family have a "*trans-trans*" stereochemistry at the bridgehead carbons (7α, 4α), *Nepeta* produce *cis-cis* and *trans-cis* in addition to *cis-trans* diastereomers of nepetalactones (fig. 2.3-B). Curiously, within the *Nepeta* genus, the distribution of the production of these isomers varies not only by species, but also by individual variety within a species. In our

collection we find several of our same-species cultivars producing different nepetalactone isomer ratios (table 2.1).

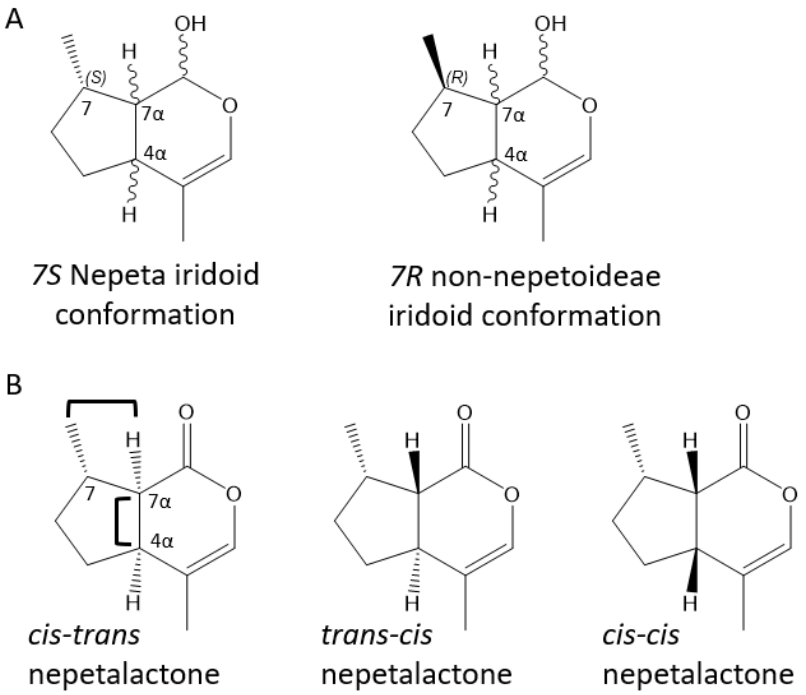


Figure 2.3: Iridoid isomers found in Lamiaceae

Basic iridoid structures found in Lamiaceae. In A), iridoid structures show the numbering and main isomeric differences between Nepetoideae (*7S*) and non-Nepetoideae (*7R*) iridoids. B) lists the various isomer structures, in relation to the bridgehead carbons 4α and 7α, and carbon 7, found in the *Nepeta* genus.

Cultivar name	Species	Metabolic profile percent total nepetalactones			
		cis-trans	trans-cis	cis-cis	trans-trans
20	*Nepeta mussinii*	90	0	10	0
21	*Nepeta mussinii*	90	0	10	0
B2E	*Nepeta cataria*	15	80	5	0
B2C	*Nepeta cataria*	10	85	5	0
HH03	*Nepeta cataria*	100	0	0	0
HH01	*Nepeta cataria*	100	0	0	0
MK 12	*Nepeta faassenii*	0	100	0	0
MK 14	*Nepeta faassenii*	0	100	0	0
B5A	*Nepeta sibirica*	15	0	0	85
B5F	*Nepeta sibirica*	15	0	0	85

Table 2.1: Nepetalactone isomer distribution in *Nepeta*

Nepeta species and cultivar inventory. Cultivar names indicate the individuals within species. The metabolic profile percent total nepetalactones is the percentage of the total nepetalactone area by each isomer. Heatmap was applied in excel, with red as low percentage and green as high percentage. Data was collected by O'Connor laboratory member Omar Kamileen.

2.1.2 Iridoid biosynthesis

Iridoids and iridoid-based compounds are produced across the plant kingdom [12], [30], [31], [33], [34], and serve a variety of ecological roles, as detailed in Chapter 1 section 1.1.4. Iridoids have been reported in Cornales, Gentianales, Olaceae and Lamiales, and often serve as the building blocks to more complex secondary metabolites, as introduced in section 1.1.9. In all characterized iridoid pathways, the early steps of all iridoid biosynthetic pathways utilize the same sequence of enzymatic reactions (fig. 2.6) [12]. In short, iridoids are derived from GPP, produced by the primary metabolic MEP pathway [11], [14]. GPP is the substrate for GES, a specialized terpene synthase (fig 2.5) [12], [39], [67]. For iridoid biosynthesis, geraniol is further oxidized, which enables the subsequent cyclization step. Hydroxylation of carbon C8 is carried out by G8H, a member of the class II cytochrome P450 enzyme family [40] and then both hydroxyl groups are oxidized to the aldehyde by the medium chain alcohol dehydrogenase hydroxy-geraniol oxidoreductase (HGO) [12]. Finally, ISY reduces, and in some cases cyclizes, these oxidized monoterpenes to produce iridoids. In some iridoid producers, such as non-Nepetoideae members of the Lamiaceae family [11],

further derivatization of nepetalactol occurs, beginning with oxidation by another class II P450 enzyme, IO to produce 7-deoxyloganetic acid (fig. 2.4) [12], [40].

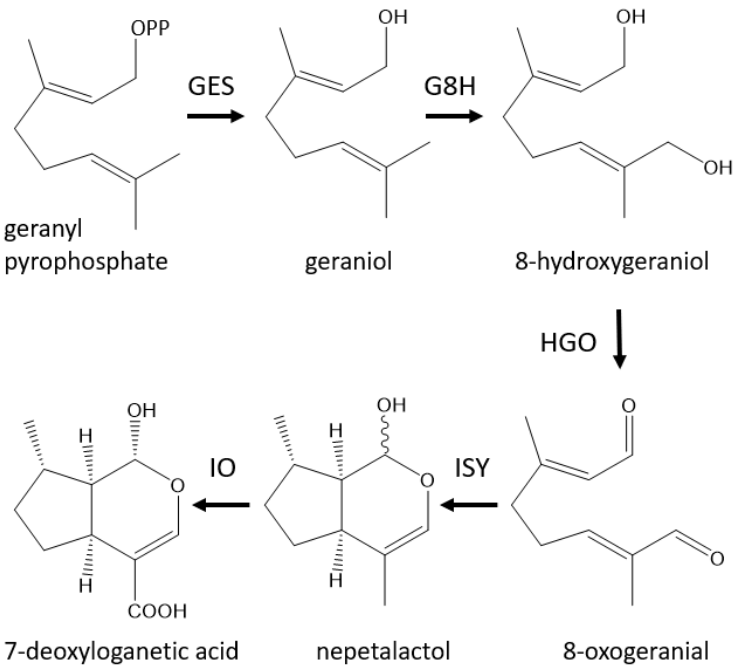


Figure 2.4: Known iridoid pathway from previously characterized plant iridoid pathways

Following up from the chemo-taxonomical results in [11], various members of the O'Connor laboratory were able to elucidate more of the pathway [23], [64]. The iridoid pathway characterized in other plant species [12], [30]–[32] depends on ISY to produce nepetalactol, an enzyme evolved from progesterone-5-beta reductase (P5βR). Both P5βR and ISY are part of the P5βR-like enzyme (PRISE) family [31]. However, phylogenetic analysis in [11] uncovered that ISY was lost in the Nepetoideae, and then re-emerged independently in the *Nepeta* genus from P5βR. Furthermore, extensive biochemical assays demonstrated that *Nepeta* ISY was only responsible for the reduction of the substrate 8-oxogeranial into 8-oxocitronellyl enol, and that cyclization into the 5-6 ring iridoid ring structure occurred spontaneously [23], [44]. Proteomic analyses of *Nepeta* trichomes led to the discovery of a novel class of enzymes: the NEPS enzymes [23], [44]. The biochemical activity of this family of enzymes was characterized *in vitro* and was found to be responsible for the stereoselective cyclization of the resulting nepetalactols and the oxidation to produce nepetalactones [64] (fig 2.6). Additionally, a novel activity for a major-latex protein-like (MLPL) enzyme was found, also key in carrying out the cyclization of iridoids, specifically

producing the *cis-trans* isomer (fig. 2.6) [65]. The *in vivo* activity of these late step enzymes (NEPS and MLPL) is explored in Chapter 4 of this thesis.

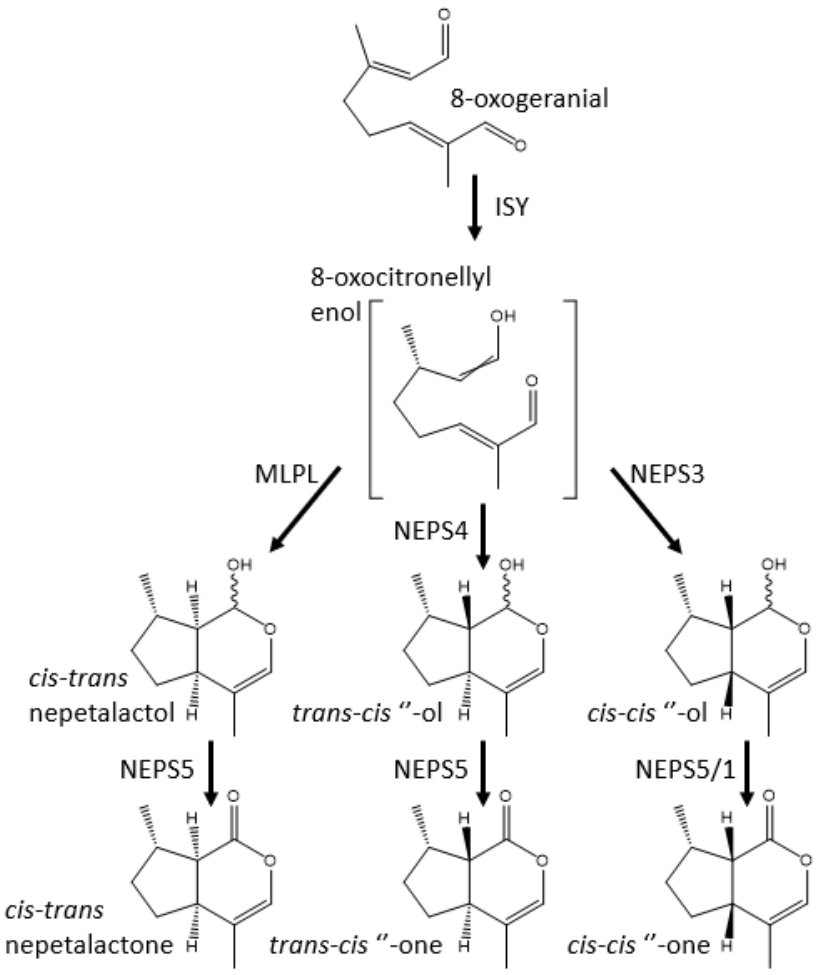


Figure 2.5: Late steps of the iridoid biosynthesis pathway in *Nepeta spp.*

As reported in Lichman *et al*. [65]

While the later steps in the *Nepeta* iridoid pathway described in Figure 2.6 and in [23], [44] had been elucidated, the earlier steps branching off from primary metabolism to 8-oxogeranial had not been found nor characterized in *Nepeta.* In this chapter, I detail the methods through which I identified candidates for the biosynthesis of the early steps in the pathway and characterized their biochemical activity.

2.2 Results and Discussion

2.2.1 Homology-based gene discovery

As mentioned in section 2.1.1, iridoids are found across the plant kingdom [12], [30], [31], [33], [34], but the early steps of biosynthesis appear to be conserved [12]. We hypothesized that iridoids found in *Nepeta* are likely to be produced by the same pathway as in *C. roseus* and other iridoid producing species, introduced in sections 1.1.5 and 1.1.9 [12]. Using the transcriptomic data from *N. cataria*, *N. mussinii* and *H. officinalis* generated in [11], I carried out a homology search with tBLASTn using the reported *C. roseus* early iridoid pathway enzymes (GES, G8H, HGOA/HGOB and IO) as a query (fig. 2.6 & table 2.2). ISY had been previously characterized by the O'Connor laboratory [11], [23], [64] and was not included in this homology search. The identity of HGOA in *C. roseus* is ambiguous; two enzymes that appear to be responsible for this reaction have been identified. Miettinen *et al.* [12] characterized an oxidoreductase HGO (referred to here as HGOA) which catalyses the formation of 8-oxogeranial in two reversible oxidation steps (via the formation of the possible two intermediates 8-hydrogeranial or 8-oxogeraniol) using NAD+. I also selected a candidate from a similar family of enzymes, here referred to as HGOB, characterized in Krithika *et al.* [68], which is specific to 8-hydroxygeraniol and the two di-hydroxy intermediates, and uses nicotinamide adenine dinucleotide phosphate (NADP+) as a cofactor.

Candidate transcriptomic sequences from the tBLASTn search were chosen based on their percent sequence identity and percent query coverage (table 2.2). Each sequence was compared between *N. cataria* and *N. mussinii* to check for orthologous transcripts. Then, each sequence candidate was also compared against the genomes of *N. mussinii* and *N. cataria*. An annotated proteome from *N. mussinii* was also used to check for enrichment in trichomes and to which orthogroup each enzyme is predicted to belong to (fig. 2.7-A).

Three enzymes, GES, G8H, HGOA, had at least one *N. cataria* and *N. mussinii* translated transcriptome sequence candidate with a sequence identity above 65% compared to amino acid sequences from *C. roseus*. Genomic and transcriptomic analysis of *N. cataria* suggests it to be a tetraploid [65] and at least two candidates of high sequence identity and query coverage were found for GES (A/B) and HGOA (1/2) (table 2.2). However, only one G8H

candidate was found (table 2.2). Several candidates for HGOB were found in *N. cataria* and the top four candidates were selected for further biochemical analyses to understand the conservation of activity amongst these highly similar sequences (further detailed in sections 2.2.4 and 2.2.5).

Notably, neither *N. cataria* nor *N. mussinii* had an iridoid oxidase candidate with a full coverage sequence of above 50% identity. In this study, I did not further characterize the resulting top IO candidates (with ~48% identity). However, glycosylated iridoids have been identified via liquid chromatography-mass spectrometry (LC-MS) analyses in *Nepeta* tissues, including 1,5,9-epi-deoxyloganic acid, and the enzymes required for production of this compound remain unknown [65]. Chapter 4 of this thesis reports on the relationship between the known iridoid pathway with 1,5,9-epi-deoxyloganic acid biosynthesis. The tBLASTn results from *H. officinalis* using *Nepeta* transcripts as queries did not yield any transcripts with high percent identity or query coverage. However, the transcripts with the highest percent identity and query coverage were selected for further downstream analyses (sections 2.2.3 and 2.2.4).

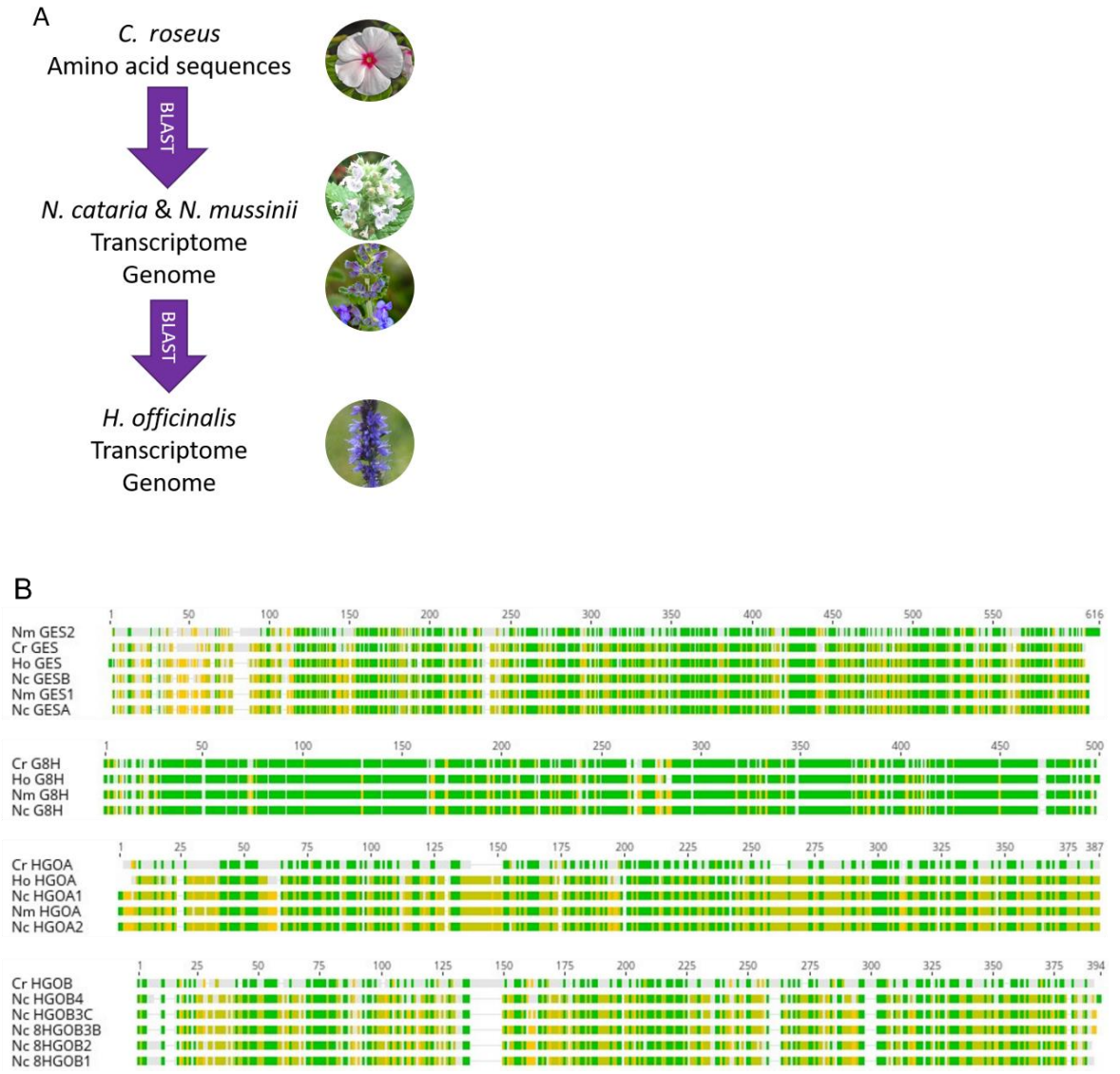


Figure 2.6: Homology based candidate selection methodology and alignment

A) is a schematic for the methodology of homology-based candidate search. B) shows amino-acid alignments of the candidates with highest degree of percent identity as compared to *C. roseus* query sequences. Nm is *N. mussini*i, Cr is *C. roseus*, Nc is *N. cataria* and Ho is *H. officinalis*. GES is geraniol synthase, G8H is geraniol-8-hydroxylase, HGO is hydroxy-geraniol oxidase. The candidates are further listed in table 2.2.

GES

Species	Name	BLAST seed	E-value	% Identity	Query Coverage
N. cataria	GESA	*C. roseus* GES	0	68.40%	66.55%
N. cataria	GESB	*C. roseus* GES	1.19E-95	37.00%	82.17%
N. mussinii	GES	*C. roseus* GES	0	67.10%	82.34%
H. officinalis	GES	*N. cataria* GES	7.48E-74	40.80%	46.97%

G8H

Species	Name	BLAST seed	E-value	% Identity	Query Coverage
N. cataria	G8H	*C. roseus* G8H	0	68.00%	95.54%
N. mussinii	G8H	*C. roseus* G8H	0	67.80%	95.54%
H. officinalis	G8H	*N. cataria* G8H	0	78.50%	48.68%

HGOA

Species	Name	BLAST seed	E-value	% Identity	Query Coverage
N. cataria	HGOA1	*C. roseus* HGOA1	1.36E-113	63.30%	78.53%
N. cataria	HGOA2	*C. roseus* HGOA2	1.26E-79	45.00%	93.50%
N. mussinii	HGOA	*C. roseus* HGOA	1.20E-152	63.80%	99.72%
H. officinalis	HGOA	*N. mussinii* HGOA	0	90.80%	52.42%

HGOB

Species	Name	BLAST seed	E-value	% Identity	Query Coverage
N. cataria	HGOB1	*C. roseus* HGOB	0	83.30%	99.72%
N. cataria	HGOB2	*C. roseus* HGOB	1.57E-179	78.60%	99.72%
N. cataria	HGOB3	*C. roseus* HGOB	1.06E-162	75.80%	98.89%
N. cataria	HGOB4	*C. roseus* HGOB	1.78E-157	71.40%	100.00%

IO

Species	Name	BLAST seed	E-value	% Identity	Query Coverage
N. cataria	IO	*C. roseus* IO	1.95E-123	43.70%	89.90%
N. mussinii	IO	*C. roseus* IO	8.00E-168	49.20%	95.73%

Table 2.2: List of candidates selected for this project

List of candidates selected after tBLASTn of *C. roseus* sequences *to H. officinalis, N. cataria* and *N. mussinii* transcriptomes. Enzymes with multiple candidates are listed in order of highest to lowest sequence identity similarity. BLAST seed is the query sequences used for the homology search. E-value denotes the number of hits that could be found by chance. % Identity denotes the percentage of amino acids that are identical to the BLAST query. Query coverage is the percentage the query length aligns with the BLAST hit.

As reported in Lichman *et al.* [65], the later steps in the *Nepeta* iridoid pathway (ISY, NEPS and MLPL) are found in metabolic gene clusters (one in diploid *N. mussinii* and two in tetraploid *N. cataria*). The early enzyme candidates found in this homology search revealed that only *N. mussinii* GES was also in the *N. mussinii* cluster. However, none of the *N. cataria*

early candidates were found in its clusters. A primordial syntenic cluster was also found in *H. officinalis*, which contains NEPS-like pseudogenes (fig. 2.8). Further information on the cluster can be found in Chapter 4 of this thesis.

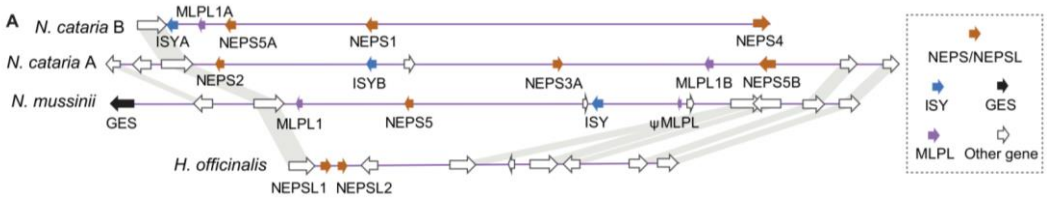


Figure 2.7: Iridoid gene cluster in *Nepeta*

Nepeta iridoid gene cluster found in *N. cataria*, *N. mussinii* and *H. officinalis*. Adapted from Lichman *et al*. [65].

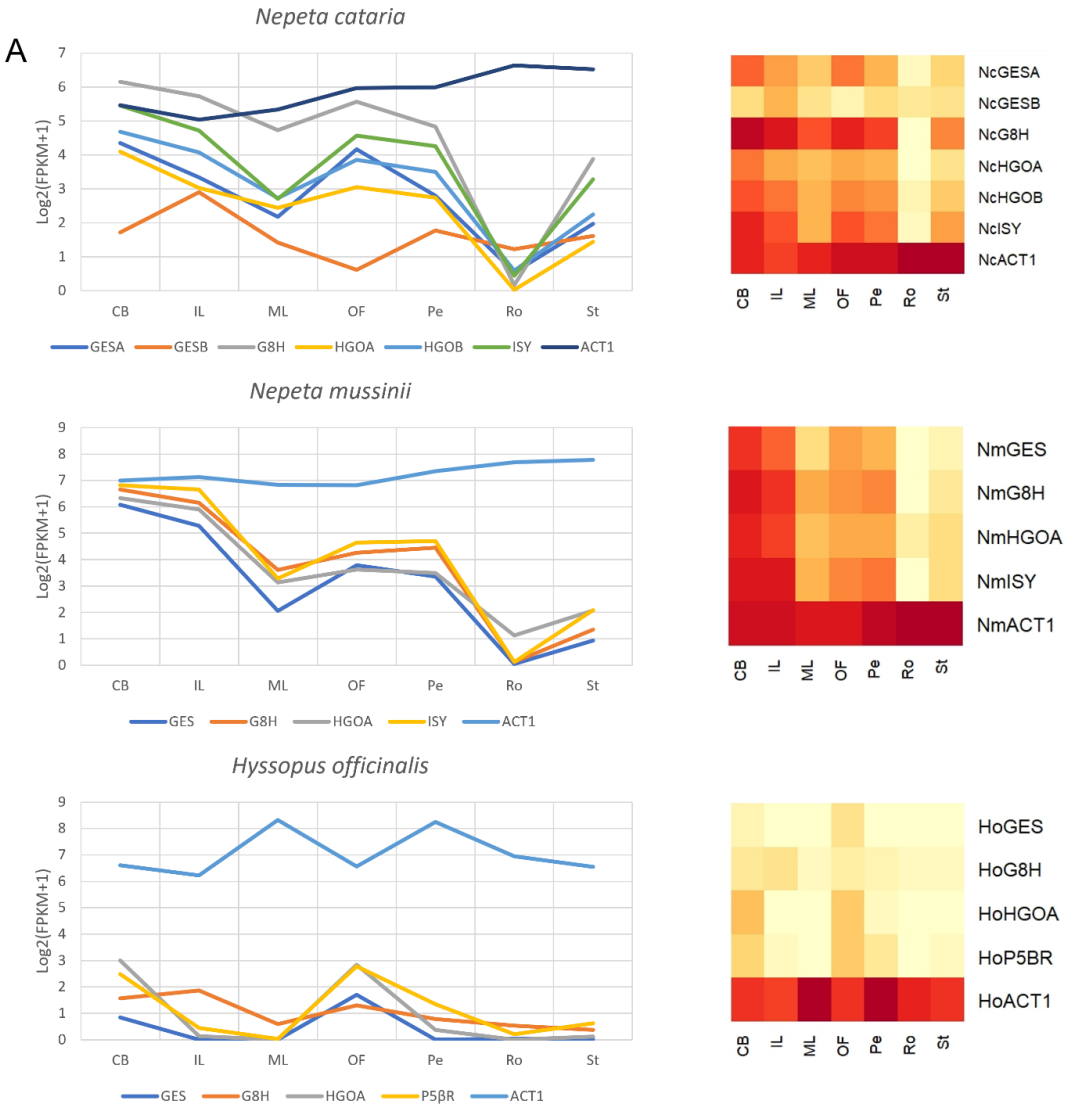
2.2.2 Gene co-expression analysis

I used transcriptomic data from root, stem, mature leaf, young leaf, petiole, flower buds and open flowers of *N. cataria*, *N. mussinii* and *H. officinalis* (previously generated by the Mint Genome Project [11]) to carry out co-expression analyses on the candidate genes: orthologues of GES, G8H, HGO and ISY. I also included P5βR from *H. officinalis*, as there is no ISY orthologue in that species. As discussed in section 2.1, P5βR and ISY belong to the PRISE enzyme family [31], and the *H. officinalis* P5βR selected was the closest relative to *Nepeta* ISY. Additionally, the ortholog of Actin1 (ACT1) was used as a baseline for gene expression. The log2 of FPKM+1 values were calculated and mapped in both a heatmap and individual graphs for each gene (fig. 2.8).

The *Nepeta* candidates were found to be expressed at the highest levels in the same tissues where iridoids are found (flower and leaf tissue) and less so in tissues where the metabolites are found in lower concentration (stem) or are not present (root) [11], [65]. This analysis also showed that the GESB orthologue in *N. cataria* had less overall expression than GESA, which may indicate the extra copies of these early genes found in *N. cataria* may already be in the process of losing function, as may happen in large scale genomic duplication events within specialized metabolism [3], [69], [70].

I also found that the gene candidates for the upstream genes GES, G8H and HGOA in *H. officinalis*, although present in the genome, were not expressed, or were expressed at

negligible levels in *H. officinalis* tissues. This corresponds to the lack of geraniol-based metabolites in the tissues of *H. officinalis* [11], [65].



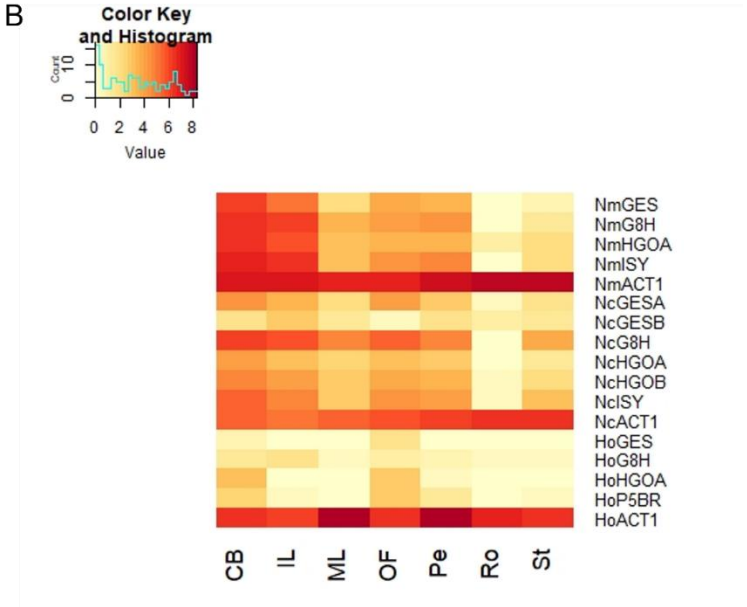


Figure 2.8: Co-expression analysis of candidate gene FPKM values

FPKM values were generated by the Buell laboratory for the Mint Genome Project. FPKM values for the early steps in the iridoid biosynthesis pathway and the homologue for Actin1 in each plant were converted to the Log2 scale before analysis. CB=closed bud, IL=immature leaf, ML=mature leaf, OF=open flower, Pe=Petiole, Ro=Root, St=Stem. Heatmaps were generated using the Heatmap package in R and were subjected to no hierarchical clustering.

The presence of potential GES, G8H and HGOA genes in *H. officinalis* that are expressed at low or non-detectable levels may contribute to an explanation of how the iridoid pathway evolved in *N. cataria* and *N. mussinii*. It was observed that while non-Nepetoideae members of the Lamiaceae have iridoids and the crucial iridoid biosynthetic gene ISY, the Nepetoideae Lamiaceae do not, apart from *Nepeta*. Careful phylogenetic reconstruction analysis strongly suggests that ISY evolved independently in *Nepeta* [11], [65]. We hypothesize that the ISY gene was lost in the Nepetoideae ancestor, and while the other genes in the early steps of the pathway remained present in the genome, changes in their transcriptional regulation meant that the corresponding proteins were not present. However, the presence of these similar upstream genes in the genomes may have facilitated the re-evolution of ISY and the fixation of an iridoid producer in *Nepeta*.

2.2.3 Design of heterologous protein expression

After the homology and transcriptomic analysis, I set out to characterize the biochemistry of these candidate enzymes. I selected *E. coli* SoluBL21 and *S. cerevisiae* (budding yeast) pep4KO for heterologous recombinant expression. GES and HGO homologues were expressed in *E. coli*, while G8H, a membrane bound cytochrome P450, was expressed in yeast.

The first attempts at expression and purification of GES included the complete coding sequence (CDS) region, expressed using the inducible pOPIN vector system, with a histidine tag at the C-terminus of the recombinant protein (pOPINF). However, expression of the complete CDS region proved to be ineffectual. However, the GES orthologue found in *O. basilicum* has been reported to contain a hydrophobic plastid transit peptide [39] on the N-terminus, which must be truncated for effective heterologous expression. An alignment of the *Nepeta* candidate genes with the reported *O. basiliscum* GES [39] sequence suggested the presence of a transit peptide in the *Nepeta* proteins. The candidate GES were re-designed without this putative transit peptide, which resulted in improved expression in *E. coli*. These GES proteins were subsequently purified using standard Ni-NTA chromatography (fig. 2.9).

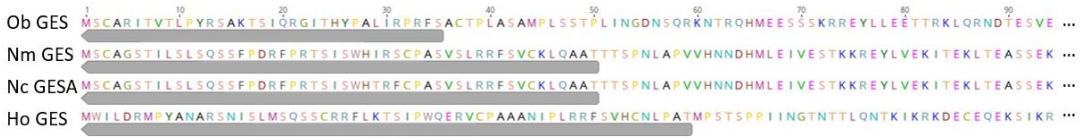


Figure 2.9: GES alignment

List of GES candidates with *O. basilicum* sequence characterized in Iijima *et al.* [39] showing truncation design for optimized expression. "Ob GES" is *O. basilicum* GES, "Nm GES" is *N. mussinii* GES, "Nc GESA" is *N. cataria* GESA and "Ho GES" is *Hyssopus officinalis* GES. The grey bar indicates the N-terminal sequences that were truncated.

As G8H is a membrane-dependent cytochrome P450 enzyme, this was cloned into the pESC-Leu vector for expression in yeast. While cloning, transformation and whole cell feeding assays proved to be successful for *Nepeta spp*. and *H. officinalis* G8H, isolation of enough pure and active microsomes was not successful (fig. 2.10). However, given the substrate's (geraniol) low cost and ease of availability, I carried out biochemical assays using whole cell feeding assays. Finally, HGO orthologues were also cloned into the pOPIN system (pOPINF

and pOPINJ), expressed, and purified with relative ease in the case of *Nepeta spp.* homologues.

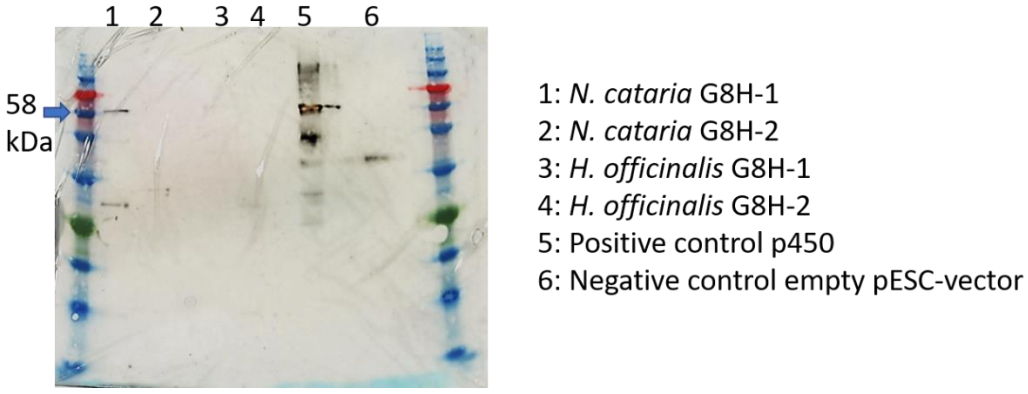


Figure 2.10: G8H western blot

Western blot showing expression of *N. cataria* and *H. officinalis* G8H heterologously expressed in *S. cerevisiae* yeast microsomes. A previously characterized *Rauwolfia serpentina* cytochrome P450 was used as a positive control, and an empty pESC-vector microsome preparation was used as a negative control. Anti-FLAG antibodies were used against the FLAG tag found on the pESC-vector.

While the *Nepeta spp*. GES and HGOA/B enzymes were purified in high yield from *E. coli*, the *H. officinalis* candidate enzymes were not. Several attempts at expression using the pOPIN vectors were made, including expression into pOPINF, which carries a histidine tag for purification, and pOPINJ, which has a GST protein tag to improve solubility, but in all cases, expression was low and subsequent purification of these enzymes failed to yield sufficient quantities required for enzymatic assays (fig. 2.11).

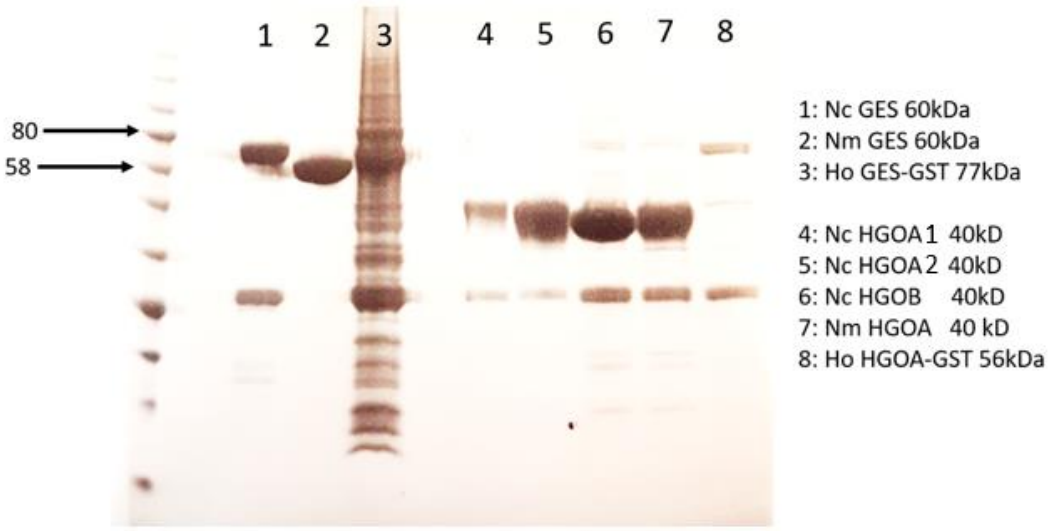


Figure 2.11: SDS-PAGE gel of GES and HGOA

SDS-PAGE showing the final purification of GES and HGOA candidates. "kDa" is Kilo Daltons; Nc GES is *N. cataria* GES, Nm GES is *N. mussinii* GES, Ho GES-GST is *H. officinalis* GES expressed in pOPINJ, Nc HGOA1 is *N. cataria* HGOA1, Nc HGOA2 is *N. cataria* HGOA2, Nc HGOB is *N. cataria* HGOB, Nm HGOA is *N. mussinii* is HGOA, and Ho HGOA-GST is *H. officinalis* HGOA expressed inn pOPINJ.

2.2.4 *In vitro* characterization of early enzyme pathway

After the successful expression and purification of the *Nepeta spp*. GES and HGOA/B enzymes (fig. 2.11), I carried out *in vitro* assays to assess their predicted activities. While microsomal preparations of G8H were not possible, as described in section 2.2.3, whole cell culture assays revealed an active enzyme as evidenced by the conversion of geraniol to geraniol-8-hydroxylase upon feeding (fig. 2.13). As all enzymatic products from these expected reactions are detectable via gas chromatography-mass spectrometry (GC-MS), I assessed activity by presence or absence of the expected products via this methodology.

Previous work has shown that GES is a specialized terpene synthase [10], [39], [67], [71]–[73]. The reported mechanism indicates that GES carries out the cleavage of the phosphate bond via the manganese ion aided formation of a carbocation to which a hydroxyl group can then be added. Previous work on recombinant *O. basilicum* GES and *C. roseus* GES used the co-factor $MnCl_2$ in *in vitro* assays. I set up the *in vitro* reaction as in Ijima *et al.* [39] and Simkin *et al.* [67]. Both *N. cataria* and *N. mussinii* GES homologues were active under these

conditions, producing geraniol after 1 hour reaction incubations, suggesting a conservation of activity in the *Nepeta spp*. GES.

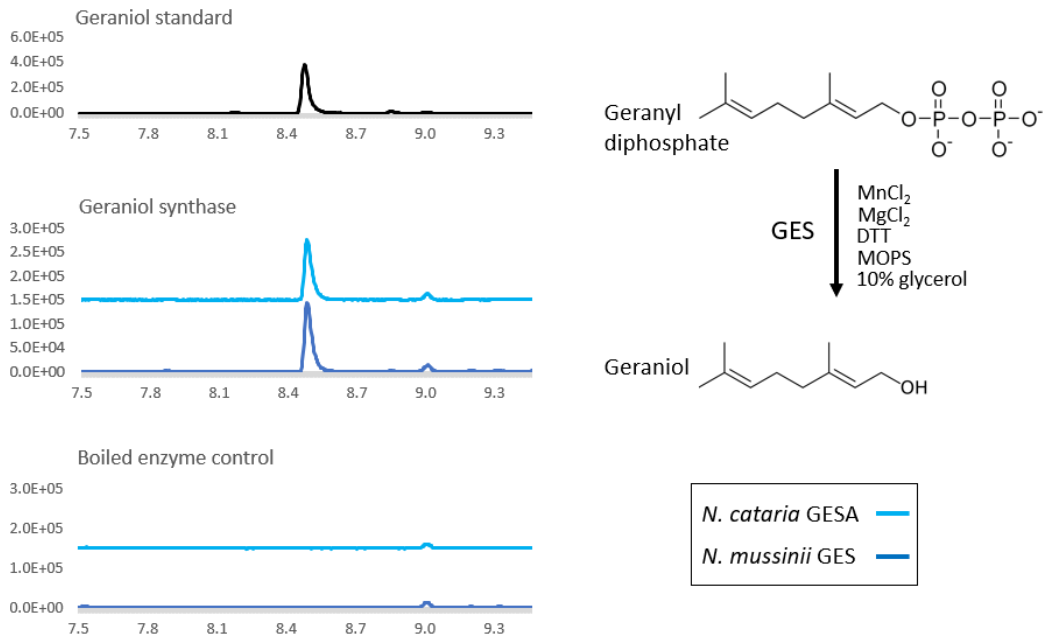


Figure 2.12: GES *in vitro* activity

Representative GC-MS data from *in vitro* reactions geraniol synthase. X-axis is retention time in minutes. Y-axis is total ion chromatogram (TIC) Individual graphs have been adjusted in the Y-axis to clearly present the data in the same graph. All reactions contain the elements listed in the chemical reaction: enzyme, magnesium chloride ($MgCl_2$), manganese chloride ($MnCl_2$), dithiothreitol (DTT), MOPS buffer and 10% glycerol.

For *N. cataria*, *N. mussinii* and *H. officinalis* G8H assays, I induced gene expression on the pESC-LEU vector via the addition of 2% galactose and added 0.5 mM of geraniol directly to the 0.5 mL cultures at pH7 for a 24 hour incubation at 30 ˚C. The cultures were then quenched with ethyl acetate (EtAc) to extract any organic hydrophobic compounds in the culture. GC-MS analysis revealed the production of 8-hydroxygeraniol in *N. cataria* and *N. mussinii* G8H (fig. 2.13). *H. officinalis* G8H cultures did not exhibit G8H activity. This may be due to low expression as observed with *H. officinalis* GES and HGOA, described in section 2.2.3, rather than the enzyme itself not being functional.

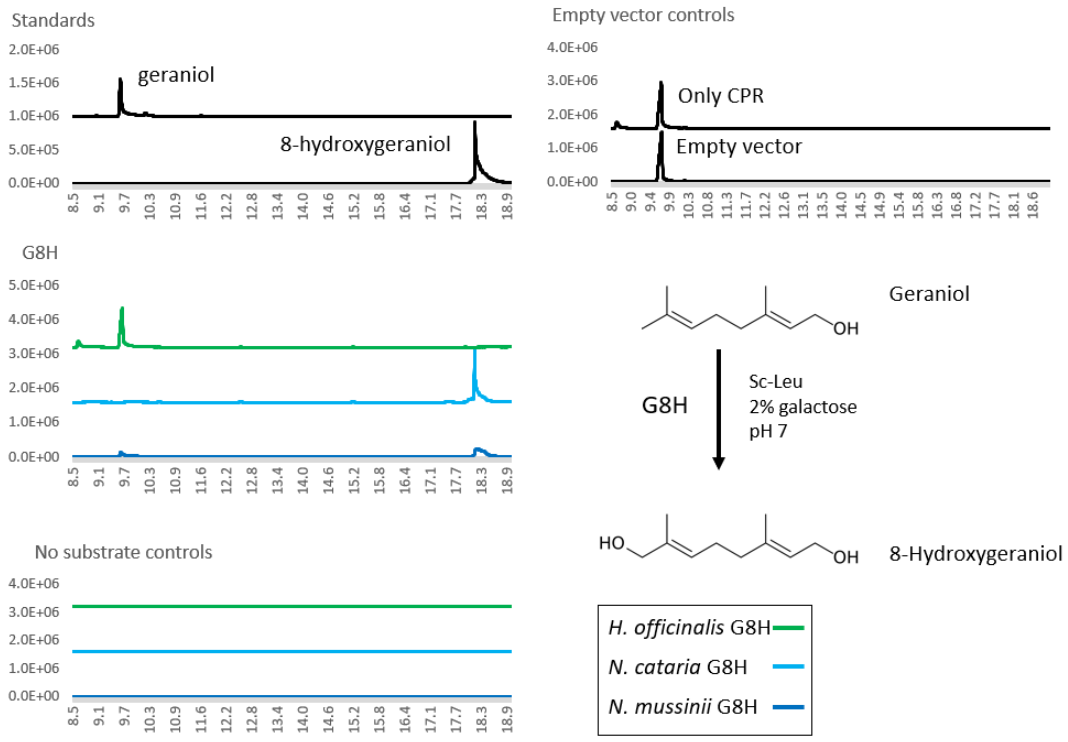


Figure 2.13: G8H whole cell culture assays

Representative GC-MS data from *in vitro* reactions of geraniol-8-hydroxylase. X-axis is retention time in minutes. Y-axis is total ion chromatogram (TIC) Individual graphs have been adjusted in the Y-axis to clearly present the data in the same graph. Feeding assay carried out with yeast grown in SC-Leucine (SC-Leu) media with 2% galactose.

As mentioned in section 2.2.1, HGO homologues have been previously characterized in Mittenien *et al.* [12] and Krithika *et al.* [68]. This enzyme carries out the oxidation of both the 1- and 8-hydroxyl groups on the molecule. While the mechanism of this reaction has not been characterized, previous work showed the complexity of this reaction, leading to formation of not only 8-oxogeranial, but also 8-hydroxygeranial and 8-oxogeraniol, oxidation intermediates (fig. 2.14). Replicating this reaction, after overnight incubation, I found the resulting products included the formation of both the intermediates 8-oxogeraniol and 8-hydroxygeraniol and the final product 8-oxogeranial, for both *N. cataria* and *N. mussinii* (fig. 2.14).

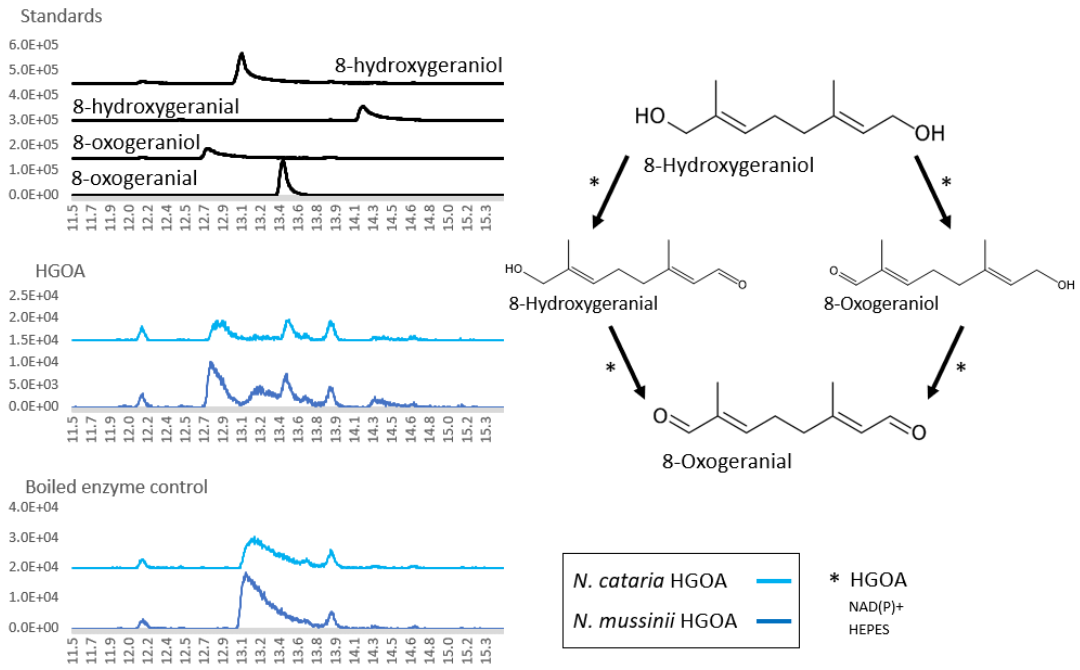
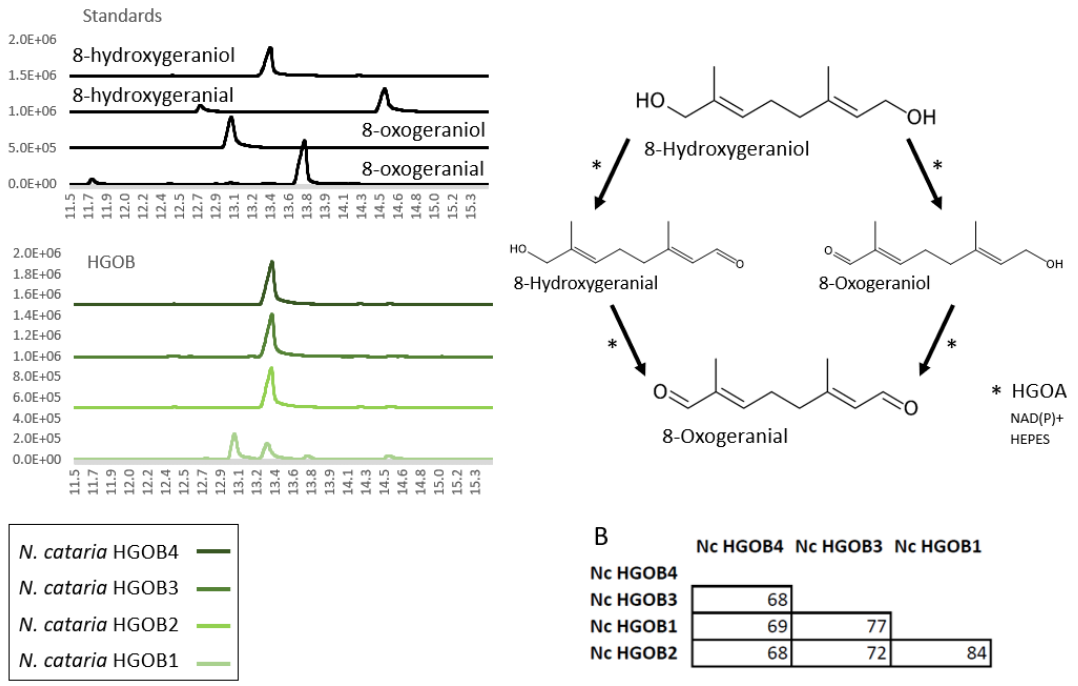


Figure 2.14: HGOA *in vitro* assays

Representative GC-MS data from *in vitro* reactions of the early iridoid pathway genes. X-axis is retention time in minutes. Y-axis is total ion chromatogram (TIC) Individual graphs have been adjusted in the Y-axis to clearly present the data in the same graph. All reactions contain the elements listed in the chemical reaction: Enzyme (HGOA/B), cofactor ($NAD(P)^+$), and buffer (HEPES), excepting standards. The poor peak shape is due to the elution properties of the dialdehyde moiety on the GC-MS.

I also tested the activity of the top four *N. cataria* HGOB candidates found in section 2.2.1 (table 2.2, fig. 2.15-A). While these candidates share high percent identity between their amino acid sequences (between 68-84%) (fig. 2.15-B), only *N. cataria* HGOB1 was able to convert 8-hydroxygeraniol to 8-oxogeranial, and the intermediates 8-hydroxygeranial and 8-oxogeraniol.



B

	Nc HGOB4	Nc HGOB3	Nc HGOB1
Nc HGOB4			
Nc HGOB3	68		
Nc HGOB1	69	77	
Nc HGOB2	68	72	84

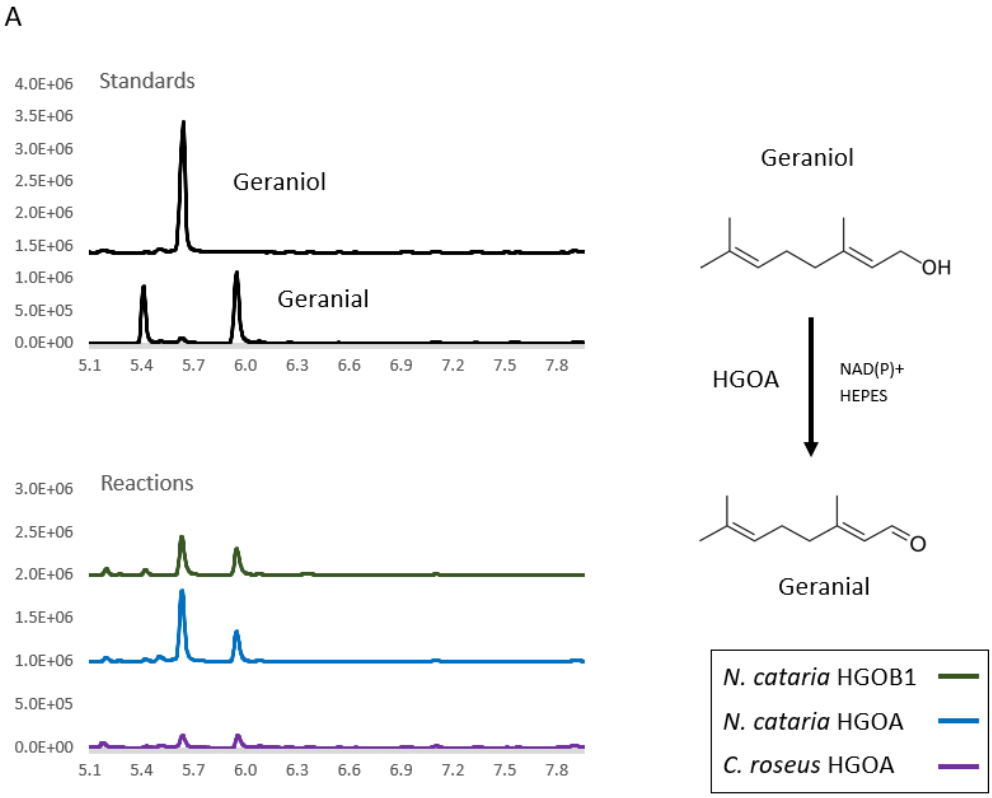
Figure 2.15: HGOB *in vitro* assays

A) is representative GC-MS data from *in vitro* reactions of *N. cataria* HGOB homologue candidates. X-axis is retention time in minutes. Y-axis is total ion chromatogram (TIC) Individual graphs have been adjusted in the Y-axis to clearly present the data in the same graph. All reactions contain the elements listed in the chemical reaction: Enzyme (HGOA/B), cofactor (NAD(P)$^+$), and buffer (HEPES), excepting standards. B) shows the percent identity matrix between the *N. cataria* (Nc) HGOB candidates.

2.2.5 HGO substrate selectivity

I chose the *N. cataria* HGOA and HGOB1 candidates to test their activity with alternative substrates. The substrate profile of these enzymes could then be compared to that of HGOA from *C. roseus* [12]. I wanted to test if any of the HGO enzymes had a more promiscuous activity. I also wanted to test for oxidation overlap as reported in other publications [68], [74]–[76]. Furthermore, testing enzyme activity on alternative substrates could reveal the potential effect on enzyme specialization from the re-emergence of the pathway in *Nepeta* after its loss in the Nepetoideae sub-family compared to the previously characterized *C. roseus* HGO [12]. The chosen alternative substrates were A) geraniol, differing from 8-hydroxygeraniol only in the missing hydroxy group at carbon 1, B) (S)-(-)-β-citronellol, which differs from geraniol by the presence of a chiral methyl group at carbon 7, and C) menthol, a cyclohexane which can be oxidized to menthone. All selected enzymes, CrHGOA, NcHGOA

and NcHGOB1, were able to oxidize geraniol into geranial. However, only *N. cataria* HGOA could turn over (S)-(-)-β-citronellol, producing trace amounts of (S)-(-)-β-citronellal. No enzyme oxidized the hydroxy group of menthol. As discussed in Krithika *et al.* [68], HGO homologues from *C. roseus* and *R. serpentina* may not be very selective for di-hydroxylated carbon chains (such as 8-hydroxygeraniol), being able to oxidize geraniol, which may contribute to the overlapping of oxidation reactions proposed by Höfer *et al.* [74]. This seems to be conserved in the *N. cataria* HGO, as evidenced by NcHGOA and NcHGOB1 abilities to oxidize geraniol (fig. 2.16 A).



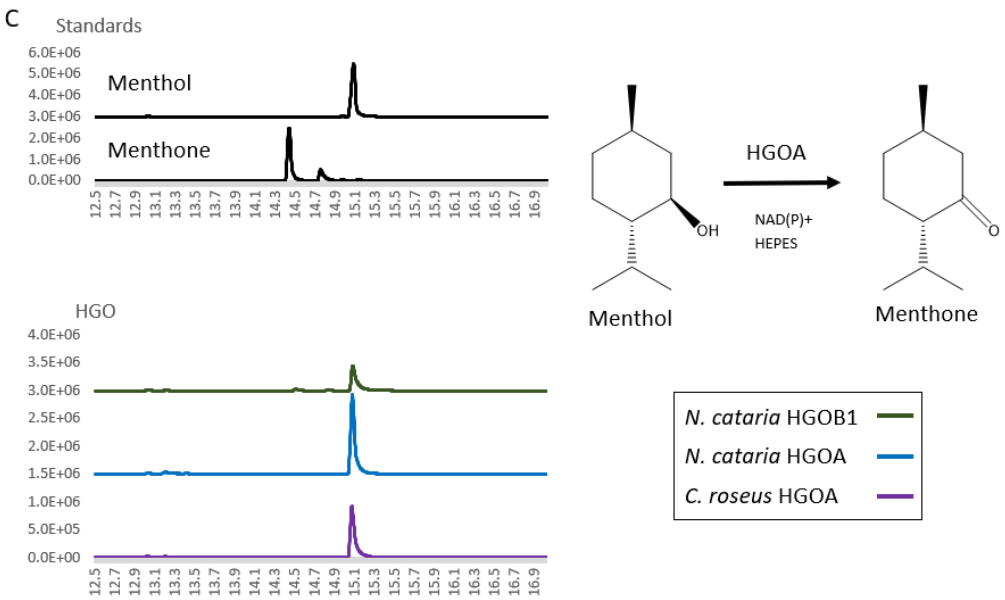


Figure 2.16: Alternative substrate in vitro reactions of *N. cataria* and *C. roseus* HGO homologues

GC-MS data showing the oxidation of various substrates. X-axis is retention time in minutes. Y-axis is total ion chromatogram (TIC) Individual graphs have been adjusted in the Y-axis to clearly present the data in the same graph. Asterisks indicate the peak corresponding to the expected product. All reactions contain the elements listed in the chemical reaction diagram in (A): enzyme (HGOA/B), cofactor ($NAD(P)^+$), and buffer (HEPES), excepting standards. (A), (B), (C) show the GC-MS data for assays carried out with geraniol (A), (S)-(-)-β-citronellol (B), and menthol (C) as substrates using *C. roseus* HGOA and *N. cataria* HGOA/B, along with the structures of the substrates and expected products.

2.3 Future Directions

2.3.1 Substrate promiscuity and gene synteny analysis

In section 2.2.5 I tested various substrates on NcHGOA, NcHGOB1 and CrHGOA to understand the oxidation promiscuity of these enzymes, as well as the effect of their evolutionary history. For the most part, the *N. cataria* homologues seemed to behave similarly as the *C. roseus* homologues, with the only exception being the trace conversion of (S)-(-)-β-citronellol into (S)-(-)-β-citronellal by NcHGOA.

GES is a specialized terpene synthase, which comprises a large family that acts upon isoprene units to form the various classes of known terpenoids [10], [71]–[73]. Plants contain several terpene synthases, not only active in secondary metabolism, but also active in essential processes such as photosynthesis [72]. This class of family has also been reported to exhibit evolutionary flexibility, with promiscuity often leading to new products or via new enzymatic activities such as cyclisation [72]. Tests on enzyme promiscuity, as those executed on the HGO homologues, would aid in understanding how GES would be maintained in non-iridoid producing Nepetoideae such as *H. officinalis*. Common terpene synthase substrates, such as those involved in the production of other monoterpenes and sesquiterpenes (such as FPP), could aid in revealing an alternative role for a GES uninvolved in iridoid production.

Furthermore, gene synteny analysis, as carried out in Lichman *et al*. [65] on the later steps of the pathway, could provide evidence whether the candidates genes of GES, G8H, and HGO from *H. officinalis* are the syntenic homologues to those found in *N. cataria*. This would shed light on the evolutionary history of the pathway.

2.3.2 Glycosylated iridoids in *Nepeta* and iridoid oxidase

Glycosylated iridoids are commonly found throughout the plant kingdom, including the Lamiaceae family. Lichman *et al.* [65] reported finding 1,5,9 epi-deoxyloganic acid, a glycosylated iridoid that would normally be synthesized via a further two-step oxidation of nepetalactol into 7-deoxyloganetic acid via IO and glycosylated via a UGT (fig. 2.4) [12]. However, searching the *Nepeta* transcriptomic datasets did not yield a candidate with high sequence identity to CrIO. While the resulting top candidates may have less identical sites, this does not rule out their activity as iridoid oxidases. *In vitro* characterization of the top IO candidates, despite their lower percent identity, would be the first step to finding the biosynthetic pathway.

2.3.3 Oxidation promiscuity

Curiously, in Höfer *et al*. [74], it was reported that G8H is able to oxidize geraniol into 8-oxogeraniol, suggesting some overlap within these early oxidation steps. Furthermore, previously identified hydroxygeraniol oxidases were non-specific to di- or mono-hydroxy linear geraniol derivatives [75], [76]. As described briefly in section 2.2.5, NcHGO homologues are also able to act on geraniol, suggesting that this oxidation promiscuity is conserved across the large species gap between the Apocynaceae *C. roseus* and *R. serpentina*

and the Nepetoideae *Nepeta spp.*. Furthermore, as an extension to the experiments proposed in section 2.3.1, replicating the oxidation experiment outlined in Höfer *et al*. [74] would reveal if the previously reported oxidation overlap is conserved in *Nepeta* iridoid biosynthesis, which may point to an inherent nature of this pathway.

2.4 Conclusion

In the work detailed in this chapter, we identified the early enzymes of the iridoid biosynthetic pathway in *N. cataria* and *N. mussinii.* Enzymes were identified by sequence similarity with previously discovered enzymes [12] from unrelated iridoid producing plants. I carried out genomic and transcriptomic analyses to narrow down the gene candidates, and carried out heterologous gene expression, protein purification and *in vitro* biochemical characterization.

While iridoids are widespread across the plant kingdom, the enzymatic steps and chemical reactions are tightly conserved amongst studied plant species [12]. In this chapter, I demonstrated that the same early steps to making iridoids, from GPP to 8-oxogeranial, were conserved in *Nepeta*, despite the loss in the Nepetoideae sub-family, and re-emergence in the *Nepeta* genus. Later steps in the pathway, as discussed in section 2.1 and reported in [23], [44], [65] show enzymatic novelty and variation, but the early steps foundation upon which *Nepeta* ISY relies, are conserved.

While the homologue candidate search did not reveal highly identical sequences (to the *Nepeta* sequence queries) for GES, G8H and HGOA in *H. officinalis*, the most highly identical candidates were selected for downstream co-expression and *in vitro* biochemical analysis. Unfortunately, heterologous expression and therefore characterization of these candidates could not be achieved. While there are assays to test for protein stability and several expression techniques that can be optimized, we did not carry these out, as our main goal in this chapter was to find and characterize the early steps in the iridoid pathway in *Nepeta*.

In this chapter we showed not only that *N. mussinii* and *N. cataria* GES, G8H and HGOA/B are conserved to their *C. roseus* homologues, but they are active in carrying out their predicted biochemical reactions. The following two chapters focus on characterizing the activity of these genes *in planta*.

2.5 Materials and Methods

2.5.1 Enzyme sequence discovery

C. roseus amino acid sequences were obtained from Uniprot accession numbers: Q8VWZ7 (G8H), W8JIS5 (IO), J9PZR5 (GES), Q6V4H0 (8HGOB), W8JWV8 (HGOA). All BLAST was carried out in Geneious, using available transcriptome databases for *Nepeta* (reference: Phytochemistry 145 (2018) 48-56), from the Mint Genomic Project http://mints.plantbiology.msu.edu/, and as described above in section 2.2.1.

2.5.2 Obtaining plasmid gene inserts

Gene inserts were obtained both by *N. cataria*, *N. mussinii* and *H. officinalis* cDNA amplification (HGOB and G8H) and ordering codon optimized for *E. coli* expression synthetic genes (GES and HGOA). Primers for cDNA amplification included a nucleotide extension which matched to the pOPINF, pOPINJ or pESC-Leu vector (oligo list). Primers were designed to fit within the coding region of the vector, and the start codon was removed for cloning in pOPINF and pOPINJ. 5' and 3' UTR primer sequences did not include pOPINF and pOPINJ extension sequences. Primers were designed to be between 15 and 25 nucleotides long and with an annealing temperature close to 55 °C. Synthetic genes were ordered from Eurofins and were codon optimized for *E. coli* gene expression.

2.5.3 cDNA production and insert amplification

RNA from young leaf tissue from *N. cataria*, *N. mussini* and *H. officinalis* was retrotranscribed using Invitrogen Superscript IV. In a 10 µL reaction, up to 500 ng (about 1 µL) of RNA, 4 µL of SSIV buffer, 1 µL of SSIV reverse transcriptase (200 U/µL) and 1 µL of 100 mM DTT. Incubated at 50 °C for 10 minutes and then at 80 °C for 10 minutes to inactivate. RNAase H (1 µL) was added and incubated at 37 °C for 20 minutes. cDNA was stored at -20 ˚C. Inserts from cDNA were amplified by PCR in a 25 µL reaction using 1 µL of cDNA, 12.5 µL of Invitrogen Platinum Superfi DNA polymerase mix, 1.25 µL of 10 µM Forward and Reverse primers each, and 5 µL of GC enhancer. Primers for each reaction can be found in the primers list.

2.5.4 Vector construction

Vectors were linearized using restriction enzymes. KpnI and HindIII were used to linearize pOPINF and pOPINJ, and Spc1-(HF) was used for pESC-LEU-2d. All enzyme digestions were

carried out with 1 µL of enzyme, 100 µg of vector, NEB Cutsmart buffer in 50 µL reactions. Reaction mixes were incubated at 37 °C for 1 hour. Linear vectors were purified by running a 1% agarose electrophoresis gel from which the DNA was purified using a Machery-Nagel NucleoSpin Gel and PCR Clean-up kit.

Ligations of the genes into pOPINF and pESC-Leu vectors were performed by InFusion reaction (ClonTech). In a 5 µL reaction, 1 µL of In-Fusion HD Plus, 20-50 ng of linear vector, and 20-50 ng of insert were added. This reaction mix was incubated at 50 °C for 15 minutes and cooled on ice. 2.5 µL of this reaction were used to transform Stellar competent cells and grown on selective media.

2.5.5 *E. coli* transformation

Stellar competent cells were used for post-infusion and ligation reaction transformations. SoluBL21 cells were used to express vector genes and purify proteins. Transformation of both cell lines followed the same protocol; cells are thawed on ice, then, 1-5 µL of transformant was added to 30-50 µL of cells and gently mixed by stirring pipette tip in mixture. Cells were left to incubate on ice for 10 minutes, heat shocked at 42 °C for 45 seconds, and cooled on ice for 2 minutes. 150 µL of SOC media was added to cells and gently mixed by stirring pipette tip in mixture and left to incubate at 37 °C for 45 minutes. 100-150 µL was plated on LB plates containing the appropriate antibody (pESC and pOPINF used carbenicillin) and was allowed to grow overnight at 37 °C.

2.5.6 Yeast transformation

A knockout strain for *pep4* was used for vectors containing P450 genes. The competent cell line was grown in 2 mL of YPAD + 2 % glucose for 40 hours at 30 °C. 1 mL of culture was added to 20 mL of YPAD + 2% glucose and incubated at 30 °C for 6 hours. Cells were washed and the transformation mix is added. Transformation mix contains 120 µL of 50% PEG, 18 µL of 1 M LiAc, and 25 µL of carrier DNA (salmon sperm). 0.1-0.5 µg of plasmid DNA was added to the cell mix, mixed gently, and incubated on ice for 1 minute. The mix was then incubated at 42 °C for 30-40 minutes, and then cooled on ice for 3 minutes. The cells were pelleted and resuspended in 200 µL of ddH2O, and 100 µL was spread on SD plates lacking Leucine. Plates were incubated at 30 °C for 3 days.

2.5.7 *E. coli* plasmid purification

E. coli cells for plasmid purification were grown overnight in 10 mL of LB media with the appropriate antibiotics at 37 °C and pelleted in the centrifuge at 2500 g for 5 minutes. Plasmid purification was carried out using a Wizard Plus SV Minipreps DNA purification kit.

2.5.8 Gene expression and protein purification

E. coli expression strain SoluBL21 carrying the plasmid of interest were grown overnight in 20 mL of LB with the appropriate antibiotics. 1 L of 2xYT medium was inoculated with the overnight culture in a 1:100 ratio and incubated in a 37 °C shaker until OD_{600} was between 0.500 and 0.600 then transferred to 18 °C shaker for 10 minutes. Isopropyl β-D-1-thiogalactopyranoside (IPTG) was added to a concentration of 250 µM and the culture was left overnight at 18 °C shaking at 200 rpm. The culture was removed, and the bacteria were pelleted at 3000 g for 5 minutes and resuspended in 50 mL of 50 mM Tris-HCl buffer pH8, 50 mM of glycine, 5% glycerol, 0.5 M of NaCl, 20 mM of imidazole (Buffer A), plus 10 mg of lysozyme and 1 M EDTA-free protease inhibitor tablet (cOmplete, Roche), then incubated on ice for 10 minutes. Cells were sonicated with a large tip for 2 minutes at 2 seconds on and 3 seconds off Amp1 40% pulses (Sonics Vibra cell) and were centrifuged at 30,000 g at 4 °C for 30 minutes. The supernatant was mixed with 2 mL of Ni-NTA that was washed with buffer and incubated on a rotator at 4 °C for 1 hour. The Ni-NTA protein mix is centrifuged at 1000 g for 2 minutes at 4 °C, the supernatant was removed, and the Ni-NTA pellet was washed twice with cold buffer A. 2.5 mL of buffer B (containing Buffer A plus an additional 240 mM of imidazole) was added to the Ni-NTA slurry and incubated on ice for 5 minutes. Ni-NTA and supernatant were separated by centrifugation at 1000 g for 5 minutes. Supernatant was filtered and buffer exchanged through PD MiniTrap G-10 columns into 3.5 mL of buffer A4 (20 mM of HEPES, 150 mM of NaCl at pH 7.5).

2.5.9 SDS-PAGE

25 µL of purified protein samples with 4 x LDS sample buffer dye and 60 mM of DTT were boiled at 70 °C for 5 minutes and cooled on ice for 5 minutes. The samples were loaded into a Sigma Aldritch TruPAGE gel which was run in TruPAGE buffer at 200 Volts, 400 mAmps for 60 minutes. 5-10 mL of Generon Quick Coomassie blue dye was used to develop the gel.

2.5.10 Western blotting and microsome preparation

Knockout *pep4* strains were grown in 2 mL of SC-Leu medium containing 2% glucose. The culture was grown overnight at 30 °C and 250 rpm. The culture was subsequently diluted 100-fold to an OD_{600} of 0.05 in SC-Leu supplemented with 2% glucose and cultured for 16 hours. Yeast was then harvested and grown for 24 hours in SC-Leu containing 2% galactose to induce recombinant protein production. Yeast cells were harvested by centrifugation and lysed in TES-B (0.6 M sorbitol in TE) using a Constant Systems cell disruptor at 35 kpsi and subsequently centrifuged at 11,000 *g* for 10 minutes at 4 °C. The supernatant was then transferred to a new tube and centrifuged at 125,000 *g* for 90 minutes at 4 °C. Finally, the pellet containing microsomes was resuspended with TEG buffer (20% glycerol in TE). Recombinant enzymes were detected by immunoblot analysis using α-FLAG M2 (Genscript) detected with SuperSignal West Pico Chemiluminescent Substrate from Thermo Fisher Scientific.

2.5.11 Whole cell culture assays

Knockout *pep4* yeast lines containing the appropriate vectors in glycerol stocks were used to inoculate 2 ml of SC-Leucine with 2% glycerol duplicates. Cultures were incubated for 48 hours in 30 °C, pooled, then pelleted at 3500 g for 5 minutes. The cultures were washed twice with 10 ml water, pelleting each time at 3500 g for 5 minutes. 2 ml of SC-Leucine with 2% galactose was added to each culture, and aliquoted 0.5 ml into four 1.5 ml Eppendorf tubes. To three of these aliquots, 0.5 mM of geraniol was added, and to one, an equal volume of analytical grade ethanol was added. The aliquots were incubated in a Eppendorf tube shaker for 24 hours at 30 °C. Products were collected by adding 200 µL of 3:1 ethyl acetate:acetone solution, vigorously vortexed, and centrifuged at top speed on a table top centrifuge for 10 minutes. 100 µL of the ethyl acetate and acetone layer was collected and filtered for GC-MS.

2.5.12 Catalytic activity assays and gas chromatography analyses

Catalytic activity of purified GES was assayed by running 100 µL reactions containing 4.43 µM of protein, 1 mM of DTT, 20 mM $MgCl_2$, 500 µM $MnCl_2$, 10% glycerol, 0.1 mM of geranyl pyrophosphate (GPP) and 10 mM of MOPS buffer at 30 °C for 1 hour. Catalytic activity of purified HGOA/B were assayed by setting up overnight reactions (16 hours) containing 7 µM protein, 2 mM $NAD(P)^+$, 0.5 mM substrate, 12.5 mM HEPES buffer and 25mM NaCl in 100 µL reactions incubated at 30 °C. Products were collected by mixing 100 µL of EtAc and centrifuging at top speed for 10 minutes and collecting 90 µL of the EtAc layer.

Samples were injected in split mode (2 µL, split ratio 5:1) at an inlet temperature of 220 °C on a Hewlett Packard 6890 GC-MS equipped with a 5973 mass selective detector (MSD), and an Agilent 7683B series injector and autosampler. Separation was performed on a Zebron ZB5-HT-INFERNO column (5% phenyl methyl siloxane; length: 35 m; diameter: 250 µm) with guard column. Helium was used as mobile phase at a constant flow rate of 1.2 mL/min and average velocity 37 cm/s. Two temperature runs were used for detection: A. After 5 minutes at 60 °C, the column temperature was increased to 180 °C at a rate of 2.5 K/min, then to 280 °C at 120 K/min, and kept at 280 °C for another 4 minutes; B. After 5 minutes at 40 °C, the column temperature was increased to 110 °C at a rate of 5 K/min, then to 280 °C at 120 K/min and kept at 280 °C for 4 minutes. A solvent delay of 5 minutes was allowed before collecting MS spectra at a fragmentation energy of 70 eV. Chemically characterised standards were used to identify compounds by retention time and electron impact spectra.

2.5.13 Heatmap of gene expression

Tissue specific gene expression analysis heatmaps were generated using the R-script package for heatmaps included in the gplot package, based on Log(FPKM+1) results. Heatmaps were generated without hierarchal clustering.

2.5.14 Primer list

Code	Gene	Sense	Sequence
LP001	NECA_8HGOB1	F	AAGTTCTGTTTCAGGGCCCGGCGAA ATCAGCAGAAACAGAG
LP002	NECA_8HGOB1	R	CTGGTCTAGAAAGCTTTAGTCGGCT TTCAGCGAAC
LP003	NECA_8HGOB2	F	AAGTTCTGTTTCAGGGCCCGGCGAA AACCCCAGAAACAGA

LP004	NECA_8HGOB2	R	CTGGTCTAGAAAGCTTTAAGCCTTCAATGATTTCTCAACA
LP005	NECA_8HGOB3 A,B,C	F	AAGTTCTGTTTCAGGGCCCGGCGAAATCAGTGAACGCC
LP006	NECA_8HGOB3 A,B	R	CTGGTCTAGAAAGCTTTAATCGGCTTTAGTCAGCGTATTC
LP007	NECA_8HGOB3 C	R	CTGGTCTAGAAAGCTTTAGTCGGGGAATACGCTGACTAAATC
LP008	NECA_8HGOB4	F	AAGTTCTGTTTCAGGGCCCGGCAAAATCAGCGGAGGC
LP009	NECA_8HGOB4	R	CTGGTCTAGAAAGCTTTAAAGTAATAATGAATTTTTGATGGTGTTGG
LP021	NECA_G8H1	F	ACCCTCACTAAAGGGCGGCCGCA-ACCATGGATTTCCTCACAATCTTCATTG
LP022	NECA_G8H1	R	GTCATCCTTGTAATCCATCGATACAAGAGCGATAGGGACAGC
LP023	NEMU_G8H1	F	ACCCTCACTAAAGGGCGGCCGCA-ACCATGGATTTCCTGACAATCTCCATTG
LP024	NEMU_G8H1	R	GTCATCCTTGTAATCCATCGATACAAGAGCGATAGGGACAGC
LP025	HYOF_G8H1	F	ACCCTCACTAAAGGGCGGCCGCA-ACCATGGATTCCATTACAACAGCAGTT
LP026	HYOF_G8H1	R	GTCATCCTTGTAATCCATCGATACGAAAGAGATGGGAACAGGTACA
LP027	CARO_G8H	F	ACCCTCACTAAAGGGCGGCCGCA-ACCATGGATTACCTTACCATAATATTAACT
LP028	CARO_G8H	R	GTCATCCTTGTAATCCATCGATACAAGGGTGCTTGGTACAGC
LP042	HYOF_GES	F	AAGTTCTGTTTCAGGGCCCGATGTGGATCCTCGATCGCATGCCT
LP046	NECA_GESA	F	AAGTTCTGTTTCAGGGCCCGACGACATCACCCAATTTGGC
LP047	NECA_GESB	F	AAGTTCTGTTTCAGGGCCCGACGACATCACCAACTTTGGC
LP048	NEMU_GES	F	AAGTTCTGTTTCAGGGCCCGACGACATCACCCAATTTGGC
LP049	HYOF_GES	F	AAGTTCTGTTTCAGGGCCCGATGCCATCAACGAGCCC
LP053	HYOF_HGOA_PopinJ	F	AAGTTCTGTTTCAGGGCCCGGCAGGAGTCATAACATGCAAAGC
LP054	HYOF_HGOA_PopinJ	R	ATGGTCTAGAAAGCTTTAGAATTTGATAATAACCTTCACACAGTC
LP055	HOYF_GES	R	CTGGTCTAGAAAGCTTTAGTCCAGCGGCGTGAAG

Chapter 3: Development of Virus-Induced Gene Silencing (VIGS) in *N. cataria*

3.1 Introduction

While the enzymes of the nepetalactone pathway have been characterized *in vitro* [23], [44], [65], the genes in this pathway had not been characterized *in vivo*. Notably, as described in Chapter 2, nepetalactone isomer content varies widely across *Nepeta* species and cultivars, and currently the mechanisms behind this chemical variation are not understood. *In planta* techniques are required to understand the mechanism behind this variation. Therefore, I developed VIGS, a gene knockdown technique, for *Nepeta.* VIGS is a transient gene silencing method that can be suitable for rapid functional genomics studies [77], [78]. Additionally, it has been adapted to a variety of plant families [45], [46], making it a more potentially reliable approach for a non-model plant with no precedence of genetic transformation.

3.1.1 Previous transformation methods in *Nepeta*

Although research in *Nepeta* phytochemistry has been an active topic since the elucidation of the iridoid compounds from *N. cataria* in McElvain *et al.* [22], little research into transformation methods has been reported for these species [59], [79], [80]. Three publications were found on *Nepeta* transformation, two in *N. cataria* [59], [80] and one for *N. pogonosperma* [79] and all three describe transformation via the hairy root method using *A. rhizogenes*. This is a plant specific bacterium that induces root growth while simultaneously introducing transfer DNA into the host plant. The roots produced via this infection are genetically and biochemically stable [58]. The reports on *A. rhizogenes* transformation in *Nepeta* were interested in the production of specialized metabolites that are produced in the roots. For example, Lee *et al.* [59] measured the production of rosmarinic acid produced in hairy root transformants. However, it is unlikely that hairy roots would be a good method to study nepetalactone production. As in Chapter 2 and reported in [11], [65], characterized *Nepeta* species produce very little to no nepetalactones in roots. The published hairy roots transformants do not report whether nepetalactones are present but since metabolite production in hairy root culture typically mimics the profile of natural roots [59], [79], [80], it is unlikely that hairy root culture would produce nepetalactones.

3.1.2 Virus-induced gene silencing

Early plant transformation methods serendipitously revealed plants' innate RNA mediated gene silencing defence systems, since transformed plants would sometimes silence endogenous genes targeted for over-expression [45], [46]. Soon this post-transcriptional gene silencing system was exploited for functional genomics studies [81], [82]. VIGS triggers this system to silence endogenous genes transiently. Viral sequences from known plant-infecting viruses such as potato virus X and tobacco rattle virus (TRV) are used to trigger the plant's post-transcriptional gene silencing system, which leads to the production of siRNA transcripts. Along with these viral particles, short (150-350 bp) homologous sequences of the targeted endogenous gene are used to guide the post-transcriptional gene silencing machinery to the target mRNA which cuts the mRNA inhibiting translation [78] (fig. 3.1-A).

While VIGS has been successfully adapted to a wide range of non-model plants, success often depends on the chosen viral vectors used on the host plant [45], [46]. Often, successful gene expression knockdown depends on the whether the viral vector comes from a virus that infects the target plant in nature. For example, vectors designed from barley stripe mosaic virus are often used on monocots and vectors coming from potato virus X are used in Solanaceae plants or closely related families [45]. Certain vectors have been shown to be successful in inducing VIGS in a wide range of plant families. The vector system adapted from tobacco rattle virus TRV, pTRV1 & 2, has been used successfully in a wide range of dicot plants and has been proven to lead to more systemic silencing [50], [54]–[57], [83]. Considering the lack of vectors adapted from viruses infecting the Lamiaceae plant family, a consistent and systemic VIGS vector used in related families within the Lamiales [57], [83], is a good candidate for use in *Nepeta*.

The pTRV vector system is made up of pTRV1, containing viral replication machinery (RdRP) and movement protein (MP) sequences, and pTRV2 containing virion formation (CP) sequences, a multiple cloning site which will house the endogenous homology sequence of the targeted gene for silencing [81], [82]. The 35S promoter induces constitutive expression of the various elements on these vectors, used in tandem during infection with *A. tumefaciens* strains. Other commonly used VIGS systems such as the cabbage leaf curl virus or potato virus x vectors contain similar structures [45], [46].

VIGS is commonly validated in a new system by targeting genes that have a visual phenotype, such as a leaf bleaching. Two commonly targeted genes used to assess gene silencing in leaf are magnesium chelatase subunit h (*ChlH*) [50], [56] and phytoene desaturase (*PDS*) [55], [81] (fig. 3.1-B). *ChlH* is a key step in the chlorophyll biosynthesis pathway, catalysing the insertion of Mg^{2+} into protoporphyrin IX, a precursor to chlorophyll [84]. Upon silencing *ChlH* produces a yellow leaf phenotype, as chlorophyll is absent from the affected areas. *PDS* is a key step in the carotenoid biosynthesis pathway [85]. Silencing *PDS* also produces a photo-bleached phenotype, as carotenoids are photo-protecting agents. These visual markers help to quickly indicate whether gene silencing has been successful, even before testing gene expression levels.

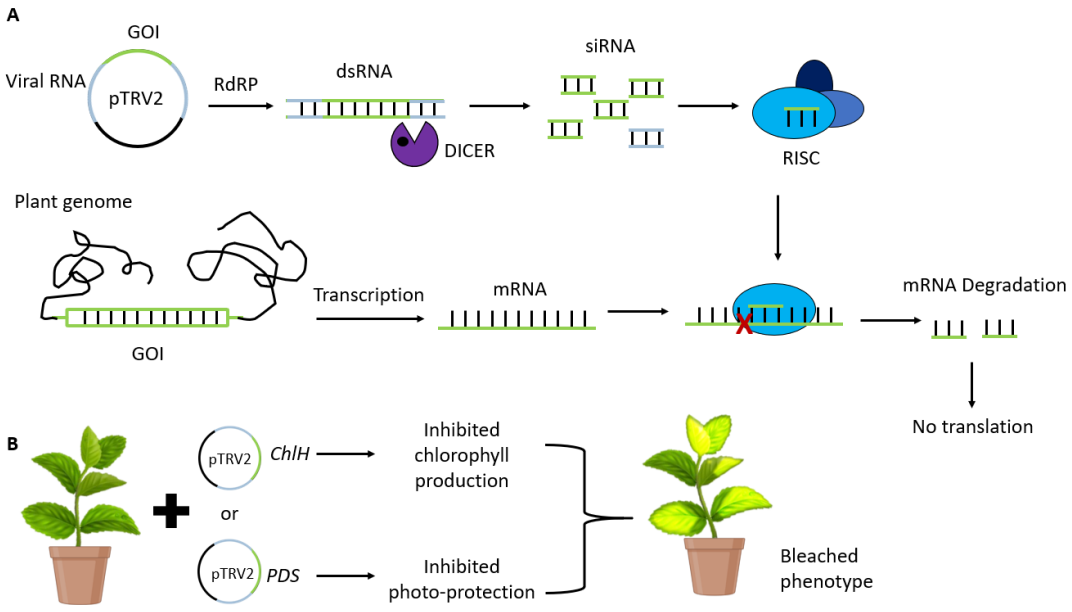


Figure 3.1: VIGS mechanism and bleaching phenotype

A) is a schematic representation of VIGS mechanism. A pTRV2 vector containing a fragment of the gene of interest and viral RNA is replicated by RNA-dependent RNA polymerases (RdRP) into double stranded RNA (dsRNA). This dsRNA is recognized by DICER, an RNA cleaving enzyme, which cleaves the dsRNA into fragments which act as small interfering RNA (siRNA). These are recognized by the RNA-induced silencing complex (RISC), an enzyme complex active in post-transcriptional gene silencing. The mRNA from the plant genome is recognized by RISC and degraded, resulting in no translation, and therefore no protein generated. B) is a simple schematic of VIGS targeting *ChlH* and *PDS* resulting in a similar bleached phenotype in the plant.

3.1.3 VIGS as a tool to study specialized metabolism

VIGS is a widely used tool to study secondary metabolism in non-model organisms [86]. For example, VIGS was adapted to the medicinal plant *C. roseus* to study MIA production [40], [50], [52], [53], [87]. As described in section 1.1.9, this class of alkaloids has several medical applications such as in cancer treatment (vinblastine) and opioid addiction (ibogaine). The biosynthetic pathways and *in planta* mechanisms involved in MIA production have been intensely studied. Liscombe *et al.* [50] used the VIGS system to test the metabolic function of 16-methoxy-2,3-dihydro-3-hydroxytabersonine N-methyltransferase (NMT). The authors infected seedlings, and in parallel to the specialized metabolism gene targets, infected seedlings with pTRV2-*ChlH* silencing cassette to use the bleaching phenotype as a marker for when to harvest seedlings. They show a decrease in the final product of the pathway, vindoline, and an increase in intermediate metabolites such as 16-methoxy-2,3-dihydro-3-hydroxytabersonine upon silencing NMT.

VIGS has also been used extensively in functional genomics studies in *S. lycopersicum* (tomato), including publications focusing on the secondary metabolism [47]–[49], [55], [81]. In Orzaez *et al.* [49], the authors developed a visual marker to study tomato metabolism to overcome the incomplete VIGS effect in tomato. The authors used a transgenic tomato line which over-expressed anthocyanins leading to purple fruits and leaves. The authors created in-tandem double knockout vectors targeting the over-expressed anthocyanin genes (Delila and Rosea1, abbreviated to Del-Ros) and a new gene of interest. The silenced tissue would lose its purple phenotype, allowing for easy identification of affected tissue. As a proof of concept, they used their tandem silencing method to target *TomloxC*, *PDS*, and *ODO1*, all genes involved in the production of different volatiles in tomatoes. Upon silencing of *TomloxC*, they found that affected tissues had a reduction of C-6 aldehydes derived from lipid 13-hydroperoxidation; silencing *PDS* yielded a decrease in geranial and 6-methyl-5-hepten-2-one volatiles; finally, silencing a *MYB* transcription factor involved in benzenoid metabolism, ODO1, yielded an increase in benzaldehyde in the fruit. This VIGS method has then been extensively applied in studying tomato metabolism, using the transgenic high-anthocyanin tomato line to achieve the visual marker silenced phenotype [47], [48].

As introduced in Chapter 2, Lamiaceae is a well-known plant family containing many economically important plants, especially valued for their specialized metabolites. While there is plenty of work exploring the metabolism of members of this family, only one

publication of VIGS in Lamiaceae was published by the time the work in this thesis was carried out. Misra *et al*. [54] adapted VIGS to *O. basilicum* using the pTRV bipartite vector system to characterize two cytochrome P450 enzymes involved in triterpene biosynthesis. When targeting *ChlH*, the authors found that vacuum infiltration at the cotyledon stage would result in the bleaching phenotype in the first leaf pair 73% of the time, allowing them to confidently collect the first leaf pair of infected cotyledons for analysis. Using this infiltration and selection method, they identified *in vivo* two new CYP716A cytochrome P450 monooxygenases involved in triterpene metabolism of *O. basilicum*.

Given VIGS adaptability to non-model organisms and its relative speed as a functional genomics tool, this system is a good candidate tool to characterize the iridoid metabolism in *Nepeta spp.* Thus, this chapter describes the work carried out to adapt this system and test its viability in studying *N. cataria* metabolism.

3.2 Results and Discussion

3.2.1 Identifying *ChlH* and *PDS* in *N. cataria* and *N. mussinii*

I used the same homology-based search method detailed in Chapter 2 to identify *N. cataria ChlH* and *PDS* genes. Using these sequences, I used tBLASTn on a *N. cataria* transcriptome provided by the Mint Genome Project to identify *N. cataria ChlH*. Two contigs with high degree of similarity to *C. roseus ChlH* [50] (77%) were identified and one contig from *N. mussinii*. The previously listed sequence for *PDS* from *A. thaliana*, obtained from TAIR (the *Arabidopsis* information resource) was used to run BLAST into the *N. cataria* transcriptome to identify *N. cataria* and *N. mussinii PDS*. One *N. cataria PDS* sequence with about 77% sequence similarity was identified.

ChlH

Species	Name	BLAST seed	E-value	% Identity	Query Coverage
N. cataria	*ChlH1*	*C. roseus ChlH*	0.00E+00	77.20%	99.81%
N. cataria	*ChlH2*	*C. roseus ChlH*	0.00E+00	77.10%	99.81%
N. mussinii	*ChlH*	*C. roseus ChlH*	0.00E+00	76.90%	99.81%

PDS

Species	Name	BLAST seed	E-value	% Identity	Query Coverage
N. cataria	*PDS*	*A. thaliana PDS*	0	76.90%	87.01%
N. mussinii	*PDS*	*A. thaliana PDS*	0	77.20%	87.01%

Table 3.1: Visual marker gene homologues candidates

List of candidates selected after tBLASTn of *C. roseus ChlH* and *A. thaliana PDS* sequences to *N. cataria* and *N. mussinii* transcriptomes. Enzymes with multiple candidates are listed in order of highest to lowest sequence identity similarity. BLAST seed is the query sequences used for the homology search. E-value denotes the number of hits that could be found by chance. % Identity denotes the percentage of amino acids that are identical to the BLAST query. Query coverage is the percentage the query length aligns with the BLAST hit.

3.2.2 Designing VIGS targets

VIGS relies on the post-transcriptional gene silencing machinery found in plants [55], [78], [81], [82]. Thus, the full gene is not needed to create the 21 nucleotide fragments used by the post-transcriptional gene silencing machinery to identify and silence the targeted gene. A region of 150-350 bp fragments are selected from the gene's CDS or untranslated region (UTR) for target design. This region must not have matching nucleotide sequences to other transcriptionally active genes. I selected regions of around 200-350 bp in the CDS region and used this as a query for running BLAST on the *N. cataria* transcriptome. If the selected region contained sequence fragments of at least 19 matching base-pairs, they were not used for VIGS to avoid off-target silencing of the matching gene (fig. 3.2). Unique target sequences were then cloned into the pTRV2 vector. It is interesting to note that *N. cataria* is reported to be an autotetraploid [65] and therefore there is usually at least two copies of each gene in the transcriptome. However, these genes tend to have highly conserved sequences and functions, and targeting the CDS regions often results in highly overlapping sequences, by-passing any redundancy that may lead to an incomplete silencing phenotype [88], [89].

ChlH gene target in *N. cataria*

Species	Name	BLAST seed	E-value	% Identity	Query Coverage
N. cataria	*ChlH1*	*C. roseus ChlH*	6.13E-180	99.10%	100.00%
N. cataria	*ChlH2*	*C. roseus ChlH*	4.77E-176	98.60%	100.00%

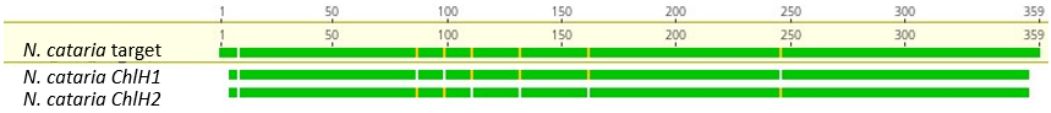


Figure 3.2: BLAST results of candidate target sequence of *ChlH* into the *N. cataria* transcriptome

A 355 bp region of the *N. cataria ChlH*1 CDS was used as a query for BLAST against the *N. cataria* transcriptome. The two listed contigs correspond to the *ChlH* homologues in *N. cataria* as found in table 3.1. No other sequences containing more than 19 matching base pairs were found.

3.2.3 *N. cataria* propagation methods

Two main methods of plant propagation were considered for VIGS infection: propagation via cuttings and via seedlings. *N. cataria* and *N. mussinii* are common garden plants and many protocols for propagating cuttings have been developed [90]. Most of these guides recommend cutting the plant just below the node from where leaf pairs emerge, removing the leaves from the bottom two nodes, and propagating in water. I tested this method, as well as rooting in coconut husk and in soil. Common garden rooting powder with indole-3-butyric acid was also tested. I found that plants rooted in water were most efficient, with most cuttings producing roots within 10 days, and that rooting hormone had no impact on rooting. Seeds were obtained by harvesting seeds from our plant stocks. These were grown in soil and allowed to grow until 2 leaf pairs had grown, about 3-4 weeks.

When considering which plants to use for infection, additional advantages and disadvantages were considered: while seedlings provide a precise developmental stage that can be targeted, which aids in recognizing leaf pairs post-infection and is often preferred in VIGS research, cuttings are genetically identical to one another. As discussed in section 2.1.1, nepetalactone content varies by *N. cataria* cultivar and the causes behind that variation are not fully known, although they were somewhat explored via transcriptome analysis in Lichman *et al.* [65]. Thus, working with genetically identical plants may insure against any genetic-based nepetalactone isomer variation between experimental replicates. However, metabolic profiles of seedlings were not assessed in this thesis, and therefore it remains

unknown how much metabolic variation occurs between generations. Furthermore, during propagation I noticed that *N. cataria* consistently rooted faster and more reliably than *N. mussinii.* For this reason, and reasons discussed in section 3.2.6, I focused on *N. cataria* VIGS for downstream studies.

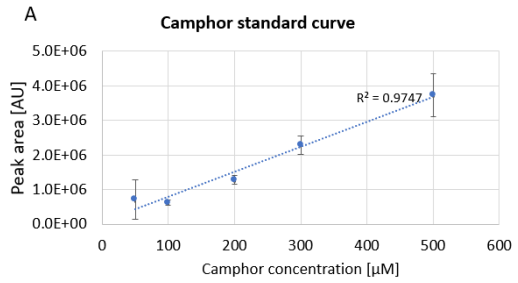
3.2.4 Optimizing nepetalactone extraction

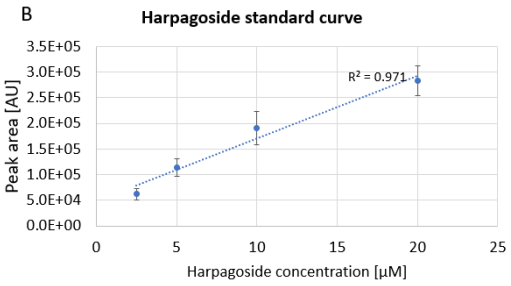
The nepetalactone extraction protocol was based on the extraction methods described in Lichman *et al.* [65] and was optimized for small tissue amounts (20-50 mg). This protocol was also streamlined so that many replicates, which are required for robust genetic studies, could be performed. Optimization of this extraction procedure included identifying GC-MS and LC-MS internal standards and their concentrations, testing the tissue *maxima* and *minima* for extraction and optimizing the sample workup using a solid phase extraction column. Furthermore, to understand nepetalactone extraction efficiency and nepetalactone concentration *in planta*, *cis-cis* nepetalactone was spiked into the HH01 cultivar (fig. 3.4) tissue sample extractions and then the recovery of this standard was quantified.

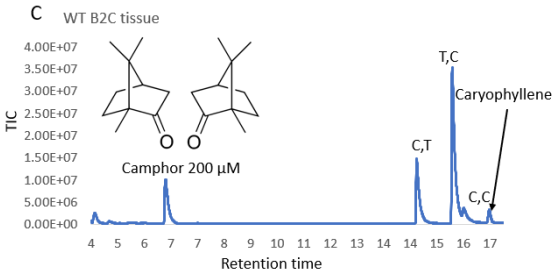
From this data, I selected camphor and harpagoside as internal standards, for their retention times relative to my compounds of interest, and in the case of harpagoside, its iridoid-ring base structure. Camphor was previously used as an internal standard in *in vitro* assays in Lichman *et al.* [64]. After generating a standard curve, I selected a concentration of 200 µM camphor and 10 µM harpagoside for WT (wildtype) tissue (fig. 3.3 A-D). A 1:20 ratio of tissue weight-to-methanol volume was selected for extraction, identifying 20 mg of tissue to 400 µL of methanol volume as being the lowest possible weight and volume for reliable extraction for both GC-MS (fig. 3.3-E) and LC-MS analysis. For further experiments, 20 mg of tissue was set as the minimum. Finally, a second wash step aided in extracting a higher concentration of nepetalactones (fig 3.3-F).

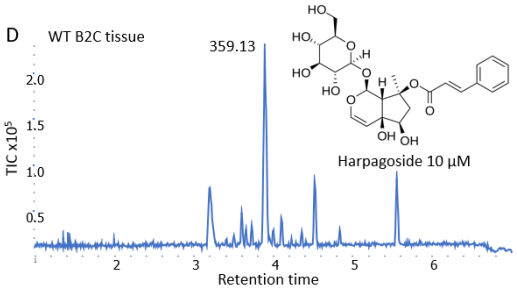
While the addition of an internal standard of *cis-cis* nepetalactone to HH01 tissue revealed the extent of loss during extraction procedure inherent to the protocol, it also demonstrated the variability amongst extraction replicates. While, as shown in figures 3.3 G-I, *cis-cis* nepetalactone content is expected to vary depending on when it was added to the extraction protocol as well as its concentration, *cis-trans* nepetalactone inherent to HH01 should be present in the same relative amount per extraction replicate; however, some variability was present (+/- 46.3%). This suggested that while this spiking experiment can aid in determining

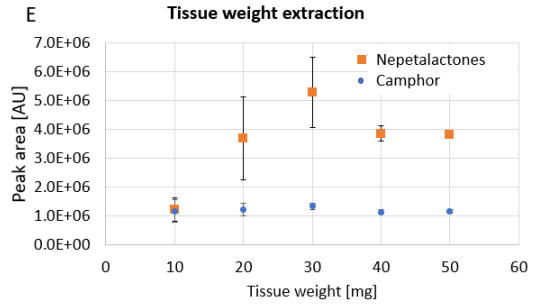
the extracted concentration of nepetalactones in any given tissue sample, nepetalactone concentration variability amongst biological replicates may obscure the extraction efficiency per sample. I decided to rely on camphor as a measure of extraction efficiency rather than the extracted concentration of nepetalactone.

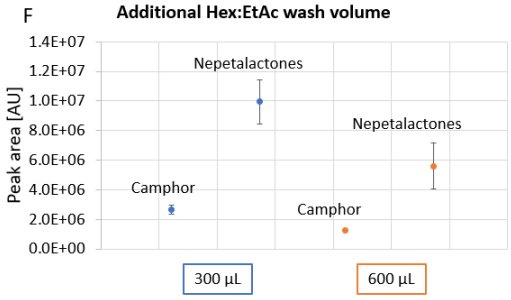

A
Camphor standard curve
Peak area [AU]
5.0E+06
4.0E+06
3.0E+06
2.0E+06
1.0E+06
0.0E+00
R² = 0.9747
0
100
200
300
400
500
600
Camphor concentration [µM]

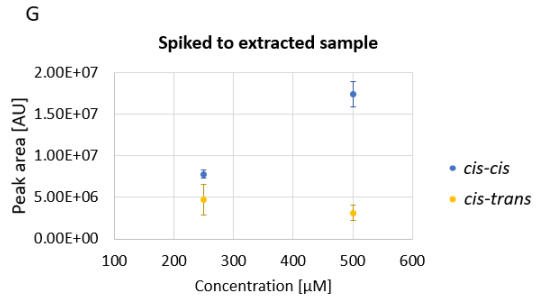

B
Harpagoside standard curve
Peak area [AU]
3.5E+05
3.0E+05
2.5E+05
2.0E+05
1.5E+05
1.0E+05
5.0E+04
0.0E+00
R² = 0.971
0
5
10
15
20
25
Harpagoside concentration [µM]

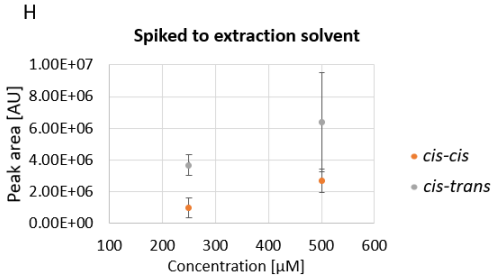

C
WT B2C tissue
TIC
4.00E+07
3.50E+07
3.00E+07
2.50E+07
2.00E+07
1.50E+07
1.00E+07
5.00E+06
0.00E+00
Camphor 200 µM
T,C
Caryophyllene
C,T
C,C
4
5
6
7
7
8
9
10
11
12
13
14
14
15
16
17
Retention time


D
WT B2C tissue
359.13
TIC x10⁵
2.0
1.5
1.0
0.5
Harpagoside 10 µM
2
3
4
5
6
Retention time


E
Tissue weight extraction
Peak area [AU]
7.0E+06
6.0E+06
5.0E+06
4.0E+06
3.0E+06
2.0E+06
1.0E+06
0.0E+00
Nepetalactones
Camphor
0
10
20
30
40
50
60
Tissue weight [mg]


F
Additional Hex:EtAc wash volume
Peak area [AU]
1.4E+07
1.2E+07
1.0E+07
8.0E+06
6.0E+06
4.0E+06
2.0E+06
0.0E+00
Nepetalactones
Nepetalactones
Camphor
Camphor
300 µL
600 µL


G
Spiked to extracted sample
Peak area [AU]
2.00E+07
1.50E+07
1.00E+07
5.00E+06
0.00E+00
cis-cis
cis-trans
100
200
300
400
500
600
Concentration [µM]


H
Spiked to extraction solvent
Peak area [AU]
1.00E+07
8.00E+06
6.00E+06
4.00E+06
2.00E+06
0.00E+00
cis-cis
cis-trans
100
200
300
400
500
600
Concentration [µM]

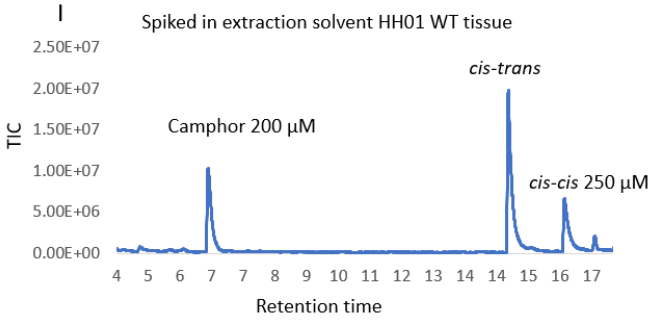

I
Spiked in extraction solvent HH01 WT tissue
TIC
2.50E+07
2.00E+07
1.50E+07
1.00E+07
5.00E+06
0.00E+00
Camphor 200 µM
cis-trans
cis-cis 250 µM
4
5
6
7
7
8
9
10
11
12
13
14
14
15
16
17
Retention time

Figure 3.3: Optimization of nepetalactone extraction from *N. cataria* tissue

Series of extraction optimization graphs. Peak area in graphs was calculated by obtaining the area under the curve of a particular metabolite of interest. AU denotes arbitrary units. TIC denotes total ion content. Averages of 3 replicates were plotted and error bars were plotted using the standard deviation. A) and B) show the standard curves for the internal standards camphor (for GC-MS analysis) and harpagoside (for LC-MS analysis). Standards were added to the initial extraction buffer and processed with 30 mg of WT *N. cataria* B2C tissue. C) and D) show example GC-MS spectra of the respective standards at the chosen working concentrations. E) shows various tissue weights tested to find a minimum weight for reliable analysis. F) describes the effect of an additional final EtAC:Hex wash volume of 0.5x (300µL) or 1x (600µL) on nepetalactone content extracted from the SPE column. G) and H) denote the effect of spiking *cis-cis* nepetalactone to HH01 WT tissue, a *cis-trans* nepetalactone exclusive producer. In G) *cis-cis* nepetalactone was spiked to the sample after extraction. In H) *cis-cis* nepetalactone was spiked to the extraction solvent pre-extraction. I) is an example GC-MS spectra of a sample from H).

3.2.5 Nepetalactone content by cultivar and leaf age

As described in 2.1.1, our cultivar collection of *N. cataria* contains individual plants presenting different nepetalactone isomer profiles. For the initial VIGS testing, I chose three varieties (B2E, B2C and HH01) presenting different metabolic profiles (fig. 3.4). Given that nepetalactone content is mostly found in leaves, which have high trichome density and the bleaching phenotype of both *ChlH* and *PDS* silencing will yield most tissue from leaves (as opposed to other tissues), I focused on characterizing the nepetalactone content by leaf age and section.

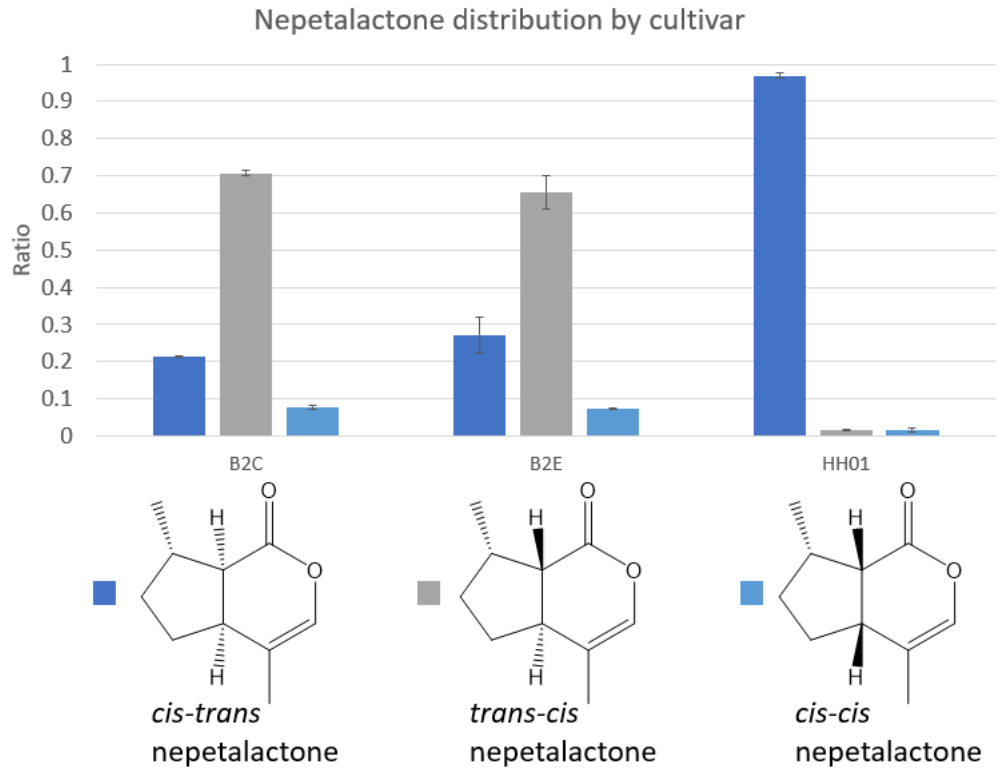


Figure 3.4: Isomer profile of *N. cataria* cultivars used in this chapter

N. cataria cultivars used in this chapter and further analysis. Cultivar name denotes our in-house naming system. Metabolic profile percent total nepetalactones represent the isomer profile breakdown in ratios of total TIC area of nepetalactone peaks present. Error bars denote standard deviation.

I measured the nepetalactone content of our in-house cultivars B2C, B2E and HH01. As confirmed in our records, B2C and B2E both produce *cis-cis* nepetalactone, *trans-cis* nepetalactone and *cis-trans* nepetalactone, while HH01 only produces *cis-trans* nepetalactone (fig. 3.4). B2C, B2E were then chosen to carry out downstream studies on the iridoid pathway, while HH01 was used for standardization experiments listed in 3.2.4.

Leaf age was determined by size and location on the cuttings since the cuttings themselves were of indeterminate age. Leaf age was divided into very young leaf, young leaf, old leaf and very old leaf. Very young leaves were those found at the tips or nodes of the cuttings and were less than 1 cm long. Young leaves were between 1 cm and 1.5 cm long and were found at tips or nodes. Old leaves were those between 1.5 cm and 3 cm long, found at nodes with branches and were not at the tip of branches. Very old leaves were those larger than 3 cm long. For B2E, younger leaves tended to yield more nepetalactone content, with higher

nepetalactone content variability in younger leaves, even though replicates were from the same pooled tissue samples (fig. 3.5-A). Finally, a pooled sample incorporating equal weights of the different ages were also extracted, to understand the average nepetalactone content for each variety regardless of leaf age. Differences between age groups were not statistically significant in either cultivar.

Leaf sections were obtained from all leaf ages and leaves were cut into three equal parts: base, middle and tip. For both B2C and B2E, the mid-section revealed a higher nepetalactone content than either the base or the tip of the leaves. There was no significance between B2C leaf sections. In B2E, the middle section was found to be significantly different as compared to the base and tips.

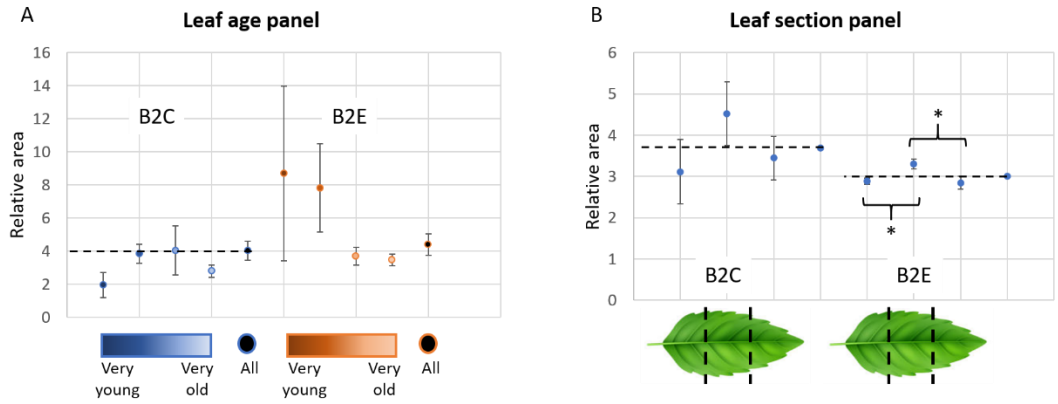


Figure 3.5: Nepetalactone content variation by leaf age and section

Relative area was calculated by dividing the nepetalactone area under the curve by the 200 µM camphor internal standard area under the curve. In A) darker to lighter hues of blue and orange denote younger to older leaf tissue. Leaf tissue was collected, flash frozen and ground into a fine powder. 30 mg of ground tissue in 3 replicates for each age group was used to extract nepetalactones. In All, designated by the black circle and the dotted line, equal weights of age group tissue powder were added, thoroughly homogenized, and 3 replicates of 30 mg tissue was processed. In B) the leaf diagram indicates the section of the leaf (base, middle or tip) processed. Leaves were cut fresh and each section flash frozen separately, and finally pulverized. Average of all sections is the right most dot from each cultivar with a dotted line. Error bars denote standard error in both A) and B). Asterisks on brackets indicate statistically significant difference (t-test p-value < 0.05).

Although there is variation amongst leaf ages and sections in both B2C and B2E leaves, and in some cases that difference is significant, no section or age is particularly devoid of nepetalactones.

3.2.6 Testing infection methodologies

VIGS infection methods vary widely depending on the plant being infected and success depends on factors such as plant hardiness, stem tissue type and as mentioned before, vector system [45], [46]. Common methods include vacuum infiltration, crown drenching and wounding the plant stem or nodes and soaking with the *A. tumefaciens* containing the vector system (fig. 3.6).

I tested several methods to establish VIGS in *N. cataria*: pinching the stem and allowing a droplet to be absorbed into the resulting wound on the stem (3.6-Droplet, [50]) crown drenching (3.6-Drenching, [50]), vacuum infiltration (3.6-Vacuum, [54]), wounding the nodes and submerging the plant (3.6-Submersion, [66]), and wounding the stem and wetting the nodes with *A. tumefaciens* (3.6-Wetting, modified from [81], [82]), as well as attempting to infiltrate the stem via applying pressure with a syringe (3.6-Syringe). *A. tumefaciens* containing pTRV2-*ChlH* was used to establish infiltration efficiency and the *chlh* bleaching phenotype was used as a visual marker for successful gene silencing. For each of these methods, I roughly assessed the time taken to infiltrate a set of 28 plants and the infection efficiency. Amongst these methods, crown drenching and drop pinching as reported in [50] yielded no plants with bleaching. Using a syringe to infiltrate *A. tumefaciens* into the nodes of the plant resulted in close to 100% plant death.

Vacuum infiltration did lead to bleaching in about 30% of plants, however, the entire process was laborious and lengthy, taking about 2 hours for 28 plants and was therefore unsuited for large-scale analyses. Wounding the plant and fully submerging into *A. tumefaciens*, as in Palmer *et al.* [66] was successful, took about 40 minutes per set, yielding about 25-30% bleaching amongst the plant set. However, the method that proved to be the fastest (20-30 minutes per set), best for scaling up and yielded consistent bleaching in about ~40% of plants was a modification of the method reported in Liu *et al.* [81], where a sterile toothpick was dipped into concentrated *A. tumefaciens* in infiltration solution and used to wound the stem of the plant, and finally used to coat the wounds with the infiltration solution.

For all these methods, plant survival was vastly improved when using cuttings as opposed to seedlings. Given the speed of propagation and the genetic consistency of cuttings, these were chosen for further analysis described in Chapter 4. Methods to infect *N. mussinii* were also tested. However, *N. mussinii* tended to have a poorer survival and bleaching rate as compared to *N. cataria*. For this reason, subsequent work characterizing the pathway *in vivo*, as explored in Chapter 4, was carried out in *N. cataria.*

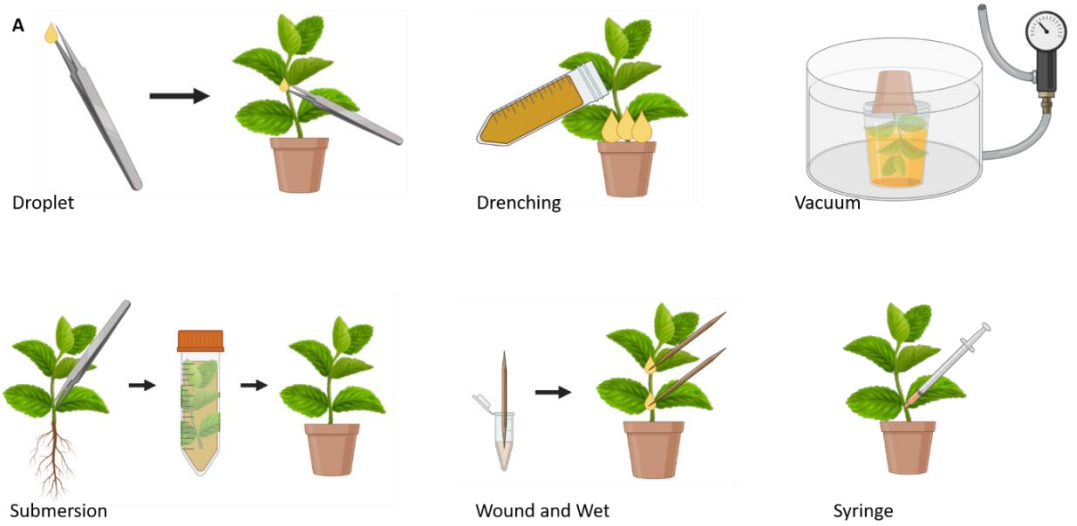


Figure 3.6: Infection methods

The diagrams of infection methods are listed. Illustrations generated in Biorender.

3.2.7 Silencing *ChlH* and *PDS* in *N. cataria*

Initially, *ChlH* was chosen as the visual marker on which to test for VIGS efficiency, as in section 3.2.6. Targeting *ChlH* revealed that although *N. cataria* was susceptible to VIGS, gene knockdown was not systemic. After allowing new leaves to grow post infection, the same individual plant would grow both green and bleached leaves, with the vast majority of leaves having both green and bleached tissue (fig. 3.7).

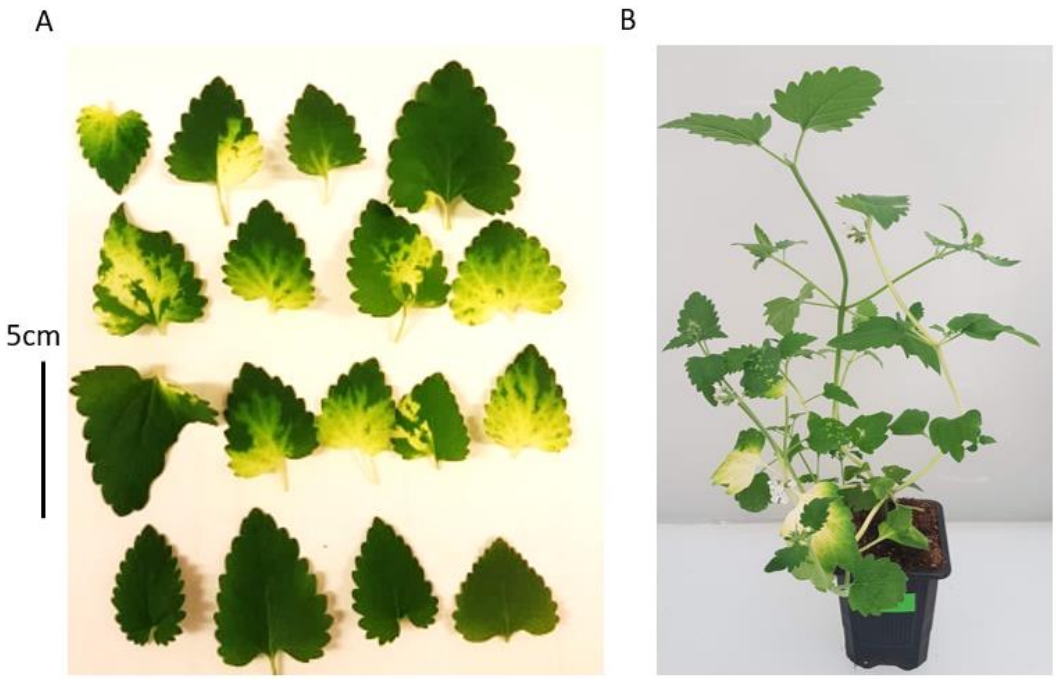


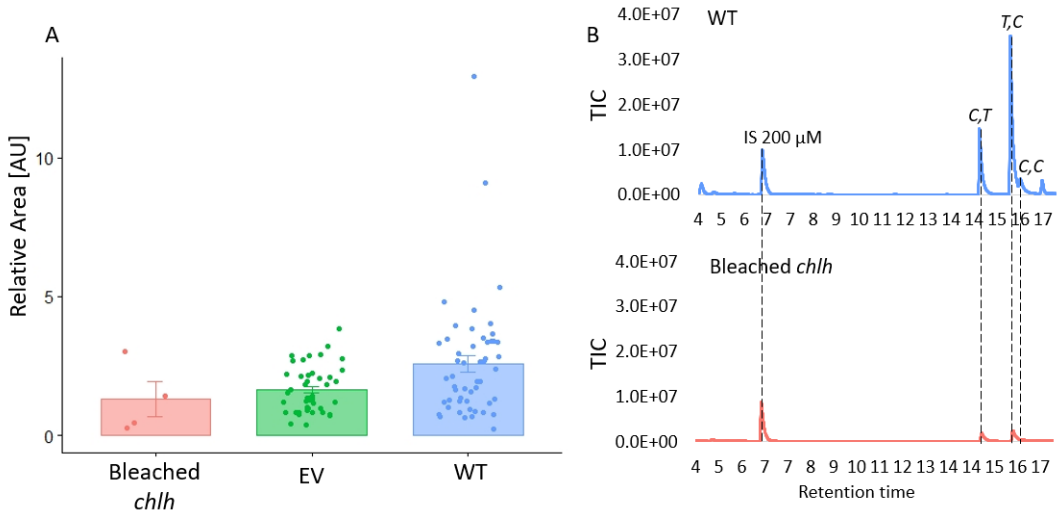
Figure 3.7: Leaves collected from a VIGS infected plant targeting *ChlH*

VIGS efficiency in *N. cataria.* A) are leaves taken from the same individual plant 6 weeks post infection. The top 3 rows show all the leaves with some bleached tissue, the bottom 4 leaves have no bleached tissue. B) shows a different plant with the *ChlH* bleaching phenotype 4 weeks post infection.

The metabolite profile of these bleached regions was compared to uninfected WT leaves and were found to have an overall reduction of nepetalactones produced in those leaf tissues, but this decrease was not significant (fig. 3.8-A,B). Given *ChlH*'s important role in chlorophyll biosynthesis, as well as the perturbance on the primary metabolism of *ChlH* mutants reported in the literature [84], this slight decrease in secondary metabolism may be a downstream effect of a compromised photosynthesis. I also wanted to understand the effect of the infection itself on nepetalactone content. I compared nepetalactone content in plants having been infected with pTRV2-*ChlH*, empty pTRV2 (EV) and uninfected leaves (fig. 3.8-A). After 4 weeks, two leaves of approximately the same size and age from each plant were taken for analysis. For the *ChlH* plants, only bleached tissue was analysed. From this experiment, nepetalactone content was significantly reduced in the EV pTRV2 *A. tumefaciens* infected plants as compared to non-infected WT, indicating the infection itself may influence the plant's ability to produce nepetalactone. However, plants infected with pTRV2-*ChlH* containing *A. tumefaciens*, while they had a reduced level of nepetalactone content, the

decrease was not statistically significant. Further experiments, as described in figure 3.8-C, also reveal that the nepetalactone content in the bleached tissue of pTRV2-*ChlH* infected plants is not significantly different compared to the nepetalactone content in plants infected with an empty pTRV2 vector. Given the decrease in nepetalactone content in pTRV2-*ChlH* infected plants, I decided to lower the concentration of Camphor to 50 µM to better match the height of nepetalactone peaks in *ChlH* silenced tissue in further experiments (including fig. 3.8-C).

Furthermore, I tested the efficiency of *PDS* as a visual marker as compared to *ChlH*, as well as the effect upon silencing on nepetalactones content. To test this, I compared the nepetalactone content from plants infected with *A. tumefaciens* containing pTRV2-*ChlH*, pTRV2-*PDS* and pTRV2-EV. From this experiment, only 5/36 of pTRV2-*PDS* infected plants showed the bleached phenotype, as well as only an average of 20 mg of bleached tissue were recovered per plant. In contrast, in this experiment, 15/36 of pTRV2-*ChlH* infected plants displayed the bleached phenotype, with an average of 45.4 mg of bleached tissue per plant. Furthermore, both *ChlH* and *PDS* silenced bleached tissue showed a non-significant decrease in nepetalactone content as compared to EV (fig. 3.8-C, D). The potential mechanism behind these difference is further discussed in section 3.2.13. Based on these results, further work was carried out with *ChlH* as a visual marker.



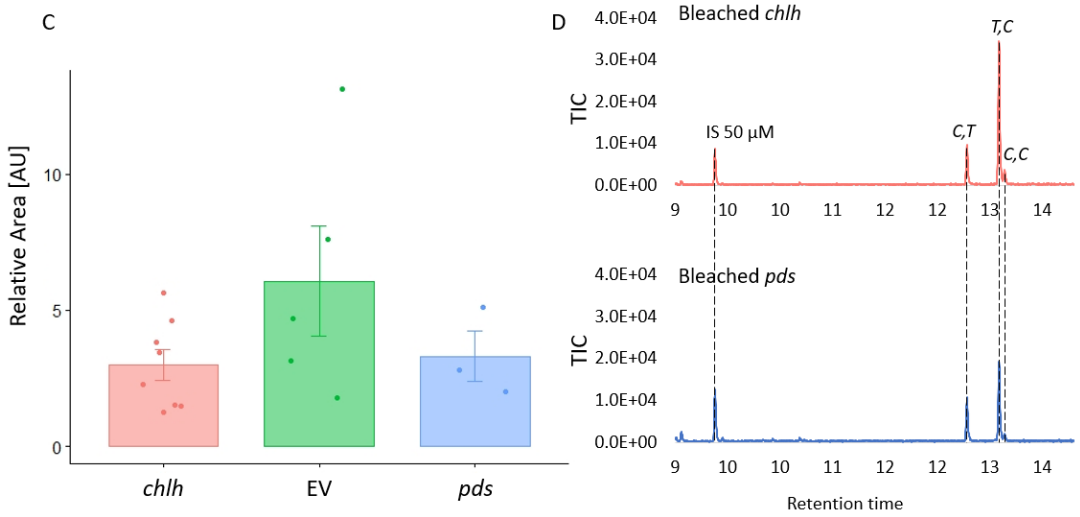


Figure 3.8: Nepetalactone content in VIGS infected tissues

Nepetalactone content of *N. cataria* B2C cultivar leaf tissue under various VIGS conditions. A) and C) are box plots indicating the average relative nepetalactone area. Relative area was calculated by dividing the total area under nepetalactone peaks by the area under the camphor peak. Each dot represents a replicate. Error bar indicates standard error. In A) EV, and WT replicates were two individual leaves from each surviving plant in an experimental replicate, in Bleached *chlh* each replicate was the pooled bleached tissue per affected plant. B) are a representative MS spectrum of WT and Bleached *chlh* samples. GC method is expanded upon in method 3.5.7.A. *C,T* is *cis-trans* nepetalactone, *T,C* is *trans-cis* nepetalactone and *C,C* is *cis-cis* nepetalactone. For A) and B) a final concentration of 200 µM camphor was used as an internal standard. In C) EV replicates were an individual leaf from each surviving plant in an experimental replicate, *chlh* and *pds* replicates were the pooled bleached tissue per affected plant. D) are a representative MS spectrum of Bleached *chlh* and Bleached *pds*. GC method is expanded upon in method 3.5.7.B. For C) and D) a lower concentration of 50 µM was added to adjust for the lower nepetalactone heights upon infection.

3.2.8 Design of dual knockdown targets to include a visual marker for metabolic studies

In Yamamoto *et al*. [52], the authors designed a dual-targeting VIGS system to study the secondary metabolism of *C. roseus*. In this system, the pTVR2 vector includes both the fragment of a gene that will produce a visual phenotype such as bleaching, and a fragment of their gene of interest (fig. 3.9). Tissue affected by the bleaching phenotype also exhibited a down regulation in expression of their gene of interest. Based on the VIGS target design

method reported in Yamamoto *et al*. [52], I designed a dual knockdown pTRV2 vector for *ChlH* and *GES* to be targeted in tandem (referred to as pTRV2-*ChlH-GES*). The *GES* target sequence was selected using the identified CDS sequence in Chapter 2 and verified against the transcriptome as in section 3.2.1. I also designed a single target pTRV2-*GES* vector only targeting *GES* and not *ChlH* to test for significant differences between metabolite concentration in tissue from plants infected with *A. tumefaciens* containing pTRV2-*ChlH-GES* and pTRV2-*GES*. *GES* is an ideal candidate in which to test the effectiveness of this approach, as it is the biosynthetic step that branches out from the MEP pathway into the specialized iridoid metabolism, as published before [11], [65] and characterized in Chapter 2. Silencing this gene should yield significantly decreased levels of nepetalactone content in the tested tissue.

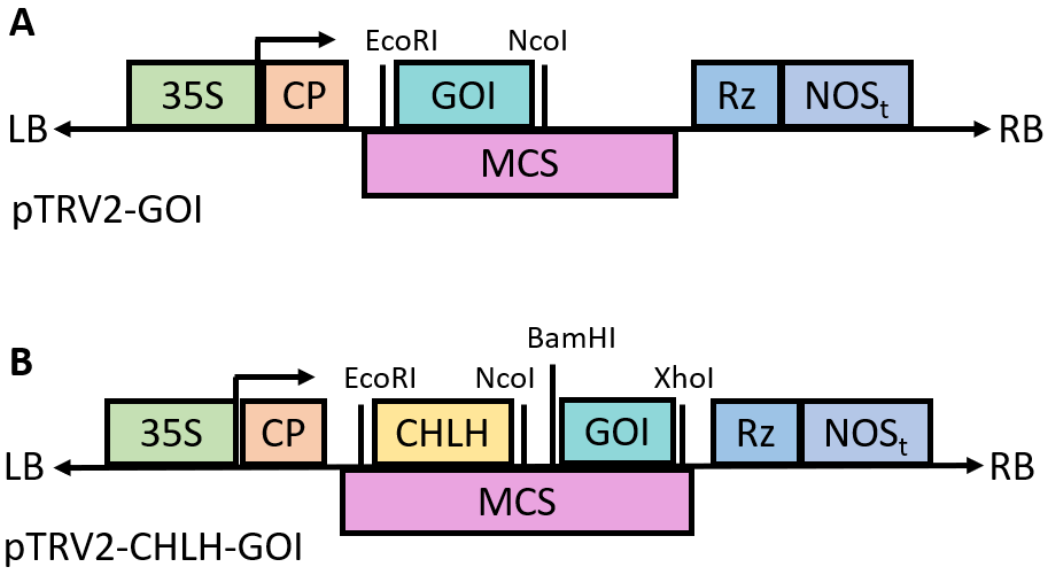


Figure 3.9: Schematic of pTRV2 system for VIGS and cloning strategy

A fragment of the gene to be silenced (gene of interest, GOI) is cloned into the MCS (multiple cloning site) of pTRV2 between BamHI and XhoI. To target two or more genes, one sequence is cloned into the MCS of pTRV2 between the EcoRI or the XmaI restriction target sites and the second (e.g *ChlH*) between BamHI and XhoI. 35S, CaMV promoter; CP, coat protein; MCS, multiple cloning site; Rz, self-cleaving ribozyme. Figure adapted from [65].

3.2.9 Metabolite and expression data from *ChlH-GES* and *GES*

Tissue collected from plants infected with pTRV2-*ChlH-GES* was limited to bleached leaf tissue from all affected plants, while tissue from plants infected with pTVR2-*GES* containing *A. tumefaciens* came from new leaves near the infection sites, collecting two leaves from randomly selected 10 plants from a 28-plant set. Silencing *ChlH* and *GES* in-tandem (*chlh-ges* in figure 3.10) and collecting the bleached part of the plant led to a statistically significant decrease of nepetalactone content (fig. 3.10-A), compared to the control in which only *ChlH* (chlh in figure 3.10) was silenced. For the samples in which only *GES* is targeted for silencing (*ges* in figure 3.10), while there is a small reduction in nepetalactone content, it is not statistically significant compared to *chlh*, presumably because the silenced tissue could not be selectively harvested. Given these results, targeting a visual marker gene such as *ChlH* along with an "invisible to the eye" metabolic pathway gene, proved to be the optimal strategy for utilizing VIGS in *N. cataria* for further *in vivo* metabolic characterization studies.

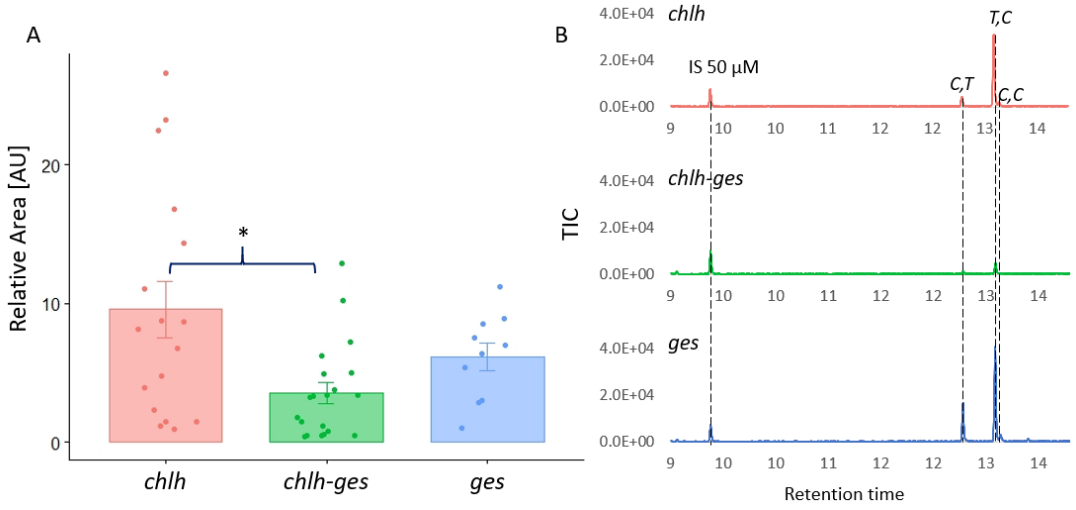


Figure 3.10: Comparison of *ges-chlh* double knockdown and *ges* single knockdown

Nepetalactone content of *N. cataria* B2C cultivar leaf tissue under various VIGS conditions. A) box plots indicating the average relative nepetalactone area. Relative area was calculated by dividing the total area under nepetalactone peaks by the area under the camphor peak. Each dot represents a replicate. Error bar indicates standard error. In A) *chlh*, *chlh-ges* and *ges* each replicate was the pooled bleached tissue per affected plant. B) are a representative GC-MS spectrum of *chlh*, *chlh-ges* and *ges* samples. GC method is expanded upon in method 3.5.7.B. *C,T* is *cis-trans* nepetalactone, *T,C* is *trans-cis* nepetalactone and *C,C* is *cis-cis* nepetalactone.

For A) and B) a final concentration of 50 µM camphor was used as an internal standard. Bracket with asterisk in A) denotes statistical significance (t-test p-value < 0.05).

3.2.10 qPCR Analysis of *ChlH* knockdown

N. cataria housekeeping genes for gene expression analysis were identified via homology-guided search as described in sections 2.2.1 and 3.2.1 of this thesis. Commonly used housekeeping genes include actin (ACT1) and ubiquitin (UBI9) [91]. These genes are required for basic molecular functions and are constitutively expressed at constant levels in different plant tissues. This provides a baseline to compare changes in gene expression of our target genes within experimental conditions. Sequences for these housekeeping genes were downloaded from TAIR and were used as queries to run BLAST on the *N. cataria* transcriptome. Homologues identified were then compared against our FPKM *N. cataria* dataset (used in section 2.2.1). Homologues with mid-range (Log2(FPKM+1) between 1 and 2) consistent expression across all tissues were selected for ACT1 and UBI9 (fig. 3.11).

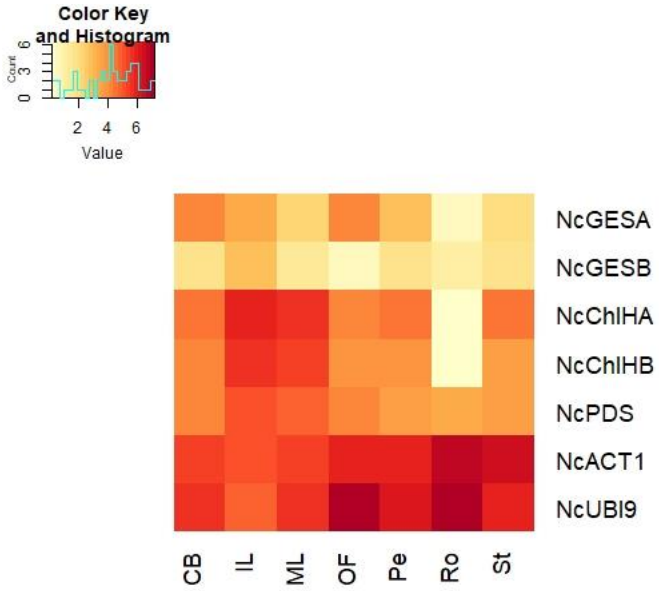


Figure 3.11: Housekeeping gene candidates expression data

Log(FPKM+1) values on candidate housekeeping genes from *N. cataria* identified via BLAST from *A. thaliana* homologues compared to the genes characterized in this chapter: *ChlH*, *PDS* and *GES*. Nc indicates *Nepeta cataria*. CB=closed bud, IL=immature leaf, ML=mature leaf, OF=open flower, Pe=Petiole, Ro=Root, St=Stem. Heatmaps were generated using the Heatmap package in R and were subjected to no hierarchical clustering.

As a confirmation of gene expression knockdown, I carried out qPCR on cDNA from experimental plants listed in section 3.2.7. A comparison between non-infected wildtype plants, plants infected with *A. tumefaciens* containing an empty pTRV2 vector and plants infected with *A. tumefaciens* containing pTRV2-*ChlH* revealed that the infection itself has a non-significant effect on gene expression, as evidenced by the broad standard error (fig. 3.12) but upon silencing of *ChlH*, gene expression is significantly reduced, including the gene expression of the selected reference house-keeping gene ubiquitin (fig. 3.12-A), parallel to the effect on metabolism revealed in section 3.2.7. A relative gene expression analysis (fig. 3.12-B) revealed that *ChlH* is significantly down regulated in tissue showing the bleached phenotype, consistent with the VIGS machinery effect. Due to the lower efficiency of *PDS* silencing, which produced fewer biological replicates and bleached tissue, I was unable to extract RNA from this tissue for qPCR analysis.

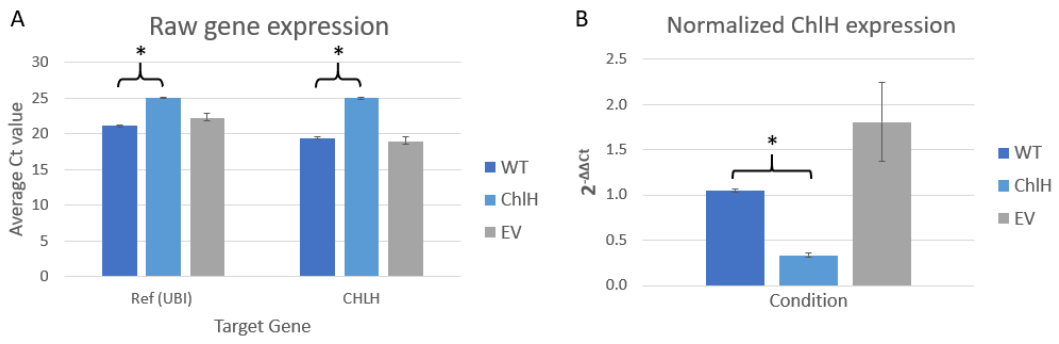


Figure 3.12: Effect of VIGS on gene expression

Quantitative PCR to assess gene expression. A) shows the average Ct values of the reference gene ubiquitin (Ref (UBI)), and *ChlH* (*CHLH*). The legend lists the VIGS infection conditions: WT is non-infected wildtype plants; *ChlH* is plants infected with pTRV2-*ChlH* containing *A. tumefaciens*; EV is plants infected with the empty pTRV2 vector containing *A. tumefaciens*. Error bars are standard error. B) shows the normalized *ChlH* gene expression ($2^{-\Delta\Delta Ct}$) in WT, *ChlH* and EV plants. Gene expression for *ChlH* was normalized to ubiquitin gene expression. Error bars denote standard error. Bracket with asterisk in A) denotes statistical significance (t-test p-value < 0.05).

3.2.11 VIGS strategies to study secondary metabolism

While VIGS in *N. cataria* proved to be successful, affected plants did not display full systemic knockdown (fig. 3.6) and not all infected plants exhibited silencing. The lack of a visual assay that validates the successful silencing of a given biosynthetic gene in a specific location greatly complicates the interpretation of silencing data. Previous publications of VIGS being used to study secondary metabolism often used indirect markers for harvesting plant tissue. In Misra *et al.* [54], *O. basilicum* cotyledons were initially infected with *A. tumefaciens* containing the pTRV2 targeting *ChlH*. The authors assessed the average probability (73%) of the first leaf pair post-infection being affected by the gene silencing. They also determined silencing via the relative transcript level of their gene of interest and the chlorophyl content of the leaf pair. Once the pTRV2 VIGS system and its limitations were established, the authors carried out VIGS on their secondary metabolism candidate genes and measured the metabolite content of their tissue, identifying the *in vivo* activities of two new enzymes in the *O. basilicum* triterpene metabolism. In Liscombe *et al.* [50], the authors carried out concurrent experiments of VIGS targeted *ChlH* on *C. roseus* seedlings to determine when to harvest tissue post-infection. Each experimental replicate had a portion of plants infected with *A. tumefaciens* containing the pTRV2-*ChlH* vector to indicate successful experimental transformation, and the first two leaf pairs post infection were harvested for metabolic study, usually at about 100% *ChlH* silencing efficiency. However, for VIGS in *N. cataria*, the low silencing rate (< 50 %) and non-systemic nature of the silencing, as well as the use of cuttings rather than seedlings, meant that a more robust indicator for silenced tissue collection is required.

As an alternative to indirectly using a visual marker to assess silencing efficiency, some studies have used direct visual markers to indicate which tissue to harvest for metabolic study. As reported in Orzaez *et al.* [49], the authors developed an in-tandem gene silencing technique in a transgenic *S. lycopersicum* cultivar overproducing anthocyanin. The authors designed pTRV2 vector constructs targeting the genes responsible for the overproduction of anthocyanin and their target genes of interest (pTRV2-Del/Ros-TomloxC, for example). This method allowed the authors precise excision of affected tissues but relied on the existence of an already transformed *S. lycopersicum* line. The Del-Ros VIGS system has been employed in further *S. lycopersicum* studies [47], [48]. In the absence of a stable transgenic cultivar for *C. roseus*, Yamamoto *et al*. [52] developed an in-tandem gene silencing technique using the pTRV2 vector system targeting *PDS* and their gene of interest serpentine synthase (pTRV2-

PDS-SS). In this method, the authors could select the *PDS*-silenced tissue exhibiting the bleached phenotype and assess the metabolic profile of this tissue.

3.2.12 *ChlH* and *PDS* as visual markers

As discussed in section 3.1.3 and 3.2.7, *ChlH* and *PDS* silencing via VIGS produces a similar "bleaching" phenotype in leaf tissue. However, their mechanism of action to produce this bleached phenotype is different. *ChlH* silencing inhibits chlorophyll formation by preventing the insertion of Mg^{2+} into protoporphyrin IX [84], leading to white, instead of green, tissue. Silencing *PDS* disrupts the formation of carotenoid biosynthesis [85] which is a photo-protector. Bleached tissue in *PDS* silenced tissue is due to photo-damage of the leaf tissue. Long term exposure in both cases can lead to "burnt" tissue; however, in the case of *N. cataria*, *PDS* silenced tissue exhibited damaged tissue faster than *ChlH* silenced tissue. Section 3.2.7 described the lessened "bleached" phenotype in *pds* tissues vs *chlh* tissue (fig. 3.13). This may not be a true representation of the silenced *PDS* effect, but rather the phenotype only becoming apparent after photo-damage, leading to a slower photo-bleaching phenotype to appear. This tissue also presents the risk of harvesting burnt tissue, which may have an unknown effect on secondary metabolic accumulation or degradation as compared to relative healthy leaves with not burnt tissue damage. For the purposes of this study, *ChlH* proved to be the best visual marker due to relatively healthier leaves which can be identified from their early growth, rather than having to wait until tissue damage occurs.

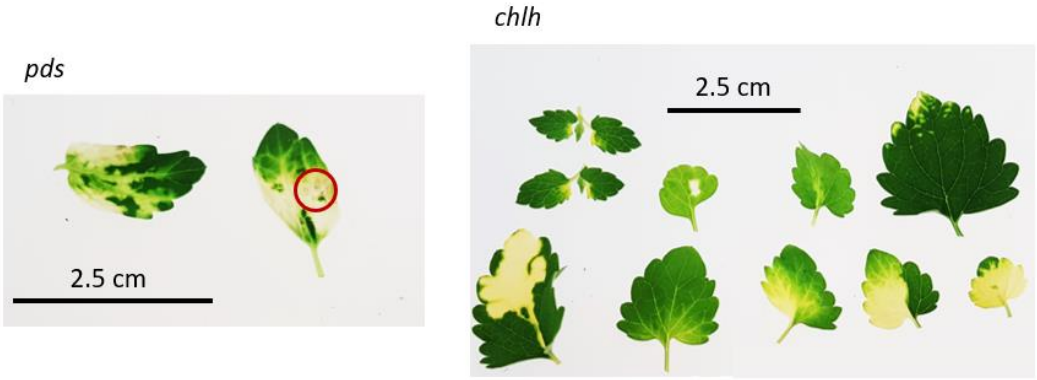


Figure 3.13: *PDS* vs *ChlH* silencing

All photobleached leaves harvested after 4 weeks of infection with *A. tumefaciens* per experimental replicate. Red circle indicates burnt tissue.

3.3 Future directions

3.3.1 A stable system for *Nepeta* transformation

As detailed in section 3.1.1, *Nepeta* has been transformed previously via the use of *A. rhizogenes* to produce stable transformants with hairy roots producing metabolites such as rosmarinic acid [59]. To study the iridoid metabolism of *Nepeta*, a transformation method must affect the aerial parts of the plant where these iridoids are produced. While VIGS has provided this functional genomics tool, qPCR analysis in section 3.2.10 revealed an incomplete knockdown of the targeted *ChlH* gene (0.335 relative to 1.05 in uninfected wild-type tissue), which may lead to some remaining gene activity which complicates functional genomics analyses (further explored in Chapter 4).

During the work for this project, I began establishing a stable transformant protocol in *N. cataria* and *N. mussinii* via the generation of callus tissue that could then be transformed with *A. tumefaciens*. This method is widely used in a wide variety of higher plants [92]–[95] and typically involves the production of a stem cell callus, whether from germ cells or shoot apical meristem cells, which are amenable to genetic transformation. Transformed callus tissue is then regenerated into plantlets and transformation is verified via the presence of a marker protein (ie. GFP) or the amplification of the new gene via PCR from genomic DNA. In *N. cataria* and *N. mussinii*, I successfully generated callus from sterilized stems and leaves by dissecting these tissues on MS media containing various concentrations of the hormones 6-benzylaminopurine, 1-naphthaleneacetic acid and 2,4-dichlorophenoxycetic acid. However, the callus line was not stable, and became non-viable after 2-3 sub-cultures. Furthermore, I was not able to re-generate a plantlet from these calluses. Transformation methods were therefore not tested, as the callus-regeneration protocol was not fully established.

3.3.2 Testing VIGS in other *Nepeta* species

As briefly mentioned in this Chapter, initial work in adapting the VIGS system into *Nepeta* included *N. cataria* and *N. mussinii.* However, due to the higher propagation and infection efficiency exhibited by *N. cataria*, work in *N. mussinii* was not continued. Adapting VIGS to other *Nepeta* species is likely to require optimization into the infection approach to improve silencing efficiency, as in section 3.2.6, and some species may be more amenable to VIGS than others. Nevertheless, establishing VIGS in *N. cataria* is a good prognostic to the possibility of expanding this tool to other species with some optimization.

Expanding VIGS into other *Nepeta* species would provide a tool to study the natural variation in secondary metabolism this genus has [24]. Work in *N. cataria* and *N. mussinii* has revealed novel enzymes such as the nepetalactol-related short-chain reductase/dehydrogenases (NEPS) enzymes and a novel function for major-latex like protein (MLPL) [64], [65] responsible for the nepetalactone isomers found in these species. However, as research across the *Nepeta* genus expands to explore more novel biochemistry, establishing a functional genomics tool such as VIGS in other *Nepeta* species may become a necessity to study the *in vivo* function of these novel genes.

3.4 Conclusion

In this Chapter, I developed a VIGS method for functional genomics studies in *N. cataria*. This method was based on those published in the literature and the tools available in our research group. I optimized the nepetalactone extraction protocol for downstream characterization studies to handle a high volume of replicates at lower tissue mass. I tested several methodologies of infection to optimize affected plants and tissues, as well as tested two commonly reported VIGS visual markers for best performance. Finally, I used an efficient dual knockdown system to use a visual marker to accurately select for affected tissue, which performs better than single knockouts targeting metabolic pathway genes.

This VIGS system in *N. cataria* provides a functional genomics tool with which I characterized the iridoid biosynthesis pathway in Chapter 4.

3.5 Materials and Methods

3.5.1 VIGS insert design

Sequences of 350 bp in length were obtained from *ChlH* and *GES* that did not contain any exact matches longer than 20 bp to other parts of the genome. Sequences were reviewed to be free from BamHI and XhoI restriction sites. Primers were designed containing the palindromic sites for BamHI (Forward primer) and XhoI (Reverse primer), to ligate to linearized pTRV2 vector, and the first and last 15 to 25 nucleotides with an annealing temperate close to 55 °C.

To construct the dual silencing vector, 200-350 bp fragments of genes of interest were selected and checked to not contain any exact 20 bp matches with other parts of the genome. Primers were designed to amplify these regions and add an overhang to the pTRV2 vector.

3.5.2 Obtaining plasmid gene inserts

Gene fragments for VIGS were obtained both by amplification of our stock of plasmids containing our genes of interest and cDNA in case if no stock being present. Primers for amplification included a nucleotide extension which matched to the pTRV2 vector. Primers were designed to be between 15 and 25 nucleotides long and with an annealing temperature close to 55 °C.

3.5.3 cDNA production and insert amplification

RNA from young leaf tissue from *N. cataria* was retrotranscribed using Invitrogen Superscript IV. In a 10 µL reaction, up to 500ng (about 1 µL) of RNA, 4 µL of SSIV buffer, 1 µL of SSIV reverse transcriptase (200 U/µL) and 1 µL of 100 mM DTT. Incubated at 50 °C for 10 minutes and then at 80 °C for 10 minutes to inactivate. RNAase H (1 µL) was added and incubated at 37 °C for 20 minutes. cDNA was stored at -20 ˚C. Inserts from cDNA were amplified by PCR in a 25 µL reaction using 1 µL of cDNA or plasmid-base (Table 2), 12.5 µL of Invitrogen Platinum Superfi DNA polymerase mix, 1.25 µL of 10 µM Forward and Reverse primers each, and 5 µL of GC enhancer. Primers for each reaction can be found in Supplemental Table 1.

3.5.4 Vector construction

BamHI, XhoI were used to linearize pTRV2, EcoRI was used to linearize pTRV2-*ChlH*. All enzyme digestions were carried out with 1 µL of enzyme, 1000 ng of vector, NEB Cutsmart buffer in 50 µL reactions. For pTRV2 an additional 1 µL of Sigma Aldritch rAPID alkaline phosphatase was added. Reaction mixes were incubated at 37 °C for 1 hour. Linear vectors were purified by running a 1% Agarose electrophoresis gel from which the DNA was purified using a Machery-Nagel NucleoSpin Gel and PCR Clean-up kit.

Insertion for plasmids pTRV2-*ChlH*H, pTRV2-*ChlH-GES* by InFusion reaction (ClonTech). In a 5 µL reaction, 1 µL of In-Fusion HD Plus, 20-50 ng of linear vector, and 20-50 ng of insert were added. This reaction mix was incubated at 50 °C for 15 minutes and cooled on ice. 2.5 µL of this reaction were used to transform Stellar competent cells and grown on selective media containing 50 mg/L of kanamycin.

Base pTRV-*ChlH* vectors were constructed via a ligation reaction. In a 20 µL reaction, 2 µL of NEB T4 DNA ligase buffer, 20-50 ng of linear vector, 60-150 ng of insert (3:1 insert:vector) and 1 µL of T4 ligase. Reaction is left at 16 °C overnight and transformed into Stellar competent cells and grown overnight on selection media containing 50 mg/L of kanamycin.

3.5.5 *E. coli* transformation

Stellar competent cells were used for post-ligation transformations. Transformation followed the protocol: cells are thawed on ice. Then, 1-5 µL of the ligation reaction were added to 30-50 µL of cells and gently mixed by stirring pipette tip in mixture. Cells were left to incubate on ice for 10 minutes, heat shocked at 42 °C for 45 seconds, and cooled on ice for 2 min. 150 µL of SOC media was added to cells and gently mixed by stirring pipette tip in mixture and left to incubate at 37 °C for 45 min. 100-150 µL was plated on LB plates containing the kanamycin and grown overnight at 37 °C.

3.5.6 *A. tumefaciens* transformation

Electrocompetent *A. tumefaciens* strain GV3101 cells were thawed on ice. 100 ng of plasmid DNA was added to 50 µL of cells and mixed gently. Cells were electrically shocked. 150 µL of SOC medium was added and mixed gently. Cells were incubated at 28 °C for 3 hours and plated on LB agar containing 50 mg/L of kanamycin, 50 mg/L of gentamycin and 50 mg/L of rifampicin. Plates were incubated at 28 °C for 2 days.

3.5.7 Gas chromatography analyses

Samples were injected in split mode (1 µL, split ratio 10:1) at an inlet temperature of 280 °C on a Thermofisher Trace1310-ISQLT GC-MS equipped with a 5973 mass selective detector (MSD), and a CTL Analytics GC PAL injector and autosampler. Separation was performed on a Zebron ZB5-HT-INFERNO column (5% phenyl methyl siloxane; length: 35 m; diameter: 250 µm) with guard column. Helium was used as mobile phase at a constant flow rate of 1.2 mL/min and average velocity 37 cm/s. Two temperature runs were used for detection: **A.** After an initial temperature at 60 °C, the column temperature was increased to 100 °C at a rate of 20 K/min, then to 160 °C at 2 K/min, then another increase to 280 °C at 120 K/min, and maintained for 4 minutes. A solvent delay of 5 minutes was allowed before collecting MS spectra at a fragmentation energy of 70 eV. **B.** After an initial temperature at 50 °C, the column temperature was increased to 170 °C at a rate of 10 K/min, then another increase to 280 °C at 50 K/min, and maintained for 4 minutes. A solvent delay of 5 minutes was allowed before collecting MS spectra at a fragmentation energy of 70 eV. Chemically characterised standards were used to identify compounds by retention time and electron impact spectra.

3.5.8 GC-MS peak integration and statistical analysis

RawMS files were converted to .csv files using Lablicate OpenChrom. Using R programming language (R Core Team 2021), areas were obtained by an in-house built script (developed by Carlos E. Rodriguez-Lopez). The script obtained peak areas from the .csv files by addition of the intensities at each time point and integrating the peak area according to a user input timeframe based on the file TIC. Nepetalactone peak areas were divided by the camphor peak area to obtain the normalized peak area. Each data set was statistically analysed using R programming language. Variances in each data set were checked for difference, and a pairwise t-test with a Bonferroni adjustment (for equal or unequal variances) was applied to obtain p-values. The R-package ggplot2 was used to generate the Normalized nepetalactone

3.5.9 Plant growth conditions and propagation method

In coordination with the greenhouse team of the Max Planck Institute of Chemical Ecology, *N. cataria* cuttings were taken from plants growing in a growth chamber with 15.5 hours of full light (23 °C), 30 minutes of dusk and 30 minutes of dawn conditions, and 7.5 hours of night (21 °C). Humidity was kept at 50%. Cuttings were taken to include 2-3 nodes. The bottom node was removed from leaves and inserted into water until rooted, about 1-2

weeks. Rooted cuttings were transplanted to a soil mix of 250 L of Klasmann TS1, 70 L of Kalsmann Tonnubstrat and 34.5 L of Raiffeisen Baustoffe sand (0.7-1.2 mm). Plants were fertilized once a week with 0.1% Ferty 3 (Planta Düngemittel GmbH) and watered as necessary.

3.5.10 Virus-induced gene silencing

A. tumefaciens cultures for pTRV1, pTRV2, pTRV2-*ChlH*, pTRV2-*PDS,* pTRV2-*ChlH-GES* were grown in 50 ml LB cultures containing 50mg/L of kanamycin, 50mg/L of gentamycin and 50mg/L of rifampicin, 10 mM MES buffer (pH 5.8) and 200 uM of acetosyringone for 24 hours at 28 °C shaking at 200 rpm. Cultures were centrifuged for 15 minutes at 3500 rpm, the supernatant discarded and resuspended in 10 mL of fresh infiltration buffer (consisting of 10 mM of MES at pH 5.8, 10 mM of $MgCl_2$ and 200 µM of acetosyringone). The OD_{600} was measured in a 1:10 dilution, and the volume of the infiltration buffer and cultures were adjusted to obtain 10 ml of a OD_{600} of 1. Cultures pTRV1 *A. tumefaciens* were mixed with each individual pTRV2 *A. tumefaciens* cultures in a 1:1 ratio (10 mL: 10 mL). Cultures were then incubated in the dark for 3-4 hours at room temperature with gentle rocking. Cultures were then centrifuged at 3500 rpm for 15 minutes, the supernatant was discarded and the pellet resuspended in 1 mL of fresh infiltration buffer. Plants for VIGS infection were cut down to 2 to 3 aerial nodes encourage new leaf growth. A sterilized toothpick was dipped into the 1 mL cultures and used to wound the stem near the nodes 2-3 times, and finally the wounds were rubbed with the toothpick with more culture.

3.5.11 VIGS tissue harvesting for metabolite analyses

Nepetalactones were extracted as detailed in Lichman *et al.* [65], with a few changes. In general, extraction procedure was as follows: *N. cataria* tissue was flash frozen and pulverized, either with a pestle and mortar or with a Qiagen TissueLyser II at 25Hz for 30 seconds twice, depending on tissue volume. Ice-cold methanol containing the internal standards (200 or 50 µM) camphor and harpagoside (10 µM) was added to the powdered *N. cataria* tissue, vortexed then sonicated at room temperature for 15 minutes. The mix was centrifuged on a table-top centrifuge at room temperature and at top speed (14,600 rpm) for 10 minutes. The methanol extract was transferred to a new 1.5 ml tube. An aliquot was taken for LC-MS analysis, leaving 400 µL of the MeOH extract. An equal volume of hexane was added, and the mix was vortexed for 2 minutes, centrifuged at top speed for 30 seconds

and the hexane layer was transferred to a solid phase extraction (SPE) column. The hexane layer was passed through the column and collected. Then an equal volume of 20:80 EtAc:Hex (ethyl acetate : hexane) mix was added to the SPE column to collect the nepetalactones. The EtAc:Hex extract is then analysed on the GC-MS. For VIGS experiments, Leaf tissue from pTRV2 empty vector was collected by cutting a full leaf from three individual plants. Tissues from plants infected with *ChlH, PDS* and *ChlH-GES*, only individuals with affected leaf tissue were harvested. Affected tissues was collected by cutting out the bleached area from the green tissue and pooling together all affected leaf tissue from one individual.

3.5.12 RNA extraction and cDNA generation

For wildtype samples, leaves of all ages were collected from a B2C cultivar plant into a falcon tube. The falcon tube was flash frozen and all leaf tissue was ground with a mortar and pestle, while cooled with liquid nitrogen to keep the tissue frozen. The ground tissue was then returned to the falcon tube and stored at -80 °C. Ground tissue powder was weight out between 25-35 mg of tissue into cooled 1.5 ml tubes. For individual experimental samples, leftover sample tissue not used for metabolic analyses, was weighed out to be between 20 and 35 mg of tissue. If tissue mass was too low to be transferred to a new tube for weighing, the tissue was not weighed. RNA was extracted using either the Qiagen® RNeasy Plant Mini Kit, or the Qiagen® RNeasy PowerPlant Kit, depending on availability. Tissue was extracted according to kit instruction, with a final incubation of 10 minutes for more extraction.

RNA concentration and A280/A260 and A280/A230 ratios were assessed using a Nanophotometer N60**.** A 2% Agarose electrophoresis gel was ran at 120 V for 1 hour to verify any genomic DNA contamination and RNA degradation. RNA samples were subjected to a DNAse treatment to remove any contaminant genomic DNA. Treatment was carried out with Sigma Amplification grade DNase I kit by adding 1000 ng of RNA, 1 µL of Sigma Amplification Grade DNase 10x buffer and 1 µL of the DNAse enzyme into a 10 µL reaction and left to incubate for 20 minutes at room temperature. Then, 1 µL of stop solution was added and the sample was incubated at 70 °C for 10 minutes. After incubation, the sample was cooled on ice for 2 minutes. The Thermo Fisher Scientific Applied Biosystems High-capacity cDNA reverse transcription kit was used to make a 10 µL mixture containing 4.2 µL of water, 2 µL of buffer, 2 µL of random hexamer primers, 0.8 µL of dNTPs and 1 µL of enzyme and was added to the RNA sample. The RT reaction was there incubated at 25 °C for 10 minutes, 37

°C for 2 hours, then at 85 °C for 5 minutes and then stored at 4 °C. Samples were stored in -20 °C.

3.5.13 qPCR primer design and analysis

Primers for qPCR analysis were designed using the CDS of the target gene in the NCBI primer design tool. The primers were restricted to replicate a product from 70 bp to 200 bp and to have a melting temperature of 60 °C. The tool's best primers, that did not lay within the targeted VIGS region, were selected to be tested.

A dilution standard curve was carried out for each primer pair. Wildtype cDNA was initially diluted to a ratio of 1:5, then a series of four half dilutions were carried out. Reaction triplicates were carried out at each dilution point. The average Cq value was then calculated and plotted against the log of the sample concentration. A trendline was plotted for these values, and the equation for the trendline calculated (y=mx+b). The slope (m) of the equation was then used to calculate the efficiency (E) in the following formula: E=2^(-1/m). The percent efficiency was then calculated by the following %E=100(E-1). Primers resulting in the E closest to 2, and the %E closest to 100% were selected to use for qPCR amplification. Primer standard curves are in Appendix 3.

3.5.14 qPCR reaction

The cDNA generated from the RT reaction was diluted by half with diWater. The reaction mixture for the qPCR contained the following: 12.5 µL of Agilent Brilliant II SYBR® green QPCR Master Mix, 1.5 µL of forward primer and 1.5 µL of reverse primer (at 6nmol each), and 7.5 µL oh diWater and 2 µL of the diluted cDNA sample. A master mix of the qPCR reaction mix was pipetted into each well of a 96 well-plate first, and the diluted cDNA sample was then added to each experimental well. Each gene targeted had 3 technical replicates and 3 biological replicates. To measure gene expression, each target gene was paired to the housekeeping gene ubiquitin (UBI). Each plate also contained a well with the reaction mixture containing diWater instead of sample as an empty control.

This reaction was then placed in a Bio-Rad CFX96 Optical Reaction Module. The qPCR reaction protocol was as follows: 3 minutes at 95 °C, then 44 cycles of 30 seconds at 95 °C, 30 seconds at 60 °C, and 30 seconds at 72 °C, and an individual incubation at 95 °C for 10

seconds before a final melt curve. A melt curve step with an increase of 5 °C every cycle from a minimum of 65 °C to a maximum of 95 °C finalized the qPCR reaction.

3.5.15 qPCR analysis

The resulting cross point (Ct) values from the qPCR reaction were averaged across technical replicates. Delta Ct (ΔCt) was calculated by substracting the average UBI Ct value from the average target gene Ct value. The DeltaDeltaCt (ΔΔCt) was calculated by subtracting a control calibrator ΔCt value from each ΔCt value. The ΔΔCt value was then used in the following equation $2^{-\Delta\Delta Ct}$ to calculate the normalized expression of each biological replicate. The average normalized expression was calculated for each experimental sample.

3.5.16 Heatmap of gene expression

Tissue specific gene expression analysis heatmaps were generated using the R-script package for heatmaps included in the gplot package, based on Log(FPKM+1) results. Heatmaps were generated without hierarchal clustering.

3.5.17 Primer list

Code	Gene	Sense	Sequence
LP029	NECA_CHLH_VIGS	F	CGAGGATCCACCAATGACATGAAGGCCAC
LP030	NECA_CHLH_VIGS	R	CGATCTCGAGACGCTGCTAACAACCCG
LP031	NECA_GES1_VIGS	F	CGAGGATCCCCTCGCTACAAAGGCGA
LP032	NECA_GES1_VIGS	R	CGATCTCGAGAATCCAACGGCTGGGAAT
LP033	NECA_GES2_VIGS	F	CGAGGATCCGCACCCGACCCTCCT
LP034	NECA_GES2_VIGS	R	CGATCTCGAGCTGAATCCAAGAGCTGTGGAATAC
LP037	NEMU_CHLH_VIGS	R	CGATCTCGAGGAGCTCACAATTTGAGGGC
LP044	NECA_CHLH-GES_VIGS	F	TTAGCAGCGTCTCGAGAATCCAACGGCTGGGAATCGG
LP045	NECA_CHLH-GES_VIGS	R	TCGGGACATGCCCGGGCCTCGCTACAAAGGCGAGG
LP050	NECA_GES_VIGS	F	ACGACAGCTGAAGAGCTAGGTTTGGTGGATAAGTT

Chapter 4: *In Vivo* Validation of the *N. cataria* Iridoid Biosynthetic Pathway

4.1. Introduction

In this Chapter, I describe the use of the VIGS methodology established in Chapter 3 to characterize and confirm key steps of the *N. cataria* iridoid biosynthesis pathway. I also explore the range of possibilities VIGS offers in the broader study of *N. cataria* secondary metabolism.

4.1.1 Iridoid metabolism in *N. cataria*

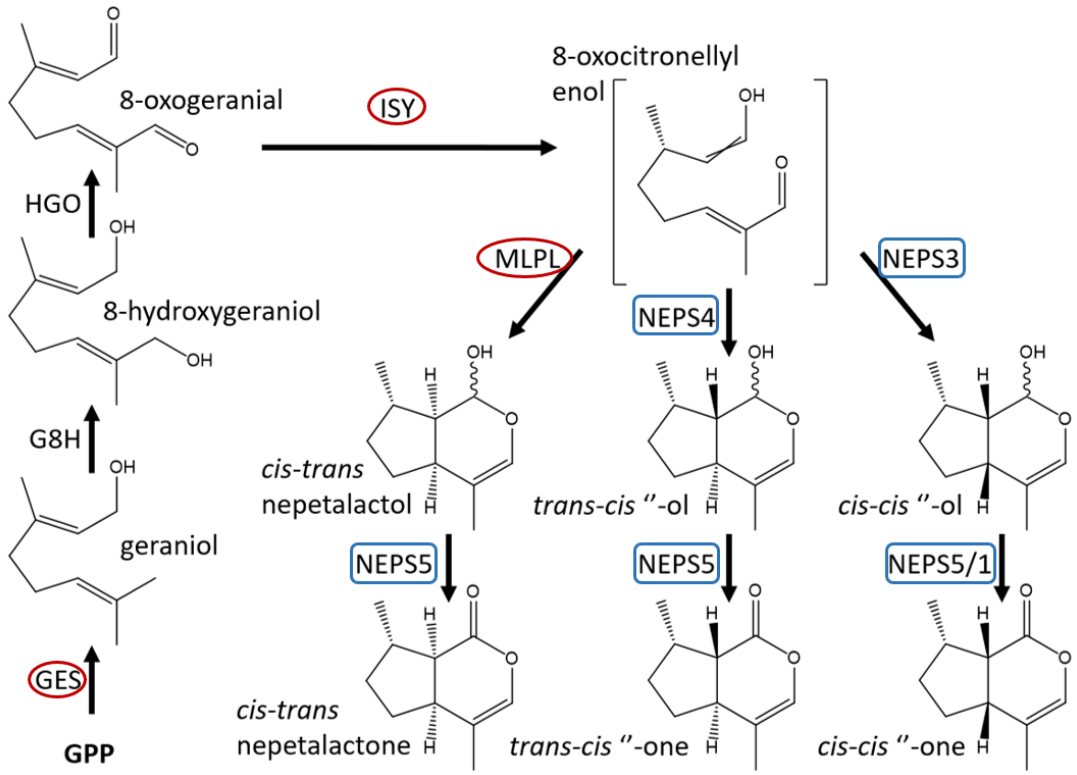


Figure 4.1: *Nepeta* iridoid pathway with VIGS targets

A schematic of the *N. cataria* iridoid pathway as described in Lichman *et al.* [65] based on *in vitro* and *in silico* studies. Enzymes in a red or blue outline were targeted for *in vivo* silencing. Enzymes in the red circles were targeted individually. Enzymes in the blue rectangle were targeted simultaneously for silencing.

Nepetalactone biosynthesis in *Nepeta* begins with GPP produced from the MEP pathway being converted to the monoterpene geraniol by the specialized terpene synthase GES [12],

[39], [65] (fig. 4.1). Geraniol is further oxidized by the CYP76 class II cytochrome P450 G8H [12], [40], [65], [74] and the oxidoreductase HGO [12], [68] to produce 8-oxogeranial. ISY then produces a reactive intermediate, 8-oxocitronellyl enol [23], [64], which in *Nepeta* is cyclized into nepetalactol by MLPL, NEPS4 or NEPS3 [64], [65]. This nepetalactol is then oxidized into nepetalactone via the action of NEPS5 or NEPS1 [64], [65]. As addressed in Chapters 2 and 3, *N. cataria* has been reported to make three isomers of nepetalactone: *cis-trans* nepetalactone, *trans-cis* nepetalactone and *cis-cis* nepetalactone (fig. 4.1) [63]. The enzyme responsible for cyclization determines the nepetalactol isomer produced [64], [65].

While the activity of GES, G8H, HGO, ISY, NEPS1, 3, 4, 5 and MLPL have been characterized *in vitro* [23], [64], [65], the *in vivo* roles of these genes and their protein products have not been explored in *Nepeta*. Of particular interest is understanding the interplay between the various cyclization and oxidation enzymes in the regulation of the nepetalactone isomers produced by *N. cataria.* Therefore, using a functional genomics tool such as VIGS can help understand the role of these enzymes *in vivo*.

4.1.2 Iridoid pathway regulation and the *Nepeta* gene clusters

Iridoids have been identified across several plant families [31], [33]. The pathways leading to their production seem to be conserved across these species [12], [14]. Previous work into the regulatory elements of these iridoid pathways has revealed two main basic-helix-loop-helix (bHLH) transcription factors, bHLH iridoid synthesis (BIS) 1 & 2 to be responsible for transactivating the iridoid portion of the larger MIA pathway [87], [96]. Other regulatory elements in iridoid production in other iridoid-producing plants have not been identified.

The physical location of metabolic pathway genes in plants is an active area of study [97]–[101]. In bacteria, non-homologous genes involved in the same metabolic pathways can often be found clustered close to each other [102], [103]. Gene clustering is not as common in plants, yet notable examples have been found. Examples in plants include the avenacin cluster in *Avena sativa* (oat) [97], and noscapine cluster in *Papaver somniferum* (poppy) [98], both examples of gene clusters which include the genes for the enzymatic steps for a secondary metabolite. In *C. roseus*, Paul *et al*. [99] reported a gene cluster of the regulatory elements controlling the production of MIA. More plant clusters are being identified as genome sequencing technologies become faster and more efficient, and specialized tools to

identify clusters in plant genomes are being developed, such as plantiSMASH, based on the bacterial antiSMASH [100].

Analysis of the *Nepeta* genomes revealed that key genes in the iridoid synthase pathway have clustered together in both *N. cataria* and *N. mussini* [65]. The compositions of the cluster include *ISY*, the *NEPS* homologues and *MLPL* in both species, and *GES* in the *N. mussini* cluster (fig. 4.2). These genes within the cluster exhibited the same tissue-specific expression patterns [65]. Along with the iridoid pathway genes, there are other open reading frames within the region of the cluster. In the *N. cataria* cluster A, annotations have been assigned to these open reading frames using sequence homology against the National Center for Biotechnology Information (NCBI) database. However, these genes have not been characterized nor reported. These assigned genes include a rosmarinate synthase, a homeobox leucine-zipper protein, a receptor-like kinase, a polygalacturonase, and a hypersensitive-induced response protein-like protein (unreported). From these genes, only the receptor-like kinase (*RLK*) can be found in both *N. cataria* clusters A and B and the *N. mussinii* cluster (fig. 4.2). Receptor-like kinases are a family of transmembrane extracellular-cytoplasmic plant proteins [104]–[107]. This large membrane protein family is divided into several subgroups, which include the Leucine-rich repeat RLK family, the *C. roseus* RLK1-like family, the LysM RLK family, the lectin RLK family and the proline-rich extension-like family [105], [106]. Many of the RLK proteins identified in plants have an unknown function and the ligands for their extracellular domains also remain unidentified. Of the RLK proteins that have been characterized, many are involved in stress responses for biotic and abiotic stresses and often play a key role in detecting microbe associated molecular patterns [104]–[107].

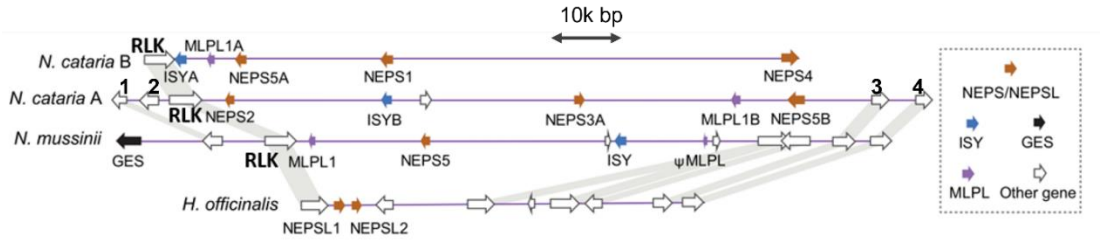


Figure 4.2: Iridoid gene cluster

Iridoid pathway gene cluster in *N. cataria* as reported in Lichman *et al.* [63]. The location of the *RLK* genes have been added to this diagram, along with the homology assigned genes in *N. cataria* cluster A. 1 is rosmarinate synthase; 2 is a homeobox leucine-zipper protein; 3 is a polygalacturonase; 4 is a hypersensitive-induced response protein-like protein. Scale bar indicates 10,000 base-pairs.

While the *Nepeta* iridoid pathway has been characterized via *in vitro* biochemical analysis [23], [64], [65], the iridoid pathway has not yet been confirmed *in vivo*. Using the VIGS method adapted to *N. cataria* described in Chapter 3, I carried out *in vivo* silencing of select genes in the pathway. Furthermore, I targeted a potential regulatory gene found in the iridoid cluster in *N. cataria, RLK*, to test for its potential involvement in regulation of the iridoid pathway.

4.2 Results & Discussion

4.2.1 Target selection and design

The work in Chapter 3 describes the development of a functional genomics tool to study secondary metabolism in *N. cataria* via VIGS. I targeted key genes in the pathway to confirm the pathway *in vivo* and to understand isomer production regulation within the plant.

In this chapter, I report the results of silencing *GES, ISY, MLPL,* the *NEPS* homologues and the cluster gene *RLK* using the visual-marker VIGS system developed in Chapter 3. I targeted *GES* as a main entry point into the pathway to confirm that *GES* was responsible for all three nepetalactone isomers produced (fig. 4.1 – red circles). As opposed to the results on *GES* silencing presented in Chapter 3, as a means-test for VIGS double knock-down efficiency, here I report on the *in vivo* function of *GES*. I then targeted *ISY*, *MLPL* (fig. 4.1 – red circles) and the *NEPS 1, 3, 4* & *5* homologues (fig. 4.1 – blue rounded rectangles) to understand their roles in the production of the three nepetalactone isomers found in *N. cataria*: *cis-cis*

nepetalactone, *cis-trans* nepetalactone and *trans-cis* nepetalactone. Furthermore, I targeted a predicted receptor like kinase (*RLK*) found on the *Nepeta* iridoid gene cluster. We theorized that this *RLK* may be playing a role in the iridoid biosynthesis pathway due to its location in the cluster.

As described in section 3.2.2, I selected short (150-350 bp) regions of the CDS of *GESA/B, ISY, MLPL* and *RLK* to clone into the double knockdown pTRV2-*ChlH* vector. The selected target regions of each gene were used as BLAST queries into the *N. cataria* transcriptome to control for off-target matching sequences. If a region had a matching site of 19 or more base-pairs, I selected another region within the gene to prevent off-target silencing. As the *NEPS* homologues (*NEPS 1, 3, 4 & 5*) are very similar to each other [64], [65], it was not possible to selectively target individual genes using the CDS region, as described in section 4.2.6. In short, highly conserved regions amongst *NEPS* homologues were targeted for the simultaneous knockdown of these homologues (further expanded in section 4.2.6).

Gene expression analysis from transcriptomic data from *N. cataria* tissues on the genes of interest compared to the housekeeping gene, UBI9, revealed consistent expression patterns amongst the iridoid pathway genes *GESA, ISY, MLPL, NEPS1, NEPS2, NEPS3A, NEPS3B, NEPS4, NEPS5* and the visual marker genes *ChlHA* and *ChlHB*, with more expression in leaf (IL and ML in figure 4.3) and flower (CB, OF and Pe) tissues, and less expression in stem (St) and roots (Ro). In contrast, the housekeeping gene UBI9 is consistently highly expressed in all tissues, and *GESB* and *RLK* are very lowly expressed in all tissues (fig. 4.3).

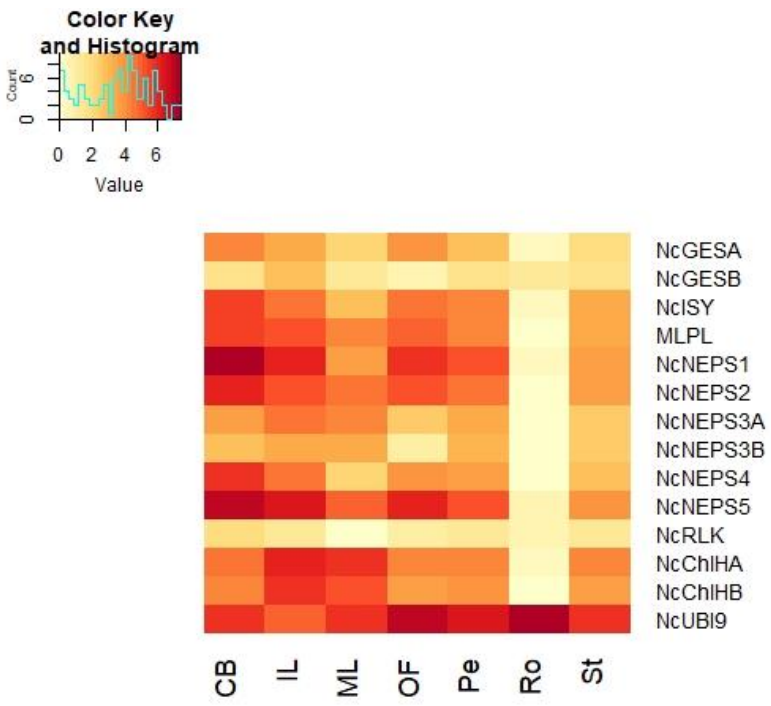


Figure 4.3: Housekeeping gene candidates expression data

Log (FPKM+1) values on the genes characterized, genes used as a visual marker (*ChlH*) and genes used in gene expression analysis (*UBI9*) in this Chapter from *N. cataria*. Nc indicates *N. cataria*. CB=closed bud, IL=immature leaf, ML=mature leaf, OF=open flower, Pe=Petiole, Ro=Root, St=Stem. Heatmaps were generated using the Heatmap package in R and were subjected to no hierarchical clustering.

4.2.2 VIGS experimental setup

Since the VIGS methodology described in Chapter 3 was not 100% efficient nor fully systemic, I needed to produce enough replicates to ensure the availability of enough samples per experiment for robust statistical analysis. I arranged for the propagation and rooting of 120 cuttings of B2C or B2E plants per experiment. These two cultivars produce all three nepetalactone isomers, *cis-trans*, *trans-cis* and *cis-cis*, found in *Nepeta* (fig. 3.4). Each experiment of the rooted and surviving plants consisted of the following: 28 plants infected with *A. tumefaciens* containing pTRV2-*ChlH* as control, 28 plants were infected with *A. tumefaciens* containing pTRV2-*ChlH-gene of interest 1* (*GOI1*), 28 plants were infected with *A. tumefaciens* containing pTRV2-*ChlH-GOI2*, and 28 plants were infected with *A. tumefaciens* containing pTRV2-*ChlH-GOI3*. Some rooted plants were discarded if they could not be used for VIGS infection. For example, some rooted plants remained in a stunted flowering phase that produced no new leaves. Others died after transfer to soil. Of the

remaining healthy plants, these were infected with an empty pTRV2 vector as an infection control.

Each gene was tested in at least three experiments. On average, experiments were harvested four to six weeks post infection, although a bleaching phenotype was observed on newly emerged leaves two to three weeks post infection. The extended harvesting time was to allow for more plant material to grow and was possible because of the double knockdown system which allowed for the precise harvesting of silenced tissue (fig. 4.3).

Figure 4.3: Tissue collection

Using the visual marker technique described in Chapter 3, affected experimental tissue was identified as bleach tissue. All tissue from an individual plant was carefully collected, pooled together, weighed and flash frozen.

4.2.3 *GES* knockdown

As described in sections 2.2.1 and 3.2.10, GES is the entry point from the MEP pathway to the specialized iridoid metabolism of *N. cataria* [11]. GES, a specialized terpene synthase, generates geraniol through a standard terpene synthase mechanism: loss of the phosphate group followed by quenching of the resulting carbocation with water [12], [39], [65] (fig. 4.4-A). Geraniol is further oxidized and cyclized to form *cis-trans*, *cis-cis* and *trans-cis* nepetalactone present in *N. cataria.* Knocking down *GES* expression is expected to downregulate the entire iridoid pathway.

As expected, tissue from plants infected with *A. tumefaciens* containing pTRV2-*ChlH-GES* (*ges* in figure 4.4) displayed a significant decrease in all nepetalactone isomers as compared to tissue from the control plants infected with *A. tumefaciens* containing pTRV2-*ChlH* (*chlh* in figure 4.4), confirming *GES* as the *in vivo* entry point into nepetalactone biosynthetic pathway.

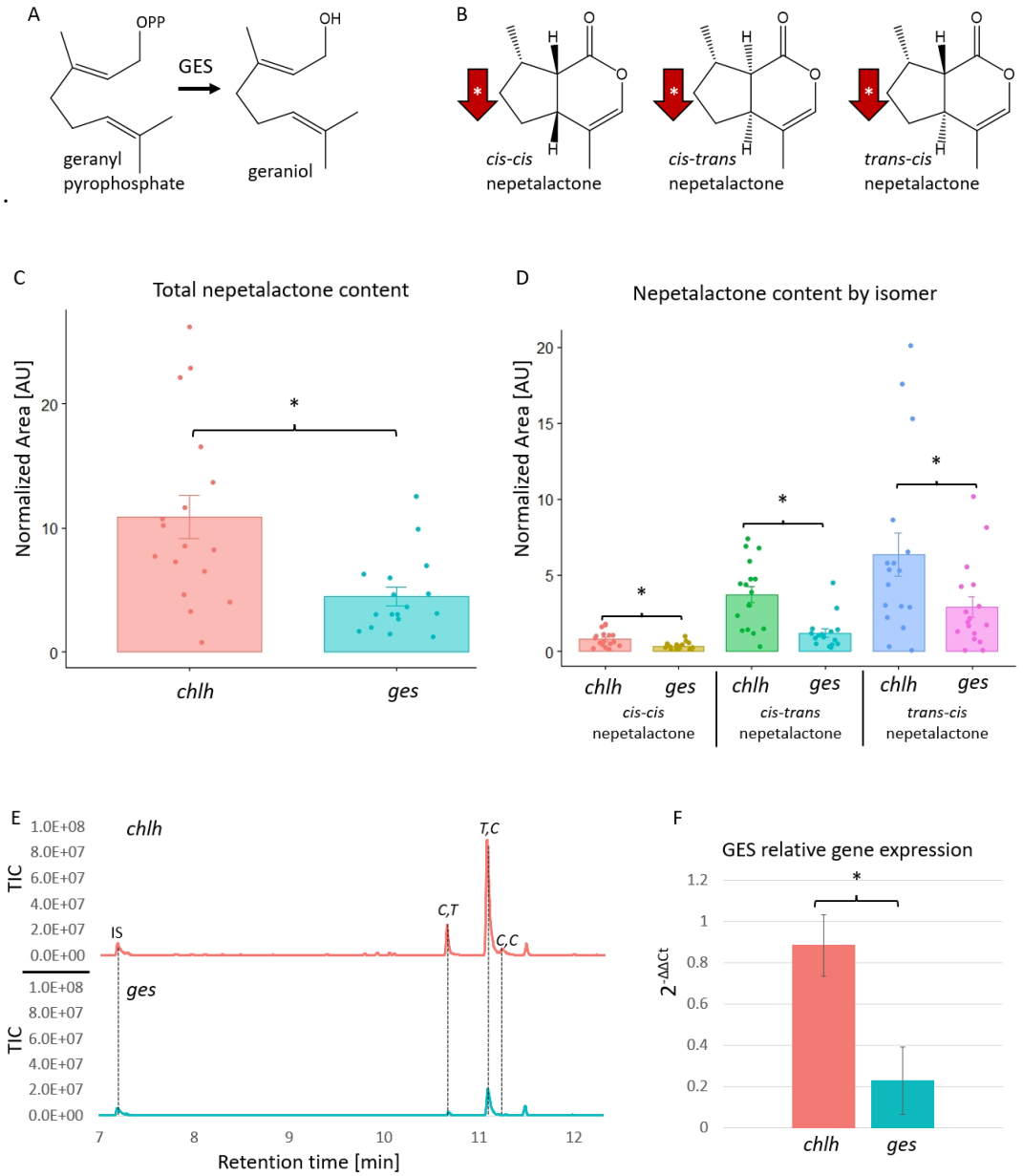


Figure 4.4: Effect of VIGS on *GES*

A) Terpene synthase reaction mechanism of *GES*. Red circle denotes the VIGS targeted gene in this reaction. B) is a summary of the changes on nepetalactone isomer production. A downward facing red arrow with a white asterisk indicates a statistically significant decrease in a given isomer. C) shows the total nepetalactone content in tissues from pTRV2-*ChlH A. tumefaciens* infected plants (*chlh*) and pTRV2-*ChlH-GES A. tumefaciens* infected plants (*ges*). Normalized nepetalactone content for each isomer is calculated by dividing the isomer peak area by the internal standard (50 µM camphor) peak area. Total nepetalactone content is the sum of each normalized isomer content. The bars indicate the average content of the sum. Each point is an individual sample. Error bars are the standard error. An asterisk above a bracket indicates a t-test p-value <0.05. AU indicates arbitrary units. D) indicates each relative nepetalactone content (nepetalactone peak area/internal standard peak area). The bars indicate the average content of each isomer. Each point is an individual sample. Error bars are the standard error. An asterisk above a bracket indicates a p-value <0.05. AU indicates arbitrary units. E) are representative GC-MS traces of *chlh* (top) and *ges* (bottom). TIC is total-ion count. Retention time is in minutes. IS is internal standard (50 µM camphor), *C,T* is *cis-trans* neptalactone, *T,C* is *trans-cis* nepetalactone and *C,C* is *cis-cis* nepetalactone. F) Relative gene expression of *GES* relative to *UBI9* in *chlh* and *ges* tissue as measured by a qPCR reaction. X-axis is relative gene expression calculated by $2^{-\Delta\Delta Ct}$. Error bars are the standard error. An asterisk above a bracket indicates a t-test p-value <0.05.

4.2.4 *ISY* knockdown

As described in Lichman *et al.* [63], *N. cataria ISY* catalyses the formation of the 8-oxocitronellyl enol intermediate (fig. 4.5-A) upon which MLPL and the NEPS enzymes cyclize and/or oxidize to form the downstream nepetalactols and nepetalactones. Upon silencing *isy*, I expected to see a decrease in all isomer content and the presence of the ISY substrate, 8-oxogeranial.

Tissue from plants infected with pTRV2-*ChlH-ISY A. tumefaciens* (*isy* in figure 4.5) displayed a significant decrease in *cis-trans* nepetalactone and *cis-cis* nepetalactone concentration as compared to tissue from plants infected with pTRV2-*ChlH A. tumefaciens* (*chlh* in figure 4.5). While *trans-cis* nepetalactone concentration was decreased in *isy* as compared to *chlh*, it was not significantly different (fig. 4.5 B-E). The total concentration of nepetalactones was significantly reduced in *isy* compared to *chlh* (fig. 4.5-C). The isomer ratios (the ratio of each normalized isomer content of the total normalized isomer content) between *chlh* and *isy* were not significantly different. However, the ratio of *cis-trans* nepetalactone was slightly decreased in *isy*, whereas the ratio of *trans-cis* nepetalactone was slightly increased in *isy* knockdown tissue (fig. 4.5 F). The GC-MS data did not reveal any new peaks that could be potential geraniol-derived intermediates or side products. However, as addressed in section 4.2.10, a putative glycosylated hydroxygeraniol derivative was upregulated upon silencing of *isy*.

While the decrease in *trans-cis* nepetalactone concentration in *isy* was not statistically significant, this does not necessarily suggest that *trans-cis* nepetalactone production is independent of *ISY* activity. Gene expression qPCR data in figure 4.4-G reveals that *ISY* has much lower expression in *isy* tissues, as compared to *ISY* in *chlh* tissues, but some relative gene expression could be detected (0.165) (fig. 4.4-G). Further discussion on the known limits of VIGS are explored in section 4.2.11. However, this result, along with the change in isomer ratios in figure 4.5-F, suggests that it is possible that *trans-cis* nepetalactone production could be produced by another *ISY*-like enzyme, though such an enzyme is not known.

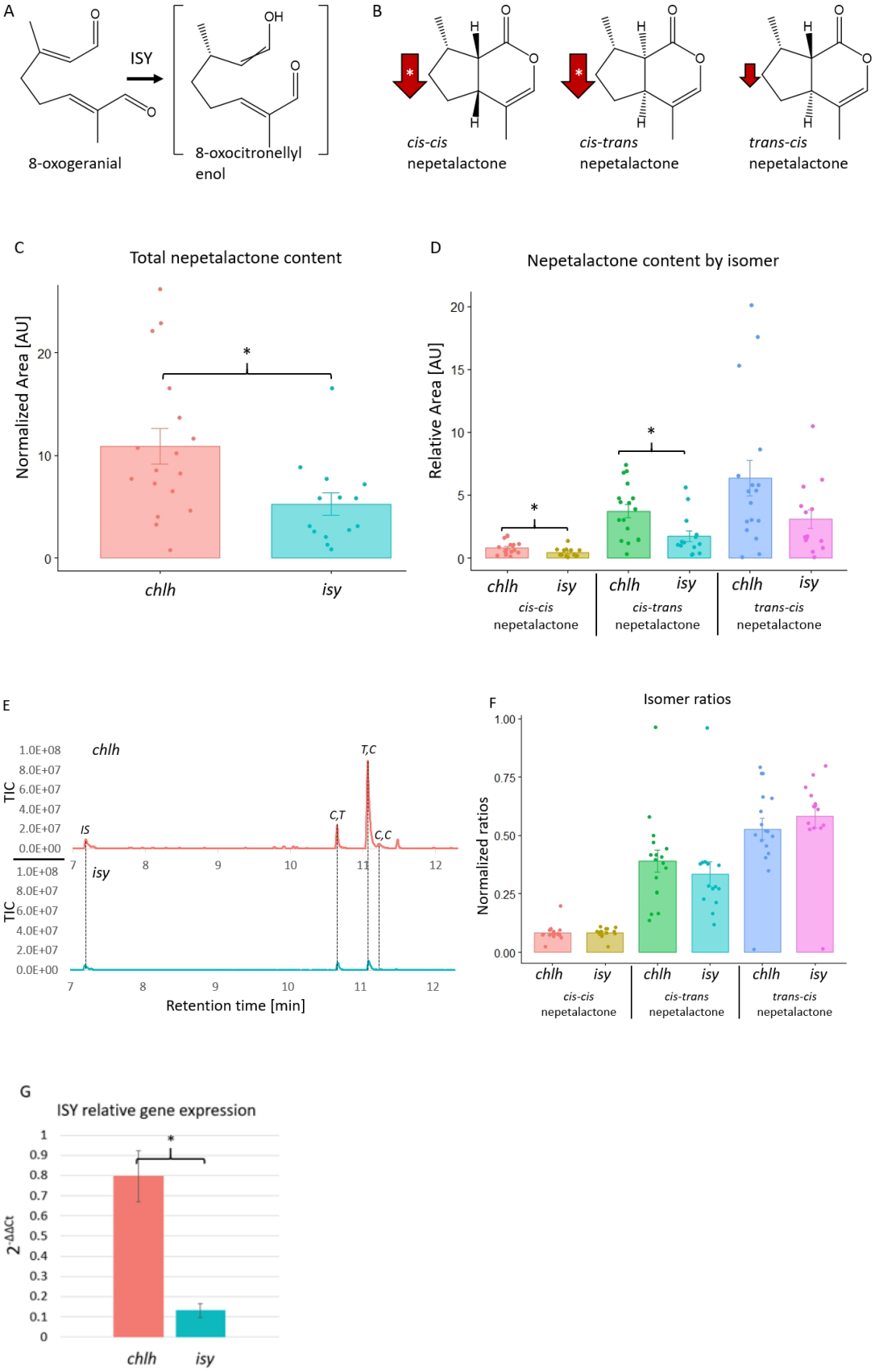

A
ISY
8-oxogeranial
8-oxocitronellyl enol
B
cis-cis nepetalactone
cis-trans nepetalactone
trans-cis nepetalactone
C
Total nepetalactone content
Normalized Area [AU]
chlh
isy
D
Nepetalactone content by isomer
Relative Area [AU]
chlh
isy
cis-cis nepetalactone
cis-trans nepetalactone
trans-cis nepetalactone
E
chlh
isy
IS
C,T
T,C
C,C
TIC
Retention time [min]
F
Isomer ratios
Normalized ratios
G
ISY relative gene expression
2-ΔΔCt
chlh
isy

Figure 4.5: Effect of VIGS on *ISY*

A) *ISY* catalysed reduction of 8-oxogeranial to 8-oxocitronellyl enol intermediate. Red circle denotes the VIGS targeted gene in this reaction. B) is a summary of the changes on nepetalactone isomer production. A downward facing red arrow with a white asterisk indicates a significant decrease in a given isomer. A smaller arrow with no asterisk indicates non-significant changes. C) shows the total nepetalactone content in tissues from pTRV2-*ChlH A. tumefaciens* infected plants (*chlh*) and pTRV2-*ChlH-ISY A. tumefaciens* infected plants (*isy*). Normalized nepetalactone content for each isomer is calculated by dividing the isomer peak area by the internal standard (50 µM camphor) peak area. Total nepetalactone content is the sum of each relative isomer content. The bars indicate the average content of the sum. Each point is an individual sample. Error bars are the standard error. An asterisk above a bracket indicates a t-test p-value <0.05. AU indicates arbitrary units. D) indicates each normalized nepetalactone content (nepetalactone peak area/internal standard peak area). The bars indicate the average content of each isomer. Each point is an individual sample. Error bars are the standard error. An asterisk above a bracket indicates a t-test p-value <0.05. AU indicates arbitrary units. E) is a representative GC-MS trace of *chlh* and *isy*. TIC is total-ion count. Retention time is in minutes. IS is internal standard (50 µM camphor), *C,T* is *cis-trans* neptalactone, *T,C* is *trans-cis* nepetalactone and *C,C* is *cis-cis* nepetalactone. F) shows the ratio of each isomer relative to the total nepetalactone content (100%). The bars indicate the average ratio. Each point is an individual sample. Error bars are the standard error. AU indicates arbitrary units. G) Relative gene expression of *ISY* relative to *UBI9* in *chlh* and *isy* tissue as measured by a qPCR reaction. X-axis is relative gene expression calculated by $2^{-\Delta\Delta Ct}$. Error bars are the standard error. An asterisk above a bracket indicates a t-test p-value <0.05.

4.2.5 *MLPL* knockdown

As previously showed *in vitro* [63], MLPL is needed for the production of *cis-trans* nepetalactone as it carries out the cyclization of 8-oxocitronellyl enol to *cis-trans* nepetalactol (fig. 4.6-A). I expected *MLPL* knockdown to decrease the production of *cis-trans* nepetalactone. Moreover, levels of *trans-cis* or *cis-cis* nepetalactones could increase and we might observe geraniol-derived intermediates.

Bleached tissue from plants infected with *A. tumefaciens* containing pTRV2-*ChlH-MLPL* (*mlpl* in figure 4.6) showed a significant decrease in *cis-trans* nepetalactone produced, as compared to plants infected with *A. tumefaciens* containing pTRV2-*ChlH* (*chlh* in figure 4.6) knockdown (fig. 4.6 B-E). The total nepetalactone content did not significantly change between *chlh* and *mlpl* (fig. 4.6 C), although *mlpl* tissue shows a slight decrease in total nepetalactone content, consistent with the lower production of *cis-trans* nepetalactone. Contrary to expectations, there was not an increase in *trans-cis* or *cis-cis* nepetalactones produced (as evidenced by the sum of these isomers in figure 4.6-D&F). As revealed in the qPCR results in figure 4.6-G, while *mlpl* gene expression is significantly downregulated as compared to *chlh*, there is still residual gene expression (0.131).

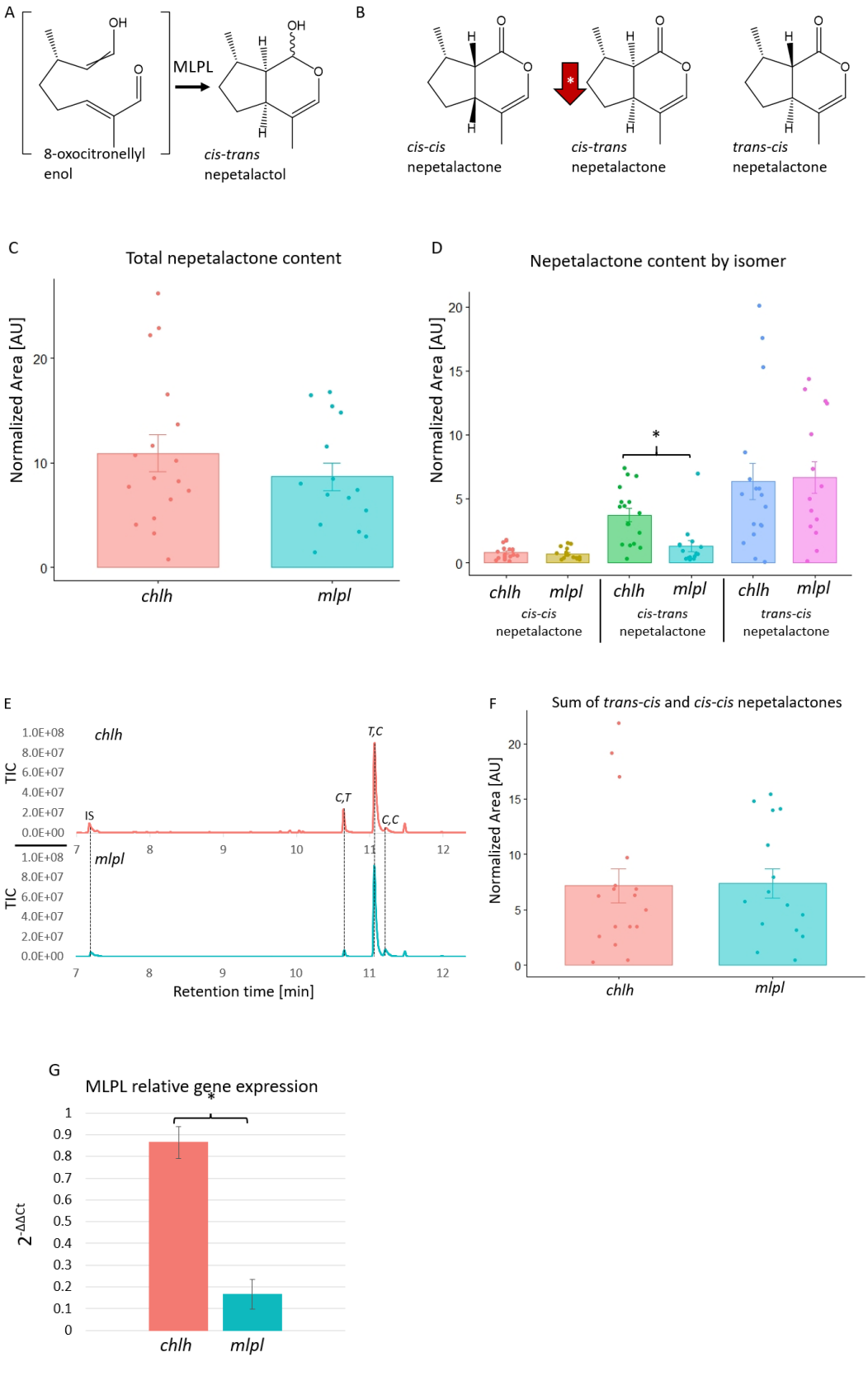

A
OH
O
MLPL
OH
H
H
O
8-oxocitronellyl enol
cis-trans nepetalactol
B
H
H
O
O
*
cis-cis nepetalactone
cis-trans nepetalactone
trans-cis nepetalactone
C
Total nepetalactone content
Normalized Area [AU]
0
10
20
chlh
mlpl
D
Nepetalactone content by isomer
Normalized Area [AU]
0
5
10
15
20
*
chlh
mlpl
cis-cis nepetalactone
cis-trans nepetalactone
trans-cis nepetalactone
E
chlh
T,C
C,T
C,C
IS
mlpl
TIC
1.0E+08
8.0E+07
6.0E+07
4.0E+07
2.0E+07
0.0E+00
7
8
9
10
11
12
Retention time [min]
F
Sum of trans-cis and cis-cis nepetalactones
Normalized Area [AU]
0
5
10
15
20
chlh
mlpl
G
MLPL relative gene expression
*
2-ΔΔCt
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
chlh
mlpl

Figure 4.6: Effect of VIGS on *MLPL*

A) *MLPL* catalysis of 8-oxocitronellyl enol into *cis-trans* nepetalactol. Red circle denotes the VIGS targeted gene in this reaction. B) is a summary of the changes on nepetalactone isomer production. A downward facing red arrow with a white asterisk indicates a significant decrease in a given isomer. C) shows the total normalized nepetalactone content for each condition. Normalized nepetalactone content for each isomer is calculated by dividing the isomer peak area by the internal standard (50 µM camphor) peak area. Total normalized nepetalactone content is the sum of each normalized isomer content. The bars indicate the average content of the sum. Each point is an individual sample. Error bars are the standard error. An asterisk above a bracket indicates at-test p-value <0.05. AU indicates arbitrary units. D) indicates each normalized nepetalactone content (nepetalactone peak area/internal standard peak area). The bars indicate the average content. Each point is an individual sample. Error bars are the standard error. An asterisk above a bracket indicates a t-test p-value <0.05. AU indicates arbitrary units. E) are two representative GC-MS traces of *chlh* (top) and *mlpl* (bottom). TIC is total-ion count. Retention time is in minutes. IS is internal standard (50 µM camphor), *C,T* is *cis-trans* neptalactone, *T,C* is *trans-cis* nepetalactone and *C,C* is *cis-cis* nepetalactone. F) shows the sum of the normalized peak area of *trans-cis* nepetalactone and *cis-trans* nepetalactone in *chlh* and *mlpl*. G) Relative gene expression of *MLPL* relative to *UBI9* in *chlh* and *mlpl* tissue as measured by a qPCR reaction. X-axis is relative gene expression calculated by $2^{-\Delta\Delta Ct}$. Error bars are the standard error. An asterisk above a bracket indicates a t-test p-value <0.05.

4.2.6 *NEPS* target design

NEPS enzymes were discovered to be necessary for the cyclization and/or oxidation in the production of the various nepetalactone isomers (fig. 4.8-A) [64], [65]. These enzymes share a 70-87% nucleotide identity, which provides both the opportunity to target all *NEPS* simultaneously to understand their collective role in the pathway, and the challenge in silencing each gene individually.

I aligned the nucleotide sequences of *NEPS 1, 3, 4* and *5* to look for regions of high and low shared nucleotide identity. To design the vector that would target all *NEPS* (*NEPS1, 3, 4* and *5*) genes simultaneously, I looked for regions with the most overlapping identical nucleotide sequences. However, there was no singular region that contained many overlapping sequences from all *NEPS*. Thus, I designed a vector that would contain a region overlapping *NEPS5/3* and another region containing overlapping *NEPS4/1* to ultimately create a vector

pTRV2-*ChlH-NEPS5/3-NEPS4/1* to increase the possibility of silencing all *NEPS* simultaneously (fig. 4.7).

For single targets, I looked for CDS regions of at least 150bp that contained unique nucleotide sequences, or at least not overlapping with other transcripts by more than 21bp. However, due to the high nucleotide identity among the *NEPS* genes, this proved this to be impossible (70-87% similarity at the nucleotide level). Early attempts at single targets yielded to a downregulation in more than one isomer, and it was impossible to confirm if this knockdown effect was due to a true single knockdown, or, more likely, the down regulation of several *NEPS* simultaneously. To counter this problem, primers were designed targeting the 3'UTR region of these genes, which proved to have much lower similarity. Work on cloning these sequences is ongoing and this strategy is reported in section 4.3.2.

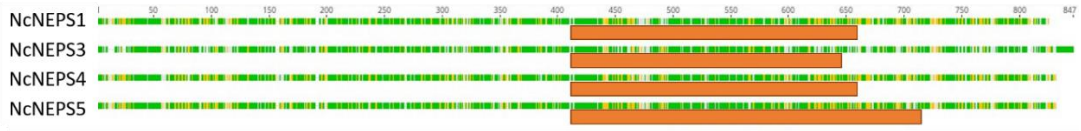


Figure 4.7: *NEPS* alignment

Nucleotide alignment of *NEPS*1/3/4/5. Green indicates 100% identical sites, and yellow indicates conserved sites in some but not all sequences. White spaces are unique nucleotides. In orange is a region of especially high conservation, and primers were designed to target this region.

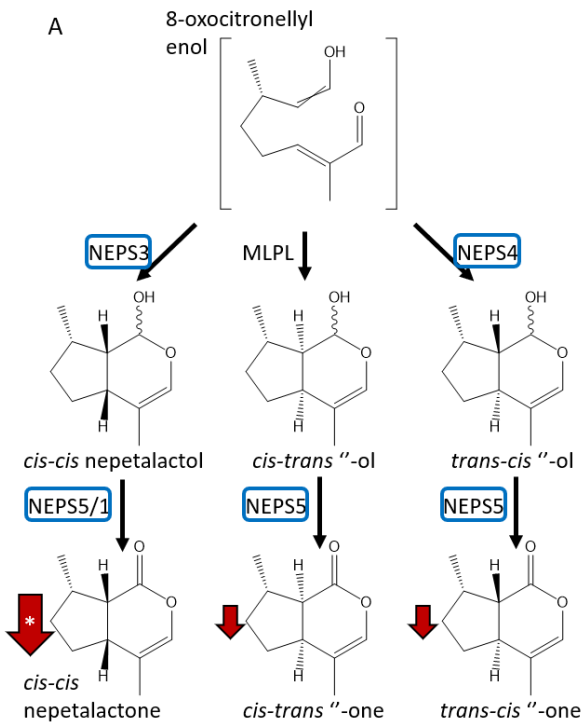
4.2.7 Tandem knockdown of *NEPS* homologues

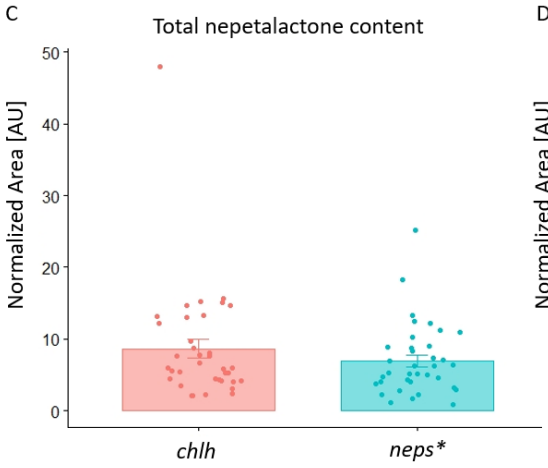
Using the multiple target construct designed in section 4.2.6, I targeted all of the *NEPS* (*NEPS 1, 3, 4,* and *5*) for simultaneous downregulation. This led to a non-significant decrease of *trans-cis* nepetalactone and *cis-trans* nepetalactone production in bleached tissues from plants infected with *A. tumefaciens* containing pTRV2-*ChlH-NEPS5/2/3/-NEPS4/1* (*neps** in figure 4.8) as compared to bleached tissues from plants infected with *A. tumefaciens* containing pTRV2-*ChlH* (*chlh* in figure 4.8). *Cis-cis* nepetalactone was significantly reduced in *neps** (fig. 4.8-C). Total nepetalactone content was non-significantly decreased in *neps** as compared to the total nepetalactone content in *chlh* (fig. 4.8-B).

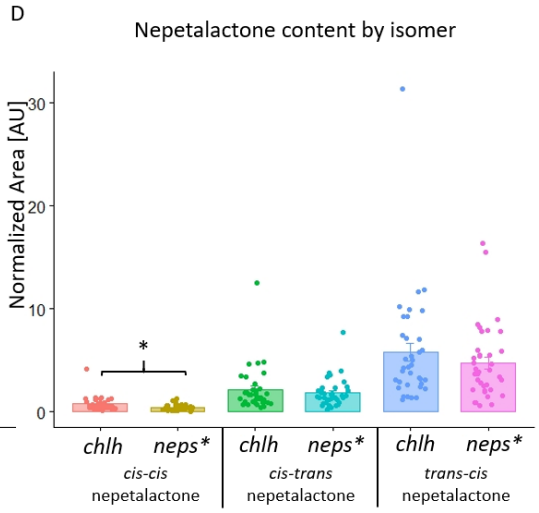
Examining the change in isomer ratios (the ratio of each normalized isomer concentration of the total normalized isomer concentration) (fig. 4.8-F) revealed a decrease in the isomer ratio

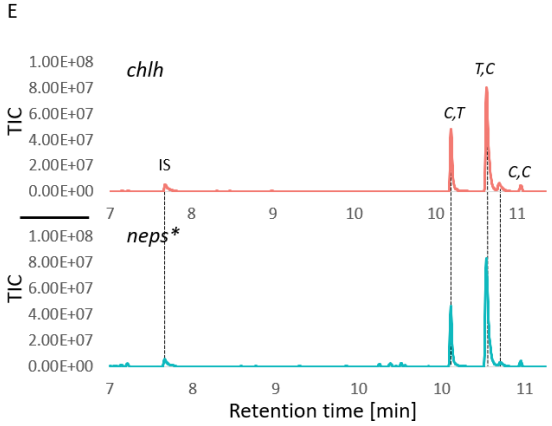
of cis-cis nepetalactone from 0.088 in *chlh* to 0.059 in *neps** (Δ*100=-2.91) of the total isomer content (the sum of all normalized isomer ratios, 1.00). Examining the change in *cis-trans* and *trans-cis* nepetalactone isomer ratios, I wanted to verify if both isomers would increase their respective ratios equally (a change of Δ*100=+1.455 per *cis-trans* and *trans-cis* nepetalactone isomer ratios) to compensate for the decrease of *cis-cis* nepetalactone in the total isomer content. However, I found that *cis-trans* nepetalactone accounted for most of the increased ratio in *neps** (Δ*100=+2.67) as compared to the change in *trans-cis* nepetalactone ratios (Δ*100=+0.24). This could be explained by *cis-trans* nepetalactone synthesis being dependent on *MLPL*, which is not affected by the pTRV2-*ChlH-NEPS5/3-Neps4/1* in tandem VIGS knockdown, while *trans-cis* nepetalactone would be more affected, as it depends on *NEPS4* and *1* homologues for its synthesis.

Unfortunately, the same issues regarding targeting single *NEPS* in the coding region also impairs examining gene downregulation via qPCR analysis. Further analysis via RNAseq will be used to study the downregulation of this particular experimental tissue. This is further addressed in section 4.3.1.


A
8-oxocitronellyl enol
NEPS3
MLPL
NEPS4
cis-cis nepetalactol
cis-trans ''-ol
trans-cis ''-ol
NEPS5/1
NEPS5
NEPS5
cis-cis nepetalactone
cis-trans ''-one
trans-cis ''-one


C
Total nepetalactone content
Normalized Area [AU]
chlh
neps*


D
Nepetalactone content by isomer
Normalized Area [AU]
chlh
neps*
cis-cis nepetalactone
cis-trans nepetalactone
trans-cis nepetalactone


E
chlh
neps*
TIC
IS
C,T
T,C
C,C
Retention time [min]

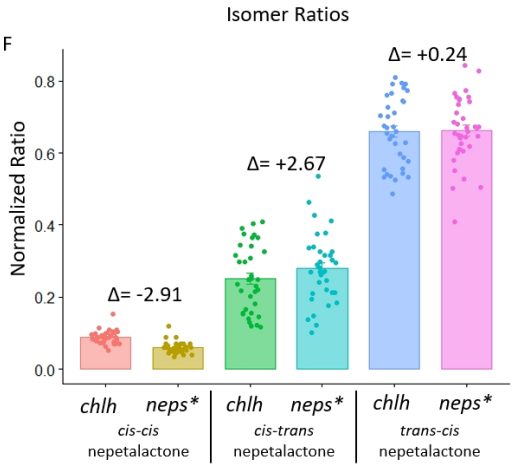

F
Isomer Ratios
Normalized Ratio
Δ= -2.91
Δ= +2.67
Δ= +0.24
chlh
neps*
cis-cis nepetalactone
cis-trans nepetalactone
trans-cis nepetalactone

Figure 4.8: Effect of VIGS on *NEPS* homologues

A) is cyclization and oxidation reactions of NEPS and MLPL. Blue rounded rectangles indicate the VIGS genes simultaneously targeted. B) Total nepetalactone content for *chlh* and *neps** plants. Normalized nepetalactone content for each isomer is calculated by dividing the isomer peak area by the internal standard (50 µM camphor) peak area. Total normalized nepetalactone content is the sum of each normalized isomer content. The bars indicate the average content of the sum. Each point is an individual sample. Error bars are the standard error. An asterisk above a bracket indicates a t-test p-value <0.05. AU indicates arbitrary units. C) Normalized nepetalactone isomer content (nepetalactone peak area/internal standard peak area). The bars indicate the average content. Each point is an individual sample. Error bars are the standard error. An asterisk above a bracket indicates a p-value <0.05. AU indicates arbitrary units. D) Two representative GC-MS traces of *chlh* and *neps**. TIC is total-ion count. Retention time is in minutes. IS is internal standard (50 µM camphor), *C,T* is *cis-trans* neptalactone, *T,C* is *trans-cis* nepetalactone and *C,C* is *cis-cis* nepetalactone. E) shows the ratio of each isomer relative to the total nepetalactone content (1.00). The difference in individual isomer ratios between *chlh* and *neps** (multiplied by 100) is summarized with the delta (Δ). The bars indicate the average ratio. Each point is an individual sample. Error bars are the standard error. AU indicates arbitrary units.

4.2.8 Iridoid pathway regulation and the iridoid gene cluster

Previously characterized bHLH transcription factors CrBIS1/2 were identified as trans-regulators for the iridoid section of the MIA biosynthesis in the medicinal plant C. roseus [87], [96]. However, orthologues of these transcription factors were not found in the *N. cataria* transcriptome upon using BLAST (fig. 4.9). While some orthologues with a short but highly conserved region were found, these are potentially other bHLH transcription factors, a large and highly conserved family across the plant kingdom [108].

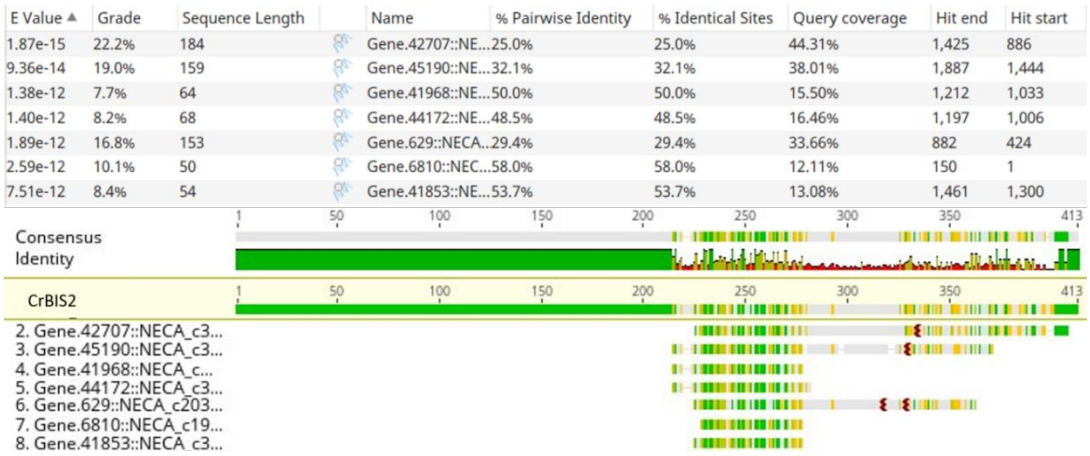
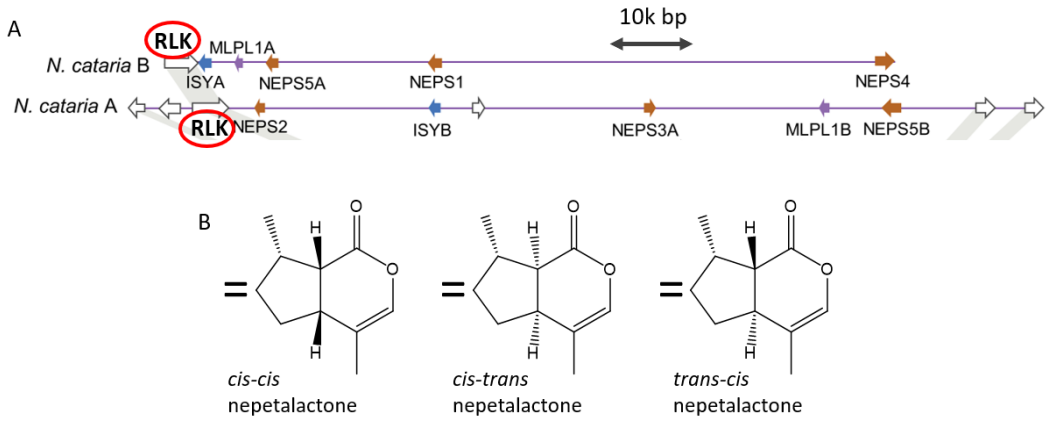


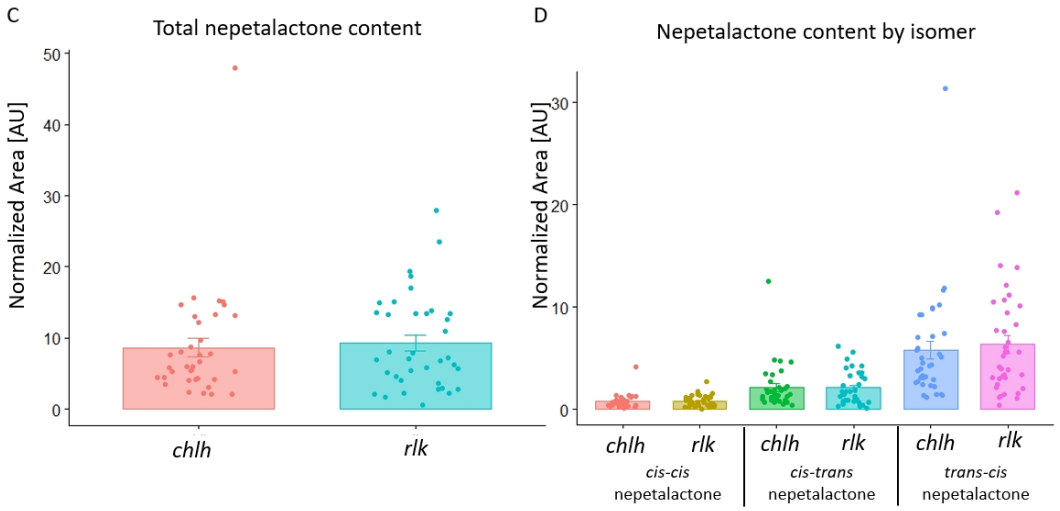
Figure 4.9: BLAST results of CrBIS1/2 homologues in *N. cataria* transcriptome

No candidates with a high nucleotide identity to CrBIS1/2, a class of bHLH transcription factors that regulate the iridoid pathway in *C. roseus*, were found in the *N. cataria* transcriptome. Some conserved regions may belong to the large number of bHLH transcription factors involved in other trans-regulatory processes.

Many genes of the *N. cataria and N. mussinii* iridoid pathway are found in a biosynthetic cluster, as evidenced by inspection of the plants' genomes [63]. Within the *N. cataria* and *N. mussinii* cluster, there is also a *RLK* gene of unknown function (fig. 4.10-A). This class of enzymes are membrane bound kinases and are often involved in signalling cascades in response to abiotic and biotic stressors. Gene expression analysis the *RLK* in *N. cataria* revealed a relatively low expression (fig. 4.3) `in plants grown under normal conditions (described in section 4.5.9). The VIGS experiments described in this Chapter was implemented in plants grown in normal growth conditions and without abiotic or biotic stressors.

Targeting *RLK* for knockdown produced no change in nepetalactone content, suggesting that this kinase does not play a role in the nepetalactone pathway under normal growth conditions (fig. 4.9 B-E). However, it is important to note that this gene is very lowly expressed in these normal growth conditions (fig. 4.3), and that VIGS had a negligible effect on gene expression (fig. 4.10-F). Further proposed work on the *in vivo* characterization in relation to the nepetalactone pathway of this *RLK* can be found in section 4.3.4, and further discussion on regulation candidates for the iridoid pathway in *N. cataria* are discussed in section. 4.3.6.


A
10k bp
N. cataria B
RLK
MLPL1A
ISYA
NEPS5A
NEPS1
NEPS4
N. cataria A
RLK
NEPS2
ISYB
NEPS3A
MLPL1B
NEPS5B
B
cis-cis
nepetalactone
cis-trans
nepetalactone
trans-cis
nepetalactone


C
Total nepetalactone content
Normalized Area [AU]
chlh
rlk
D
Nepetalactone content by isomer
Normalized Area [AU]
chlh
rlk
cis-cis
nepetalactone
cis-trans
nepetalactone
trans-cis
nepetalactone

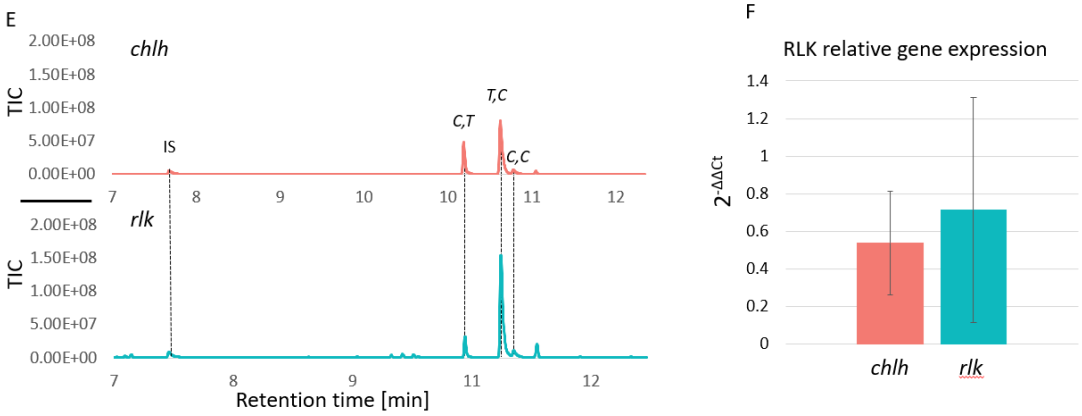

E
chlh
TIC
IS
C,T
T,C
C,C
rlk
Retention time [min]
F
RLK relative gene expression
chlh
rlk

Figure 4.10: Effect of VIGS on *RLK*

A) Location of *RLK* on the *Nepeta* iridoid gene cluster. B) is a summary of the changes on nepetalactone isomer production. Equal signs signify no change. C) shows the total normalized nepetalactone content for each condition. Normalized nepetalactone content for each isomer is calculated by dividing the isomer peak area by the internal standard (50 µM camphor) peak area. Total nepetalactone content is the sum of each relative isomer content. The bars indicate the average content of the sum. Each point is an individual sample. Error bars are the standard error. An asterisk above a bracket indicates a t-test p-value <0.05. AU indicates arbitrary units. D) indicates each relative nepetalactone content (nepetalactone peak area/internal standard peak area). The bars indicate the average content of the sum. Each point is an individual sample. Error bars are the standard error. An asterisk above a bracket indicates a p-value <0.05. AU indicates arbitrary units. E) Two representative GC-MS traces of *chlh* and *rlk*. TIC is total-ion count. Retention time is in minutes. IS is internal standard (50 µM camphor), *C,T* is *cis-trans* neptalactone, *T,C* is *trans-cis* nepetalactone and *C,C* is *cis-cis* nepetalactone. F) Relative gene expression of *RLK* relative to *UBI9* in *chlh* and *rlk* tissue as measured by a qPCR reaction. X-axis is relative gene expression calculated by $2^{-\Delta\Delta Ct}$. Error bars are the standard error. An asterisk above a bracket indicates a t-test p-value <0.05.

4.2.9 Semi-untargeted metabolomics

While GC-MS analysis on the extracted tissue samples were instrumental in measuring the decrease in nepetalactone content as a result of VIGS, these traces revealed no new peaks of the expected substrates, precursors, or intermediate molecules that would accumulate upon knockdown of a metabolic pathway genes. We hypothesized that these missing intermediates or precursors may be modified in the plant, as has been reported previously [109]–[112]. Dudley *et al*. [109] and Dong *et al*. [111], [112] reported the production of derivatised geraniol-based metabolites upon the expression of iridoid pathway genes in *Nicotiana benthamiana*, theorized to be catalysed by endogenous enzymes in the host plant.

In metabolomic analyses, targeted analysis is carried out to detect and quantify specific target molecules, such as the GC-MS analysis of nepetalactones in this Chapter; untargeted analysis is used to explore the unknown metabolic landscape in a sample and is hypothesis generating [113]. Semi-targeted metabolic analysis to detect unspecified geraniol-derived

adjuncts was carried out on the methanol extracts from the collected VIGS samples listed in sections 4.2.4-6, 8, 9 on a non-volatile liquid chromatography-quadrupole time of flight mass spectrometer (LC-QTOF-MS). These methanol extracts were run on negative ion mode as previous work [109], [111], [112] has reported higher resolution on geraniol-based glycosylated or oxidized compounds with negatively charged ions. I set out to identify geraniol-based monoterpenes that may be modified precursors as a consequence of disrupting the iridoid pathway via VIGS. All runs were then processed through the Bruker Metaboscape software (2021-B) for analysis in metabolite peak changes above the arbitrary cut off of 100,000 ($10x^5$) counts.

This metabolomic analyses revealed the upregulation of a compound with a retention time of 4.22 minutes and a mass to charge ratio (*m/z*) of 345.1561, found only in *isy* tissue (fig. 4.11-A, B). We tentatively assigned this compound to be a formic acid adduct of a pentose conjugated hydroxygeranial (C15H24O6-COOH) which has an exact mass of 345.1549 (Δppm of 3.48) (fig. 4.11-B). In *neps** knockdown tissue, an increase in two compounds, one with retention time 3.99 and *m/z* of 359.1353, and one with retention time 4.03 and *m/z* 719.2777 were significantly increased (fig. 4.10-C, D). We suspect these to be the monomer and dimer forms of the same metabolite. We tentatively assigned this compound to be either a formic acid adduct of a pentosyl hydroxy-dioxogeranial (C15H22O7-COOH) with an exact mass of 359.1342 (Δppm of 2.78) or, less likely, a hexyl carboxygeranic acid (C16H24O9) with exact mass of 360.1420 (Δppm of -2795.3). Rigorous characterization is impaired due to the lack of standards available for some of these compounds, and the amount produced in each plant is too low for collection, isolation, and characterization.

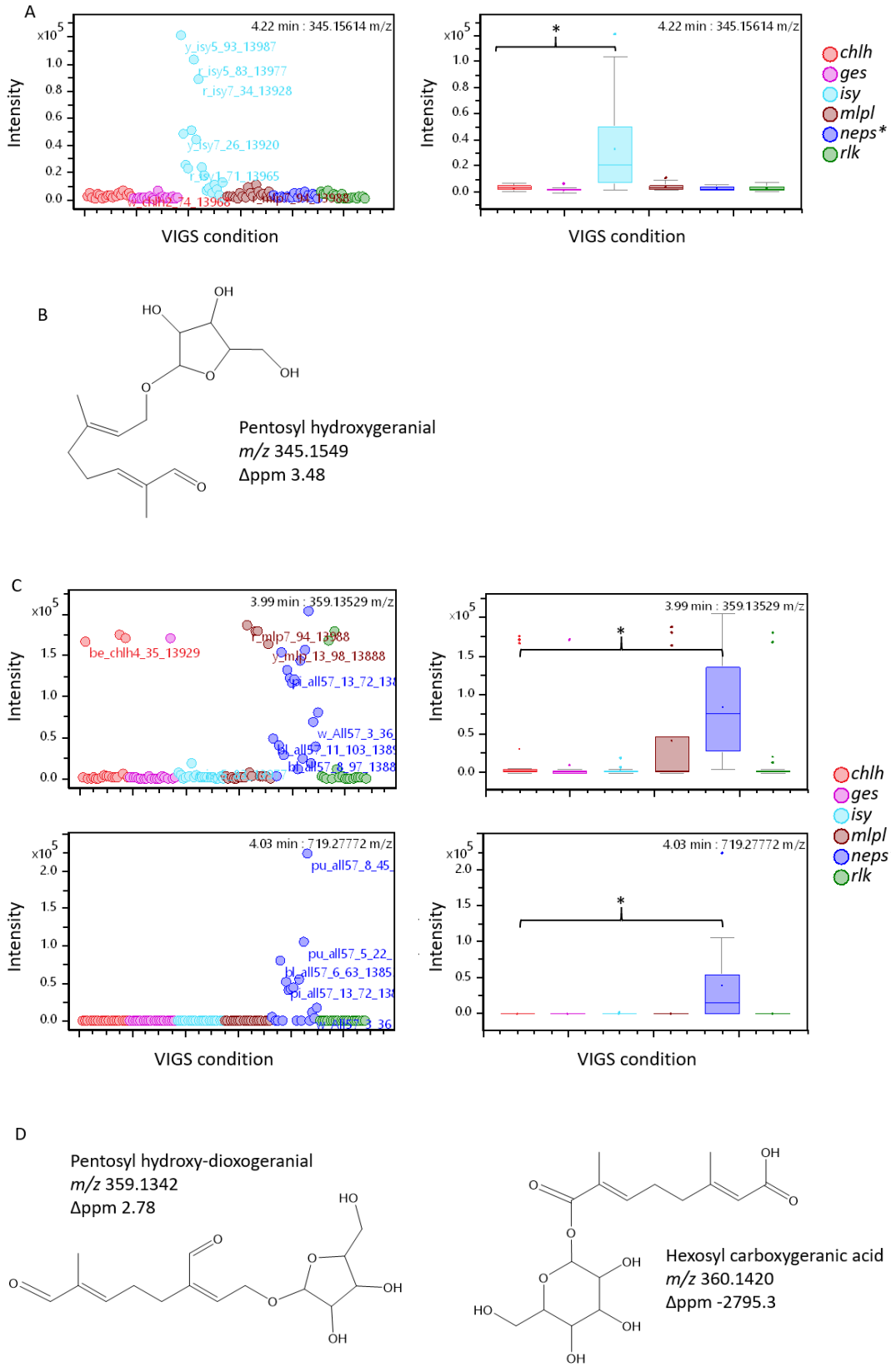

A
4.22 min : 345.15614 m/z
Intensity
VIGS condition
chlh
ges
isy
mlpl
neps*
rlk
B
Pentosyl hydroxygeranial
m/z 345.1549
Δppm 3.48
C
3.99 min : 359.13529 m/z
4.03 min : 719.27772 m/z
D
Pentosyl hydroxy-dioxogeranial
m/z 359.1342
Δppm 2.78
Hexosyl carboxygeranic acid
m/z 360.1420
Δppm -2795.3

Figure 4.11: Semi-untargeted analysis of VIGS samples

Intensity distribution by VIGS sample and boxplots of bucket statistics from Metaboscape (2021-B). Two high intensity (>$10x^5$) molecules with a significantly different intensity change are shown. Left plots in A and C show the distribution of intensities (Y-axis) by gene targeted (legend) in each VIGS condition (X-axis) and right plots are boxplots of these intensities. Boxplots show median (line) and average (dot) values, and 25% and 75% quartiles. Grey whiskers indicate the lower and upper extremities. Coloured circles outside boxes are outliers. Asterisk over a bracket indicates a statistically significantly difference (t-test p-value <0.05). Information in upper right corner indicates the retention time and the observed mass-to-charge ratio for the metabolite studied. A) shows a metabolite significantly upregulated in *isy* samples. B) shows a possible structure for the tentative pentosyl hydroxygeranial (formic acid adduct is not shown). Exact mass (345.1549) was calculated using the proposed molecular formula ($C_{15}H_{24}O_6$-COOH). Difference in parts per million (Δppm) was calculated by the following formula: $\Delta ppm = 10^6((M_{observed} - M_{theoretical})/M_{theoretical})$. C) shows the monomer (top) and tentative dimer (bottom) molecule that is significantly upregulated in *neps** samples. D) shows a possible structure for the tentative pentosyl hydroxy-dioxogeranial (formic acid adduct is not shown) (left) and hexosyl carboxygeranic acid (right). Exact masses (359.1342 & 360.1420) were calculated using the proposed molecular formulas ($C_{15}H_{24}O_6$-COOH & $C_{16}H_{24}O_9$). Difference in parts per million (Δppm) was calculated by the following formula: $\Delta ppm = 10^6((M_{observed} - M_{theoretical})/M_{theoretical})$. Tentative identification was carried out with the aid of Lorenzo Caputi.

4.2.10 The glycosylated iridoid 1,5,9 epi-deoxyloganic acid

One aspect of this pathway that has not been fully explored is the presence of glycosylated iridoids in *Nepeta*, as reported in Lichman *et al*. [63], and reported in other Lamiaceae plants [11], [114]–[116]. Previously studied iridoid pathways in other plants [12], [41], [42] report that an iridoid oxidase/7-deoxyloganetic acid synthase is required to oxidize nepetalactol into 7-deoxyloganetic acid which is then glycosylated by a glucosyl transferase (UGT). As described in Chapter 2, section 2.2.1, no iridoid oxidase candidate was found in *Nepeta*. However, Omar Kamileen in our group has identified several UGTs that may be involved in the production of glycosylated iridoids such as 1,5,9 epi-deoxyloganic acid [unreported], and which may add the sugar moiety to the hydroxyl group of nepetalactol, or onto a further oxidized form of the molecule.

Methanol extracts from the above described (4.2.9) VIGS samples were visualized on the LC-QTOF-MS to probe for changes in the production of 1,5,9 epi-deoxyloganic acid (fig. 4.12). LC-QTOF-MS analysis of *ges* knockdown samples revealed a significant decrease in the production of the tentative dimer (*m/z* 719.27776) of the glycosylated iridoid 1,5,9 epi-deoxyloganic acid (fig. 4.12-E). While the monomer (*m/z* 359.13522) showed a decrease, this was non-significant (fig. 4.12-D). Upon knockdown of the other genes in the iridoid pathway (*ISY, MLPL* & all *NEPS*), no downregulation of 1,5,9 epi-deoxyloganic acid was found.

This does not necessarily suggest the nepetalactone iridoid pathway is not involved, or involved to a lesser extent, in the production of glycosylated iridoids. The volatile nepetalactones are produced in trichomes, as reported in Lichman *et al.* [64]. Research into trichomes and secondary metabolites report that these plant specialized organs will often sequester secondary metabolites, especially volatile metabolites. Release from the trichome is often towards the exterior of the plant, rather than into transportation within the plant [117], [118]. As such, there is not much predicted transfer of these nepetalactones between tissues or within the plant. Hence, using the bleaching visual marker *ChlH* to collect specific affected trichome rich leaf-tissue can result in clear downregulation of the downstream products of these pathways synthesized or stored in the trichomes [43]. However, glycosylated metabolites are soluble and often glycosylation is used to mediate transportation of secondary metabolites throughout the plant [13], [119], [120]. Therefore, it is possible that non-affected tissue may be contributing to the levels of 1,5,9 epi-deoxyloganetic acid seen in bleached tissue with *GES, ISY, MLPL* or *NEPS 1, 3, 4* & *5* knockdown. Stable transformants knocking out *GES* or *ISY*, for example, would provide better evidence of whether the nepetalactone pathway is responsible for the formation of glycosylated iridoids found in *Nepeta* tissue, or if another pathway may be responsible.

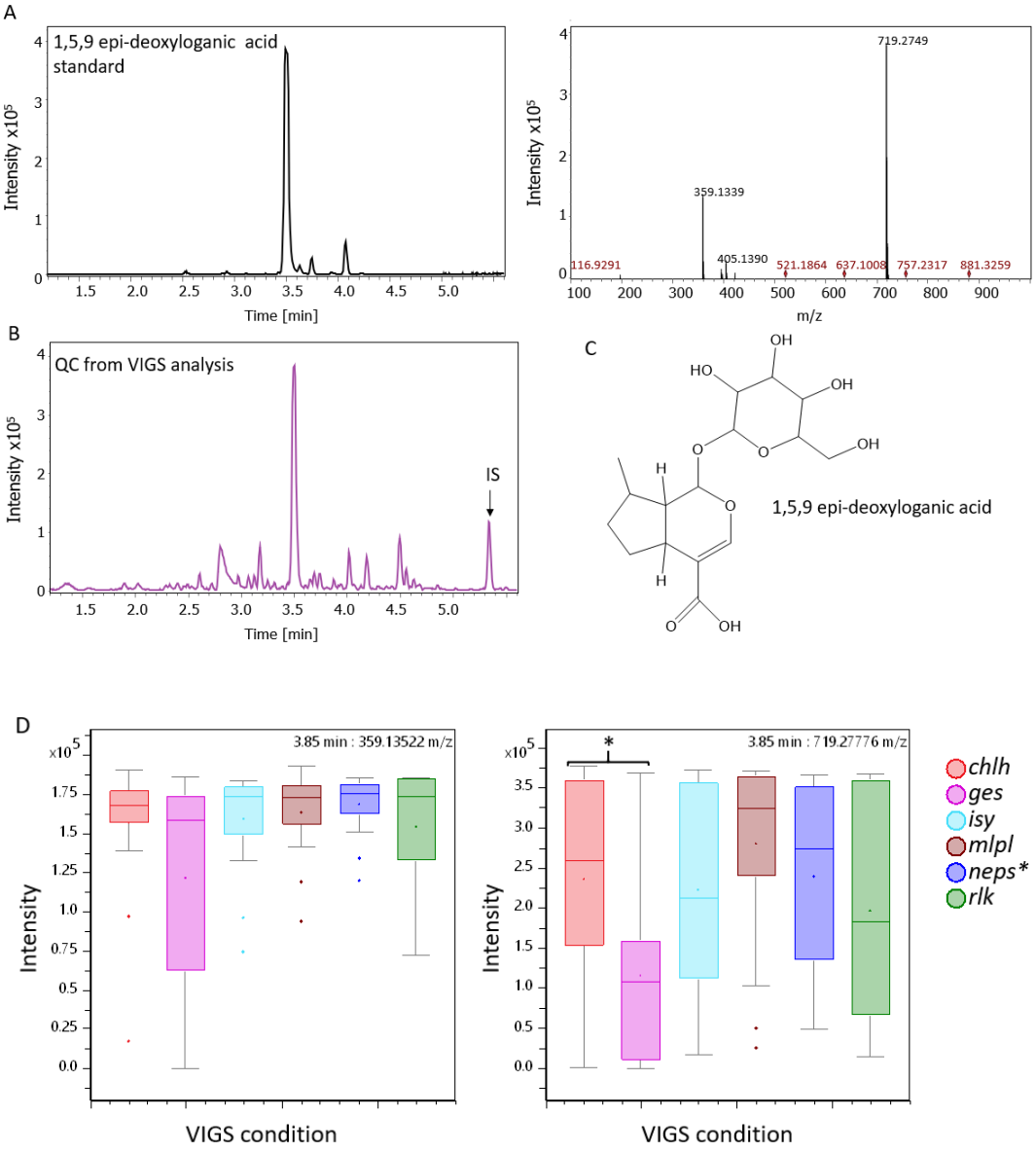


Figure 4.12: Nepetalactone biosynthetic genes and 1,5,9 epi-deoxyloganic acid concentration

A) shows the LC-QTOF-MS trace and *m/z* distribution of 1,5,9 epi-deoxyloganetic acid standard. Standard was isolated by Omar Kamileen from *Nepeta cataria* and reported in [63]. B) shows a quality control (QC) sample of the VIGS methanol extract. QC was made by taking a 10 µL aliquot of each methanol sample in the VIGS experiments. IS is the internal standard 10 µM harpagoside. C) shows the structure of 1,5,9 epi-deoxyloganic acid as reported in [63]. D) is the intensities (Y-axis) by gene targeted (legend) in each VIGS condition (X-axis). Boxplots show median (line) and average (dot) values, and 25% and 75% quartiles. Grey whiskers indicate the lower and upper extremities. Coloured circles outside boxes are outliers. Asterisk over a bracket indicates a statistically significantly difference (t-test p-value <0.05).

Information in upper right corner indicates the retention time and the observed mass-to-charge ratio for the metabolite studied.

4.2.11 Drawbacks of VIGS as a functional genomics tool to study the iridoid pathway

As revealed in some of the above results, namely the non-significant reduction on nepetalactone content upon knocking down *ISY* or all *NEPS* coupled with qPCR data revealing a knockdown and not complete silencing of the targeted genes suggest that although VIGS is a useful tool to study nepetalactone metabolism, its silencing is incomplete, which may lead to results that are difficult to interpret as compared to a fully stable gene knockout. This incomplete silencing leads to certain ambiguities that lay open further questions, such as whether there are alternative contributors to a certain metabolite (in section 4.3.4) or transport of metabolites between non-affected and affected tissue (in section 4.2.10). As described in Chapter 3, *N. cataria* and *N. pogonosperma* have been showed to be susceptible to the use of *A. rhizogenes* [59], [79], [80] as a transformation method on excised stems and leaves, as well as co-cultured seedlings. Preliminary work into establishing protocols to generate *Nepeta* calluses amenable to gene editing transformation techniques, and regeneration of said calluses into plants, is being carried out in the group, as addressed in Chapter 3.

4.3 Future work

4.3.1 Transcriptomics

While qPCR is a useful tool for measuring the gene expression for individual genes and serves to corroborate the VIGS knockdown effect on nepetalactone concentration in bleached tissues, it is not practical to use for measuring gene expression of genes with a high degree of nucleotide identity, as is the case with the *NEPS* homologues. Transcriptomic analysis on *neps** samples would yield gene expression data on individual *NEPS*, identifying the extent of which each *NEPS* gene is downregulated.

A further advantage to transcriptomic analysis on these experimental tissues is to aid in identifying other genes that may be involved in the pathway, for example, those that may be involved in regulating the pathway expression, and to understand how downregulating one gene may affect the rest in the pathway. For example, upon the knockdown of *MLPL*, there was not an increase in *trans-cis* and *cis-cis* production as initially expected. We want to test

if this is a result of a loss of the unstable substrate available for the other enzymes, or if there is a compensation in downregulating the genes involved in producing *trans-cis* and *cis-cis* nepetalactones.

As described in Sections 4.2.9 and 4.2.110, downregulation of certain genes such as *ISY* and the *NEPS* led to the formation of putative glycosylated geraniol-based monoterpenes. As reported before [109]–[112], plants may deploy glucosyl-transferases and other modifying enzymes to deal with unexpected metabolites. By generating the transcriptome of *isy* sample mRNA, for example, we can begin to understand some of the mechanisms in which plants deal with unexpected metabolites by a targeted search into upregulated glucose transferases.

4.3.2 Silencing individual *NEPS* homologues via VIGS

Knocking down *NEPS1, 3, 4* and *5* in-tandem yielded some understanding of their role in nepetalactone isomer production. Nepetalactone isomers were differentially affected by the downregulation of the *NEPS* genes, such as *cis-cis* nepetalactone being significantly decreased, and while both *trans-cis* and *cis-trans* were non-significantly downregulated, the ratio between these isomers did differ from the control *chlh*, suggesting that the roles NEPS enzymes have collectively behind each isomer are unequal. Knocking down individual *NEPS* genes would clarify the individual roles each enzyme has in the pathway. As of the most recent publication on this pathway [63], the roles of NEPS5 and 1, dehydrogenases responsible for producing nepetalactone from nepetalactol, have not been clearly established. While we know from *in vitro* data that both NEPS1 and NEPS5 oxidize nepetalactol to nepetalactone for the *cis-cis* isomer, *cis-trans* isomer and *trans-cis* isomer, gene expression data from *N. mussinii* favoured NEPS5 as the dehydrogenase for *cis-trans* nepetalactol and NEPS1 as the dehydrogenase for *trans-cis* nepetalactol [63]. We therefore do not know if these enzymes specifically act on certain isomers over the others, or if they are more promiscuous *in planta*. This is further discussed in section 5.2.1.

Many VIGS constructs target the 3' untranslated region (3' UTR) region of a gene rather than the CDS region [47], [52], [53]. This region plays a key role in gene-expression and is usually non-homologous to other CDS regions, even when the coding sequences are nearly identical. Predicted UTR regions of the *NEPS* genes were identified using the *N. cataria* genome generated by the Mint Genome Project, aligned against each other to exclude any regions

similar to each other. Target sequences were verified against the genome and transcriptome to ensure no off-target silencing would incur, as described in Chapter 3. Primers were designed to amplify these regions. However, amplification under various permissive PCR conditions proved difficult, and only one out of the four *NEPS* genes was amplified. This may be due to differences between the reported genome and our cultivars. Work is still on-going to obtain verified UTR target sequences for the *NEPS* genes in order to silence them individually.

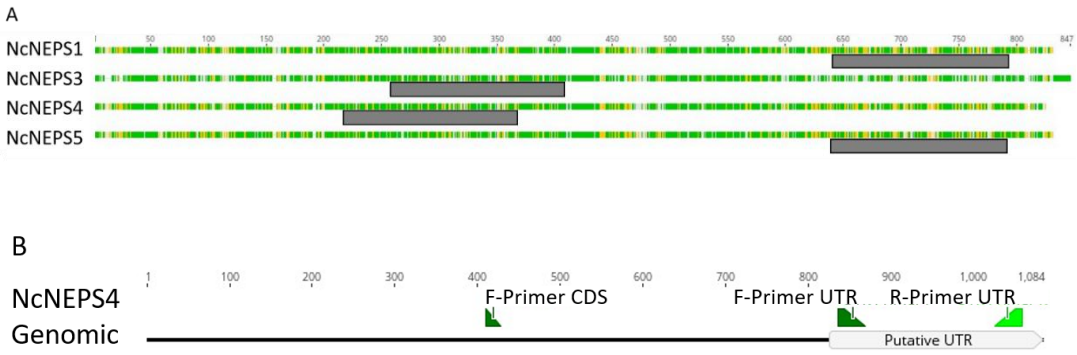


Figure: 4.13: *NEPS* homologues single VIGS silencing strategies

A) shows the nucleotide alignment of *NEPS1/3/4/5* with an early unsuccessful attempt at identifying relatively low areas of conserved nucleotide identity. Green indicates 100% identical sites, and yellow indicates conserved sites in some but not all sequences. White spaces are unique nucleotides. In grey are the regions selected for single targeting. B) shows the cloning strategy to amplify the putative 3'UTR region for single UTR targeting. An initial PCR used F-Primer CDS paired with R-Primer UTR. A second PCR was then performed using F-Primer UTR and R-Primer UTR as primer pairs. Only amplification of the *NEPS4* UTR region was possible with his technique. Further UTR regions will be synthetically produced.

4.3.3 Iridoid synthase and progesterone-5-beta reductase

As reported in [11] and in section 2.1.1, *ISY* evolutionary history involved the duplication and specialization of a P5βR ancestor, with some iridoid synthase activity, into a fully specialized iridoid synthase. Further predicted ancestral enzyme activities in [63] revealed the steps taken to fully specialize into becoming an iridoid synthase, losing P5βR activity. While *Nepeta ISY* became fully specialized, the P5βR homologues in *N. cataria* retained some iridoid synthase activity (fig. 4.14). Knocking out P5βR activity in *N. cataria* to test for its effect on nepetalactone content would reveal the influence, albeit small, this enzyme may have on

nepetalactone production. Furthermore, while silencing *ISY* resulted in a general downregulation of nepetalactone end product, its knockdown effect was not as severe as knocking down *GES* activity. One potential hypothesis would be a redundancy on behalf of P5βR enzymes. Comparing the extent of knockdown from a pTRV2-*ChlH-ISY-P5βR* construct against the effect of knockdown of the pTRV2-*ChlH-ISY* and pTRV2-*ChlH-P5βR* constructs would aid in this understanding.

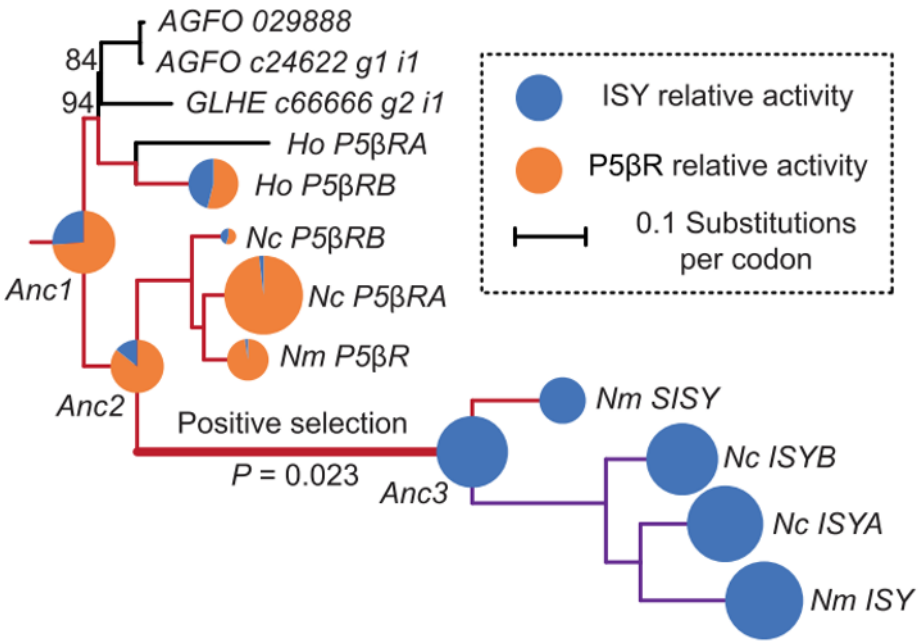


Figure: 4.14: The evolution of *ISY* and *P5βR* activity in Nepeta

Figure adapted from Lichman *et al.* [63]. Ancestrally predicted and extant *ISY* and *P5βR* homologues were tested for their relative respective activities.

4.3.4 RLK activity under stress conditions related to nepetalactone biosynthesis

Reported in section 4.2.8, targeting *N. cataria RLK* via VIGS did not lead to either a significant downregulation of the gene, which is already lowly expressed, nor to any changes in the nepetalactone metabolism. Many reported RLKs play a role in plant immunity, often becoming active in situations of abiotic or biotic stress [104]–[107]. It may be that this *Nepeta* cluster *RLK* is only activated in stress conditions. To test the function of this gene, gene expression should be measured under stress conditions, for example, upon infection with a common Lamiaceae pathogen [121]. If upregulated in stressful conditions, VIGS could then be used to silence the activated *RLK* gene, and the nepetalactone content of these

stressed plants could be then analysed to test for the involvement of the iridoid pathway under biotic stress conditions.

It is also possible that this *RLK* is uninvolved in the iridoid pathway. The *RLK* found in the *Nepeta* cluster is a Leucine-rich repeat RLK, according to its homology-based assignment. From the NCBI database, it has an 88% identity with *Salvia splendans* LRR-RLK FEI 1. FEI proteins (named after the Chinese word for fat) were found to play a role in cell-wall biosynthesis [122].

4.3.5 Trichome morphology and VIGS impact

As nepetalactones are specifically produced in Nepeta trichomes [64], we wondered if downregulating the pathway would influence trichomes morphology or number. Using light microscopy, I examined trichome structure and relative density on leaves of several ages from a *neps** knockdown and a WT leaf. *Nepeta* trichomes seem to be glandular multicellular trichomes.

Neither trichome morphology nor trichome density seemed to change on the preliminary data set of microscopy photos that were acquired from WT and *neps** leaves (fig. 4.15). However, further work into defining leaf quadrants, acquiring replicates and robustly testing for morphology and density change in different knockdown samples is still needed.

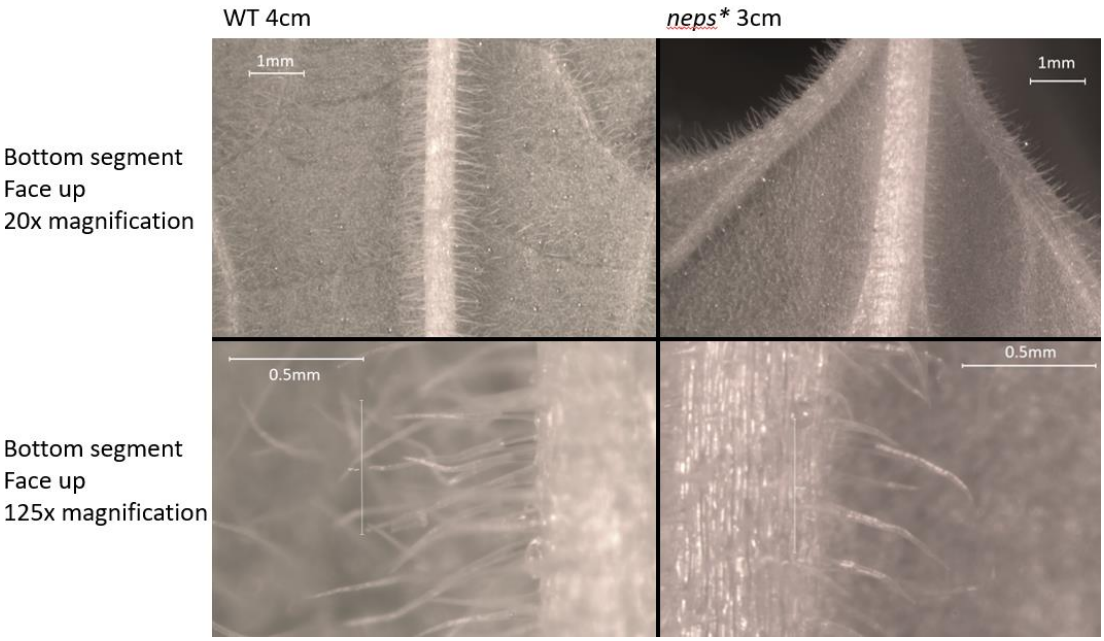


Figure 4.15: *N. cataria* trichome morphology

Microscopy images of WT and bleached *neps** tissue. Leaves were taken 6 weeks post infection. Images taken in stereomicroscopy (M205FA, Leica).

4.3.6 Candidates for iridoid pathway regulation in *N. cataria*

As addressed in section 4.2.8, a homology search for the previously characterized iridoid pathway transcription factors BIS1/2 [87], [96] in the *N. cataria* transcriptome did not yield any candidates with high query coverage nor high percent sequence identity. However, regulation of other monoterpene pathways have been reported, including MYB transcription factors responsible for monoterpene regulation in *Mentha spicata* (spearmint) and *Lavandula spp.* (lavender) and the WRKY transcription factor in *Vitis vinifera* (grape vine) (fig.4.16) [123]–[125]. It is possible that *N. cataria* has co-opted another monoterpene regulation machinery for iridoid production. These mentioned transcription factors have been shown to work as trans-regulators. A homology-based search using the Lamiaceae transcription factors as a query to find *N. cataria* homologues that would be co-expressed with the iridoid pathway genes would be good candidates.

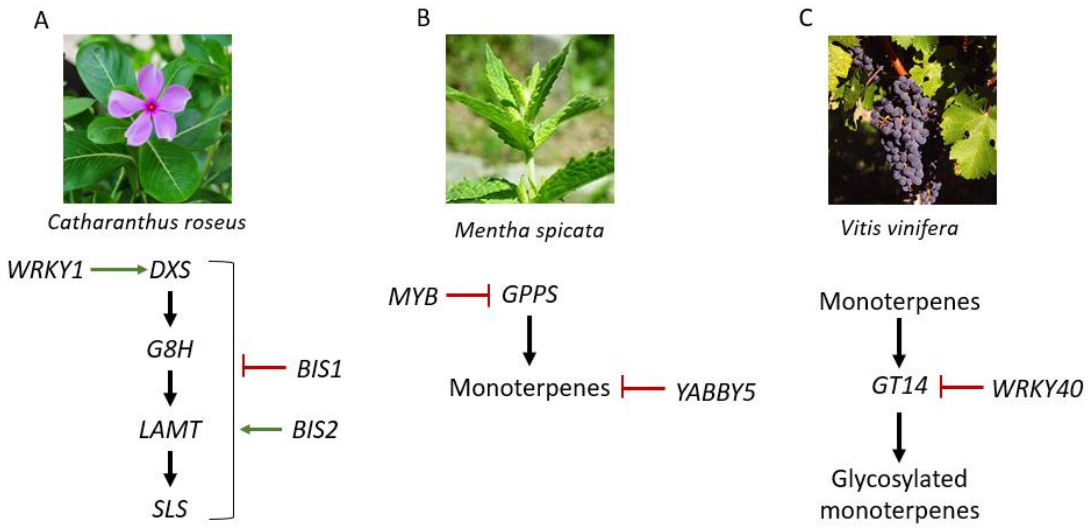


Figure 4.16: Monoterpene regulation systems in example plants

Basic schematic of some known regulatory systems for monoterpene biosynthesis. In A), *C. roseus*, WRKY1 regulates the MEP pathway (deoxyxylulose 5-phosphate synthase, (*DXS)*) leading up to iridoid biosynthesis, and BIS1/2 regulate the iridoid part of the MIA pathway (G8H, loganic acid O-methyltransferase (LAMT) and secologanin synthase (SLS)). B) shows what is known about the regulation of monoterpene production in *Mentha spicata*. In C), the relationship between the transcription factor WRKY40 and the production of glycosylated monoterpenes found in *Vitis vinifera*. Plant images are from Wikimedia commons.

As described in Chapter 3, VIGS has been previously used to study the secondary metabolism in various non-model plant systems (10, 39-41, 49)[47], [52]–[54], [87]. While many of these reports focus on the enzymes involved in the biochemical steps of the pathway, there are also reports which target the regulatory elements of these pathways, using VIGS as a functional genomics tool to characterize them. In Patra *et al*. [126] and Van Moerkercke *et al*. [87], the authors identified several key regulatory genes, including the transcription factor BIS1 [85] and the upstream regulator coronatine insensitive 1 (COI1) F-box protein [124], controlling the MIA pathway in *C. roseus* and used VIGS to partially validate their gene functions *in vivo*. Other examples of VIGS as a functional genomics tool to study regulatory elements include the identification of the transcription factor nuclear factor Y's (NF-Y) role in pathogen resistance in *Manihot esculenta* (cassava) [127], the role of the transcription factor Teosinte Branched 1, Cycloidea, and Proliferating4 Cell Factors (TCP) in *Fragaria vesca* (woodland strawberry) in fruit ripening [128] and the characterization of the RLKs Receptor-like protein RE02 from *Nicotiana benthamiana* [104] and the cysteine rich receptor like kinase 2 (CRK2) from *Triticum aestivum* (wheat) [129] and their roles in pathogen triggered

immune responses. Based on these reports, and the validation of VIGS as a tool to study the nepetalactone metabolism in Chapter 3 and 4 of this thesis, VIGS could be used in *N. cataria* to quickly screen for iridoid pathway regulator candidates.

4.4 Conclusion

VIGS has proven to be a useful and quick tool to study iridoid metabolism in *N. cataria. GES, ISY*, *MLPL* and the *NEPS* homologues knocked down using VIGS proved to be a part of the iridoid metabolism as expected from their *in vitro* roles by the effect their silencing had on the iridoid content in the infected plants [64], [65].

Silencing of *GES* by VIGS led to a significant decrease in the production of all nepetalactone isomers and the 1,5,9 epi-deoxyloganic acid dimer consistent with an early shut down of the pathway. Silencing *ISY* led to a significant decrease in *cis-cis* nepetalactone and *cis-trans* nepetalactone, and a non-significant decrease of *trans-cis* nepetalactone, the most abundant isomer in the *N. cataria* cultivars used in this work. This effect is consistent with *ISY* being further down the pathway, where residual activity could still produce sufficient 8-oxogeraniol to supply the final pathway steps. Furthermore, silencing *ISY* produced a tentatively assigned glycosylated hydroxygeranial, possibly derived from excess 8-oxogeranial in the affected tissue. Silencing *MLPL* lead to the expected decrease of *cis-trans* nepetalactone content. Finally, silencing all *NEPS* (*NEPS1, 3, 4* and *5*) homologues simultaneously led to a significant decrease in *cis-cis* nepetalactone, the least abundant isomer in the cultivars used, and to a statistically non-significant decrease in *cis-trans* nepetalactone and *trans-cis* nepetalactone. Levels of trans-cis were reduced more than cis-trans nepetalactone. This is consistent with the fact that cis-trans is also produced by MLPL and therefore the depletion of NEPS enzymes may have less of an impact on its accumulation. Silencing the member of the iridoid gene cluster *RLK* led to non-significant alteration of the nepetalactone content produced under normal growth conditions. However, the role *RLK* may play in abiotic or biotic stress conditions and its effect on nepetalactone content was not a part of this thesis.

This system provides an example of how plants neutralize unexpected metabolites, such as the pathway intermediates formed upon silencing *ISY* and the *NEPS* enzymes, an area of interest in plant metabolic engineering [109]–[112].

The results from these VIGS experiments also raise the question: why is *trans-cis* nepetalactone seemingly less affected by iridoid pathway genes (*ISY* and *NEPS* homologues) knockdown as compared to the other isomers found in *N. cataria*? Given this isomer's abundance in the plant, it could be that VIGS may not be a powerful enough off-switch to have a significant effect on the downregulation of *trans-cis* nepetalactone production. Gene expression analysis on *ges, isy* and *mlpl* plants revealed an incomplete silencing of *GES*, *ISY* and *MLPL* respectively. This is further addressed in section 4.2.11. Furthermore, given the similarity between the *NEPS* sequences making it difficult to initially target individual *NEPS* homologues, I was not able to directly test the effect downregulating *NEPS4*, the enzyme shown to be responsible for cyclizing 8-oxocitronellyl enol into *trans-cis* nepetalactol. Approaches to target individual *NEPS* genes for silencing is further detailed in section 4.3.2.

4.5 Materials and Methods

4.5.1 VIGS insert design

Sequences of 350 bp in length were obtained from *ChlH*, *GES, ISY, MLPL and RLK* that did not contain any exact matches longer than 20 bp to other parts of the genome. Sequences for the *NEPS* homologues were designed as described in section 4.2.7. Sequences were reviewed to be free from BamHI and XhoI restriction sites. Primers were designed containing the palindromic sites for BamHI (forward primer) and XhoI (reverse primer), to ligate to linearized pTRV2 vector, and the first and last 15 to 25 nucleotides with an annealing temperate close to 55 °C.

To construct the dual silencing vector, 200-350 bp fragments of genes of interest were selected and checked to not contain any exact 20 bp matches with other parts of the genome. Primers were designed to amplify these regions and add an overhang to the pTRV2 vector.

4.5.2 Obtaining plasmid gene inserts

Gene fragments for VIGS were obtained both by amplification of our stock of plasmids containing our genes of interest and cDNA in case if no stock being present. Primers for amplification included a nucleotide extension which matched to the pTRV2 vector. Primers were designed to be between 15 and 25 nucleotides long and with an annealing temperature close to 55 °C.

4.5.3 cDNA production for insert amplification

RNA from young leaf tissue from *N. cataria* was retrotranscribed using Invitrogen Superscript IV. In a 10 µL reaction, up to 500ng (about 1 µL) of RNA, 4 µL of SSIV buffer, 1 µL of SSIV reverse transcriptase (200 U/µL) and 1 µL of 100 mM DTT. Incubated at 50 °C for 10 minutes and then at 80 °C for 10 minutes to inactivate. RNAase H (1 µL) was added and incubated at 37 °C for 20 minutes. cDNA was stored at -20 ˚C. Inserts from cDNA were amplified by PCR in a 25 µL reaction using 1 µL of cDNA or plasmid-base, 12.5 µL of Invitrogen Platinum Superfi DNA polymerase mix, 1.25 µL of 10 µM Forward and Reverse primers each, and 5 µL of GC enhancer. Primers for each reaction can be found in primer list.

4.5.4 Vector construction

BamHI, XhoI were used to linearize pTRV2, EcoRI was used to linearize pTRV2-*CHLH*. All enzyme digestions were carried out with 1 µL of enzyme, 1000 ng of vector, NEB Cutsmart buffer in 50 µL reactions. For pTRV2 an additional 1 µL of Sigma Aldritch rAPID alkaline phosphatase was added. Reaction mixes were incubated at 37 °C for 1 hour. Linear vectors were purified by running a 1% Agarose electrophoresis gel from which the DNA was purified using a Machery-Nagel NucleoSpin Gel and PCR Clean-up kit.

Insertion for plasmids pTRV2-*CHLH*, pTRV2-*CHLH-NEPS5* by InFusion reaction (ClonTech). In a 5 µL reaction, 1 µL of In-Fusion HD Plus, 20-50 ng of linear vector, and 20-50 ng of insert were added. This reaction mix was incubated at 50 °C for 15 minutes and cooled on ice. 2.5 µL of this reaction were used to transform Stellar competent cells and grown on selective media containing 50mg/L of kanamycin.

Base pTRV-*CHLH* vectors were constructed via a ligation reaction. In a 20 µL reaction, 2 µL of NEB T4 DNA ligase buffer, 20-50 ng of linear vector, 60-150 ng of insert (3:1 insert:vector) and 1 µL of T4 ligase. Reaction is left at 16 °C overnight and transformed into Stellar competent cells and grown overnight on selection media containing 50 mg/L of kanamycin.

4.5.5 *E. coli* transformation

Stellar competent cells were used for post-ligation transformations. Transformation followed the protocol: cells are thawed on ice. Then, 1-5 µL of the ligation reaction were added to 30-50 µL of cells and gently mixed by stirring pipette tip in mixture. Cells were left to incubate on ice for 10 minutes, heat shocked at 42 °C for 45 seconds, and cooled on ice for 2 min. 150 µL of SOC media was added to cells and gently mixed by stirring pipette tip in mixture and left to incubate at 37 °C for 45 min. 100-150 µL was plated on LB plates containing the kanamycin and grown overnight at 37 °C.

4.5.6 *A. tumefaciens* transformation

Electrocompetent *A. tumefaciens* strain GV3101 cells were thawed on ice. 100 ng of plasmid DNA was added to 50 µL of cells and mixed gently. Cells were electrically shocked. 150 µL of SOC medium was added and mixed gently. Cells were incubated at 28 °C for 3 hours and plated on LB agar containing 50 mg/L of kanamycin, 50 mg/L of gentamycin and 50 mg/L of rifampicin. Plates were incubated at 28 °C for 2 days.

4.5.7 Gas chromatography analyses

Samples were injected in split mode (1 µL, split ratio 10:1) at an inlet temperature of 280 °C on a Thermofisher Trace1310-ISQLT GC-MS equipped with a 5973 mass selective detector (MSD), and a CTL Analytics GC PAL injector and autosampler. Separation was performed on a Zebron ZB5-HT-INFERNO column (5% phenyl methyl siloxane; length: 35 m; diameter: 250 µm) with guard column. Helium was used as mobile phase at a constant flow rate of 1.2 mL/min and average velocity 37 cm/s. Two temperature runs were used for detection: **A.** After an initial temperature at 60 °C, the column temperature was increased to 100 °C at a rate of 20 K/min, then to 160 °C at 2 K/min, then another increase to 280 °C at 120 K/min, and maintained for 4 minutes. A solvent delay of 5 minutes was allowed before collecting MS spectra at a fragmentation energy of 70 eV. **B.** After an initial temperature at 50 °C, the column temperature was increased to 170 °C at a rate of 10 K/min, then another increase to 280 °C at 50 K/min, and maintained for 4 minutes. A solvent delay of 5 minutes was allowed before collecting MS spectra at a fragmentation energy of 70 eV. Chemically characterised standards were used to identify compounds by retention time and electron impact spectra.

4.5.8 GC-MS peak integration and statistical analysis

RawMS files were converted to .csv files using Lablicate OpenChrom. Using R programming language (R Core Team 2021), areas were obtained by an in-house built script (developed by Carlos E. Rodriguez-Lopez). The script obtained peak areas from the .csv files by addition of the intensities at each time point and integrating the peak area according to a user input timeframe based on the file TIC. Nepetalactone peak areas were divided by the camphor peak area to obtain the normalized peak area. Each data set was statistically analysed using R programming language. Variances in each data set were checked for difference, and a pairwise t-test with a Bonferroni adjustment (for equal or unequal variances) was applied to obtain p-values. The R-package ggplot2 was used to generate the Normalized nepetalactone graphs.

4.5.9 Plant growth conditions and propagation method

In coordination with the greenhouse team of the Max Planck Institute of Chemical Ecology, *N. cataria* cuttings were taken from plants growing in a growth chamber with 15.5 hours of full light (23 °C), 30 minutes of dusk and 30 minutes of dawn conditions, and 7.5 hours of night (21 °C). Humidity was kept at 50%. Cuttings were taken to include 2-3 nodes. The

bottom node was removed from leaves and inserted into water until rooted, about 1-2 weeks. Rooted cuttings were transplanted to a soil mix of 250 L of Klasmann TS1, 70 L of Kalsmann Tonnubstrat and 34.5 L of Raiffeisen Baustoffe sand (0.7-1.2 mm). Plants were fertilized once a week with 0.1% Ferty 3 (Planta Düngemittel GmbH) and watered as necessary.

4.5.10 Virus-induced gene silencing

A. tumefaciens cultures for pTRV1, pTRV2, pTRV2-*ChlH*, pTRV2-*ChlH-GOI* were grown in 50 ml LB cultures containing 50 mg/L of kanamycin, 50 mg/L of gentamycin and 50 mg/L of rifampicin, 10 mM MES buffer (pH 5.8) and 200 µM of acetosyringone for 24 hours at 28 °C shaking at 200 rpm. Cultures were centrifuged for 15 minutes at 3500 rpm, the supernatant discarded and resuspended in 10 mL of fresh infiltration buffer (consisting of 10 mM of MES at pH 5.8, 10 mM of $MgCl_2$ and 200 µM of acetosyringone). The OD_{600} was measured in a 1:10 dilution, and the volume of the infiltration buffer and cultures were adjusted to obtain 10 ml of a OD_{600} of 1. Cultures of pTRV1 *A. tumefaciens* were mixed with each individual pTRV2 *A. tumefaciens* cultures in a 1:1 ratio (10 mL: 10 mL). Cultures were then incubated in the dark for 3-4 hours at room temperature with gentle rocking. Cultures were then centrifuged at 3500 rpm for 15 minutes, the supernatant was discarded and the pellet resuspended in 1 mL of fresh infiltration buffer. Plants for VIGS infection were cut down to 2 to 3 aerial nodes encourage new leaf growth. A sterilized toothpick was dipped into the 1 mL cultures and used to wound the stem near the nodes 2-3 times, and finally the wounds were rubbed with the toothpick with more culture. Two to three weeks post infection, *ChlH*-only targeted plants were re-infected to ensure enough control replicates could be harvested from each experiment.

4.5.11 VIGS tissue harvesting for metabolite analyses

Nepetalactones were extracted as detailed in Lichman *et al.* [63], with a few changes. In general, extraction procedure was as follows: *N. cataria* tissue was flash frozen and pulverized, either with a pestle and mortar or with a Qiagen TissueLyser II at 25Hz for 30 seconds twice, depending on tissue volume. Ice-cold methanol containing the internal standards (200 or 50 µM) camphor and harpagoside (10 µM) was added to the powdered *N. cataria* tissue, vortexed then sonicated at room temperature for 15 minutes. The mix was centrifuged on a table-top centrifuge at room temperature and at top speed (14,600 rpm)

for 10 minutes. The methanol extract was transferred to a new 1.5 ml tube. An aliquot was taken for LC-MS analysis, leaving 400 µL of the MeOH extract. An equal volume of hexane was added, and the mix was vortexed for 2 minutes, centrifuged at top speed for 30 seconds and the hexane layer was transferred to a solid phase extraction (SPE) column. The hexane layer was passed through the column and collected. Then an equal volume of 20:80 EtAc:Hex (ethyl acetate : hexane) mix was added to the SPE column to collect the nepetalactones. The EtAc:Hex extract is then analysed on the GC-MS. For VIGS experiments, Leaf tissue from pTRV2 empty vector was collected by cutting a full leaf from three individual plants. Only individuals with affected leaf tissue were harvested. Affected tissues was collected by cutting out the bleached area from the green tissue and pooling together all affected leaf tissue from one individual.

4.5.12 RNA extraction and cDNA generation

For wildtype samples, leaves of all ages were collected from a B2C cultivar plant into a falcon tube. The falcon tube was flash frozen and all leaf tissue was ground with a mortar and pestle, while cooled with liquid nitrogen to keep the tissue frozen. The ground tissue was then returned to the falcon tube and stored at -80 °C. Ground tissue powder was weight out between 25-35 mg of tissue into cooled 1.5 ml tubes. For individual experimental samples, leftover sample tissue not used for metabolic analyses, was weighed out to be between 20 and 35 mg of tissue. If tissue mass was too low to be transferred to a new tube for weighing, the tissue was not weighed. RNA was extracted using either the Qiagen® RNeasy Plant Mini Kit, or the Qiagen® Rneasy PowerPlant Kit, depending on availability. Tissue was extracted according to kit instruction, with a final incubation of 10 minutes for more extraction.

RNA concentration and A280/A260 and A280/A230 ratios were assessed using a Nanophotometer N60**.** A 2% Agarose electrophoresis gel was ran at 120 V for 1 hour to verify any genomic DNA contamination and RNA degradation. RNA samples were subjected to a DNAse treatment to remove any contaminant genomic DNA. Treatment was carried out with Sigma Amplification grade Dnase I kit by adding 1000 ng of RNA, 1 µL of Sigma Amplification Grade Dnase 10x buffer and 1 µL of the DNAse enzyme into a 10 µL reaction and left to incubate for 20 minutes at room temperature. Then, 1 µL of stop solution was added and the sample was incubated at 70 °C for 10 minutes. After incubation, the sample was cooled on ice for 2 minutes. The Thermo Fisher Scientific Applied Biosystems High capacity cDNA reverse transcription kit was used to make a 10 µL mixture containing 4.2 µL of water, 2 µL

of buffer, 2 µL of random hexamer primers, 0.8 µL of dNTPs and 1 µL of enzyme and was added to the RNA sample. The RT reaction was there incubated at 25 °C for 10 minutes, 37 °C for 2 hours, then at 85 °C for 5 minutes and then stored at 4 °C. Samples were stored in -20 °C.

4.5.13 qPCR primer design and analysis

Primers for qPCR analysis were designed using the CDS of the target gene in the NCBI primer design tool. The primers were restricted to replicate a product from 70 bp to 200 bp and to have a melting temperature of 60 °C. The tool's best primers, that did not lay within the targeted VIGS region, were selected to be tested.

A dilution standard curve was carried out for each primer pair. Wildtype cDNA was initially diluted to a ratio of 1:5, then a series of four half dilutions were carried out. Reaction triplicates were carried out at each dilution point. The average Ct value was then calculated and plotted against the log of the sample concentration. A trendline was plotted for these values, and the equation for the trendline calculated (y=mx+b). The slope (m) of the equation was then used to calculate the efficiency I in the following formula: E=2^(-1/m). The percent efficiency was then calculated by the following %E=100(E-1). Primers resulting in the E closest to 2, and the %E closest to 100% were selected to use for qPCR amplification.

Primer standard curves are in Appendix 3.

4.5.14 qPCR reaction

The cDNA generated from the RT reaction was diluted by half with diWater. The reaction mixture for the qPCR contained the following: 12.5 µL of Agilent Brilliant II SYBR® green QPCR Master Mix, 1.5 µL of forward primer and 1.5 µL of reverse primer (at 6nmol each), and 7.5 µL oh diWater and 2 µL of the diluted cDNA sample. A master mix of the qPCR reaction mix was pipetted into each well of a 96 well-plate first, and the diluted cDNA sample was then added to each experimental well. Each gene targeted had 3 technical replicates and 3 biological replicates. To measure gene expression, each target gene was paired to the housekeeping gene *UBI9*. Each plate also contained a well with the reaction mixture containing diWater instead of sample as an empty control.

This reaction was then placed in a Bio-Rad CFX96 Optical Reaction Module. The qPCR reaction protocol was as follows: 3 minutes at 95 °C, then 44 cycles of 30 seconds at 95 °C, 30 seconds at 60 °C, and 30 seconds at 72 °C, and an individual incubation at 95 °C for 10 seconds before a final melt curve. A melt curve step with an increase of 5 °C every cycle from a minimum of 65 °C to a maximum of 95 °C finalized the qPCR reaction.

4.5.15 qPCR analysis

The resulting cross point (Ct) values from the qPCR reaction were averaged across technical replicates. Delta Ct (ΔCt) was calculated by substracting the average UBI Ct value from the average target gene Ct value. The DeltaDeltaCt (ΔΔCt) was calculated by subtracting a control calibrator ΔCt value from each ΔCt value. The ΔΔCt value was then used in the following equation $2^{-\Delta\Delta Ct}$ to calculate the normalized expression of each biological replicate. The average normalized expression was calculated for each experimental sample.

4.5.16 LC-MS method

Metabolite analysis. UPLC/MS analysis was performed on an Impact II qTOF mass spectrometer (Bruker) coupled to an Elute UPLC (Bruker) chromatographic system. Chromatographic separation was carried out on a Phenomenex Kinetex column XB-C18 (100 × 2.10 mm, 2.6 µm particle size) kept at 40 °C and the binary solvent system consisted of solvent A (H_2O + 0.1% formic acid) and solvent B (acetonitrile + 0.1% formic acid). Flow rate was 600 µL/min. The column was equilibrated with 99% A and 1% B. During the first minute of chromatography, solvent B reached 5%. Then a linear gradient from 5% B to 40% B in 5 min allowed the separation of the compounds of interest. The column was then washed at 100% B for 1.5 min and re-equilibrated to 1% B. Injection volume was 2 µL.

Mass spectrometry was performed in negative ion mode with a scan range *m/z* 100–1000. The mass spectrometer was calibrated using sodium formate. The source settings were the following: capillary voltage 3.5 kV, nebulizer 2.5 Bar, dry gas 11.0 L/min, dry temperature 250 °C. Data analysis was performed using the Bruker Data Analysis Metaboscape (Version 2021-B) software.

4.5.17 Heatmap of Gene Expression

Tissue specific gene expression analysis heatmaps were generated using the R-script package for heatmaps included in the gplot package, based on Log(FPKM+1) results. Heatmaps were generated without hierarchal clustering.

4.5.18 Primer list

Code	Gene	Sense	Sequence
LP056	NECA_E-N_VIGS_CHLH	F	TAAGGTTACCGAATTACCAATGACATGAAGGCCAC
LP057	NECA_E-N_VIGS_CHLH	R	TACCGGATCCCCATGGAGCTCACAATTTGAGGGCCAC
LP058	NECA_B-X_VIGS_GES1	F	GCCTCCATGGGGATCTCGCCTTTGTAGCGAGGTGG
LP059	NECA_B-X_VIGS_GES1	R	ATGCCCGGGCCTCGATTTGAGTACTTCGGGTCGGG
LP069	Neca_EcoR1_VIGS_GES	F	TAAGGTTACCGAATTAATCCAACGGCTGGGAATCGG
Lp070	Neca_Nco1_VIGS_GES	R	GGTGGATCCCCCATGTCGCTACAAAGGCGAGGTGC
Lp071	Neca_EcoR1_VIGS_GES	R	GCCTTCTAGAGAATTTCGCTACAAAGGCGAGGTGC
Lp072	Neca_EcoR1_VIGS_ISY	F	TAAGGTTACCGAATTATGCTGCCGCAGGCGGACACCC
Lp073	Neca_EcoR1_VIGS_ISY	R	GCCTTCTAGAGAATTGGTCTGCAAGCAGATGTGCTTC
LP076	Neca_EcoR1_VIGS_Neps5,3,2	F	TAAGGTTACCGAATTGCGGCGCGTAAGATGGTGGAGC
LP077	Neca_EcoR1_VIGS_Neps5	R	GCCTTCTAGAGAATTGTAGAAATCATCGGGCGTGGCAA
Lp078	Neca_EcoR1_VIGS_Neps3	R	GCCTTCTAGAGAATTCTGTACATCAGCAGGCGTCGAAA
Lp079	Neca_EcoR1_VIGS_Neps2	R	GCCTTCTAGAGAATTCTCGACGTCGGCGGGGGTGGCAA
Lp080	Neca_Xma1_VIGS_Neps4,1	F	GGTCTCGAGGCCCGGGCAGCGCGTAAGATGGTGG
Lp081	Neca_Xma1_VIGS_Neps4	R	TCGGGACATGCCCGGATAGAAATCATCGGGCGTGGC
Lp082	Neca_Xma1_VIGS_Neps1	R	TCGGGACATGCCCGGATGGAAATCATCGGGTGTGGC
Lp091	Neca_EcoR1_VIGS_MLP1	F	TAAGGTTACCGAATTACCAAATTATTCCCCAAGGCTT
Lp092	Neca_EcoR1_VIGS_MLP1	R	GCCTTCTAGAGAATTTGCCTTTTCAAATTCTCCTGAA
Lp131	Neca_RLK_VIGS_EcoR1	F	TAAGGTTACCGAATTGATGCTACATTTGTGCAGAGAGG
lp132	Neca_RLK_vigs_EcoR1	R	GCCTTCTAGAGAATTCCGAATGTGACTCATAC
lp228	qRT_PCR_NC_CHLH_1	F	TCAGACCACGCAAGAAGTCC
lp229	qRT_PCR_NC_CHLH_1	R	CCAAGTAGCCCACAAGCTCA

lp230	qRT_PCR_NC_CHLH_2	F	GAGCTTGTGGGCTACTTGGT
lp231	qRT_PCR_NC_CHLH_2	R	CTTCTCCACAGCAGCCTTGA
lp232	qRT_PCR_NC_GESA_1	F	CGCTTCCCAAGGACCTCAAT
lp233	qRT_PCR_NC_GESA_1	R	TCGTTGTTGTGGACTACGGG
lp234	qRT_PCR_NC_GESA_2	F	GTGTAACCGACCAAAACGCC
lp235	qRT_PCR_NC_GESA_2	R	CGAGATCATCCCAAAGGCGA
lp236	qRT_PCR_NC_ACT1_1	F	TCGTGTTGGTCCTGAAGAGC
lp237	qRT_PCR_NC_ACT1_1	R	GAGAGAACGGCCTGGATAGC
lp238	qRT_PCR_NC_ACT1_2	F	GCTATCCAGGCCGTTCTCTC
lp239	qRT_PCR_NC_ACT1_2	R	CTCACACCATCACCGGAGTC
lp240	qRT_PCR_NC_UBQ9_1	F	AGGTCGAGAGCTCGGATACT
lp241	qRT_PCR_NC_UBQ9_1	R	AGCCTTTGCTGATCTGGTGG
lp242	qRT_PCR_NC_UBQ9_2	F	GAGGATGGGAGAACCTTGGC
lp243	qRT_PCR_NC_UBQ9_2	R	CTCAAACGCAGCACCAGATG
lp248	qRT_PCR_NC_MLPL_1	F	ATGAACAAGTCCCTGATCCTGT
lp249	qRT_PCR_NC_MLPL_1	R	TGACATGTGTGGTTCATGCCA
lp250	qRT_PCR_NC_MLPL_2	F	GCAGCAGCAAAAATGGTGGA
lp251	qRT_PCR_NC_MLPL_2	R	TGACATGTGTGGTTCATGCCA
lp252	qRT_PCR_NC_RLK_1	F	GTCGGGTGAAGTCCCAGATG
lp253	qRT_PCR_NC_RLK_1	R	ATCACTTTCGGCATGAGGCA
lp254	qRT_PCR_NC_RLK_2	F	TGCCTCATGCCGAAAGTGAT
lp255	qRT_PCR_NC_RLK_2	R	TTTTTCGCAGCCCTCTCCTT
lp272	qrt_pcr_ncisy_3	F	GGTGCGGTTTTGAGGTTTCC
lp273	qrt_pcr_ncisy_3	R	GCGGCCCATATTTGATGCTC
lp274	qrt_pcr_ncisy_4	F	TTCAACGTCAGCAATGGGGA
lp275	qrt_pcr_ncisy_4	R	TTGCTGGCCTTCCTCGTATC

Chapter 5

5.1 Conclusion

Previous work into the Nepeta iridoid metabolism has highlighted its evolutionary history and interesting chemistry [11], [23], [64], with a focus on the downstream part of the pathway, providing a case study for the evolutionary dynamics of plant secondary metabolism. However, while the downstream genes in the pathway had been characterized *in vitro*, the early steps of the pathway had not been identified. Given Nepeta's unique chemotaxonomy [11] and the convergent evolution of ISY [65], identifying the early steps of the pathway was important to place the evolutionary history of this pathway into context.

In Chapter 2, I identified and characterized the early steps of the iridoid pathway, GES, G8H, and HGO, in *N. cataria* and *N. mussinii* (fig 5.1). While iridoids can be found in a wide array of plants [12], [31], [33], [34], the enzymatic steps leading to them have not been characterized in many organisms [12]. Therefore, finding the enzymatic steps to be conserved across different plant families shed light on how conserved, and how varied, these secondary metabolic pathways can be. Furthermore, these early genes were found to be present in the genome, but low or not expressed, of a non-iridoid producer of the Nepetoideae sub-family, *Hyssopus officinalis*, closely related to *Nepeta*. This finding suggests that pathways can remain inactive in the genome but may be able to re-emerge when other evolutionary events, as discussed in section 5.3.1, for example, the re-emergence of ISY, occur. The early genes identified in *N. cataria* and *N. mussinii* and their *in vitro* activities were reported in Lichman *et al.* [65].

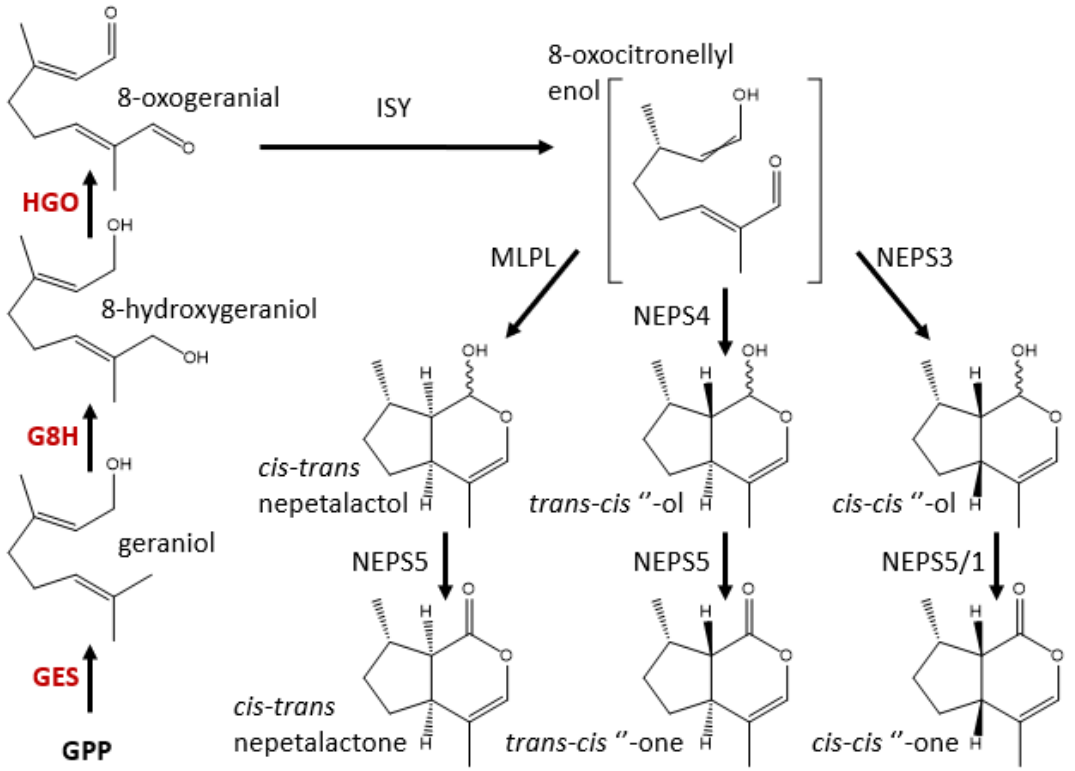


Figure 5.1: Nepeta iridoid pathway

The iridoid pathway as most recently reported in Lichman et al. [63]. Enzymes highlighted in red were identified and characterized from *N. cataria, N. mussinii* and *H. officinalis* in Chapter 2 of this thesis.

While the pathway had been characterized *in vitro* and *in silico*, it was not validated *in vivo*. This is the case with many iridoid biosynthetic genes identified, as very little literature exists on how these enzymes function *in planta*, with the exception of the iridoid producer *C. roseus*, which most iridoid metabolism studies have used as a model system [32], [42], [43]. With the aim to characterize the *Nepeta* iridoid pathway *in vivo*, I developed a functional genomics tool. In Chapter 3, I established and optimized the VIGS technique into *N. cataria* (fig. 5.2). VIGS has been shown to be adaptable to a wide range of species, including *O. basilicum*, a Nepetoideae [54]. I optimized this functional genomics tool to work with *N. cataria* cuttings to ensure full genetic replicability, rather than seedlings, given the variation in nepetalactone isomer content amongst *N. cataria* cultivars. While *N. cataria* was susceptible to VIGS, the infection was not fully systemic. I designed a VIGS vector that would silence two genes simultaneously, based on the system in Yamamoto *et al.* [52]. This allowed for the silencing of a visual marker gene which would indicate which tissue had been affected by VIGS (fig. 5.2). Dual-knockdown was shown to be more successful at demonstrating a true

silencing phenotype as opposed to only targeting the specialized metabolic gene of interest. This methodology was published in Palmer *et al.* [66].

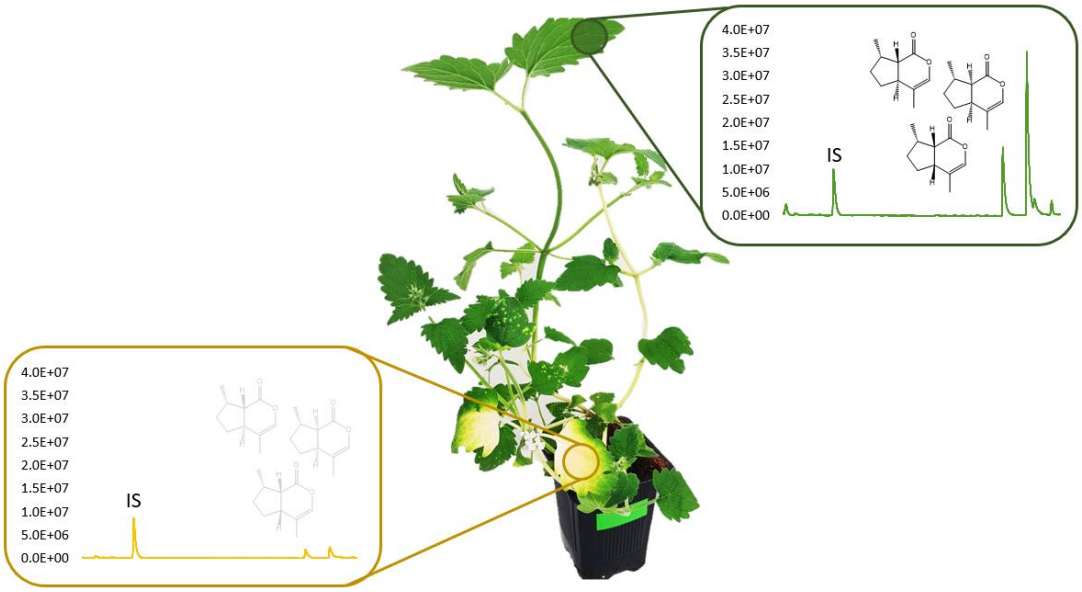


Figure 5.2: Use of *ChlH* as a visual marker in *N. cataria* VIGS

VIGS in *N. cataria* was developed using a visual marker, such as the silencing of *ChlH* which leads to inhibited chlorophyll biosynthesis, to study "invisible to the eye" phenotypes, including the effect of silencing key genes in the iridoid biosynthesis pathway. Leaf tissue without chlorophyll, yellow or white in the example picture above, also showed a significant decrease in iridoid production when key genes of the pathway (i.e. *GES*) were silenced. Chapter 3 described the methodology behind the VIGS double knockdown strategy. IS stands for internal standard.

In Chapter 4 I report the results of silencing the iridoid pathway genes *GES, ISY, MLPL* and the general function of the *NEPS* enzymes, confirming their *in vitro* roles (fig. 5.3). I established *GES* as the entry point enzyme into the pathway, as its silencing showed a significant decrease of all nepetalactones produced. *ISY* silencing resulted in a significant decrease of nepetalactone content, to a lesser extent than *GES* silencing, consistent with a more downstream effect. Furthermore, I identified the accumulation of glycosylated hydroxy-geraniol-based intermediate in *ISY* silenced tissues. I confirmed the *in vitro* data for *MLPL* being necessary for the production of *cis-trans* nepetalactone. I targeted all of the *NEPS* homologues simultaneously, revealing their collective roles behind the various nepetalactone isomers found in *N. cataria*.

Furthermore, while VIGS characterization of the *Nepeta* iridoid pathway provides more context on iridoid pathways *in planta*, it would also clarify some of the *in vivo* roles for the downstream enzymes MLPL and NEPS identified and characterized *in vitro* in [64], [65]. The NEPS enzymes were found to have somewhat overlapping roles in the cyclization of the ISY product 8-oxocitronellyl enol and further oxidation of nepetalactol into nepetalactones, synthesizing the *cis-trans, trans-cis* and *cis-cis* nepetalactones found in *N*. *cataria* and *N. mussinii*. While transcriptomic data suggested compartmentalized activities, *in vitro* analysis revealed ambiguities in enzymatic function *in planta*. The data reported in this thesis further suggests a more nuanced role for these enzymes, highlighting the need for individual *in vivo* characterization.

In Chapter 4 I also briefly explored a possible regulatory aspect of the pathway. Previously identified transcription factors BIS1 and BIS2 were found to regulate the iridoid part of the MIA pathway in *C. roseus* [87], [96] were not found in *Nepeta spp.* To test for any involvement in the iridoid pathway, I targeted a RLK found to be conserved on the *Nepeta* iridoid metabolic gene cluster that Lichman *et al*. [65] reported on. However, gene expression analysis revealed the low expression of this gene in normal growth conditions, and upon VIGS, did not reveal any perturbation of the iridoid pathway. However, this gene family is known to be functional in stress conditions, primarily as a receptor to pathogenic effectors [104]–[107]. This opens the question to the role of nepetalactones in stress responses, and whether this RLK will be functional under biotic stress conditions and involved in the iridoid pathway.

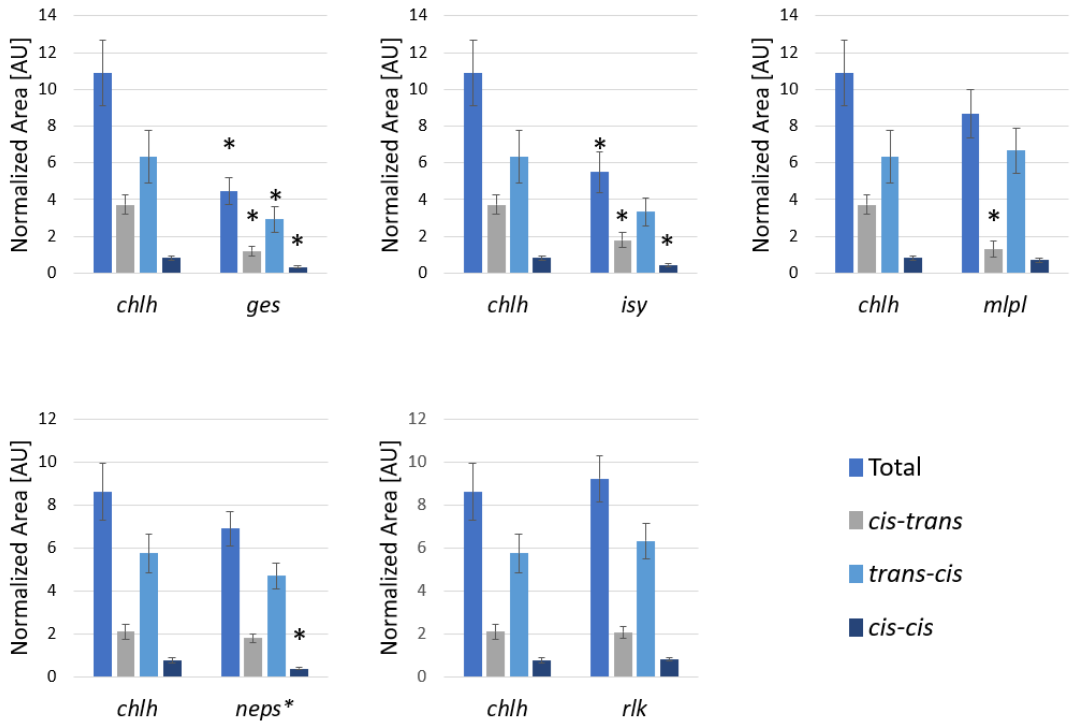


Figure 5.3: Normalized nepetalactone content under various VIGS conditions

Summary of the results in Chapter 4. Normalized nepetalactone content for each isomer is calculated by dividing the isomer peak area by the internal standard (50 µM camphor) peak area and reported as Normalized Area in arbitrary units [AU]. Each bar is the average normalized nepetalactone content. Error bars indicate standard error. Each condition is the gene targeted for silencing: *chlh* indicates silenced *ChlH* only; *ges* indicates *ChlH* and *GES* silenced; *isy* indicated *ChlH* and *ISY* silenced; *mlpl* indicates *ChlH* and *MLPL* silenced; *neps** indicates *ChlH* and all *NEPS* homologues silenced; *rlk* indicates *ChlH* and *RLK* silenced. An asterisk above a bar indicates it is statistically significant (t-test p-value <0.05) compared to the *chlh* condition.

5.2 Future Directions

5.2.1 NEPS enzymes as single targets

An exciting experiment that will complement the *in vitro* and *in silico* information in Lichman *et al.* [65] will be silencing the *NEPS* genes individually. As the pathway stands, MLPL controls the flux of 8-oxocitronellyl enol into *cis-trans* nepetalactones, as corroborated by my VIGS results in Chapter 4. *In vitro* activity for NEPS4 and NEPS3 revealed precision in making *trans-cis* nepetalactone and *cis-cis* nepetalactones respectively *in vitro* (table 5.1). However, the

precise role of NEPS1 and NEPS5 were ambiguous. Their activity as nepetalactol oxidizers was verified, acting on all nepetalactols isomers, raising the question of whether they had overlapping functions *in planta* (table 5.1). Furthermore, their activity into cyclization of the 8-oxocitronellyl enol was left unverified [65].

If the *in vitro* results were to be verified *in vivo*, I would expect to see a decrease in *trans-cis* nepetalactone upon the silencing of *NEPS4* and a decrease in *cis-cis* nepetalactone upon silencing *NEPS3*. For silencing *NEPS1* and *NEPS5*, I may see a significant decrease in nepetalactones, and perhaps an increase in nepetalactol content. It is important to note that nepetalactols can decay into iridodials, which may be glycosylated in the plant, resulting in new intermediates. As described in section 4.3.2, work is ongoing to target the UTR region of the single *NEPS* homologues.

Species	Enzyme	Cyclization			Dehydrogenation			*in vivo*		
		cis-trans	*cis-cis*	*trans-cis*	*cis-trans*	*cis-cis*	*trans-cis*	*cis-trans*	*cis-cis*	*trans-cis*
N. cataria	NEPS1	?			*	*	*			
N. cataria	NEPS2									
N. cataria	NEPS3A		*			*			*	
N. cataria	NEPS3B		*			~				
N. cataria	NEPS4			*	*	?	~			
N. cataria	NEPS5	?			*	*	*			
N. cataria	MLPLA	*						*		
N. cataria	MLPLB	*						*		

Table 5.1: *In vitro* and *in vivo* activities of the NEPS and MLPL enzymes

Summary of *in vitro* enzymatic activities of NEPS and MLPL enzymes, adapted from Lichman *et al.* [63]. Asterisk (*) indicates observed activity; question mark (?) indicates possible activity, but not verified; tilde (~) indicates trace activity. Cyclization indicates the formation of nepetalactols from 8-oxocitronellyl enol and Dehydrogenation refers to the formation of nepetalactones from nepetalactol. *In vivo* indicates the effect of silencing via VIGS as reported in Chapter 4: asterisks indicate statistically significant downregulation of a given isomer.

5.2.2 Identifying potential regulators of the iridoid pathway

As explored in sections 4.2.9 and 4.2.10, the regulatory systems for the iridoid pathway in *Nepeta* remain undiscovered. Characterized regulatory networks from other iridoids producers do not seem to be conserved in *Nepeta*, as evidenced by the results of a homology-based query using BIS1/2 sequences into the *N. cataria* transcriptome. This is not unexpected, since the iridoid regulatory elements that have been discovered come from plants phylogenetically distant from *Nepeta*. However, it may be that the regulation

machinery associated with monoterpene pathways in other Nepetoideae, such as the MYB transcription factors identified in *Mentha spicata* and *Lavandula spp*. [123], [124] might be conserved and co-opted into regulating the iridoid pathway in *Nepeta spp.*. Searching for homologues of Nepetoideae monoterpene transcription factors using homology may provide some candidates to test. A gene co-expression analysis which includes the nepetalactone pathway genes expression patterns would aid in narrowing down potential candidates.

Upon assembly of a candidate list, these candidate transcription factors can be knocked down via VIGS. Previously, regulation machineries from other secondary metabolisms have been studied via VIGS [126], [128]. Based on the literature, very little is known about the nepetalactone pathway response to stress. Iridoids are known to be anti-herbivory compounds [26], and recent research shows an antagonistic effect on mosquitoes and other insects' neural pathways [25], [27]. Some research indicates that iridoids also play a role against microorganisms [26]. However, the change in the *Nepeta* pathway upon abiotic or biotic stressors has not been tested. In Chapter 3, plants infected with an empty vector pTRV2 displayed a significant decrease in nepetalactone content, as compared to non-infected wild-type plants. This may indicate that bacterial infection itself may lead to a down regulation in the pathway, perhaps in favour of increasing the production of other defensive compounds. Comparing the concentration of 1,5,9 deoxyloganic acid in tissues infected by pTRV2 compared to WT tissues for an upregulation of the glycosylated iridoid would help determine the role of the pathway upon biotic stressors. Furthermore, as explored in section 4.3.1, RLK expression could be tested upon infection with a generalist pathogen, or a known *Nepeta* pathogen, such as *Xanthomonas campestris* [121], as well as the response of the nepetalactone pathway upon infection.

5.2.3 Developing a stable transformation method

Most transformation methods developed in the Lamiaceae family involve the generation of a callus from explant tissues, such as stem and leaves. In section 3.4.2, I briefly explored the work carried out to develop a reliable protocol for callus induction from explants and regeneration into plantlets (fig. 5.4). Upon a reliable protocol, calluses are typically co-cultured with *Agrobacteria tumefaciens* containing a desired trans-gene.

Work on this type of methodology is ongoing. While stable transformants would provide answers to some of the ambiguities found from the VIGS experiments in Chapter 4, such as the non-significant decrease in *trans-cis* production upon ISY silencing, they would also take several months to produce a final stable transformant. Ideally, both VIGS and stable transformations could be used in tandem, with VIGS as an initial screener phase, and stable knockouts to confirm or remove ambiguities from VIGS results.

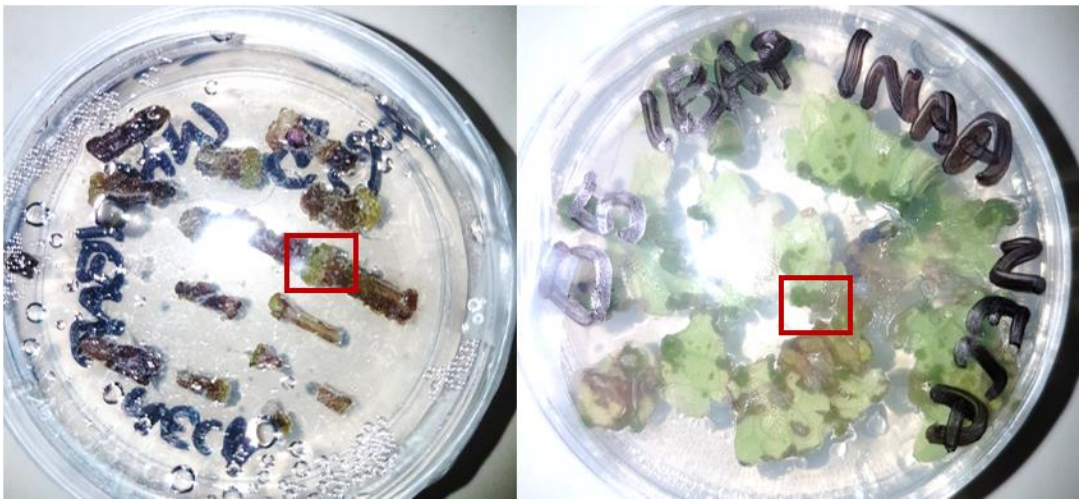


Figure 5.4: Calluses from *N. cataria* explant tissue

Two example calluses formed from stem and leaf tissue from *N. cataria*. Calluses were formed from surface-sterilized explant tissue on MS media with the hormones 6-benzylaminopurine, 1-naphthaleneacetic acid and 2,4-dichlorophenoxycetic acid at various concentrations.

5.3 Perspectives and Outlook

5.3.1 Iridoid pathway evolutionary history

As reported in Lichman *et al*. [65] *Nepeta* ISY independently evolved from P5βR as compared to other characterized ISYs. The authors also explored the evolutionary history of GES, G8H and HGOA using orthogroups and transcriptomes reported in [11]. The authors reported that these genes are all orthologous to the previously characterized *C. roseus* homologues, indicating no further cases of convergent evolution.

Curiously, orthogroup data and transcripts from non-iridoid producing Nepeteoideae species also had candidate genes for the early iridoid steps (GES, G8H and HGOA); however, as seen in the homologues from *H. officinalis* in section 2.2.2, these genes can be found on the

genome with low or no gene expression. This reveals some interesting insights into the evolution of secondary metabolic pathways. Are there pathways that are functionally "turned off" but are kept in reserve for novel evolutionary events, for example with the rise of new enzymes such as MLPL, the NEPS homologues or a new ISY? In the -omics era, metabolomics, phylogenomics and genomics can be simultaneously explored to discover the dynamic paths plant secondary pathways can evolve, including their appearance, disappearance and even re-appearance.

5.3.2 VIGS: disadvantages and advantages

As described in section 4.4.1, VIGS has certain advantages and disadvantages as a functional genomics tool. Gene expression analysis via qRT-PCR in Chapter 4 revealed genes silenced via VIGS were not fully silenced, with some relative gene expression remaining. It is therefore likely there was some enzymatic function within the tissues assayed, leading to downstream metabolites being produced. This may explain the non-significant decrease in nepetalactone production found when *ISY* and the *NEPS* homologues were silenced. While there may be an alternative explanation, for example, that *trans-cis* neptalactone is not fully controlled by this pathway, this cannot be ascertained with the current results.

Furthermore, while the work carried out in section 4.2.11 also meant to study the glycosylated iridoid metabolism in *N. cataria*, which includes the 1,5,9 deoxyloganic acid reported in Lichman *et al.* [65], I was only able to establish a tenuous link between GES and 1,5,9 deoxyloganic acid, as the dimer of this compound was significantly downregulated, but the monomer, although decreased, was not significantly decreased. As explained in section 4.2.11, this may be due to glycosylated iridoids being able to move throughout the plant. Indeed, as reported previously, secondary metabolites are often glycosylated to be transported between sub-cellular and cellular compartments [13], [119], [120]. This may mean that the non-fully systemic nature of VIGS in *N. cataria* (hence the need for a visual marker) may impede characterization of the iridoid glucoside pathways. This may be ameliorated if VIGS was performed in young plantlets and the first leaf pair post infection is harvested, depending on the efficiency of silencing in that leaf-pair, as reported in [50], [54].

While the VIGS system developed in this thesis is particularly suited to studying leaf and stem tissues, the *ChlH* visual marker does not indicate the extent of infection into non-chlorophyl containing tissues such as roots or flowers. To determine the effect of VIGS on these non-

chlorophyl tissues an alternative visual marker would have to be employed. Exploring floral tissue may be of particular interest for the nepetalactone metabolism in *N. cataria*, as *NEPS3A/B* are most highly expressed in the flowers (fig. 2.8). Furthermore, as reported in Lichman *et al*. [63] and in figure 2.2, flowers contain high nepetalactone content, and may be useful for the study of nepetalactone pathways.

While there are some disadvantages to the VIGS in *N. cataria* developed in this thesis, there are many advantages. Secondary metabolism not only relies on the enzymes that carry out the catalytic chemical steps to create these metabolites, such as ISY and GES, but also depend heavily on spatial and temporal compartmentalization. Therefore, *in vivo* work on secondary metabolic pathways often discovers the regulatory mechanisms behind these pathways, including mechanisms behind gene expression [87], [96], metabolite transportation [130], [131] and storage [53]. The VIGS proposed in this thesis would be an ideal tool to study the regulatory mechanisms behind the iridoid metabolism that are yet to be uncovered in *N. cataria*.

One significant advantage over stable transformants is the quick turnover time of VIGS. From a few weeks post-infection, as early as two weeks to see the bleaching phenotype, tissue can be harvested. This tissue can be tested for gene expression, nepetalactone content, or even physiological changes, for example, in probing the morphology of glandular trichomes, an important organ for secondary metabolite biosynthesis. Compared to a stable mutant, which may take months to establish and requires the optimization of a reliable breeding method for *N. cataria*, VIGS can provide quick answers to help screen for candidates suitable for long term studies in a stable transformant.

Finally, *in vivo* studies such as the work presented in Chapter 4 can help refine the roles of these chemically catalytic enzymes. While the results from the *MLPL* knockdown were as expected, with the significant decrease of *cis-trans* nepetalactone, the results from knocking down the *NEPS* homologues simultaneously were not as straightforward. From *in vitro* analyses [23], [64], [65], a simultaneous knockdown of *NEPS* would be expected to lead to a significant decrease of *trans-cis* nepetalactone and *cis-cis* nepetalactone. However, upon simultaneous knockdown of the *NEPS* homologues, only *cis-cis* nepetalactone was significantly decreased. While transcriptomic analysis on this experimental tissue would reveal the extent of knockdown for each individual NEPS enzyme, these results conflicting with the previously reported *in vitro* results [23], [64], [65] is puzzling. Single *NEPS*

knockdowns can lead to a better understanding of each individual role, which may serve for iterative *in vitro* studies on these enzymes to understand the relationship between their chemistry and their *in planta* activity.

5.4 Conclusion

The work presented in this thesis advanced the knowledge of plant secondary metabolism from a biochemical and *in vivo* perspective. A key contribution was the development of an *in vivo* functional genomics tool for *N. cataria*, which was applied to the study of secondary metabolism. However, this tool can be applied for several areas of research, and can be potentially adapted to other *Nepeta*, or even Nepetoideae. Building on the foundational knowledge of iridoid metabolisms in other species (section 2.2.1) to identify and understand the gene expression (section 2.2.2) and biochemical activity (2.2.4-5) of the early steps in the pathway contributed to the full understanding of how iridoids are biosynthetically produced in *Nepeta spp*., uniquely for the Nepetoideae sub-family. Furthermore, in this thesis, I presented the development (Chapter 3) and application (Chapter 4) of the *in vivo* tool VIGS to study the gene function of key genes in the iridoid biosynthesis pathway. Work in this thesis also touched upon the dynamics of secondary pathway evolution (section 2.2.1-2) and regulation (4.2.9), the flexibility in developing *in vivo* tools to study secondary metabolisms (section 3.2.9) and the effect of perturbing the pathway *in vivo* (section 4.2.10). Work presented in this thesis has laid the foundation for the opportunity for further studies into the broad field of plant secondary metabolism, from the development of a stable transgenic method for the *Nepeta* genus, evolutionary studies in pathway degradation and re-emergence and novel metabolic regulatory systems.

References

[1] H. A. Maeda, “Evolutionary Diversification of Primary Metabolism and Its Contribution to Plant Chemical Diversity,” *Front. Plant Sci.*, vol. 10, no. July, pp. 1–8, 2019.

[2] M. Erb and D. J. Kliebenstein, “Plant Secondary Metabolites as Defenses, Regulators, and Primary Metabolites: The Blurred Functional Trichotomy,” *Plant Physiol.*, vol. 184, no. September, pp. 39–52, 2020.

[3] G. D. Moghe and R. L. Last, “Something Old, Something New: Conserved Enzymes and the Evolution of Novelty in Plant Specialized Metabolism,” *Plant Physiol.*, vol. 169, no. November, pp. 1512–1523, 2015.

[4] A. Kessler and A. Kalske, “Plant Secondary Metabolite Diversity and Species Interactions,” *Annu. Rev. Evol. Evol. Syst.*, vol. 49, pp. 115–138, 2018.

[5] J. Kroymann, “Natural diversity and adaptation in plant secondary metabolism,” *Curr. Opin. Plant Biol.*, vol. 14, no. 3, pp. 246–251, 2011.

[6] “Ethnobotany,” *US Forest Service, USDA*. [Online]. Available: https://www.fs.fed.us/wildflowers/ethnobotany/index.shtml.

[7] M. Wink, “Introduction: Biochemistry, Physiology and Ecological Functions of Secondary Metabolites,” in *Biochemistry of Plant Metabolism: Second Edition*, 2010, pp. 1–19.

[8] S. E. O’Connor, “Engineering of Secondary Metabolism,” *Annu. Rev. Genet.*, vol. 49, pp. 71–94, 2015.

[9] F. Zhou and E. Pichersky, “More is better: the diversity of terpene metabolism in plants,” *Curr. Opin. Plant Biol.*, vol. 55, pp. 1–10, 2020.

[10] P. S. Karunanithi and P. Zerbe, “Terpene Synthases as Metabolic Gatekeepers in the Evolution of Plant Terpenoid Chemical Diversity,” *Front. Plant Sci.*, vol. 10, no. October, pp. 1–23, 2019.

[11] B. Boachon *et al.*, “Phylogenomic Mining of the Mints Reveals Multiple Mechanisms Contributing to the Evolution of Chemical Diversity in Lamiaceae,” *Mol. Plant*, vol. 11,

no. 8, pp. 1084–1096, Aug. 2018.

[12] K. Miettinen *et al.*, “The seco-iridoid pathway from Catharanthus roseus,” *Nat. Commun.*, vol. 5:3606, 2014.

[13] J. Le Roy, B. Huss, A. Creach, S. Hawkins, and G. Neutelings, “Glycosylation is a Major Regulator of Phenylpropanoid Availability and Biological Activity in Plants,” *Front. Plant Sci.*, vol. 7, no. May, 2016.

[14] L. K. Henry *et al.*, “Contribution of isopentenyl phosphate to plant terpenoid metabolism,” *Nat. Plants*, vol. 4, no. September, pp. 721–729, 2018.

[15] D. Tholl, “Biosynthesis and Biological Functions of Terpenoids in Plants,” in *Biotechnology of Isoprenoids*, J. Schrader and J. Bohlmann, Eds. Cham: Springer International Publishing, 2015, pp. 63–106.

[16] B. Li *et al.*, “A large-scale chloroplast phylogeny of the Lamiaceae sheds new light on its subfamilial classification,” *Sci. Rep.*, vol. 6:34343, pp. 1–18, 2016.

[17] F. Zhao *et al.*, “An updated tribal classification of Lamiaceae based on plastome phylogenomics,” *BMC Biol.*, vol. 19, no. 2, pp. 1–27, 2021.

[18] “Global Lavender Oil Industry,” 2020.

[19] C. Yoo, “12 Mosquito Repellent Plants,” *Garden Design*. [Online]. Available: https://www.gardendesign.com/plants/mosquito-repellent.html.

[20] M. Petruzzello, “List of economically important members of the family Lamiaceae,” *Britannica,*. [Online]. Available: https://www.britannica.com/topic/list-of-economically-important-members-of-the-family-Lamiaceae-2035852.

[21] “Plant families: Lamiaceae,” *BBC Gardeners’ World*, 2020. [Online]. Available: https://www.gardenersworld.com/plants/the-mint-family-lamiaceae/.

[22] S. M. McElvain, R. D. Bright, and P. R. Johnson, “The constituents of the volatile oil of catnip. I. Nepetalic acid, nepetalactone and related compounds,” *J. Am. Chem. Soc.*, vol. 63, no. 6, pp. 1558–1563, 1941.

[23] N. H. Sherden *et al.*, "Identification of iridoid synthases from Nepeta species : Iridoid cyclization does not determine nepetalactone stereochemistry," *Phytochemistry*, vol. 145, pp. 48–56, 2018.

[24] C. Formisano, D. Rigano, and F. Senatore, "Chemical Constituents and Biological Activities of Nepeta Species," *Chem. Biodivers.*, vol. 8, pp. 1783–1818, 2011.

[25] N. Melo *et al.*, "The irritant receptor TRPA1 mediates the mosquito repellent effect of catnip," *Curr. Biol.*, vol. 31, no. 9, pp. 1988-1994.e5, 2021.

[26] S. Dobler, G. Petschenka, and H. Pankoke, "Coping with toxic plant compounds – The insect's perspective on iridoid glycosides and cardenolides," *Phytochemistry*, vol. 72, no. 13, pp. 1593–1604, 2011.

[27] R. Uenoyama *et al.*, "The characteristic response of domestic cats to plant iridoids allows them to gain chemical defense against mosquitoes," *Sci. Adv.*, vol. 7, no. January, pp. 1–15, 2021.

[28] S. Bol, J. Caspers, L. Buckingham, G. D. Anderson-shelton, and C. Ridgway, "Responsiveness of cats (Felidae) to silver vine (Actinidia polygama), Tatarian honeysuckle (Lonicera tatarica), valerian (Valeriana officinalis) and catnip (Nepeta cataria) Responsiveness of cats (Felidae) to silver vine (Actinidia polygama)," *BMC Vet. Res.*, vol. 13, no. 70, 2017.

[29] S. J. Partridge *et al.*, "Iridoid Sex Pheromone Biosynthesis in Aphids Mimics Iridoid-Producing Plants," *Chem. Eur.*, vol. 27, pp. 7231–7234, 2021.

[30] H. Kries, F. Kellner, and M. O. Kamileen, "Inverted stereocontrol of iridoid synthase in snapdragon," *J. Biol. Chem.*, vol. 292, no. 35, pp. 14659–14667, 2017.

[31] T. Nguyen and S. E. O'Connor, "The Progesterone 5β-Reductase/Iridoid Synthase Family: A Catalytic Reservoir for Specialized Metabolism across Land Plants," *ACS Chem. Biol.*, vol. 15, pp. 1780–1787, 2020.

[32] F. Geu-flores *et al.*, "An alternative route to cyclic terpenes by reductive cyclization in iridoid biosynthesis," *Nature*, vol. 492, pp. 138–142, 2012.

[33] B. Dinda, D. Roy Chowdhury, and B. C. Mohanta, “Naturally Occurring Iridoids, Secoiridoids and Their Bioactivity. An Updated Review, Part 3,” *Chem. Pharm. Bull.*, vol. 57, no. August, pp. 765–796, 2009.

[34] D. Martins and C. V. Nunez, “Secondary Metabolites from Rubiaceae Species,” *Molecules*, vol. 20, pp. 13422–13495, 2015.

[35] K. Okada, “The Biosynthesis of Isoprenoids and the Mechanisms Regulating It in Plants,” *Biosci. Biotechnol. Biochem.*, vol. 75, no. 7, pp. 1219–1225, 2011.

[36] H. Kries, F. Kellner, and M. O. Kamileen, “Inverted stereocontrol of iridoid synthase in snapdragon,” vol. 292, pp. 14659–14667, 2017.

[37] Y. Iijima, D. R. Gang, E. Fridman, E. Lewinsohn, and E. Pichersky, “Characterization of Geraniol Synthase from the Peltate Glands of Sweet Basil,” *Plant Physiol.*, vol. 134, no. January, pp. 370–379, 2004.

[38] T. Kuzuyama and H. Seto, “Two distinct pathways for essential metabolic precursors for isoprenoid biosynthesis,” *Proc. Jpn. Acad.*, vol. 88, no. 3, pp. 41–52, 2012.

[39] Y. Iijima, D. R. Gang, E. Fridman, E. Lewinsohn, and E. Pichersky, “Characterization of Geraniol Synthase from the Peltate Glands of Sweet Basil 1,” vol. 134, no. January, pp. 370–379, 2004.

[40] C. Parage *et al.*, “Class II Cytochrome P450 Reductase Governs the Biosynthesis of Alkaloids,” *Plant Physiol.*, vol. 172, no. November, pp. 1563–1577, 2016.

[41] C. E. Rodrıguez-Lopez *et al.*, “Two bi-functional cytochrome P450 CYP72 enzymes from olive (Olea europaea) catalyze the oxidative C-C bond cleavage in the biosynthesis of secoxy-iridoids – flavor and quality determinants in olive oil,” *New Phytol.*, no. 229, pp. 2288–2301, 2020.

[42] V. Salim, B. Wiens, S. Masada-atsumi, F. Yu, and V. De Luca, “7-Deoxyloganetic acid synthase catalyzes a key 3 step oxidation to form 7-deoxyloganetic acid in Catharanthus roseus iridoid biosynthesis,” *Phytochemistry*, vol. 101, pp. 23–31, 2014.

[43] K. Asada *et al.*, “A 7-Deoxyloganetic Acid Glucosyltransferase Contributes a Key Step in

Secologanin Biosynthesis in Madagascar Periwinkle," *Plant Cell*, vol. 25, no. October, pp. 4123–4134, 2013.

[44] B. R. Lichman *et al.*, "Uncoupled activation and cyclization in catmint reductive terpenoid biosynthesis," *Nat. Chem. Biol.*, vol. 15, no. 1, pp. 71–79, Dec. 2018.

[45] M. Lange, A. L. Yellina, S. Orashakova, and A. Becker, "Virus-Induced Gene Silencing (VIGS) in Plants: An Overview of Target Species and the Virus-Derived Vector Systems BT - Virus-Induced Gene Silencing: Methods and Protocols," A. Becker, Ed. Totowa, NJ: Humana Press, 2013, pp. 1–14.

[46] D. Robertson, "VIGS Vectors for Gene Silencing: Many Targets, Many Tools," *Annu. Rev. Plant Biol.*, vol. 55, pp. 495–519, 2004.

[47] P. D. Sonawane *et al.*, "Plant Cholesterol biosynthetic pathway overlaps with phytosterol metabolism," *Nat. Plants*, vol. 3, no. December, 2016.

[48] P. Pankaj Sahu, S. Puranik, M. Khan, and M. Prasad, "Recent advances in tomato functional genomics: utilization of VIGS," *Protoplasma*, vol. 249, pp. 1017–1027, 2012.

[49] D. Orzaez *et al.*, "A Visual Reporter System for Virus-Induced Gene Silencing in Tomato Fruit Based on Anthocyanin Accumulation," *Plant Physiol.*, vol. 150, no. July, pp. 1122–1134, 2009.

[50] D. K. Liscombe and S. E. O. Connor, "A virus-induced gene silencing approach to understanding alkaloid metabolism in Catharanthus roseus," *Phytochemistry*, vol. 72, no. 16, pp. 1969–1977, 2011.

[51] Y. Qu, M. L. A. E. Easson, J. Froese, R. Simionescu, T. Hudlicky, and V. De Luca, "Completion of the seven-step pathway from tabersonine to the anticancer drug precursor vindoline and its assembly in yeast," *PNAS*, vol. 112, no. 19, pp. 6224–6229, 2015.

[52] K. Yamamoto *et al.*, "Improved virus-induced gene silencing allows discovery of a serpentine synthase gene in Catharanthus roseus," *Plant Physiol.*, vol. 187, pp. 846–857, 2021.

[53] L. Caputi *et al.*, “Missing enzymes in the biosynthesis of the anticancer drug vinblastine in Madagascar periwinkle,” *Science (80-.).*, vol. 4100, no. May, pp. 1–8, 2018.

[54] R. C. Misra, S. Sharma, Sandeep, A. Garg, C. S. Chanotiya, and S. Ghosh, “Two CYP716A subfamily cytochrome P450 monooxygenases of sweet basil play similar but nonredundant roles in ursane- and oleanane-type pentacyclic triterpene biosynthesis,” *New Phytol.*, vol. 214, no. 2, pp. 706–720, Apr. 2017.

[55] A. C. Velásquez, S. Chakravarthy, and G. B. Martin, “Virus-induced Gene Silencing (VIGS) in Nicotiana benthamiana and Tomato,” *J. Vis. Exp.*, vol. 28, pp. 20–23, 2009.

[56] H. Xu *et al.*, “Tobacco rattle virus-induced PHYTOENE DESATURASE (PDS) and Mg-chelatase H subunit (ChlH) gene silencing in Solanum pseudocapsicum L.,” *PeerJ*, vol. 20;6:e4424, pp. 1–22, 2018.

[57] J. C. Preston, L. L. Barnett, M. A. Kost, N. J. Oborny, and L. C. Hileman, “Optimization of Virus-induced Gene Silencing to Facilitate Evo-devo Studies in the Emerging Model Species Mimulus guttatus (Phrymaceae),” *Ann. Missouri Bot. Gard.*, vol. 99, no. May, pp. 301–312, 2014.

[58] S. Desmet *et al.*, “Rhizogenic agrobacteria as an innovative tool for plant breeding : current achievements and limitations,” *Appl. Microbiol. Technol.*, vol. 104, pp. 2435–2451, 2020.

[59] S. Y. Lee, C. Y. Lee, S. H. Eom, Y. K. Kim, N. Il Park, and S. U. Park, “Rosmarinic acid production from transformed root cultures of Nepeta cataria L.,” *Sci. Res. Essays*, vol. 5, no. May, pp. 1122–1126, 2010.

[60] S. S. Mahmoud, S. Maddock, and A. M. Adal, “Isoprenoid Metabolism and Engineering in Glandular Trichomes of Lamiaceae,” *Front. Plant Sci.*, vol. 12, no. July, 2021.

[61] N. Navet and M. Tian, “Efficient targeted mutagenesis in allotetraploid sweet basil by CRISPR/Cas9,” *Plant Direct*, no. March, pp. 1–13, 2020.

[62] B. Li, G. Cui, G. Shen, Z. Zhan, L. Huang, and J. Chen, “Targeted mutagenesis in the medicinal plant Salvia miltiorrhiza,” *Nat. Publ. Gr.*, vol. 7, no. 43320, pp. 1–9, 2017.

[63] Z. Zhou *et al.*, "CRISPR/Cas9-mediated efficient targeted mutagenesis of RAS in Salvia miltiorrhiza," *Phytochemistry*, vol. 148, pp. 63–70, 2018.

[64] A. B. R. Lichman *et al.*, "Uncoupled activation and cyclisation in catmint reductive terpenoid biosynthesis," vol. 5, pp. 1–16, 2018.

[65] B. R. Lichman *et al.*, "The evolutionary origins of the cat attractant nepetalactone in catnip," *Sci. Adv.*, vol. 6:eaba0721, no. May, 2020.

[66] L. Palmer and S. E. O'Connor, "Virus-Induced Gene Silencing in Nepeta," in *Virus-Induced Gene Silencing: Methods and Protocols*, vol. 2172, Springer, 2020, pp. 111–121.

[67] A. J. Simkin *et al.*, "Characterization of the plastidial geraniol synthase from Madagascar periwinkle which initiates the monoterpenoid branch of the alkaloid pathway in internal phloem associated parenchyma," *Phytochemistry*, vol. 85, pp. 36–43, 2013.

[68] R. Krithika, P. L. Srivastava, B. Rani, S. P. Kolet, and M. Chopade, "Dehydrogenase from Catharanthus roseus Reveals Cascaded Enzymatic Activity in Iridoid Biosynthesis," *Sci. Rep.*, vol. 5:8258, pp. 1–6, 2015.

[69] D. J. Kliebenstein, "A Role for Gene Duplication and Natural Variation of Gene Expression in the Evolution of Metabolism," *PLoS One*, vol. 3, no. March 3, 2008.

[70] G. D. Moghe and S. Shiu, "The causes and molecular consequences of polyploidy in flowering plants," *Ann. N. Y. Acad. Sci.*, vol. 1320, no. The Year in Evolutionary Bioligy, pp. 16–34, 2014.

[71] J. D. Rudolf and C.-Y. Chang, "Natural Product Reports structure, and opportunities of non-canonical terpene synthases," *Nat. Prod. Rep.*, vol. 37, pp. 425–463, 2020.

[72] F. Zhou and E. Pichersky, "The Complete Functional Characterisation of the Terpene Synthase Family in Tomato," *New Phytol.*, vol. 226, pp. 1341–1360, 2020.

[73] S. R. Johnson, W. W. Bhat, R. Sadre, G. P. Miller, and A. S. Garcia, "Promiscuous Terpene Synthases from Prunella vulgaris Highlight the Importance of Substrate and Compartment Switching in Terpene Synthase Evolution," *New Phytol.*, vol. 223, pp. 323–335, 2019.

[74] R. Höfer *et al.*, “Geraniol hydroxylase and hydroxygeraniol oxidase activities of the CYP76 family of cytochrome P450 enzymes and potential for engineering the early steps of the (seco) iridoid pathway,” *Metab. Eng.*, vol. 20, pp. 221–232, 2013.

[75] N. Esaki, S. Nakai, and S. Uesato, “Acyclic Monoterpene Rauwolfia serpentina Monoterpene Alcohols Primary Alcohol:NADP+ Oxidoreductase of Rauwolfia serpentina Cells: The Key Enzyme in Biosynthesis of Monoterpene Alcohols,” *J. Biochem.*, vol. 109, no. 2, pp. 341–347, 1991.

[76] D. L. Hallahan *et al.*, “Purification and Characterization of an Acyclic Monoterpene Primary Alcohol:NADP+ Oxidoreducatse from Catmint (Nepeta racemosa),” *Arch. Biochem. Biophys.*, vol. 318, no. April, pp. 105–112, 1995.

[77] H. Changjun, Q. Yajuan, L. I. Zhenghe, and Z. Xueping, “Virus-induced gene silencing and its application in plant functional genomics,” *Sci. China. Life Sci.*, vol. 55, no. 2, pp. 99–108, 2012.

[78] A. Purkayastha and I. Dasgupta, “Virus-induced gene silencing: A versatile tool for discovery of gene functions in plants,” *Plant Physiol. Biochem.*, vol. 47, no. 11–12, pp. 967–976, 2009.

[79] S. Valimehr, F. Sanjarian, H. sohi Hashemi, A. Sharafi, and S. Farzaneh, “A reliable and efficient protocol for inducing genetically transformed roots in medicinal plant Nepeta pogonosperma,” *Physiol. Mol. Biol. plants*, vol. 20, no. September, pp. 351–356, 2014.

[80] N. M. Arafa, A. M. M. Gabr, M. M. Ibrahim, S. Y., and I. Smetanska, “Study the effect of hairy root transformation on rapid growth (growth morphology) of Nepeta cataria in vitro cultures,” *J. Innov. Pharm. Biol. Sci.*, vol. 2, no. November, pp. 439–450, 2015.

[81] Y. Liu, M. Schiff, and S. P. Dinesh-Kumar, “Virus-induced gene silencing in tomato,” *Plant J.*, vol. 31, no. 6, pp. 777–786, Sep. 2002.

[82] Y. Liu, M. Schiff, and R. Marathe, “Tobacco Rar1, EDS1 and NPR1/NIM1 like genes are required for N-mediated resistance to tobacco mosaic virus,” *Plant J.*, vol. 30, no. 4, pp. 415–429, 2002.

[83] K. Koudounas, M. Thomopoulou, E. Angeli, D. Tsitsekian, S. Rigas, and P. Hatzopoulos,

"Virus-Induced Gene Silencing in Olive Tree (Oleaceae) BT - Virus-Induced Gene Silencing in Plants: Methods and Protocols," V. Courdavault and S. Besseau, Eds. New York, NY: Springer US, 2020, pp. 165–182.

[84] J. Papenbrock, H. Mock, R. Tanaka, E. Kruse, and B. Grimm, "Role of Magnesium Chelatase Activity in the Early Steps of the Tetrapyrrole Biosynthetic Pathway 1," *Plant Physiol.*, vol. 122, no. April, pp. 1161–1169, 2000.

[85] G. Qin *et al.*, "Disruption of phytoene desaturase gene results in albino and dwarf phenotypes in Arabidopsis by impairing chlorophyll, carotenoid, and gibberellin biosynthesis," *Cell Res.*, vol. 17, pp. 471–482, 2007.

[86] A. B. Dommes, T. Gross, D. B. Herbert, K. I. Kivivirta, and A. Becker, "Virus-induced gene silencing : empowering genetics in non-model organisms," *J. Exp. Bot.*, vol. 70, no. 1, pp. 757–770, 2019.

[87] A. Van Moerkercke, P. Steensma, F. Schweizer, J. Pollier, I. Gariboldi, and R. Payne, "The bHLH transcription factor BIS1 controls the iridoid branch of the monoterpenoid indole alkaloid pathway in Catharanthus roseus," *PNAS*, vol. 112, no. 26, pp. 1–6, 2015.

[88] S. R. Scofield and A. S. Brandt, "Virus-Induced Gene Silencing in Hexaploid Wheat Using Barley Stripe Mosaic Virus Vectors," in *Antiviral Resistance in Plants: Methods and Protocols*, vol. 894, 2012, pp. 93–112.

[89] J. Zhao, H. Jiang, G. Wang, Z. Wang, J. Dong, and J. Song, "Virus-Induced Gene Silencing in Diploid and Tetraploid Potato Species," in *Virus-Induced Gene Silencing: Methods and Protocols*, vol. 2172, Springer, 2020, pp. 39–50.

[90] T. Spengler, "How To Root Catnip Cuttings," *Gardening Know How*, 2021. [Online]. Available: https://www.gardeningknowhow.com/edible/herbs/catnip/how-to-root-catnip-cuttings.htm.

[91] W. Liang *et al.*, "Selection and evaluation of reference genes for qRT - PCR analysis in Euscaphis konishii Hayata based on transcriptome data," *Plant Methods*, vol. 14, no. 42, pp. 1–9, 2018.

[92] H. H. Hwang *et al.*, "Effective Agrobacterium-mediated transformation protocols for

callus and roots of halophyte ice plant (Mesembryanthemum crystallinum)," *Bot. Stud.*, vol. 60, no. 145, pp. 1–15, 2019.

[93] T. Zhao *et al.*, "An efficient method for transgenic callus induction from Vitis amurensis petiole," *PLoS One*, vol. 12, no. 6, pp. 1–13, 2017.

[94] X. U. Kun *et al.*, "A callus transformation system for gene functional studies in soybean," *J. Integr. Agric.*, vol. 16, no. 9, pp. 1913–1922, 2017.

[95] R. K. Singh and M. Prasad, "Advances in Agrobacterium tumefaciens-mediated genetic transformation of graminaceous crops," *Protoplasma*, vol. 253, no. Cell Biology in Agricultural and Food Science, pp. 691–707, 2016.

[96] S. K. Singh, S. Pattanaik, B. Patra, P. Paul, and Y. Liu, "BHLH IRIDOID SYNTHESIS 3 is a member of a bHLH gene cluster regulating terpenoid indole alkaloid biosynthesis in Catharanthus roseus," *Plant Direct*, no. July 2020, pp. 1–13, 2021.

[97] X. Qi, S. Bakht, M. Leggett, C. Maxwell, R. Melton, and A. Osbourn, "A gene cluster for secondary metabolism in oat: Implications for the evolution of metabolic diversity in plants," *PNAS*, vol. 101, no. 21, pp. 8233–8238, 2004.

[98] D. DellaPenna and S. E. O'Connor, "Plant Gene Clusters and Opiates," *Science (80-.).*, vol. 336, no. June, pp. 11–13, 2012.

[99] P. Paul, S. K. Singh, B. Patra, X. Sui, S. Pattanaik, and L. Yuan, "A differentially regulated AP2/ERF transcription factor gene cluster acts downstream of a MAP kinase cascade to modulate terpenoid indole alkaloid biosynthesis in Catharanthus roseus," *New Phytol.*, vol. 213, pp. 1107–1123, 2017.

[100] S. A. Kautsar, H. G. S. Duran, K. Blin, A. Osbourn, and M. H. Medema, "plantiSMASH: automated identification, annotation and expression analysis of plant biosynthetic gene clusters," *Nucleic Acids Res.*, vol. 45, no. April, pp. 55–63, 2017.

[101] H.-W. Nützmann, A. Huang, and A. Osbourn, "Plant metabolic clusters - from genetics to genomics," *New Phytol.*, vol. 211, no. 3, pp. 771–789, 2017.

[102] G. Prakash, M. Surabhi, R. Suman, N. Sushanto, and G. Gitishree, "Metagenomics

Approaches in Discovery and Development of New Bioactive Compounds from Marine Actinomycetes," *Curr. Microbiol.*, vol. 77, no. 4, pp. 645–656, 2020.

[103] R. H. Baltz, "Gifted microbes for genome mining and natural product discovery," *J. Ind. Microbiol. Biotechnol.*, vol. 44, no. 4, pp. 573–588, 2017.

[104] J. Nie, W. Zhou, J. Liu, N. Tan, J.-M. Zhou, and L. Huang, "A receptor-like protein from Nicotiana benthamiana mediates VmE02 PAMP-triggered immunity," *New Phytol.*, vol. 229, pp. 2260–2272, 2020.

[105] S. Shiu and A. B. Bleecker, "Plant Receptor-Like Kinase Gene Family: Diversity, Function, and Signaling," *Sci. STKE.*, vol. 113, no. December, 2001.

[106] X. Liang and J. Zhou, "Receptor-Like Cytoplasmic Kinases: Central Players in Plant Receptor Kinase – Mediated Signaling," *Annu. Rev. Plant Biol.*, vol. 69, pp. 267–299, 2018.

[107] J. Dufayard *et al.*, "New Insights on Leucine-Rich Repeats Receptor-Like Kinase Orthologous Relationships in Angiosperms," *Front. Plant Sci.*, vol. 8, no. April, pp. 1–18, 2017.

[108] Y. Qian, T. Zhang, Y. Yu, L. Gou, J. Yang, and J. Xu, "Regulatory Mechanisms of bHLH Transcription Factors in Plant Adaptive Responses to Various Abiotic Stresses," *Front. Plant Sci.*, vol. 12, no. June, 2021.

[109] Q. M. Dudley, S. Jo, D. A. Serna Guerrero, S. E. O. Connor, L. Caputi, and N. J. Patron, "Reconstitution of monoterpene indole alkaloid biosynthesis in Nicotiana benthamiana," *Preprint*, 2021.

[110] R. D. La Peña and E. S. Sattely, "Rerouting plant terpene biosynthesis enables momilactone pathway elucidation," *Nat. Chem. Biol.*, vol. 17, no. February, p. 205212, 2021.

[111] L. Dong *et al.*, "Characterization of two geraniol synthases from Valeriana offi cinalis and Lippia dulcis: Similar activity but difference in subcellular localization," *Metab. Eng.*, vol. 20, pp. 198–211, 2013.

[112] L. Dong, E. Jongedijk, H. Bouwmeester, and A. Van Der Krol, “Monoterpene biosynthesis potential of plant subcellular compartments,” *New Phytol.*, vol. 209, pp. 679–690, 2016.

[113] L. D. Roberts, A. L. Souza, R. E. Gerszten, and C. B. Clish, “Targeted Metabolomics,” *Curr Protoc Mol Biol.*, pp. 1–34, 2013.

[114] H. Fu, Y. Ma, S. Ma, and Z. Zhou, “Two new iridoid glycosides from Callicarpa nudiflora,” *J. Asian Nat. Prod. Res.*, vol. 0, no. 0, pp. 1–7, 2018.

[115] C. Long, K. S. Tojibaev, E. Alshammari, and M. L. Ashour, “The Genus Lagochilus (Lamiaceae): A Review of Its Diversity, Ethnobotany, Phytochemistry, and Pharmacology,” *Plants*, vol. 10, pp. 1–21, 2021.

[116] A. Venditti *et al.*, “Phytochemistry, micromorphology and bioactivities of Ajuga chamaepitys (L.) Schreb. (Lamiaceae, Ajugoideae): Two new harpagide derivatives and an unusual iridoid glycosides pattern,” *Fitoterapia*, vol. 113, pp. 35–43, 2016.

[117] R. Schuurink and A. Tissier, “Glandular trichomes: micro-organs with model status?,” *New Phytol.*, vol. 225, pp. 2251–2266, 2020.

[118] A. Tissier, J. A. Morgan, and N. Dudareva, “Plant Volatiles: Going ‘In’ but not ‘Out’ of Trichome Cavities,” *Trends Plant Sci.*, vol. 22, no. 11, pp. 930–938, 2017.

[119] K. Yazaki, A. Sugiyama, M. Morita, and N. Shitan, “Secondary transport as an efficient membrane transport mechanism for plant secondary metabolites,” *Phytochem Rev*, vol. 7, pp. 513–524, 2007.

[120] U. Gani, R. A. Vishwakarma, and P. Misra, “Membrane transporters: the key drivers of transport of secondary metabolites in plants,” *Plant Cell Rep.*, vol. 40, no. 1, pp. 1–18, 2021.

[121] S. T. Koike, H. R. Azad, and D. A. Cooksey, “Xanthomonas Leaf Spot of Catnip: A New Disease Caused by a Pathovar of Xanthomonas campestris,” *Plant Dis.*, vol. 85, no. Nov, pp. 1157–1159, 2001.

[122] S. Xu, A. Rahman, T. I. Baskin, and J. J. Kieber, “Two Leucine-Rich Repeat Receptor Kinases Mediate Signaling, Linking Cell Wall Biosynthesis and ACC Synthase in

Arabidopsis," *Plant Cell*, vol. 20, no. November, pp. 3065–3079, 2008.

[123] V. A. Reddy *et al.*, "Spearmint R2R3-MYB transcription factor MsMYB negatively regulates monoterpene production and suppresses the expression of geranyl diphosphate synthase large subunit (MsGPPS . LSU)," pp. 1105–1119, 2017.

[124] L. S. Sarker, A. M. Adal, and S. S. Mahmoud, "Diverse transcription factors control monoterpene synthase expression in lavender (Lavandula)," *Planta*, vol. 251, no. 1, pp. 1–5, 2020.

[125] X. Li, L. He, X. An, K. Yu, N. Meng, and C. Duan, "Glycosylated Monoterpenoid Production by VviGT14," *Genes (Basel).*, vol. 11, no. 485, 2020.

[126] B. Patra, S. Pattanaik, C. Schluttenhofer, and L. Yuan, "A network of jasmonate-responsive bHLH factors modulate monoterpenoid indole alkaloid biosynthesis in Catharanthus roseus," *New Phytol.*, vol. 217, pp. 1566–1581, 2017.

[127] X. He *et al.*, "Functional analysis of the heterotrimeric NF-Y transcription factor complex in cassava disease resistance," pp. 1185–1197, 2019.

[128] Y. Xie *et al.*, "Transcription factor FvTCP9 promotes strawberry fruit ripening by regulating the biosynthesis of abscisic acid and anthocyanins," *Plant Physiol. Biochem.*, vol. 146, no. November 2019, pp. 374–383, 2020.

[129] J. Gu *et al.*, "A novel cysteine-rich receptor-like kinase gene, TaCRK2, contributes to leaf rust resistance in wheat," *Mol. Plant Pathol.*, vol. 21, no. February, pp. 732–746, 2020.

[130] K. Yamamoto *et al.*, "Cell-specific localization of alkaloids in Catharanthus roseus stem tissue measured with Imaging MS and Singl-cell MS," *PNAS*, vol. 113, no. 14, pp. 3891–3896, 2016.

[131] R. M. E. Payne, D. Xu, E. Foureau, M. Ines, and S. Teto, "An NPF Transporter Exports a Central Monoterpene Indole Alkaloid Intermediate from the Vacuole," *Nat. Plants*, vol. 3, 2017.

Appendix

Appendix 1: List of gene sequences

C. roseus GES

ATGGCAGCCACAATTAGTAACCTTTCTTTCTTAGCAAAATCTAGGGCACTTTCAAGGCCTTCTTCTTC
TTCACTTTCATGGCTAGAAAGGCCTAAAACTTCATCGACTATTTGCATGTCTATGCCATCATCTTCAT
CATCATCATCATCTTCATCCATGTCTCTGCCTTTGGCAACTCCATTGATCAAAGACAATGAATCTCTC
ATCAAGTTCTTGCGCCAACCCCTGGTGCTTCCTCATGAGGTTGATGACAGCACAAAAAGGAGGGAA
TTGTTGGAAAGAACAAGAAAAGAACTAGAATTAAATGCAGAAAAACCATTGGAGGCCTTGAAGAT
GATAGATATAATTCAAAGATTGGGATTATCATATCATTTTGAAGATGATATTAATTCAATTCTCACA
GGATTTTCAAATATTAGCAGCCAAACTCATGAAGATCTCCTCACTGCTTCACTTTGTTTTCGATTGCT
TCGACACAATGGGCATAAGATCAATCCTGATATATTCCAAAAATTCATGGACAACAATGGAAAGTT
TAAAGATTCATTAAAGGATGACACATTAGGCATGTTAAGCTTATATGAAGCTTCATATTTGGGAGCC
AATGGAGAAGAAATATTGATGGAAGCCCAAGAATTCACCAAAACTCACCTGAAAAACTCATTGCCA
GCCATGGCACCATCTCTTTCTAAGAAGGTTTCTCAAGCTTTAGAGCAACCAAGACATAGAAGAATGT
TGAGGTTAGAAGCTAGAAGATTTATTGAAGAATATGGTGCTGAAAATGACCATAATCCAGACCTTC
TTGAGCTTGCAAAATTGGATTATAACAAAGTCCAATCTCTACACCAAATGGAATTGTCTGAGATAAC
AAGGTGGTGGAAACAATTAGGGCTTGTGGATAAACTCACCTTTGCTCGAGATCGACCCCTTGAATG
CTTTCTTTGGACAGTGGGATTATTACCAGAGCCTAAGTATTCAGGTTGCAGAATTGAGCTTGCAAAA
ACCATAGCCATTTTGCTTGTCATTGATGATATCTTTGATACTCATGGTACCCTAGATGAGCTTCTTCT
ATTCACTAATGCCATTAAAAGATGGGATCTTGAGGCCATGGAAGATTTACCAGAATATATGAGAAT
TTGTTACATGGCATTGTACAATACTACTAATGAAATTTGCTATAAAGTTCTTAAGGAAAATGGTTGG
AGTGTTCTTCCTTACCTAAAGGCAACGTGGATTGATATGATTGAAGGATTCATGGTTGAAGCAGAA
TGGTTCAATTCTGATTATGTACCAAACATGGAAGAATATGTAGAAAATGGAGTTAGAACAGCAGGA
TCATATATGGCCTTAGTCCATTTGTTCTTTCTAATAGGGCAAGGTGTCACTGAAGATAATGTGAAAT
TACTGATTAAACCCTATCCAAAGCTCTTTTCCTCCTCAGGAAGAATCCTTCGCCTTTGGGATGATTTG
GGAACTGCAAAGGAGGAACAAGAAAGAGGAGACTTGGCATCAAGCATACAATTGTTTATGAGAGA
GAAAGAGATAAAATCAGAAGAAGAAGGAAGGAAAGGAATATTGGAAATTATAGAGAATTTATGG
AAAGAATTGAATGGAGAATTAGTTTATAGAGAAGAAATGCCTCTTGCAATAATCAAGACAGCATTC
AACATGGCAAGAGCTTCCCAAGTTGTGTATCAACATGAAGAAGACACCTATTTTTCAAGTGTAGAT
AATTATGTAAAAGCTTTGTTTTTTACACCTTGTTTTTAA

N. cataria GESA

ATGTCGTGTGCAGGGAGCACCATCTTATCTTTGTCACAATCATCATTTCCCGATCGCTTCCCAAGGA
CCTCAATTTCATGGCATACAAGATTTTGCCCCGCCTCCGTCTCCCTTCGTCGCTTCTCCGTCTGCAAG
CTGCAGGCGGCAACCACGACATCACCCAATTTGGCCCCCGTAGTCCACAACAACGATCATATGCTT
GAAATTGTGGAGAGCACAAAGAAGAGAGAATATTTGGTTGAGAAGATTACAGAAAAGCTCACAGA
AGCATCATCAGAGAAACTCAAACTCATCGATGAAATCCAACGGCTGGGAATCGGCTACCACTTTGA
AGATGTCATCAACAGCATACTGCAGGTCCAGTGCTCCGCTTTCTCTACTGAAGAAGACCTCTTCACC
ACTGCTCTGCGCTTCCGTCTGCTCCGCCAGGCTGGCTTCCACGTCACCCCCGAGGTGTTTACGAGAG
TCAAGGACAAAAGTGGAAATTTCAAAGAATCCTTAGGCGGGGACACACTTGGGTTATTGAGTTTAT
ATGAAGCGTCGCATATAGGAGTTGATGGAGAGAGAATGTTGGAGGATGCTAGACAATTTACAGAG
TCTTATTTGAGACAATCACTAGCTAAGCTGGCACCTCGCCTTTGTAGCGAGGTGGGCCAGGCCCTA
GAGGTCCCGAGGCATCATAGGATGGTTAGATTGGAAGCAAGAAGATTCGTTCAGGAGTATGGTAA
GCAAAGTGGCCATGATCGAGACCTTTTGGAGCTAGCAATATTGGATTATAACCAAGTTCAAGCTCA
ACACAACATGGAACTCACTGAAATTGTAAGATGGTGGAAAGAGCTAGGTTTGGTGGATAAGTTAA

GTTTTGCACGAGATAGACCGTTGGAGTGTTTTTTATGGTCGGTTGGACTTCTTCCCGACCCGAAGTA
CTCAAATTGCAGAATAGAGCTTGCCAAAAATATAGCTATCTTATTAGTGATCGATGATATTTTCGAC
ACTTATGGCAAGATGGATGATCTAATCCTCTTCACCCATGCAATTCGAAGATGGGATCTTGAAGCAA
TGGAAACCTTACCCGAGTACATGAAAATATGCTATATGGCATTGTACAATACAACGAACGAGACAT
GCTATCGAGTGCTTAAGGAAACTGGACGAACTGTCTTCCCATATTTCAAATCCACGTGGATAGACAT
GATCGAAGGATTCATGGTGGAGGCGAAGTGGTTCAATGGTGGAATTGAACCTAATTTGGAAGAAT
ATATAGAGAATGGAGTGTCTACTGCGGGTGCATACATGGCTTTGGTGTACCTCTTCTTTCTAATTGG
AGAAGGTGTAACCGACCAAAACGCCTCACTTTTGATTAGAAAACCCTATCCTAAGCTCTTCTCTGCC
GCCGGTCGAATTCTTCGCCTTTGGGATGATCTCGGAACTGCTAAGGAGGAAGAAGAGCGTGGTGA
TCTGGCGTCAAGCACCCACTTATTCATGAAAGAGGAGAATTTATCAACAGAGGAAGATGCTAGAAG
TTGCATTTTGGACGAAATTTTCCGATTGTGGAAAGATTTGAATGGGGAGCTCATATCCGATAATAAA
GTGTTGCCATTGTCCATAATCAAAGTCGCACTTGACATGGCACGATCTTCCCAAGTTGTGTACAAGC
ACGAAGGCGACACTTATTTTTCAAGCGTGGACAATTATGTCGAAGCCCTATTTTTCACTCCTCTTGTT
TCATCTTAA

N. mussinii GES

ATGTCGTGTGCAGGGAGCACCATCTTATCTTTGTCACAATCATCATTTCCCGATCGCTTCCCAAGGA
CTTCAATTTCATGGCATATAAGATCTTGCCCCGCCTCCGTCTCCCTTCGTCGCTTCTCCGTGTGCAAG
CTGCAGGCGGCAACCACGACATCACCCAATTTGGCCCCCGTAGTCCACAACAACGATCATATGCTT
GAAATTGTGGAGAGCACAAAGAAGAGAGAATATTTGGTTGAAAAGATTACAGAAAAGCTCACAGA
AGCATCATCAGAGAAACTCAAACTTATCGATGAAATCCAACGGCTGGGAATCGGCTACCACTTTGA
AGATGTCATCAAGAGCATAATGCAGGTCCAGTGCTCCGCTTTCTCTACTGAAGAAGACCTCTTCACC
ACTGCTCTGCGCTTCCGTCTGCTCCGCCAGGCTGGCTTCCACGTCATCCCCGAGGTGTTTACGAGAT
TCAAGGACAAAAGTGGAAATTTCAAAGAATCCTTAGGCGGGGACACACTTGGGTTATTGAGTTTGT
ATGAAGCGTCGCATATAGGAGTTGATGGAGAGAGAATGTTGGAGGATGCTCGGCAATTTACAGAG
TCTTACTTGAGACAATCACTAGCTAAGCTGGCACCTCGCCTTCGTGGCGAGGTGGGCCAGGCCCTA
GAGGTCCCGAGGCATCATAGGATGGTTAGATTGGAAGCAAGAAGATTCATTCAGGAGTATGGTAA
GCAAAGTGGCCATGATCGAGACCTTTTGGAGCTAGCAATATTGGATTATAACCAAGTTCAAGCTCA
ACACAACATGGAACTCACTGAAATTGTAAGATGGTGGAAAGAGCTAGGTTTGGTGGATAAGTTAA
GCTTTGCACGAGATAGACCGTTGGAGTGCTTTTTATGGTCGGTTGGACTTCTTCCCGACCCGAAGTA
CTCTAATTGCAGAATAGAGCTTGCCAAAAATATAGCTATCTTATTAGTGATCGATGATATTTTCGAC
ACTTATGGCAAGATGGATGATCTAATCATCTTCACCCATGCAATTCGAAGATGGGATCTTGAAGCA
ATTGAAACCTTACCCGAGTACATGAAAATATGCTACATGGCATTGTACAACACAACGAACGAGACA
TGTTATCGAGTGCTTAAGGACACTGGACGAAGTGTCTTCCCATATTTCATATCTACGTGGGTAGACA
TGATCGAAGGATTCATGGTGGAGGCAAAATGGTTCAATGGTGGAATTGCACCTAATTTGGAAGAA
TATATAGAGAATGGAGTGTCTACTGCGGGTGCATACATGGCTTTGGTGTACCTCTTCTTTCTAATTG
GATATGGTGTAACCGACCAAAACGCCTCACTTTTGATTAGAAAACCGTATCCTAAACTCTTCTCCGC
TGCCGGTCGAATTCTTCGCCTTTGGGATGATCTAGGAACTGCTAAGGAGGAAGAAGAGCGTGGTG
ATCTGGCGTCAAGCATACACTTATTTATAAAAGAGAAGAATTTATCAACAGAAGAAGATGCTAGAA
GTCGCATTTTGGAAGAAATTTTCCGATTATGGAAAGATTTAAATGGGGAGCTCATATCCAATAATAA
AGTGTTGCCGTTGTCCATAATCAAAGTCGCACTTGACATGGCACGATCTTCCCAAGTTGTGTACAAA
CACGAAGGCGACAATTATTTTTCAAGCGTGGACAATTATGTCGAAGCCCTATTTTTCACTCCTCTTGT
TTCATCTTAA

H. officinalis GES

ATGTGGATCCTCGATCGCATGCCTTATGCCAACGCGCGCAGTAATATCAGCCTGATGAGCCAAAGC
TCCTGCCGTCGCTTCCTCAAAACGTCGATCCCGTGGCAGGAACGCGTTTGCCCGGCGGCGGCCAAT
ATTCCACTGCGCCGCTTTAGCGTTCACTGCAACCTGCCTGCCACCATGCCATCAACGAGCCCGCCGA
TTATTAATGGCACCAATACCACTTTGCAGAACACCAAAATCAAACGCAAAGATGAATGCGAACAGG

AGAAATCTATCAAGCGTCGTGATTACCTGGTCGAGCAAATTGCCCGTAAATTACAGATGATCACCA
CGGAATCCAGCGAAGAGTTAAAACGTATTGACGAAATTCAGCGTCTCGGTATTGGTTATCACTTTG
AAGATGTGATCGACAGCATCTTACGTGTGCAGTGCAGCGCGTTTTCCACAGAAGAAGACCTCTTCA
CCACCGCGTTGCGTTTCCGCCTGTTGCGCCATGCGGGCTTCCACGTTTCGCCGGAAGTGTTTATGCG
TTTTAAAGAAAAATCCGGCAAGTTCAAAGAAAGCGATACGCTGGGCCTGCTGTCGCTGTATGAAGC
TAGCAATATGCGCGCGGAAGGTGAGGAGATCCTTGATGAAGCCAAAGAGTTTACCGAAAGTCGGA
TGAAGCGCAGCCTGGATGAATTAGCGCCGCATCTGCGTGAACAAGTCGGTCAGGCACTGGAAGTT
GCCCGTCATATGCGTATGGGCCGTCTGGAGGCGCGTCGTTTTATCGAAGAATACAGCAAACAGAGC
GGTCATGACCGTGACCTGCTGGAACTGGCGACCCTGGATTATAACCAGGTTCAGGCGCAGCATAAA
ATGGAACTGGCTGAAATCATCCGCTGGTGGAAAGAGCTGGGTCTGGTTGATAAACTGAAATTTGG
CCGCGACCGCCCGCTTGAGTGCTTCCTGTGGACGGTGGGGCTGTTACCGGATCCGAAATATTCAAC
CTGCCGCATTGAACTGGCAAAGACCATCGCCCTGTTGCTGGTGCTCGACGATATTTTTGATACCTAC
GGCAAAATGGATGAGCTGATTCTGTTTACCCACGCCATTCGCCGTCGCGATCTGGAAGCGATGAAA
ACCTTGCCGGAGTACATGAAAATTTGCTATATGGCGCTGTACAACACGACCAACGAAATTGGCTAT
CGCGTGCTGAAAGACACCGGTCGCACCGTTCTGCCGTATCTGAAAAGCACCTGGATTGATATGATT
GAGGGTTTTATGGTGGAAGCAAAATGGTTTGGTGGCGGCATTTCGCCACCGAATCTGGAGGAATA
TATTGAAAACGGTGTTTCCACCGCCGGTGCCTATATGGCACTGGTGCATTTGTTCTTTCTGATCGGT
GAAGGCGTGACCGATCAAAACGCCAGTCTGCTCAGTCAGAAACCGTATCCGAAGCTGTTCTCCGCC
GCGGGTCGTATCCTGCGTTTATGGGATGATCTCGGCACGGCGAAAGAAGAGCAGGAACGTGGCGA
TCTGGCGAGCAGCATTCCGCTGTTTATTAAGGAAAAAAACCTCGCCACTGAAGAAGAAGTCACCAG
CCATATTCTCGAAGAGATTCTGCGCTTGTGGAAAGATATGAACGGCGAATTGATCAGTAAAAATAA
ATTCATGCCGCTCTCCATCATTAAAGTGGCGCTGAACATGGCCCGCGCCAGCCAGGTGGTTTATAA
ACACGATCATCATACCTATTTCAGCAACGTCGATAACTATGTTGAAGCGATTTTCTTCACGCCGCTG
GACTAA

C. roseus G8H

ACAAATTGTACACTTCCATTCCATGGATTACCTTACCATAATATTAACTTTACTATTTGCCTTGACTCT
CTATGAAGCCTTCAGCTACCTATCCAGAAGAACCAAAAACCTTCCTCCAGGACCATCGCCATTGCCG
TTCATCGGAAGCCTCCATTTATTAGGCGACCAACCACACAAATCCTTAGCAAAACTTTCCAAAAAAC
ACGGTCCAATTATGAGTCTCAAATTAGGCCAGATCACTACAATCGTCATATCTTCATCAACAATGGC
GAAAGAAGTTCTTCAAAAACAGGATTTAGCATTTTCAAGCAGATCAGTTCCAAACGCACTCCACGCT
CACAATCAATTCAAATTCTCCGTTGTATGGCTTCCGGTAGCCTCACGATGGAGAAGTCTTCGAAAAG
TTTTGAATTCTAATATATTTTCCGGCAATCGGCTCGACGCTAATCAACATTTGAGAACTAGAAAAGT
ACAGGAACTAATTGCGTATTGCCGGAAAAATAGCCAGAGCGGAGAAGCGGTTGACGTCGGCCGA
GCTGCTTTTAGAACTTCGTTGAATTTGTTGTCGAATTTGATTTTTTCAAAGGATTTGACGGATCCTTA
TTCGGATTCTGCCAAGGAATTCAAGGATTTGGTTTGGAATATAATGGTTGAGGCGGGGAAACCTAA
TTTGGTCGATTTTTTTCCCCTGCTTGAAAAAGTTGATCCTCAAGGTATACGACATCGTATGACGATTC
ACTTTGGGGAAGTTCTTAAGCTTTTTGGTGGACTTGTTAATGAAAGATTGGAGCAAAGAAGATCAA
AAGGGGAAAAAAATGATGTGTTGGATGTACTTCTAACTACCAGCCAAGAAAGCCCTGAGGAAATC
GATAGAACTCACATTGAGCGAATGTGCTTGGACCTGTTTGTAGCAGGGACGGACACAACATCAAGC
ACATTAGAATGGGCAATGTCAGAAATGCTTAAAAACCCAGACAAAATGAAGAAAACCCAAGATGA
ACTTGCACAAGTAATCGGCAGAGGAAAAACAATAGAAGAATCCGATATTAACCGCTTACCTTACTT
AAGATGCGTTATGAAAGAAACCTTAAGGATACATCCACCAGTTCCCTTCTTAATTCCTCGCAAAGTG
GAACAAAGTGTTGAGGTTTGTGGATACAATGTCCCTAAAGGATCACAAGTTCTTGTGAATGCTTGG
GCAATTGGACGTGATGAAACTGTTTGGGATGATGCTTTGGCATTCAAACCCGAGAGATTTATGGAA
TCTGAATTGGATATCCGTGGAAGAGATTTCGAGCTGATTCCGTTCGGTGCTGGCCGAAGAATTTGC
CCAGGGTTGCCATTGGCACTAAGGACTGTGCCTTTGATGCTTGGTTCTTTGTTGAACTCTTTTAATTG
GAAGCTTGAAGGTGGGATGGCTCCAAAAGATTTGGATATGGAGGAGAAGTTTGGTATTACACTGC
AGAAGGCTCATCCTTTGCGTGCTGTACCAAGCACCCTTTAAAATTCTTGCCCTTTTTTTTTTTAAGCTT

GCATAAAATAAGGGCAGTGCCAAACTGTCCTATTTCGCGTTATGTACTCTCTCCCTCATGGTTGGAA
TACGTTTAATATATGTATTCATCTTTTTTTTTTTAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA

N. cataria G8H

ATGGATTTCCTCACAATCTTCATTGGACCATTACTTGCCATCATCGCCTTATTCTATGCACGCTTAAC
AATCTCCACGAGGGGCAAAAACCTGCCCCCGGGGCCGACGCCGCTTCCCTTCATCGGAAACCTCCA
TTTGCTCGGCCACCACCCCCACAAGTCCCTCGCCCGTCTGGCCAAGACTTACGGCCCACTCATGCTC
CTCCGCCTCGGCTCCGAAAACACCGTCGTTGTCTCCTCCGCCGCCGTCGCTAAAGAAGTCCTCCAGA
AGCATGACCTCGCCTTCTCCAGCAGGAGCATCCCCGACGCCGTCCATGCTCACGACCAGTTCAAGTA
CTCCGTCGTTTGGCTCCCGGTGGCCTCGCGGTGGCGCAGCCTCCGCAAAACCATGAACTCCAACAT
CTTCTCCGGCAGCTGCCTCGACGCCAATCAGCATCTTAGGAGCAGGAAGGTGGAGGAGCTCATTGC
CTACTGCAAGGAAAGCTGCCGGAGGGCGGAAGCCGTCGACGTCGGCCGCGCCGCCTTCAGAACCG
CCTTGAACTCGCTCTCCAACACCATATTTTCGAGGGATTTGGCCGATCCGTTGTCGGATTCCGGTAA
GGAATTCAAGGATGTGGTGTCGAGTATCATGATGGAGGCAGGGAAGCCCAACTCCGTGGATTACT
ATCCTTTTTTATCAAGATTTGATCCACAAGGAATACGACGTCGCACCGCTATCTATTTCGACAAATTG
ATCAATATTTTCAAAGGTTTGATCGATGAAAGGTTGGAGAAAAGAAGATCTCAATACGAAGAAGCA
ACTGATGTTATCGATATCCTCCTCAACACCATCGAGCAAAACCCCGAAGAGATCGACCGAACTCACA
TCGAACGAATGTTTTTGGATCTGTTCTTGGCGGGAACGGACACAACTTCGAGCACAGTGGAATGGG
CAATGGCGGAGGCGATGAAAAATCCAGAAACAATGAAGAAAGCCAAGGCAGAGCTAGCACAAGT
GATCGGAAAAGGAAAAATAGTACAAGAAACCGACATACCGCATCTGCCATATCTGCAATGCATTGT
TAAAGAGACGCTGAGACTACACCCGCCGGTTCCCTTCCTCATTCCACGCAGAGTCCACGAAGACGT
CGACATCTTGGGCTACGTCGTGCCCAATAACTCGCAGGTCTTCGTGAACATGTGGGCCATCGGGCG
GGACCCCGATATGTGGAAAAACCCTCTACAGTTCGACCCCGACCGCTTCCACGATTCGGAGGTGGG
GTTGAGGGGGAAAGATTTCGAGCTGATTCCATTCGGCGCGGGCAGAAGAATCTGCCCGGGGCTGC
CGCTGGCGATGAGAATGGTGCCCGTAGTGCTGGGGTCGCTTCTGAATTCATTCGACTGGAAAGTG
GAAGGAGGGAAGGAATTAGAGATGAATGAGAAGTTTGGAATCACGTTGGAGAAGGCTCAACCAC
TTATGGCTGTCCCTATCGCTCTTTAG

N. mussinii G8H

ATGGATTTCCTGACAATCTCCATTGGATCATTACTTGCCATCATCGCCTTATTCTATGCACGCTTAAC
AATTTCCACGAGGGGCAAAAACCTGCCTCCGGGGCCGACGCCGCTGCCCTTCATCGGAAACCTCCA
CTTGCTCGGCCACCACCCCCACAAGTCCCTCGCCCGTCTCGCCAAGACTTACGGCCCACTCATGCTC
CTCCGCCTCGGCTCCGAAAACACAGCCGTCGTCTCCTCCGCCGCCGTCGCTAAAGAAGTCCTACAGA
AGCATGACCTCGCCTTCTCCAGCAGGAGCATCCCCGACGCCGTCCATGCTCACGACCAGTTCAAGTA
CTCCGTCGTCTGGCTCCCGGTGGCCTCGCGCTGGCGCAGCCTCCGCAAGACCATGAACTCCAACAT
CTTCTCCGGCAGCTGCCTCGACGCCAACCAGCATCTTAGGAGTAGGAAGGTGGAGGAGCTCATTGC
CTACTGCAAGGAAAGCTGCCGGAGAGCGGAAGCCGTCGACGTCGGCCGCGCCGCCTTCAGAACCG
CCCTGAACTCGCTCTCCAACACCATATTTTCGAGGGATTTGGCCGATCCGTTGTCGGATTCCGGTAA
GGAATTCAAGGAAGTGGTGTCGAGTATCATGATGGAGGCCGGGAAGCCCAACTCCGTGGATTACT
ATCCTTTTTTATCAAGATTTGATCCACAGGGAATACGACGTCGCACCGCTTTCTATTTCGATAAATTG
ATTAATATCTTCAAAGGTTTGATCGATGAAAGGTTGGAGAAAAGAAGATCTCAACACGAAGAAGC
AACTGATGTTATCGATATCCTCCTCAACAGCATCGAGCAAAACCCCGAAGAGATCGACCGAACTCA
CATCGAACGAATGTTTTTGGATCTATTCCTGGCGGGAACAGACACAACTTCGAGCACAGTAGAATG
GGCAATGGCGGAGGCGATGAAAAATCCAGAAACAATGAAGAAAGCCAAGGCAGAGCTGGCACAA
GTGATCGGAAAAGGAAAAATAATACAAGAAAGCGACATACCGCATCTGCCGTATCTGCAATGCATT
GTTAAAGAGACGCTGAGACTACACCCGCCGGTTCCCTTCCTCATCCCACGCAGAGTCCACGAAGAC
GTCGACGTCTTGGGCTACGTCGTGCCCAAGAACACGCAAGTCTTCGTGAACATGTGGGCCATCGGG
AGGGACCCCGATATGTGGAAAAACCCTCTAGAGTTCGACCCCGACCGCTTCCGCGATTCGGAAGTG
GGGTTGAGGGGGAAAGATTTCGAGCTGATTCCGTTCGGCGCGGGAAGAAGAATCTGCCCGGGCCT

GCCGCTGGCGATGAGAATGGTGCCCGTAGTGTTGGGGTCGCTGCTGAATTCATTTGACTGGAAAG
TGGAAGGAGGGAAGGAACTAGAGATGAATGAGAAGTTTGGAATCACGTTGGAGAAGGCTCAACC
ACTTATGGCTGTCCCTATCGCTCTTTAG

H. officinalis G8H

ATGGATTCCATTACAACAGCAGTTGGATTAGTATTTGCCCTTACTTTCTTATTCTATGCATTCCAAAC
ACTTTCCTTTAGTCGTCGGAAAAACCTGCCGCCGGGGCCGACGCCGCTGCCCGTAATCGGAAACCT
CCATCTGCTCGGCGATCAGCCCCACAAATCCCTCGCTCGTCTCGCCAAGATTTACGGGCCCCTCATG
CGGCTCCGCCTTGGCTCCGTCAACACCGTCGTCGTCTCCTCCGCCACCGTCGCCAGAGAAGTCCTGC
AGAAGCAAGACATGGCCTTTTCCAGCAGGAGCGCCCCCAATGCCGTCCATGCACACGACCAGTTCA
AGTACTCCGTCGTCTGGCTTCCGGTGGCGTCGCGGTGGCGCAGCCTCCGCAAGACTATGAACTCCA
ACATCTTCTCCGGCAGCCGCCTCGACGCCAACCAGCACGTCAGGAGTAGGAAAGTGGAGGAGCTC
ATTGCCTACTGCCGGGAAAGCAGTCGAAAGGCGGAGGCCGTCGATGTCAGCCACGCCGCCTTCAG
AACCTCCATGAACTCGCTGTCCAACACCATTTTTTCCAAAGATTTGGCCCACCCCCTTTCGGATTCCG
ATAAAGAATTCAAACAAGTGGTGTCCAAAATCATGATTGAGGCCGGCAAGGCCAACCTCGTGGATT
TCTTTCCCTTTTTAGCAAGATTTGATCCCCAGGGAATACGACGTCGCACCACTCTCTACTTCGGGAA
AGTTATCGACCTCTTCAGTGGCCTCATCGACGAAAGGTTGGAGAAAAGAAAATCACAACACGGAG
AAGAATCTGCAAGTGACGTGATCGACGTTCTCCTCAACGCCAGCGATCAAGAGATCGACCGAACTC
ACATCGAAAGAATGTGTCTGGATCTATTCGTAGCTGGAACAGACACAACTTCGAGCACGGTGGAAT
GGGCAATGGCGGAGGCGATAAAAAATCCAGAAACGATGAAGAAAGCGAAAGCGGAGCTAGAACA
AGTGGTCGGGAAAGGTAAAATAATAAGAGAAAGCGACATATCCCGGCTGCCGTATGTACGATGCA
TGGTGAAGGAGACGTTAAGACTCCATCCGCCCGTTCCCTTCCTCATCCCACGCAGAGTGGACCAAG
ACGTCGAGGTCATGGGCAGCATAGTGCCCAAAAACTCACGGGTCTTAGTTAACACGTGGGCCATCG
GGCGTGACCCCGATGTTTGGAAAGACCCTCTAGAGTTCAAACCCCACCGCTTCCTTGATTCTGAGGT
GGATGTACGGGGCACGGACTTTGAGCTGATTCCGTTTGGGGCGGGGAGAAGAATATGCCCGGGG
CTGCCGCTGGCGATGAGAATGGTGCCGGTAATGTTGGGGTCAATGTTGAATTCATTTGATTGGAAG
ATGGAAGGAGGGATGGAATTAGAGATGGAGGACAAGTTTGGAATTACTTTGGTCAAGGCTCACCC
GCTCAAAGCTGTACCTGTTCCCATCTCTTTCTAG

C. roseus HGOA

ATGACCAAGACCAATTCCCCTGCTCCTTCTGTCATCACTTGCAAAGCTGCTGTGGTATGGAAATCAG
GGGAGCCACCAAAGGTGGAAGAGATACAAGTTGATCCACCAAAGGCCTCAGAAGTTAGGATTAAG
ATGCTTTGTGCCAGTTTGTGCCACACTGATTTCCTTGCCTGCAATGGCCTTCCTGTTCCATTGTTCCCT
CGCATTCCTGGACACGAAGGAGTCGGAATGATCGAGAGCGTTGGAGAAAATGTTACGAACCTAAA
AGAAGGAGACATAGTGATGCCACTTTACTTGGGAGAATGTGGGGAATGCTTGAATTGCAAATCAG
GAAGGACAAATTTGTGCCACAAATATCCTTTAGGTTTTAGTGGATTATTGCTTGATGGAACATCAAG
AATGTCAATTGGAGAGCAAAAAGTATATCACCATTTCAGCTGTTCGACATGGTCAGAGTATATAGT
GATTGAAGCAGCTTATGCAGTGAAGGTAGATCCAAGGGTTTCTCTTCCACATGCTAGCTTTCTTTGC
TGTGGATTCACCACTGGTTTTGGTGCTACTTGGAGAGATGTCAATGTTGTCAAGGGCTCTACTGTG
GCTGTTCTAGGCCTTGGTGCTGTTGGACTTGGGGCTGTGCAAGGAGCTAAATCACAAGGAGCATCA
AGAATTATAGGTTTGGATATCAACGACAAGAAACGTGAAAAGGGAGAAGCATTTGGAATGACTGA
ATTTATAAATCCAAAAGGTTCAAACAAATCAATTTCTGAACTAATAAATGAAGCAACTGGTGGATTA
GGACTTGACTATGTCTACGAATGCACTGGAGTTCCAGCTTTACTCAATGAAGCCATTGAATCCTCTA
AAGTGGGACTTGGAACTGCAGTATTGATTGGGGCAGGACTTGAAACAAGTGGAGAAATCAAATTC
ATTCCTCTTCTGTGTGGTAGAACTGTTAAAGGTTCAATTTATGGTGGAGTAAGACCTAAATCAGACC
TCCCCACTTTAATTGAAAAGTGCATAAACAAGGAAATTCCAATGGATGAACTAATGACTCATGAAG
TTTCATTGTCTGAGATAAACAAAGGATTTGAGTACCTTAAGCATCCTGATTGTGTCAAAGTTGTTAT
CAAGTTCTAA

N. cataria HGOA

ATGGCCGACAACACCAATCCGGGAGTCATAACGTGCAAAGCCGCCGTGGCATGGAAGCCTCTGGA
GCCGTTGAAGGTGGAGGAGATTCAAATACAACCACCAAAACCAACAGAAGTTAGGATCAAGATGT
TGTATGCCAGCATTTGTCACACAGATACAGTGTGCTGGACAGGCATCCTTTATGACATGTTCCCTCG
GGTTCTAGGACATGAAGGTGTAGGTGTGATTGAAAGTGTTGGAGAAGAAGTGACAGACCTGAAG
GTGGGAGATACTGTGATGCCACTCTACGTTGGAGAGTGTGGAAAATGCCCCAACTGCGCCTCGGG
GAAGTCGAATTTCTGCTCCAATTACCCTCTAACATTGGGCCTGATGGCTGATGGCACAACAAGAAT
GTCTGCTAAGGGGCAGCAACTGTACCAATTGTTTAGCTGCTCGACTTGGTCCGAATACGCGGTTATT
GAATCAAACTACGTGGTTAAGGTGGATCCCAAGCTGTCGCTTTCCGGAGCCAGTCTCCTTACCTGTG
GTTTCACAACAGGTTATGGAGCAGTATGGAAAGTAGTCAATTTTGAAAAAGGTTCAACTGTTGCTG
TAGTAGGCCTTGGTGCTGTTGGATTTGGAGCAGTTAAGGCAGCTCAAATCTTAGGAGCCTCGAAGA
TAATCGGGGTGGACATTAATGAAACGAGACGCGATAAGGCAAAAGCTCTTGGGGTTACAGACTTC
ATAAATCCTAAGAGTTCTGATAAAACTATGTCTCAACTTATCCAAGAAGCCACCGGAGGACTAGGT
GTGGATTGCTGCATTGAGTGCACGGGCGTTCCCTCTCTCCTCAACGAAGCCATTGCAAGCACAAAA
GTGGGATCCGGGGAGGTAGTGTTGATCAGTGCAGGAGACGAGCAGAATACGGAGCTTAACTACG
TTCCTCTGCTGTTGGGGAGGACTGTCAAGGGAACAATCTACGGTGGTGTGAGGATTCATTCAGATC
TTCATAAAATTGTTGACAAATGTATCAGCAAGGAAATCGATCTCGATGAAATCATAACTCATGAGGT
TTCACTTGCTGATGTTAACAAGGGATTTCTGGAATACATGAAGCAGCCTGATTGTGTCAAGGTTATA
ATCAAGTTCTGA

N. mussinii HGOA

ATGGCCGACAACACCACTCCGGGAGTTATAACATGCAAAGCCGCCGTGGCATGGAAGCCGCTGGA
GCCGTTGAAGGTGGGGGAGATTCAAGTACATCCACCAAAACCAACAGAAGTTAGGATCAAGATGT
TGTATGCCAGCATTTGCCACACAGATACAGTGTGCTGGACAGGCATCCTTTATGACATGTTCCCTCG
GGTTCTAGGACATGAAGGTGTAGGTGTGATTGAAAGTGTTGGGGAAGAAGTGAGAGACCTGAAG
GTGGGAGATACTGTGATGCCACTCTACGTTGGAGAGTGTGGAAAATGCCCGAACTGCACCTCGGG
GAAGTCGAATTTCTGCTCCAAATACCCTCTAACATTGGGCCTGATGGCTGATGGCACAACAAGAAT
GTCTGCTAAGGGGCAGCAACTGTACCAATTGTTTAGCTGCTCAACTTGGTCCGAATATGCGGTTATT
GAATCAAACTACGTGGTTAAGGTGGATCCGAAGCTGTCGCTTTCCGGAGCCAGTCTCCTTACCTGT
GGTTTCACAACAGGTTATGGATCAGTATGGAAAGTAGTCAATTTTGAAAAAGGTTCAACTGTTGCT
GTAGTAGGCCTTGGTGCTGTTGGTTTTGGAGCAGTGAAGGCAGCTCAAATATTAGGAGCCTCGAA
GATAATTGGGGTGGACATTAATGAAACGAGACGCGATAAGGCAAAAGCTCTTGGGGTTACAGACT
TCATAAATCCTAAGAGTTCTGATAAAACTATGTCTCAACTTATCCAAGAAGCCACCGGAGGAGTAG
GTGTGGATTGCTGCATCGAGTGCACGGGCGTTCCCTCCCTCCTCAACGAAGCCATTGCAAGCACAA
AAGTGGGATCCGGGGAGGTAGTGTTGATTAGTGCAGGAGACGAGCAAAATACGGAGCTTAACTAC
GTTCCTCTGCTATTGGGGAGGACTGTCAAGGGAACAATCTATGGTGGTGTGAGGATTCATTCGGAT
CTTCATAAAATTGTTGACAAATGTATCAGCAAGGAAATCGATCTCGATGAAATCATAACTCATGAG
GTTTCACTTGCTGATGTTAACAAGGGATTTCTGGAATACATGAAGCAGCCTGATTGTGTCAAGGTTA
TTATCAAGTTCTAA

H. officinalis HGOA

ATGGCAGGAGTCATAACATGCAAAGCCGCGGTGGTGTGGAAGGCGTCGGAGCCGTTGAAGGTGG
AGGAAATTCAAGTAGAACCACCAAAATCAACAGAAGTCAGGATCAAGATGTTGTATGCCAGCATGT
GCCATACAGATACCGTATGTTGGGCGAACTCCTTCTATAATATGTTTCCTCGAATCCTAGGCCACGA
AGGGGCGGGAGTGATTGAAAGTATTGGAGAAGAAGTGGCAAACCTCAAGGTGGGAGACACTGTG
ATGCTACTCTACGTCGGAGAGTGTGGAAAATGCCCGAACTGTGAGTTGGGGAAGACGAATTTTTGC
TCCAAATATCCTATGACATTTGGCCTAATGGTGGATGGCACGACAAGAATGTCTGCTAGGGGCCAA
CAAATGTTCCAATTGTCTAGCTGCTCGCCTTGGTCTGAATATGCGGTTATTGAATCAAACTACGTGG

TTAAGATAGACCCGAGAATATCGCCTTCCGAAGCCAGCCTCCTTACCTGTGGTTTCACAACAGGTTA
TGGAGCAGTATGGAAAGAACTCAAAGTTGGACAGGGTTCAACTGTTGCTGTAATAGGCCTTGGAG
CTGTTGGATTAGGAGCTGTGAAGGCATCACAAATCTTAGGAGCATCGAAGATAATTGGGATCGATA
TAAATGAAATGAAACGAGACAAGGCAAAAGCTTTTGGGGCTACAGACTTCATAAATCCCAAGGATT
CTGATAAAACTATGTCTGAACTGATCCAAGAAGCCACCGGAGGACTAGGTGTGGATTACTGTGTTG
AGTGTACGGGCGTACCCTCCCTCCTCGATGAAGCCATTGCAAGCACAAAAATGGGAGTCGGAGAG
GTAGTGTTGATCAGTGCCGGAGATGAGCCGAAGACGGAGCTTAGCTACATGCCTCTGTTGCTAGG
GAGGACTCTTAAGGGGACAACCTACGGTGGAGTAAGAACTCACTCGGATCTTCCTAAAATCATCGA
GAAATGTATCAACAAGGAAATTGAGCTCGGTGATCTCATAACTCATGAGGTTTCACTTGCTGATGTT
AACAAAGGATTTATGGAGTACATGAAGCAGCCTGACTGTGTGAAGGTTATTATCAAATTCTGA

N. cataria HGOB1

ATGGCGAAATCAGCAGAAACAGAGCACCCAGTGAAAGCTGTTGGTTGGGCTGCCAGAGATACATC
TGGAGTCTTCTCCCCTTTCAAATTCTCAAGAAGGGCAACTGGTGAGCGTGATGTTCAGTTCAAAGTT
CTTTACTGCGGCGTCTGCCATTCTGATCTTCATATGATCAAGAATGAGTGGGGTTTCACACAGTACC
CTATTGTTCCTGGGCATGAGATTGTGGGCATAGTAACTGAGGTGGGAAACAAGGTGGAGAAAGTC
AAGGTGGGAGACAAAGTAGGCGTTGGTTGCTTGGTCGGATCGTGCCGCCAATGTGACCAATGCTC
CGATGATCTCGAAAATTACTGCCCCAAACAGATTCTCACCTACAGCGCACCTTACTTGGATGGAACC
ATCACTTACGGTGGTTACTCCGATCTCATGGTTGCTGACGAGCACTTTGTCGTCCATTGGCCCGAGA
ATTTTCCCATGGACCTCGGTGCTCCTCTTCTCTGTGCCGGCATTACTACTTACAGCCCCTTGAGGTAC
TTTGGACTCGACAAGCCCGGGATGAATGTTGGGATTGTGGGTCTTGGTGGCCTCGGTCATGTTGCT
GTCAAGTTTGCCAAGGCTTTTGGGACTAAAGTCACCGTCATCAGTACTTCTGCAAGCAAGAAGGAG
GAAGCGCTCGGGAAGCTTGGTGCTGATGAGTTCTTGGTTAGCCGTGACCCCGAACAGATGCAGGC
AGCAGCTGGCACATTGGATGGCATCCTTGATACTGTCGCAGCAGTTCATCCCCTGCCGCCACTGCTA
AGCTTACTGAAGCCTCATGGGAGGCTTATCTTGGTCGGGGTTCCAGAAAAGCCTCACGAGCTTCCA
GCCTTTCCTCTGATTATGGGGAGAAAGTCGATTTCGGGCAGTGGGATCGGAGGGTTGAAGGAGAC
GCAGGAGATGGTGGATTTTGCAGCGAAGCACAACATACTGCCGGATGTGGAGATGATTCCCATAG
ACTACGTCAACACGGCTATGGAGCGTCTCGAGAAATCCGATGTCAAGTACCGGTTTGTCATCGACG
TTGCCGGTTCGCTGAAAGCCGACTAG

N. cataria HGOB2

ATGGCGAAAACCCCAGAAACAGAGCACCCAGTGAAGGCCTTTGGATATGCGGCCACTGACAATTCT
GGCACCTTCTCTCCCTTCAAGTTTTCTAGGAGGGCTACTGGTGAGCGTGATGTGCAGTTCAAGGTGT
TGTATTGTGGTGTTTGCCATTCAGATCTGCACATGGTCAAGAATGAGTGGGGTGTCACTCACTACCC
CATCGTTCCCGGGCATGAAATTGTGGGAGTTGCAACTGAAGTTGGTAGCAAGGTTGAAAAGGTTA
AGGTTGGCGACAAAGTTGGTGTCGGAGTGATTGTGGGTTCATGCCGCCAGTGTGATCAGTGCAGC
AATGATCTTGAAAACTACTGCTCCAAGCAAATACTCACCTACGGTGCCCCTTATCTAGATGGCACCA
TTACTCGCGGAGGCTACTCTGATATCATGGTGGCAGACGAGCATTTCATCATCCGTTGGCCTGAAA
ACTTCCCCCTTGATGCTGGAGCTCCTCTCCTTTGCGCTGGAATCACAACTTACAGTCCACTTAGGTAC
TTTGGACTCGACAAACCTGGCCTAAGAGTTGGTGTCAACGGTCTTGGTGGGCTAGGCCACGTAGCT
GTCAAATTCGCCAAAGCTTTTGGTACTAAGGTCACAGTTATCAGCTCATCTCTCCGCAAGAAGAAGG
AAGCTGTTGAAAAACTCGGTGTAGATGAGTTTTTGGTCAGTACTGATCCAGAGCAGATGCAGGCTG
CAATAGGCACACTAGATGGCATCATTGACACAGTCTCCGCGGAACATTCAGTTGTGTCTTTGCTGAG
TTTGTTGAAGCCACACGGCAGGCTGATTGTGGTTGGTGCACCGGAAAAGCCCCATCAGCTCCCTGC
TTTCCCCTTGATCACTGGTAGGAAGTCCATATCTGGGAGTGCAATCGGAGGGATGAAGGAGACGC
AAGAAATGATTGATTTCGCGGCCAAGCACAACATTCTGCCGGATGTGGAGGTCATACCGATAGATT
ACATAAACACGGCTATGGAACGGCTTCTGAAATCGGATGTGAAGTATAGGTTCGTGATTGATGTTG
AGAAATCATTGAAGGCTTGA

N. cataria HGOB3

ATGGCGAAATCAGTGAACGCCGTCGGCTGGGCCGCCCGCGACACCTCCGGCCTCCTCTCCCCTCTC
AAATTCGCCCGCCCCGCCGCCGGCGACCGCGACGTCCAGTTCAAGGTCCTCTACTGCGGCATCTGC
CACTCCGATCTCCACAGCATCAAGAACGAGTGGGGCAACACCCAATACCCCATCGTCCCCGGCCAC
GAGATCGTCGGCGTCGTCACCGCCGTCGGCCCCAAGGCCACCAAATTCAAGGCCGGCGACAAAGT
CGGCGTCGGCTGCCTCGTCAATTCCTGCCGGCATTGCGACCAATGCGCCGCCGATCTCGAGAATTA
CTGCCCGCAGCTCGTCCAAACCTACAATAGCGTCCTCGCCGACGGCTCCGTCACCTACGGCGGCTAC
TCCGATGTGATGACGGCGGACGAAGACTTCGTCATCCGCTGGCCGGAGAATTTCCCCATGGATAAG
GGCGCCCCGCTATTGTGCGCCGGGGTGACGACGTACAGCCCGTTGCGATACTACGGGCTCGACAA
GCCGGGCCTCCACATCGGCGTCGTCGGGCTGGGCGGGCTGGGCCATGTGGCGGTCAAGTTCGCCA
AGGCCTTCGGGACGAAGGTGACGGTGATCAGCACCTCCGCCGGAAAGAAGAAGGAGGCGATTGA
GACGCTAGGCGCCGACGCGTTCCTAATAAGCAAGGACGCGTCGGAGATGCAGGCGGCGGCGGGG
ACGCTGGACGGGATGATTGATACGGTCTCGGCCCAGCATCCCCTGCCGCCGCTTTTGAGCCTGTTG
AAGCCGCATGGGAAGCTGATTTTGGTCGGCGCGCCGGATAAGCCGTTGGAGCTGCYGGCATTCCC
CCTCATCATGGGGCGAAAGATGGTGGCGGGGAGTATTATCGGAGGGATGAAGGAAACGCAGGAG
ATGATCGATTTCGCGGCGAAGCATAATATATTGCCGGACGTGGAGATGATTCCGATGGACTACGTG
AACACGGCGATGGAACGGCTGGCTAAAGCCGACGTCAAGTATCGCTTTGTGATTGATGTGGGGAA
TACGCTGACTAAAGCCGATTGA

N. cataria HGOB4

ATGGCAAAATCAGCGGAGGCAGTGGAGGCCTTTGGATACGGTGCGACCGACTCAGCCGGCCATCT
CTCCCCCTTCCGCTTCTCCAGAAGGGCCACCGGCGACCAAGACGTGCAGTTCAAAGTTCTCTACTGC
GGCATCTGCCACACCGATCTTCACTACATCAACAACGAATGGGGCTTCTCTCACTACCCTCTCGTCCC
AGGGCATGAGATCGTCGGCCTAGTCACACAGGTAGGAACCAGAGTCCACAACTTCAAACCCGGAG
ACACAGTCGGCGTCGGATGCATCGTCGGATCCTGCCGATCATGCCACAGCTGCGCCAATCATCTCG
AGAATTACTGCCCCAACGTCATACCCACCTACAGCGCCCCTTACTACGACGGAACCATCACCTACGG
CGGTTACTCCGATATAATGGTCGTCCACCACCATTTCATTGTTCGTATCCCGGACAACATGCCTCTCG
ACGCAGCTGCCCCGCTTCTATGCGCCGGGATCACCACTTACAGCCCCTTGAGATACTTCGGACTCGA
CAAACCTGGCCTACACATCGGCGTTGTTGGCCTCGGCGGGTTAGGCCATCTTGCCGTTAAATTCGCC
AAGGCTTTTGGCGCTAAGGTCACCGTCATTAGCACCTCCCCTAACAAAAAAGACCAAGCACTCTCAC
ATTTGGGAGCCGATTCCTTCTTGATCAGTAAAAATGCTTGTGAAATGAAGGGTGCTATGGGAACAA
TGGATGGTATTCTTGATACGGTATCAGCGGTTCATCCTGTGATGGAGTTGGTCGGGTTGTTGAAGG
CTCAGGGGAAGCTGATAGTGCTGGGTGCACCAGAGAAGCCGCTGGAGGTGACAGCGCTTCCTCTG
ATAGCAGGAAGGAAGATGATAGCGGGGAGTGGGATCGGAGGGATGAAGGAGACGCAGGAAATG
ATTGATTTTGCAGCCAAACATAATATCAAGGCAGATATCGAACTCATTCCCATGGACTACGTTAACA
CTGCCATGGACCGTCTGGCCAAAGCGGATGTCAATTATCGTTTCGTCATAGATGTTGCCAACACCAT
CAAAAATTCATTATTACTTTAG

N. cataria ChlH

TTGCAGAGTTTCTCATCAATGGCTGCTTCTTTGGTTTCATCGCCATTTACACTTCCAAAATCCAAAAC
AGTAAACCTTTCATCACTCTCACAAAAGCATTACCTCCTCCACTCTTTCCTTCCCAAGAGAATCACCA
AAACCAACATAAATTCATCCCAGAAATTCAAATGCAATGCCATTGGCAACGGTCTCTTCACTCAGAC
CACGCAAGAAGTCCGTAGAATCGTGCCCGAGAAATCGAACCTCACAACGGTTAAGGTTGTCTACGT
GGTGCTAGAAGCTCAATACCAATCATCCCTCACAGCTGCAGTCCAAGCTCTCAACAAAAATGGAGA
GTATGCTTCATTTGAGCTTGTGGGCTACTTGGTTGAGGAGCTGAGAGATAAGGACAACTACGAGA
GCTTCTGCAAGGATCTCGAAGACGCCAATATCTTCATCGGATCACTCATTTTCGTGGAGGAGTTGGC
GTTGAAGGTCAAGGCTGCTGTGGAGAAGGAGAGAGAGAGGCTTGATGCTGTGCTGGTTTTCCCAT
CAATGCCTGAGGTGATGAGATTGAACAAGCTTGGATCTTTCAGCATGTCGCAGTTGGGGCAGTCGA

AAAGCCCCTTCTTTCAGCTGTTCAAGAAGAAGGGCAAATCTAATGCAGCCTTTGCAGACAGCATGTT
GAAGCTTGTAAGAACTTTACCTAAAGTCTTGAAGTATTTGCCCAGTGATAAGGCTCAAGATGCAAG
GTTGTACATTCTCAGCCTCCAGTTTTGGCTTGGTGGATCGCCTGATAACCTGATGAACTTCTTGAAG
ATGATCTCTGGATCATATGTGCCAGCGTTGAAACAGGCGAAAATCGAGTATTCTGATCCAGTTCTGT
ATTTTGATAATGGGATTTGGCACCCTTTGGCTCCCTGTATGTATGATGATGTGAAGGAGTATTTGAA
TTGGTATGCAACGAGAAGGGATGCTAATGAGCAGCTCAAGAGCAAGAATGCACCTGTGGTTGGGC
TTGTTTTGCAGAGGAGTCATATTGTGACTGGTGATGATAGTCACTATGTTGCTGTGATCATGGAGCT
CGAAGCTAAAGGGGCGAAGGTGATCCCGATCTTTGCTGGTGGCCTTGATTTCTCTGGGCCGGTGG
AGAAGTACTTCATTGATCCGATCACCAAGAAACCAATGATAAACTCAGTTGTGTCGTTGACTGGTTT
TGCTCTTGTGGGAGGGCCCGCTAGGCAGGACCATCCGAGGGCAATCGAGGCCTTGATGAAGCTAG
ATGTGCCTTACATTGTGGCACTGCCGTTGGTGTTCCAGACAACGGAAGAGTGGCTGAACAGCACGT
TGGGGTTGCATCCCATTCAGGTTGCTCTGCAAGTGGCTCTCCCTGAGCTTGATGGAGGCATGGAGC
CTATTGTTTTCTCGGGGCGAGATCCAAGAACAGGGAAATCACATGCTCTTCACAAGAGGGTGGAGC
AGCTCTGCACCAGAGCTATCAGATGGGCTGAACTCAAGAGGAAAACAAAGGCGGAAAAGAAGCTA
GCAATCACTGTATTTAGCTTCCCACCAGACAAAGGCAATGTAGGAACTGCTGCTTACCTCAATGTCT
TCTCTTCCATCTACTCTGTGCTCAAAGAGCTCAAGAAAGATGGCTACAACGTTGATGGCCTTCCAGA
AACTTCAGAAGCCTTGATCGAAGAAGTCATTCACGATAAGGAGGCCCAATTCAACAGCCCCAACCT
CAACGTAGTTCACAAGATGAGCGTCCGGGAATACCAGAATCTGACTCCTTACTCTACTGCTCTAGAA
GAAAACTGGGGAAAGCCTCCAGGCAACTTGAACTCCGATGGTGAAAATCTTCTTGTCTATGGCAAA
CAGTATGGAAACATCTTCATTGGTGTTCAGCCTACTTTTGGCTACGAGGGTGATCCTATGCGTCTTC
TCTTCTCCAAATCCGCCAGCCCACACCACGGATTTGCTGCATACTACTCATATGTGGAGAAGATCTTT
AAGGCTGACGCAGTTCTGCACTTTGGAACTCATGGTTCTCTTGAATTCATGCCTGGGAAACAAGTTG
GCATGAGTGATGCCTGCTTCCCCGACAGTCTTATTGGAAACATCCCCAACATCTACTACTACGCTGC
TAACAACCCGTCTGAGGCCACTGTTGCAAAGAGGCGTAGCTATGCGAATACTATCAGTTACCTGAC
TCCTCCAGCTGAGAATGCAGGTCTCTACAAAGGTCTGAAGCAGTTGAGCGAGCTTATTGCTTCTTAC
CAATCTCTCAAAGATTCGGGCCGTGGCCCTCAAATTGTGAGCTCTATTATCAGTACTGCTAGGCAGT
GCAATCTTGATAAAGATGTGGTTCTTCCTGAAGATGGAGTCGAGCTCTCTCAAAAAGAGCGTGACT
CTGTAGTGGGACAAGTGTATTCCAAGATTATGGAGATCGAATCGAGATTGCTACCTTGTGGCCTTC
ATGTCATTGGTGAGCCTCCAACAGCAATGGAGGCAGTGGCTACACTGGTCAATATCGCAGCATTGG
ATCGTGAAGAGGAACAGATTTCGTCTCTTCCTTCTATATTAGCGCAGACTGTGGGGAGAGAAATTG
AAGATATCTACAGAGGAAGTGATAAGGGTGTTCTGCGCGATGTGGAGCTCCTTAAACAGATCACTG
AAGTATCTCGAGGTGCAATCAGTGCTTTTGTGGAAAGAAGCACCAACAGCAAAGGGCAAGTTGTT
GAAGTAGCTGATAAGCTCACATCAATCCTTGGTTTTGGTGTAAACGAGCCATGGATTCAATACTTGT
CGAACACCAAGTTTTACCGGGCTGACAGAGAAAAACTCAGAGTCTTGTTCGGATACTTAGGTGAAT
GCTTGAGACTGATTGTAACTGATAACGAGGTGGGAAGCTTGAAACAAGCTTTGGAAGGGAAGTAT
GTCGAGCCAGGGCCAGGTGGGGACCCGATCAGAAACCCGAAGGTGCTGCCTACAGGAAAGAACA
TCCATGCCTTGGACCCGCAGTCCATTCCAACAACTGCAGCTATGCAGAGTGCTAAAGTGGTGGTGG
AGAGGCTACTCGAGAGGCAGAAGATCGATAATGGTGGAAAATATCCCGAGACAGTTGCTTTAGTA
CTATGGGGTACTGATAACATCAAGACATATGGTGAATCACTAGCTCAAGTTCTTTGGATGATTGGA
GTTACACCGGTAACTGATGGGCTTGGGCGTGTTAACCGTGTGGAGATCGTGAGCCTTGAAGAGCT
CGGAAGGCCGAGAATTGATGTCGTTGTCAACTGCTCCGGGGTCTTCAGAGACCTCTTCATCAATCA
GATGAATCTTCTTGATAGGGCAGTGAAGATGGTAGCTGAGCTGGACGAGCCGGAGGAGCAGAACT
ACGTGAGGAAACATGCATTGGAGCAAGCCAAAGAGCTGGGAGTTGAAGTCAGAGAAGCTGCCTC
ACGTATTTTCTCAAACGCCTCCGGCTCATACTCCTCCAACATCAATCTTGCTGTTGAGAACTCATCTT
GGAATGATGAAAAGCAGCTACAGGACATGTACTTGAGCAGAAAGTCGTTTGCATTCGACTCTGATG
CCCCCGGCGTAGGCATGACGGAGAAACGAAAGATCTTTGAGATGGCTCTCAGCACAGCAGATGCC
ACCTTCCAGAATCTTGATTCCTCAGAGATATCTCTCACCGATGTCAGTCACTACTTCGATTCAGACCC
AACTAATCTCGTACAAGGCCTCAGGAAGGACGGGAAGAAGCCGAGTGCATACATTGCAGACACCA
CCACAGCTAATGCACAGGTGCGGACGCTGTCTGAGACAGTGAGGCTCGATGCAAGGACCAAGCTG
TTGAATCCCAAGTGGTACGAGGGCATGCTGTCCAGTGGCTATGAGGGTGTTCGTGAGATTGAGAA
ACGTCTCACAAACACTGTGGGTTGGAGCGCAACTTCAGGACAAGTCGACAACTGGGTGTACGAGG

AGGCCAACACAACATTCATCCAAGACGAGGAGATGCTGAACAAGCTGATGAGCACGAACCCGAAC
TCGTTCAGGAAATTGCTTCAGACTTTCTTGGAGGCTAATGGAAGAGGATACTGGGAGACTAGTGAA
GAAAACATTGAGAGGTTAAAGCAGTTGTACTCTGAAGTTGAAGACAAGATTGAAGGAATCGATCG
TTAG

N. cataria PDS

TTGGCCAAAGAGAATCTTCAATTTTGGAAGATGTCTCAGTTTGGACATGTTTCTGCAGTCAACTTGA
GTAGACAGGGTAACATAACCACTGTCTCGACTCATGGGTGCTCCATTAGCAATGGCCTATCATTTGC
AGTTAGTGATGCCATGGGTCACAAATTGAGGATTCCAACTGCACGTGTTCTTTCAACAAGATCTAAT
AAGAATGTGCTTCCTTTAAAGGTTGTTTGCGTTGATTATCCGAGACCCGAGCTTGACAATACAGTTA
ACTATTTGGAAGCTGCTTACTTATCATCAACATTTCGCACTTCTCCCCGTCCTAACAAGCCGTTGGAA
ATTGTGATTGCCGGTTCAGGTCTGGCTGGTTTGTCCACTGCGAAGTATTTAGCAGATGCAGGTCAC
AAACCAATATTGTTGGAAGCGAGAGATGTTCTCGGTGGAAAGGTGGCTGCATGGAAAGACGGGG
ATGGAGATTGGTATGAGACTGGCTTGCACATATTCTTTGGGGCTTACCCAAATATACAGAACCTGTT
TGGAGAGCTGGGAATAAATGATCGGTTGCAATGGAAAGAGCATTCTATGATATTTGCCATGCCAAG
CAAGCCAGGAGAATTCAGCAGATTCGACTTCCCTGAAATCTTACCTGCACCGTTCAACGGTATTTTG
GCAATATTGAAGAACAACGAAATGCTTACTTGGCCAGAGAAAGTCAAGTTTGCCATCGGGCTTTTA
CCAGCCATAGTTGGGGGTCAATCTTATGTCGAGGCTCAAGATGGGATAACTGTTAAAGACTGGATG
AAAAAGCAAGGTGTACCAGAGAGGATAACCGACGAAGTGTTCATTGCCATGTCAAAGGCACTGAA
TTTCATAAACCCGGATGAGCTTTCGATGCAGTGTATTTTGATTGCATTGAACAGATTTCTACAGGAG
AAGCATGGCTCAAAGATGGCTTTTCTGGATGGTAATCCACCAGAAAGACTTTGCATGCCTATTGTCG
AACATATTCAATCACGTGGAGGCGAAGTCCGTCTGAACTCACGAATACAAAAGATTGACTTAAATG
AAGATGGAAGTGTCAAAAGCTTCATACTAAATAATGGAAATGTGATTAAAGGAGATGCATATGTTT
TTGCAACTCCAGTTGATATCCTAAAGCTTCTTCTGCCCGAAGAATGGAAGGACATTCCATATTTCAA
GAAGTTGGATAAATTAGTCGGAGTTCCAGTTATAAACGTTCACGTATGGTTTGACAGAAAACTGAA
GAATACGTACGATCATCTACTTTTCAGCAGGAGTCCTCTCCTCAGTGTCTACGCCGACATGTCCGTG
ACTTGTAAGGAATATTACGACCCTAATAAATCCATGTTGGAGCTGGTTTTTGCTCCCGCAGAGGAAT
GGATCTCGCGTAGTGACGAGGAAATCATTGACGCCACAATGAAGGAACTCGCAAAACTCTTCCCGG
ATGAGATCTCTGCTGATCAGAGCAAAGCAAAAATCTTGAAGTATCGTGTTGTCAAAACTCCAAGGT
CTGTATATAAAACTGTACCTGGCACGGAGCCTGCTCGTCCCTTGCAAAAATCTCCCATCGAGGGATT
CTATTTAGCCGGCGACTACACGAAGCAGAAATACTTGGCTTCGATGGAAGGCGCTGTGCTATCAGG
AAAGCTTTGTGCTCAGTCAATTGTAAAGGATGTTGAATTGCTGTCTGAGAGAGGGCAGGAAAATGT
GACAGAGGCAAGCCTTGTTGTCTAG

N. cataria UBI9

CTGATCTTCGCTGGGAAGCAGCTTGAAGACGGCAGAACGTTGGCGGATTACAACATACAGAAGGA
ATCCACGCTGCACTTGGTTTTGAGGCTTCGCGGCGGGATGCAGATTTTTGTGAAAACTTTGACTGG
AAAGACGATAACGCTGGAAGTCGAGAGCTCGGATACCATCGACAATGTGAAGGCTAAGATACAGG
ACAAGGAAGGCATCCCACCCGATCAGCAGAGGCTGATTTTCGCCGGCAAGCAGCTGGAAGATGGG
AGGACTTTGGCGGACTACAATATTCAAAAGGAGTCGACGTTGCATTTGGTGCTGAGGCTGCGCGG
CGGGATGCAGATCTTTGTCAAGACTTTGACGGGAAAGACCATAACTTTGGAGGTAGAGAGTTCCG
ACACTATTGATAATGTGAAGGCGAAGATCCAGGACAAAGAGGGCATTCCCCCGGATCAGCAGAGG
CTTATCTTTGCTGGGAAGCAGCTCGAAGATGGGCGGACCCTTGCTGATTATAATATCCAAAAGGAA
TCGACTCTGCACCTCGTCCTCCGTCTCCGTGGTGGCGATGATCTCTGA

N. cataria ACT1

ATGGCCGAAGCCGAGGATATTCAGCCCCTTGTCTGTGACAATGGAACTGGAATGGTCAAGGCTGG
ATTTGCTGGAGATGATGCTCCAAGGGCTGTCTTCCCTAGCATTGTTGGTCGCCCTCGTCACACTGGT

GTGATGGTTGGCATGGGCCAAAAGGATGCCTATGTTGGTGACGAGGCTCAATCCAAGAGAGGCAT
TTTGACACTGAAATATCCCATCGAGCATGGAATTGTCAGCAACTGGGATGATATGGAGAAAATATG
GCATCACACATTCTACAACGAGCTTCGTGTGGCCCCAGAAGAGCATCCAATTCTCTTGACAGAGGCT
CCTCTTAATCCTAAGGCCAATCGTGAAAAAATGACTCAGATAATGTTTGAGACATTTAACGCCCCTG
CTATGTATGTTGCCATTCAGGCAGTTCTGTCCCTTTACGCCAGTGGTCGTACAACTGGTATTGTTCTC
GACTCTGGTGATGGTGTCAGCCATACAGTCCCGATCTACGAAGGTTATGCACTCCCTCATGCAATCC
TCCGCCTTGATCTGGCTGGGCGTGATCTGACCGACCACCTCATGAAGATATTGACAGAAAGAGGTT
ACTCATTCACCACCACTGCTGAGCGGGAAATTGTAAGGGACATCAAAGAGAAGCTGGCCTACATTG
CTTTGGACTACGAGCAAGAGATAGAGACGGCAAAGACCAGCTCCGCTGTGGAGAAGAACTACGAG
CTTCCCGATGGACAGGTCATCACTATTGGTGCTGAGAGATTCCGCTGCCCTGAGGTCCTCTTCCAGC
CATCTATGATCGGAATGGAAGCTGCTGGAATTCACGAGACAACTTACAATTCTATCATGAAGTGTG
ATGTGGATATCAGGAAGGATCTGTATGGAAACATCGTCCTCAGTGGTGGGTCGACTATGTTCCCAG
GTATTGCTGACAGGATGAGCAAGGAGATCACTGCTCTAGCACCAAGCAGCATGAAGATCAAGGTG
GTGGCCCCGCCTGAGAGGAAGTACAGTGTCTGGATCGGAGGTTCCATCTTGGCATCTCTCAGCACC
TTCCAGCAGATGTGGATTGCTAAGGCAGAGTACGATGAATCTGGGCCGTCCATCGTCCACAGAAAG
TGCTTCTAA

N. cataria ISY

ATGAGCATGAGCATGAGCATGAGCTGGTGGAGGGCTGGAGCTGCAAAGAAAAGAATGGATGATG
AGGAGTGGTCTTCAGTTGTAAAGCAGCAGCAGCAGCAGCAGCAGCAATGCGTAGCTCTGATC
ATTGGGGTGACCGGACTGGTGGGCAACAGCCTAGCGGAGATGCTGCCGCAGGCGGACACCCCGG
GGGGGCCATGGAAGGTATACGGGGTGGCGCGGCGGGCCCGTCCGTCGTGGAACGAGGACCACCC
CATCACGTACATCTCATGCGACGTGAGCAACACCGCGGAGGTGGAGGCGAAGCTGTCCCCTCTGA
GCGACGTGACGCACATCTTCTACGCCACGTGGACGAGCCGATCGAGCGAGGAGGAGAACTGCGA
GGCCAACGGGAAGATGCTGAAGAATGTGGTGGACGCAATGATCCCCAACTGCCCCAACCTGAAGC
ACATCTGCTTGCAGACCGGGCGGTTCCACTACGTTGCCTCAGTTGTGGACTGGAAGATCAACGGCA
GCCACGACACCCCGTTGACGGAGGATTTACCTCGGCTGAAGACGAAAAACTTCTACTACACGCAGG
AGGACATTCTGTTGGAGGCCGTTAAGAGGAAGGAGGGGCTGACATGGTCCGTGCATCGCCCGGG
GACTATATTCGGCTTCTCGCCCTATAGCATGATGAATCTGGTGGGGACGCTGTGCGTGTATGCAGC
TATCTGTAAGCAGGAGGGTGCGGTTTTGAGGTTTCCTGGTTGTAAGGGTGCATGGGATGGGTACT
CAGACTGTGCGGATGCCGACTTGATTGCAGAGCATCAAATATGGGCCGCCCTCCATCCTCATGCCA
AGAACCAAGCATTCAACGTCAGCAATGGGGACGTTTTCAAATGGAAACATCTCTGGAAGGTGTTGG
CCGACCAGTTCGGCGTCGAATGCCCCGGGGGATACGAGGAAGGCCAGCAAGTGAGGCTGCAGGA
TGTGATGAAGGGCAAAGGTCCGGTGTGGGACAAAATCGTTAGGGAGAATGGGCTGTCCAATACCA
AATTGGAGGATGTGGGGAAATGGTGGTTTAGTGATACTATCTTGTGGAATGAGTGTAGGTTGGAT
AGTATGAATAAGAGCAAGGAGCATGGCTTTCTTGGCTTTAGGAATTCCAACACTTGCTTTCTTTATT
GGATTCATAAGGTCAAGGCTTACAACATTGTTCCTTCTACCTAG

N. cataria NEPS1

ATGGCAAGCATTGTAAATCCGGTGCAGGTGATGAAGAAGAAGCTCGAAGGCAAAGTTGTGATAGT
AACAGGAGGGGCGAGCGGCATCGGGCAGACGGCAGCGCATGTGTTTGCACAACATGGCGCGCGT
GCAGTGGTGATCGCTGACATCCAATCTGAAGTTGGGAAGTCCGTGGCGGAGTCCATCGGGAAGCA
ATGCAGCTACGTCCAGTGTGACGTCTCGCACGAGGAGGAGGTTAAGTCGATGATAGAATGGACGG
CCAGCACGTACGGCGGGCTAGACGTGATGTTCTCCAATGTGGGAATCATGAGCAACTCTGCTCAAA
CAGTAATGGACCTCGACCTTTTGGAATTCGATAAAGTGATGCGTGTGAACGCGCGCGGGACGGCC
GCGTGCTTGAAGCACGCAGCGCGTAAGATGGTGGAGCTAGGAACAAGAGGCACTATTATCTGCAC
GACCACCCCACTCTCATCGAAGGGCGGATCAAGCATGACGGACTATGCGATGTCGAAGCACGCAG
TGTTGGGGCTGGTCCGGTCGGCCAGCATACAGCTGGGGGCCCACGGGATTAGAGTTAACTGCGTG
ACGCCGTCTGTGGTGCTTACGCCGCTCGCCCAAAGGATGGGGCTTGCCACACCCGATGATTTCCAT

ACTCATTTTGGCAACTTCACCAGTCTCAAAGGGGTCTACCTCACCGCCGACCAAGTCGCCGAAGCTG
TCACCTTTCTCGCTTCCGATGATGCCGCCTTCATCACTGGACATGATTTGATCCTCGATGGTGGACTG
CTTTGTTTACCATTCTTTGCTCCTTCATGA

N. cataria NEPS3

ATGGCTAACAATTCAGTGATTATGAAGAAGAAGCTGGAAGGCAAAGTAGCCATTATAACCGGCGG
CGCCAGCGGTATCGGCGAGGCCACGGCCCGCATGTTCGTGAGGTACGGGGCGCGTGCGGTGGTG
ATCGCTGATATTCAGTCGGAATTAGGACGGTCTGTGGCGGAATCGATCGGGAGGGAGCGGTGCAG
CTTCGTGCAGTGCGACGTCACCGACGAGGAGCAGGTGAAGTCCATGATAGAGTGGACGGCCACCA
CGTACGGCGGCCTGGACGTAATGTTCAGCAACGCCGGCGTCTTGAACAGCGCCGCGCAAACCGTG
ATGGACTTGGACCTGCCGCTTTTCGATAAGGTCATGCGTGTGAACACGCGCGGCTCGGCTGTGTGC
GTGAAGCAGGCGGCGCGTAAGATGGTGGAGCTGGGAAGGGGCGGCTCCATCATATGCAACGCCG
GCTCGTCGGCGGTGAGGGGCGCACATAACGTGACGGACTACGTGATGTCGAAGCACGCGGTGATA
GGGCTGGTGCGGTCGGCCAGCATGCAGCTTGGGGCCCACGGGATTAGGGTTAACAGCGTGTCGCC
GATGGCCGTGGTGACGCCGCTCACCCGGAACGTAGGGATTTCGTCGCCGGCTGATGTACAGAATG
TTTTGATGCCGTTCATTAGCTTGAAAGGGGTGCCGCCCACGGCGGAGCAGGTGGCGGAAGGGGCG
GCGTTCCTGGCTTCCGACGAGGCGGCGTTCGTTACGGGGATTGATCTGCCGGTGGATGGCGGCGT
GCTCTGTATGCCATTTCTCCTCGCTTCAGCTTCAGCTTCAGCTTCAGCATAA

N. cataria NEPS4

ATGGCAAGCATGGTAGATGCGGTGCAGGTGATTAAGAAGAAGCTAGAAGGCAAAGTTATGGTAG
TAACAGGCGGGGCGAAAGGCATCGGGCAGACGGCAGCGCGCGTGTTTGTGGAGCATGGCGCGCG
TGCGGTGGTGATAGCTGACATCCAATCTATAGTTGTTGGGAAGTCCGTGGCGGAGTCCATCGGGG
AGCGGTGCAGCTACGTCCAGTGCGATGTCTCTGAAGAGGAGCAGGTTAAGTCTATGATAGAATGG
ACGGCCAACACATATGGCGGGCTGGACGTGATGTTCTGCAATGCGGGCATCATGAGCTACTCCGCT
CAAACCGTAATGGACCTGGACTTCTCGCAATTTGATAAGGTGATGCGTGTGAACGCACACGGGACG
GCCGTGTGCGTGAAGCAGGCAGCGCGTAAGATGGTGGAGCTAGGAACGAAAGGCACTATTATCTG
CACGACTAGCGCGACAGCATCAAAGGGCGGACAAAACATGACGGACTATGCGATGTCAAAGCACG
CGGTGATGGGGCTGGTGCGGTCGGCAAGCATTCAGCTGGGGGCCCATGGGATTAGGGTTAACTGT
GTGTCGCCCTTGGTCGTGCTCACGCCGCTTGCCCAAAGGATGGGGCTTGCCACGCCCGATGATTTC
TATACTCATTTTGGCAACTTCACTAGCCTCAAAGGAGTCTACCTCACCGCCGACCAAGTCGCCGAAG
CCATTACCTTTCTCGCTTCCGACGATGCCGCTTTCATCACCGGACATAATTTGGACCTCGATGGTGG
ACTGCTTTGTTTACCATTCGTTGCACCTTCAACAACATGA

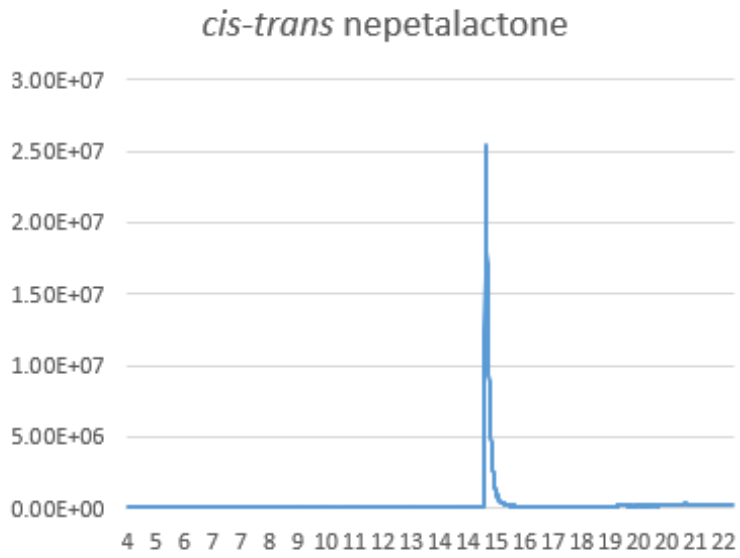
N. cataria NEPS5

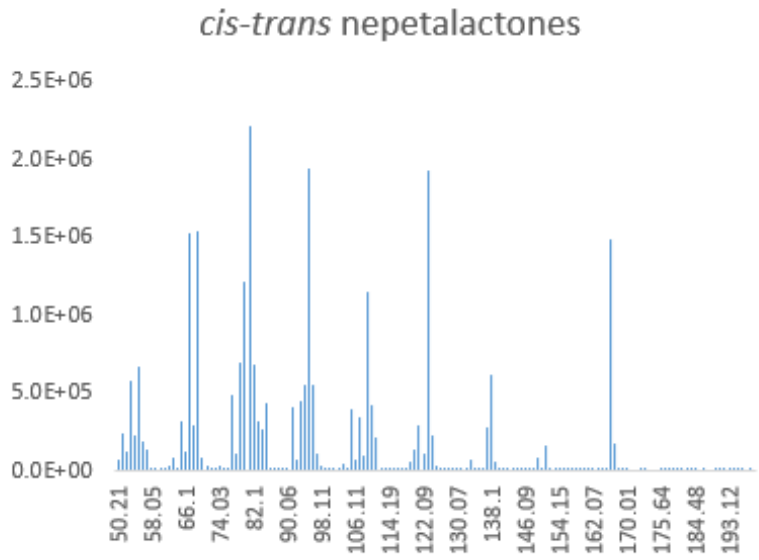
ATGGCAAACATTAGAAATTCGGAGCAGGTGATGAAGAAGAAGCTGGAAGGCAAAGTAGCAATAA
TAACAGGCGGGGCGAGCGGCATCGGGAAGACGGCAGCGCGCGTGTTTGCGGAACATGGCGCGCG
TGCGGTGGTGATAGTTGATATCCAATCTGGAAATGGGCAGTCCGTGGCGAAGTCCATCGGGGAGG
TGTGCAGCTACGTCCAGTGCGACGTCTCGGACGAGGAACAGGTTAAGTCGATGATTGAATGGACA
GCCCGCACATACGGCGGGTTGGACGTGATGTTCTGCAATGTGGGCATCCTAAGCCACACCCCTCAA
ACCGTGATGGACCTCGACTTCTCGCAGTACGATAAGTTGATGCGTGTGAACGCGCGCGGTACGGCC
GCGTCCGTGAAGCATGCGGCGCGTAAGATGGTGGAGCTGGGAACGAGAGGCACTATTATCTGCAC
GACCAGCCCGGCTTCGACTATGGGCGGACACAACCTGACGGACTATGTGATGTCCAAGCACGCGG
TATTAGGGCTTGTCCGGTCGGCCAGCATGCAGCTCGGGGGTCACGGGATTAGGGTTAACTGTGTG
TCGCCCTCAGTAGTGATCACGCCGCTCGCCCAAAGGATGGGGCTTGCCACGCCCGATGATTTCTAC
ACTCAGTTTGGCAACTTCACTAGCATCAAAGGGGTCCACCTCACGCTCGAGAATGTCGCCGAAGCC
GTCGTCTATCTCGCTTCCGATGATGCCGCTTTCATCTCCGGACATGATTTGAAGCTCGATGGTGGAC
TGCTTTGTTTACCATTCGTTATGCCCGCTTCAACATCATAA

N. cataria RLK

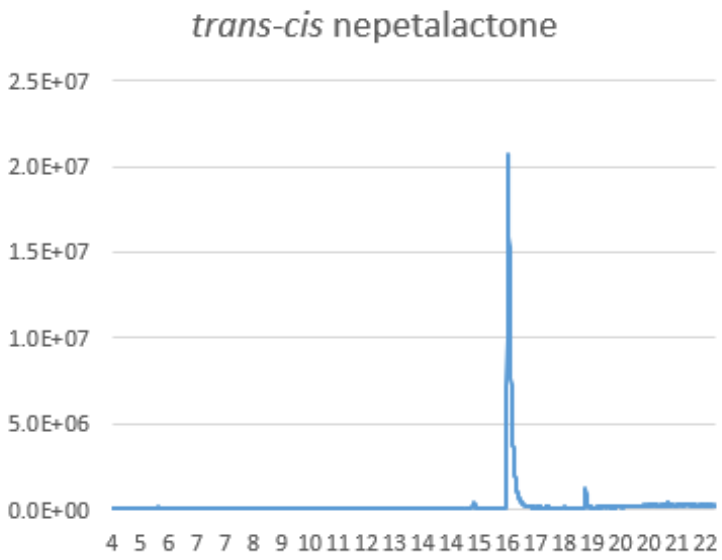
ATGAAAATGGGCGTTTCAATTTTGGTGTCCCTAGCAGCCTTGGCAACATCCCTTTTGCTGGATTCCT
CTCTCGCTCTCACTCCAGATGGTCTTGCACTTTTGGAAGTGAAAACCAGTTTGAATGACAGTAAGAA
CTTTCTGAGTAACTGGAAAGATATTGATGAATATCCCTGTAACTGGACTGGTATTTTGTGCAGCCCA
CAAGACCAAAGAGTGATATCTATAAACCTACCATATATGCAGCTTGGAGGAATAATTTCCCCCAGC
ATTCGCAAGCTTGATAAATTACAAAGACTGGCACTTCATCAGAATAGCCTTCATGGTTTCATTCCTA
ATGAGATTGCACAGTGTTCTGAACTCAGAGCTCTGTATCTTCGAGCTAATTATTTTCAAGGAGGAGT
ACCTTCAACAATTGGAAACCTTTCTATGTTGACTATATTGGATTTATCAAGCAATACACTAAAAGGT
GCCATACCTTCTTCTATAGGCCGTTTAACACGTCTATCATATCTAAATTTGTCGTCAAACTTCTTGTCG
GGTGAAGTCCCAGATGTTGGAGTTCTAAGCAAGTTTGGAAGCAAGTCGTTTATTGGTAATTTAGAT
CTGTGTGGCCAACAAGTAAACAAGCCTTGTCGAACCTCTCTTGGTTTTCCTGCTGTATTGCCTCATGC
CGAAAGTGATGAAGCAACAGTGCACAAGCGATCATCTCGCTACATCAAAGGTGTTGTGATTGGTGC
AGTATCCATGTTAGGCTTTGGGATTATCTTCCTTCTTGGCTTTCTCTGGATTTGGATGCTGACGAAGA
AGGAGAGGGCTGCGAAAAAATACATAGAAGTGAAAAAACAAGTTCATCAGGATGCAAGAGCCAA
GTTAATTACCTTCCATGGTGATCTTCCATACCCTTCGTCCGAGCTTGTAGAGAAAATCGAGTCTCTTG
ATGAGGAAGACATTGTTGGAGCTGGAGGCTTTGGAACTGTATATCGAATGGTTATGAATGATTGTG
GAACTTTTGCTGTTAAAAAGATTGACCGGACGAGAGAAGGTTCTGATCAAGTATTCGAGAGAGAG
TTAGAGATCTTAGGCAGTGTCAAGCACATAAATCTAGTAAACCTACGAGGTTACTGTCGGCTTCCTA
CAGCGAAGCTTCTTATATATGACTACTTAGCTAAGGGCAGTTTGGACTATTTCTTGCACGATAATAT
GTCAGAAGAACGTTTGCTGAATTGGAATGCTCGATTAAGAATAGCTATTGGTTCTGCTAGAGGAGT
AGCATATCTGCACCATGACTGCTCTCCGAAGATAGTTCATCGTGATATAAAGTCGAGTAATATTCTC
CTTGATGAAAACCTCGAGCCTCATGTTTCAGATTTCGGCCTTGCTAAGCTACTCGTGGATGAGGATG
CTCATGTCACAACTGTGGTTGCTGGCACATTCGGCTATCTTGCTCCAGAGTACTTGCAGAGCGGTAG
GGCCACTGAGAAGACAGACGTGTACAGCTTTGGCGTGCTGCTGCTAGAACTGGTAACCGGTAAAA
GACCAACTGATGCTACATTTGTGCAGAGAGGACTAAATGTGGTTGGATGGGTGAACACACAGAGT
CGAGTTGAGGACATAGTGGACAAGAGGTGCAGAGATGCAGAGGTGGAGACTGTGGAGGTGCTCG
TGGAGATAGCAGCGAGATGCACGGATGCAAATCCGGAGGAACGTCCCTCGATGCAGCAGGTTCTG
CAGTTCTTGGAGCAAGAGGTCATGTCACCCTGTGTTAGTGAGTTGTATGAGTCACATTCGGATAAA
TGTTGA

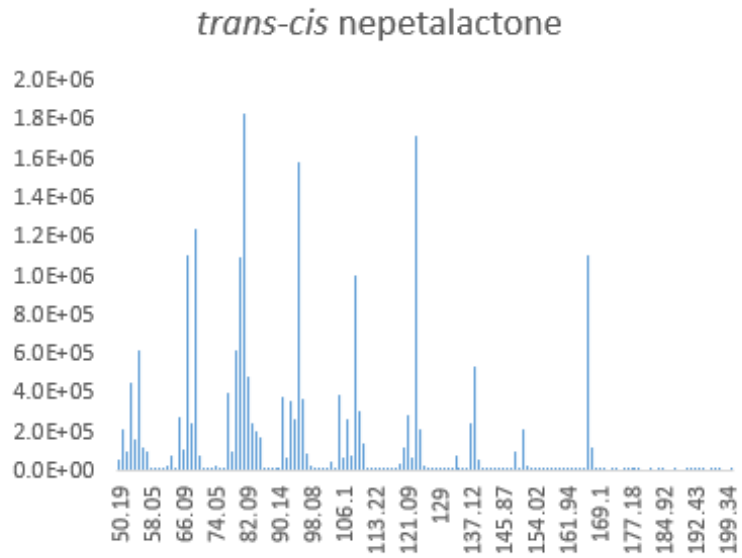
Appendix 2: GC-MS chromatograms and fragmentation patterns of nepetalactone standards (GC-MS method 1)



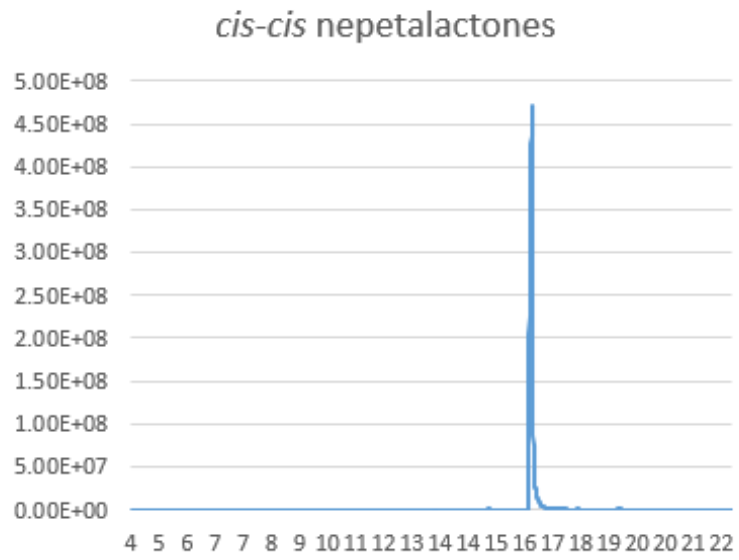


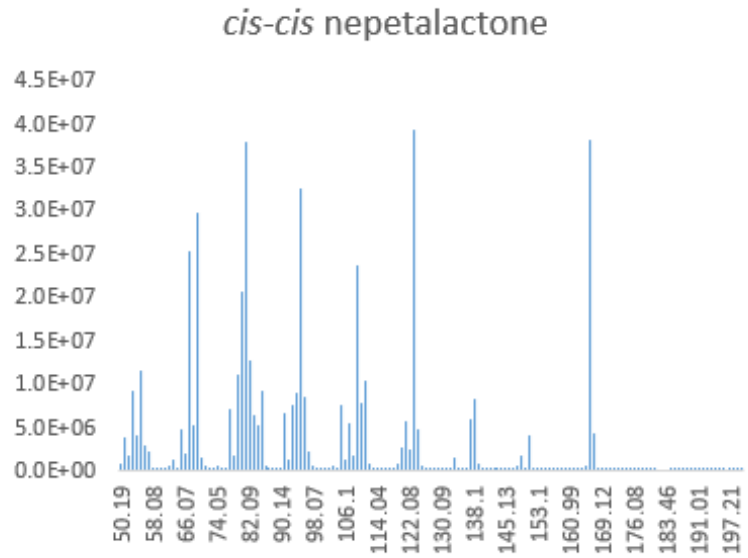
500µM of *cis-trans* Nepetalactone. Standard purified from *Nepeta* tissue as reported in Sherden *et al.* (2018).





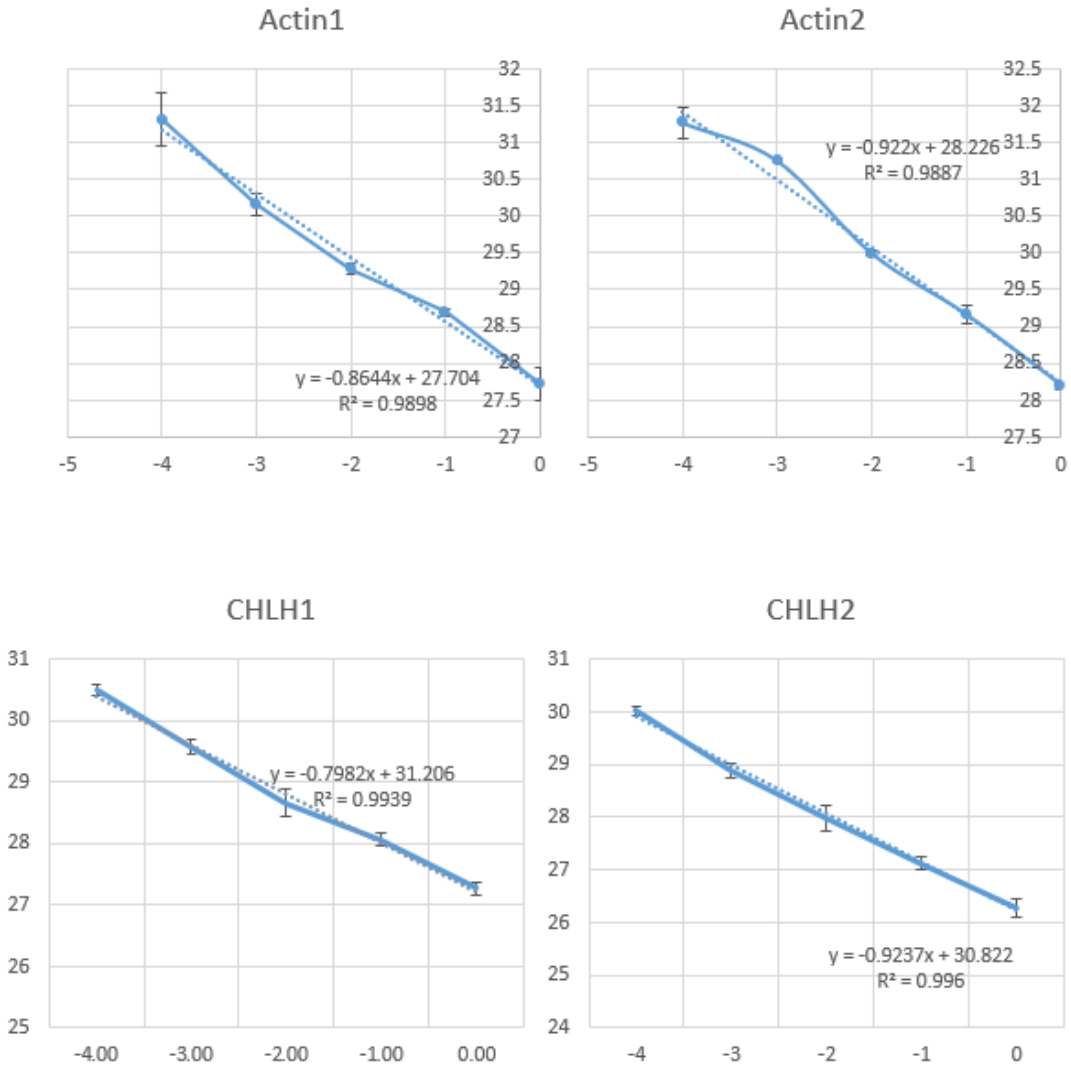
500µM of *trans-cis* Nepetalactone. Standard purified from *Nepeta* tissue as reported in Sherden *et al.* (2018).



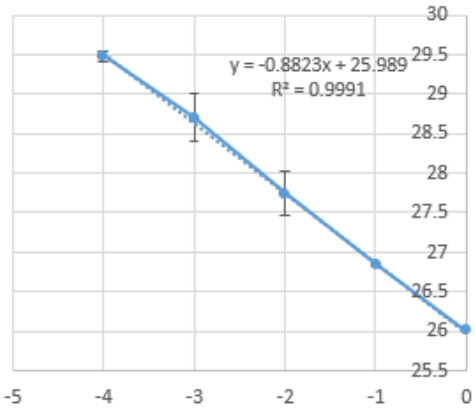


500µM of *cis-cis* Nepetalactone. Standard purified from *Nepeta* tissue as reported in Sherden *et al.* (2018).

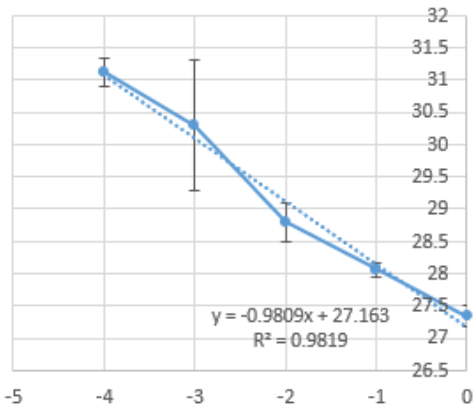
Appendix 3: qPCR primer efficiency curves



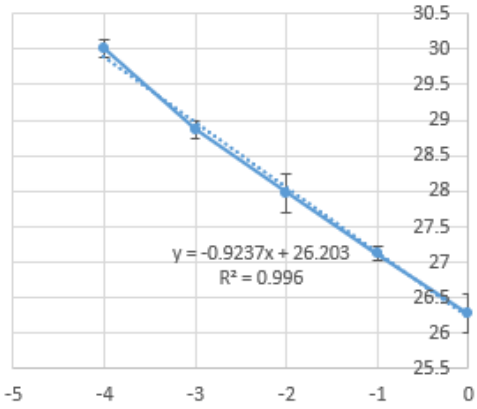
UBI1



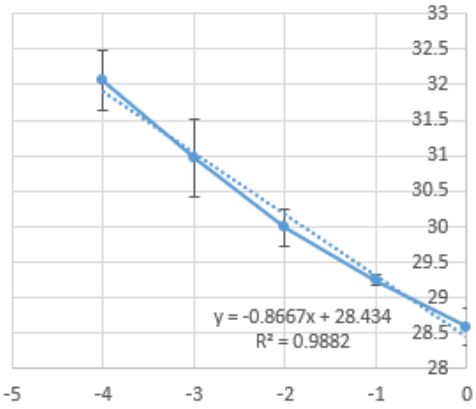
UBI2

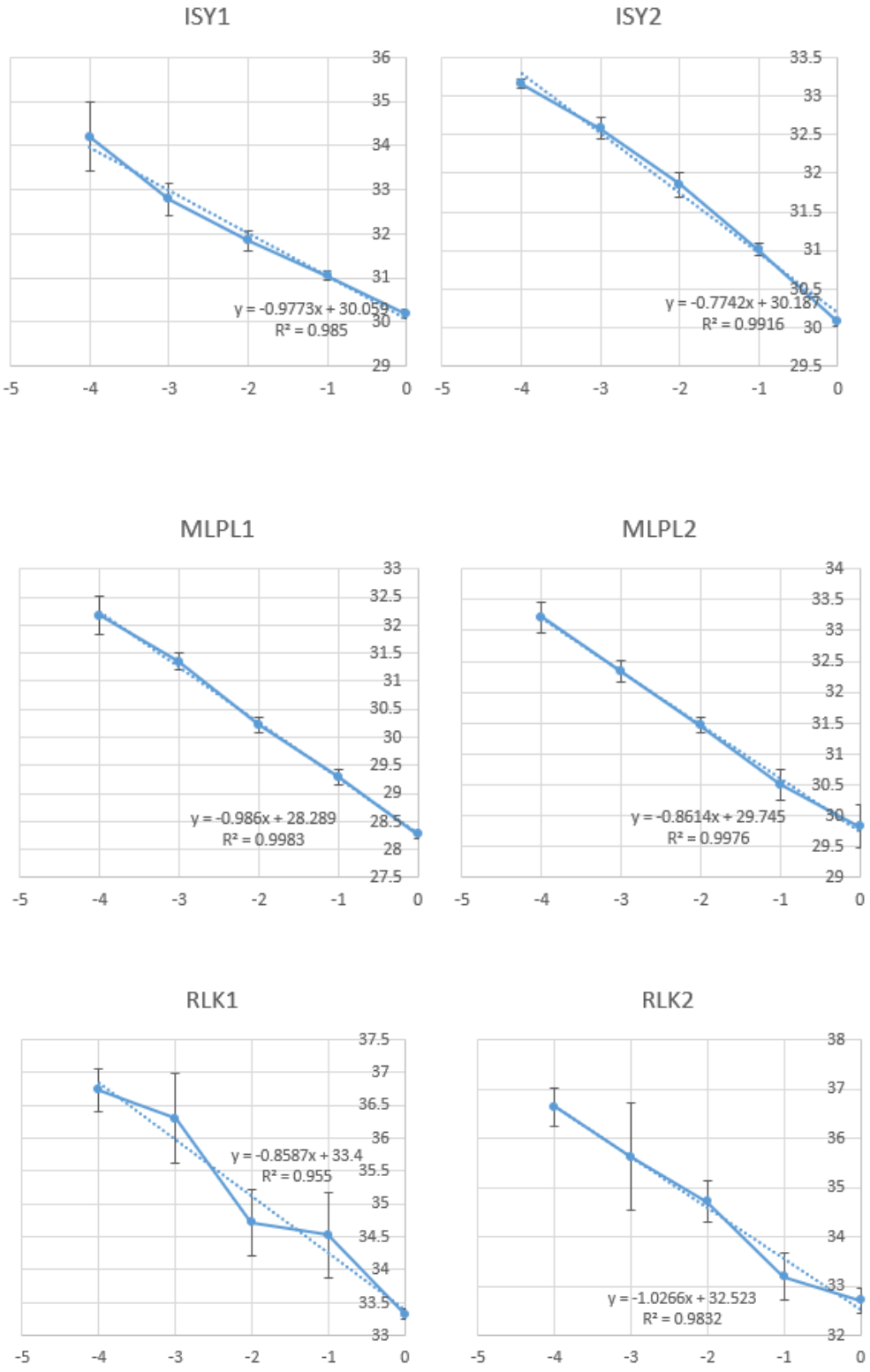


GES1



GES2





Primer pairs with R closest to 1 were selected for qPCR analysis. *N. cataria* WT cDNA was diluted 5x for standard curve formation.