Testing of candidate genes for compound biosynthesis in Arnica montana

Mrakovčić, Nikolina

University of Zagreb
Faculty of Science
Department of Biology

Nikolina Mrakovčić

Testing of candidate genes for compound biosynthesis in *Arnica montana* L.

Master thesis

Zagreb, 2022.

Sveučilište u Zagrebu
Prirodoslovno-matematički fakultet
Biološki odsjek

Nikolina Mrakovčić

Testiranje gena kandidata uključenih u biosintezu metabolita biljke *Arnica montana* L.

Diplomski rad

Zagreb, 2022.

This master thesis was completed in Professors Dirk Prüfer's group at the Institute of Plant Biology and Biotechnology of University of Münster, Germany, under supervision of Associate Professor Gundula Noll and co-supervision of Associate Professor Nataša Bauer. The thesis was submitted for assessment to the Department of Biology at the Faculty of Science, University of Zagreb to obtain master's degree in molecular biology.

Firstly, I would like to thank Prof. Dirk Prüfer and my mentor Prof. Gundula Noll for the opportunity to do a master thesis research in the group. I cannot say enough thanks to Dr. Lena Grundmann for everything she taught me, for the immense support and record-fast replies to my e-mails. Big thanks go to Michael who taught me a lot with a smile and lots of patience. I greatly appreciate all the shared knowledge, fun moments and good everyday atmosphere shared with members of AG Prüfer. This was a priceless experience, through which I have learned a lot and become more independent. Thank you!

I am thankful to my co-mentor Prof. Nataša Bauer for all the advice and help with my thesis and the knowledge passed on through courses.

To my friends, who have always been there, thank you.

Najveće hvala ide mojoj obitelji bez koje ništa od ovoga ne bi bilo moguće. Oni su mi sve omogućili, vjerovali u mene i neumorno pružali ljubav i potporu.

TEMELJNA DOKUMENTACIJSKA KARTICA

Sveučilište u Zagrebu
Prirodoslovno-matematički fakultet
Biološki odsjek Diplomski rad

Testiranje gena kandidata uključenih u biosintezu metabolita biljke *Arnica montana* L.

Nikolina Mrakovčić

Rooseveltov trg 6, 10000 Zagreb, Hrvatska

Arnica montana L. je biljka koja se stoljećima koristi u tradicionalnoj medicini. Za njezina ljekovita svojstva odgovorni su seskviterpenski laktoni, helenalin i dihidrohelenalin. Arnika se tradicionalno ubire s prirodnog staništa i postoji potreba za alternativnim i održivijim načinima sinteze bioaktivnih sastojaka. Kako bi alternativna proizvodnja bioaktivnih seskviterpenskih laktona bila moguća, potrebno je razjasniti put biosinteze. Provedena je funkcionalna karakterizacija gena kandidata iz *A. montana*, koji su potencijalno uključeni u biosintetski put seskviterpenskih laktona, u heterolognom ekspresijskom sustavu, kvascu *Saccharomyces cerevisiae*. Metaboliti sintetizirani u kvascu analizirani su plinskom kromatografijom i masenom spektrometrijom. Također, analizirani su ekspresijski obrasci gena kandidata te distribucija i sadržaj seskviterpenskih laktona u različitim tkivima biljke *A. montana*. Potvrđeno je da geni kandidati *AmGAS* i *AmGAO* kodiraju sintazu germakrina A, odnosno oksidazu germakrina A te je potvrđena njihova enzimska aktivnost. Geni *AmCYP* su djelomično funkcionalno okarakterizirani, a gen *AmCYP7* pokazao je najveći potencijal za daljnje istraživanje. Korelacija ekspresijskih obrazaca gena kandidata i sadržaja bioaktivnih sastojaka u tkivima biljke *A. montana* pokazala je da je gen *AmCYP7* vjerojatno uključen u biosintezu seskviterpenskih laktona i pružila preliminarne dokaze za biosintezu seskviterpenskih laktona u različitim razvojnim stadijima biljke *A. montana*.

(61 stranica, 13 slika, 13 tablica, 43 literaturnih navoda, jezik izvornika: engleski)
Rad je pohranjen u Središnjoj biološkoj knjižnici
Ključne riječi: helenalin, seskviterpenski laktoni, biosintetski put, heterologna ekspresija, *Saccharomyces cereviasiae*, germakrin A, germakrinska A kiselina
Voditeljica: Izv. prof. dr. sc. Gundula Noll
Suvoditeljica: Izv. prof. dr. sc. Nataša Bauer

Ocjenitelji: Izv. prof. dr. sc. Nataša Bauer
Prof. dr. sc. Mirta Tkalec
Prof. dr. sc. Biljana Balen
Doc. dr. sc. Marin Ježić (zamjena)
Rad prihvaćen: 10. veljače 2022.

BASIC DOCUMENTATION CARD

University of Zagreb
Faculty of Science
Department of Biology

Master Thesis

Testing of candidate genes for compound biosynthesis in *Arnica montana* L.

Nikolina Mrakovčić

Rooseveltov trg 6, 10000 Zagreb, Hrvatska

Arnica montana L. is a traditional medicinal plant used for centuries. It owes its medicinal properties to sesquiterpene lactones, helenalin and dihydrohelenalin. Traditionally, the plants are harvested from its natural habitat, but there is need for alternative ways to sustainably obtain bioactive compounds. To achieve alternative production of *A. montana* bioactive sesquiterpene lactones, their biosynthesis pathway must be elucidated. Functional characterization of candidate genes, that might be involved in sesquiterpene lactones biosynthesis, has been conducted in the heterologous expression system yeast *Saccharomyces cerevisiae* and compounds analysis via gas chromatography-mass spectrometry. Additionally, analysis of potential candidate genes expression patterns and sesquiterpene lactones content distribution in plants was carried out. Candidate genes *AmGAS* and *AmGAO* were verified to encode respective germacrene A synthase and germacrene A acid oxidase, and their enzymatic activity was confirmed. Candidate genes *AmCYP*s were functionally analysed to a certain extent, with candidate *AmCYP7* showing great potential for further investigation. Correlation of expression patterns and content of bioactive compounds in *A. montana* tissues revealed that among tested, *AmCYP7* gene has potential to be part of sesquiterpene lactones biosynthesis pathway and provided preliminary evidence for their biosynthesis at different developmental stages during the plant´s life cycle.

(61 pages, 13 figures, 13 tables, 43 references, original in: English)
Thesis is deposited in Central Biological Library.
Keywords: helenalin, sesquiterpene lactones, biosynthesis pathways, heterologous expression, *Saccharomyces cereviasiae,* germacrene A, germacrene A acid
Supervisor: Assoc. Prof. Gundula Noll, PhD
Co-supervisor: Assoc. Prof. Nataša Bauer, PhD

Reviewers: Assoc. Prof. Nataša Bauer, PhD
Prof. Mirta Tkalec, PhD
Prof. Biljana Balen, PhD
Asst. Prof. Marin Ježić, PhD (substitution)
Thesis accepted: 10th February 2022

Table of Contents

Abbreviations

AMA – *Arnica montana* ssp. *atlantica*

AMM – *Arnica montana* ssp. *montana*

CYP – cytochrome P450

DH – dihydrohelenalin

DMAPP – dimethylallyl diphosphate

EtOAc- ethyl acetate

FPP – farnesyl diphosphate

GAA – germacrene A acid

GAO – germacrene A acid oxidase

GAS – germacrene A synthase

GC-MS – gas chromatography–mass spectrometry

H – helenalin

IPP – isopentenyl diphosphate

IPP – isopentenyl diphosphate

LB – lysogeny broth

LC-MS – liquid chromatography–mass spectrometry

LiAc – lithium acetate

MeOH – methanol

MVA – mevalonate

SD – supplement dropout

SLs – sesquiterpene lactones

YPDA – yeast extract peptone dextrose adenine

1. Introduction

Throughout history, humans have relied on medicinal plants to treat various illnesses and conditions. The oldest written evidence for the preparation of medicinal drugs from plants is a Sumerian clay slab from approximately 5000 years ago (Petrovska, 2012). Even today, medicinal plants and their active compounds are of great significance for traditional and modern medicine. In European traditional medicine, one of the plants which is extensively used is *Arnica montana* L. or mountain Arnica. Healing properties of Arnica owe mostly to sesquiterpene lactones (SLs), helenalin and dihydrohelenalin. These bioactive compounds and their mechanism of action are partly elucidated; however, their biosynthesis pathway remains unclear until today. Insight in the biosynthesis pathway offers an opportunity for more efficient and sustainable handling of plants and alternatives for production of *A. montana* bioactive substances.

1.1 Medicinal plants

Vascular plants are estimated to be 300.000 to around half a million species, of which ten percent are used for medicinal purposes. Some of the plant species are cultivated, while rest of them are harvested from nature, to acquire substances of medical and pharmaceutical importance (Salmerón-Manzano et al., 2020). According to worldwide research trends on medicinal plants review by Salmerón-Manzano et al. (2020), family Asteraceae is being one of the most investigated, with highlight on genus Vernonia with research focus on drug efficacy. Medicinal plants are of great importance for cultures that mostly rely on traditional medicine, but cultures that rely on modern medicine, also use herbal medicines due to their effectiveness and little side-effects. A great role of phytopharmaceuticals is prevention of diseases, they are often taken as dietary supplements, which makes interest in, and utilization of medicinal plants grow. In line with the high demand for medicinal plants, of which in Europe 90% are harvested from wild populations, there is need for conservation strategies. Conservation strategies include *in situ* conservation, for example in natural reserves, *ex situ* conservation, e.g., botanical gardens and seed banks, and enhanced cultivation practices (Kumar Srivastava, 2018). Cultivation of medicinal plants ensures more stable supply of plant material than collection in wild, attributes to decreases in the harvest volume resulting in the recovery of wild populations, and decreases their prices (Kumar Srivastava, 2018). Cultivation alternatives like secondary

metabolites production in tissue culture, protoplasts or microbial fermentation, in addition to benefits of cultivation, enable controllable conditions, higher yields, and steady quality of plant-derived substances (Kulagina et al., 2021). However, those strategies require more research and knowledge on the plant propagation possibilities, and metabolism of compounds of interest including their biosynthesis genes and enzymes.

Aside from their roles in plants defence and communication systems, as well as use of crude plant extracts and active compounds in medical treatments, plant secondary metabolites have more to offer. They are often used as templates for the development of novel medicaments or as source of material for semi-synthesis of medical substances (Douglas Kinghorn & Seo, 1996). Considering that, most of research on medicinal plants focus on their pharmaceutical benefits and utilization, while agronomic, genetic or biotechnological research on medicinal plants is scarce (Salmerón-Manzano et al., 2020). The great potential of phytopharmaceuticals still needs to be given attention, and biotechnology could be used to generate novel or enhanced bioactive compounds. Hence, the importance of improved production of secondary metabolites and research on biosynthesis pathways of bioactive substances is stressed as important for future of medicinal plants research in recent review by Shafi and Zahoor (2021).

1.2 *Arnica montana*

Arnica montana is a perennial plant belonging to the Asteraceae family, spread in mountain areas of Europe and the Northern hemisphere (Kriplani et al., 2017). Owing to its medicinal properties, the plant has been used in traditional medicine for centuries, and as consequence of exploration of its wild populations and habitat loss, it is included in the European Red List of Vascular Plants (Petrova et al., 2021). Alcoholic extracts of *A. montana* flowerheads, in form of tinctures and ointments, show antibacterial, anti-tumour, antioxidant, anti-inflammatory, antifungal and immunomodulatory properties (Kriplani et al., 2017; Matos et al., 2021; Wiesner & Kairies, 2014). Two subspecies of *A. montana* have been described, namely *Arnica montana* ssp. *montana* (AMM) and *Arnica montana* ssp. *atlantica* (AMA), and a commercially available cultivar "Arbo" has been established in the 1990s (Petrova et al., 2012; Schmiderer et al., 2018). Arbo variety is considered to represent central European AMM populations, while the AMA plants are popularly called "Spanish" due to their presence concentrated on Iberian Peninsula (Schmiderer et al., 2018). *Arnica montana* growth,

population and secondary metabolites production greatly varies with changing environmental conditions such as altitude, drought, temperature, nutrient levels and soil acidity (Hollmann et al., 2020; Petrova et al., 2012). *A. montana* wild populations require low pH and nutrient-poor soils for growth and it is reported that the level of SLs content in *A. montana* populations showed significant correlation of lower temperatures and higher rainfall, both connected to higher altitude and higher SLs content (Seemann et al., 2010). Nonetheless, constant growth in usage of *A. montana* for its medicinal properties has led to depletion in its natural population and there is a need for more sustainable solution to foraging wild plants. Potential solutions include field or greenhouse cultivation, *in vitro* cultivation as well as biotechnological approaches like tissue or cell cultures.

1.2.1 Strategies for obtaining *A. montana* material and secondary metabolites

There are some studies that investigated ways to achieve a more sustainable production of *Arnica montana* plants and production of its metabolites, including induction of hairy root growth mediated by *Agrobacterium rhizogenes,* cultivation in field and cultivation *in vitro* as well as optimisation of micropropagation conditions (Surmacz-Magdziak & Sugier, 2012; Leoni et al., 2021; Petrova et al., 2013; Petrova et al., 2021).

Requirement for growth conditions similar to natural ones and no protocol for indoor cultivation available makes common methods to grow and propagate Arnica more difficult (Petrova et al., 2012). Internally communicated results showed that *A. montana* flowering strongly depends on vernalization and only happens on the second spring after sowing, which can be attributed to it being a perennial plant, and is important to note since the bioactive substances are mostly concentrated in the flowerheads. *Arnica montana* can be successfully grown in field using wild seeds with comparable chemical composition to plants grown using Arbo cultivar seeds (Leoni et al., 2021). Cultivation of Arnica in greenhouse conditions requires specific soil conditions, more specifically: well-aeriated nutrient-poor acidic soil with high sand content (Schmiderer et al., 2018; Sugier et al., 2013). Petrova et al. (2013) established a system for *A. montana* hairy root induction mediated by *A. rhizogenes* wild-strain with the maximum efficiency as high as 12% by immersing leaf explants in bacterial suspension for 40 minutes. Optimized hairy root induction protocol provides basis for potential secondary metabolites production as well as platform for elucidating biosynthesis pathway genes in the future. More recently, techniques and conditions required for rapid clonal multiplication and slow-growth

storage of *A. montana* which offer a base for commercial propagation and conservation of the species were elaborated (Maria Petrova et al., 2021).

Other strategies are yet to be applied to *A. montana*, and include over-production of secondary metabolites, which are naturally present in plants in low abundance, in microbial hosts which could be optimized to ensure sufficient amounts of bioactive compounds (Gou et al., 2018). To be able to apply these strategies, knowledge of biosynthesis genes and pathways is required, more about which is reviewed in the chapter 1.2.3 *Biosynthesis pathway of sesquiterpene lactones*.

1.2.2 Arnica as source of sesquiterpene lactones

The numerous bioactive compounds of *A. montana* include sesquiterpene lactones (SLs), phenolic acids, and flavonoids. Many sesquiterpenoids have been in focus in the past years for their use in cosmetic and pharmaceutical industry (Frey, 2020; Liu et al., 2021; Zhou & Pichersky, 2020). Sesquiterpenoids belong to terpenes, the largest plant secondary metabolite group, and are characterized by a 15 carbon atoms backbone structure and SLs have an additional lactone moiety. A conserved α-methylene-γ-lactone group is a characteristic of SLs which allows them to react with biological nucleophilic thiol groups present in enzymes and other functional proteins forming covalent adducts nucleophiles, a reaction called Michael-type addition (Frey et al., 2018). SLs are poorly dissolved in water which makes them less available as therapeutic agents, but their solubility increases by forming amino-adducts through Michael-type reaction while their bioactivity stays unchanged (Valkute et al., 2018). In plants, SLs have a range of protective activities including acting as anti-herbivory and antimicrobial substances, inhibiting growth of competing plants, as well as other physiological roles such as meditating plant communication (Padilla-Gonzalez et al., 2016). SLs in *A. montana* are mainly represented by helenalin and dihydrohelenalin, and respective derivates (Figure 1). AMM/Arbo is a representative of central European chemotype containing predominantly helenalin, AMA/Spanish is a chemotype containing predominately dihydrohelenalin (Matos et al., 2021; Perry et al., 2009)

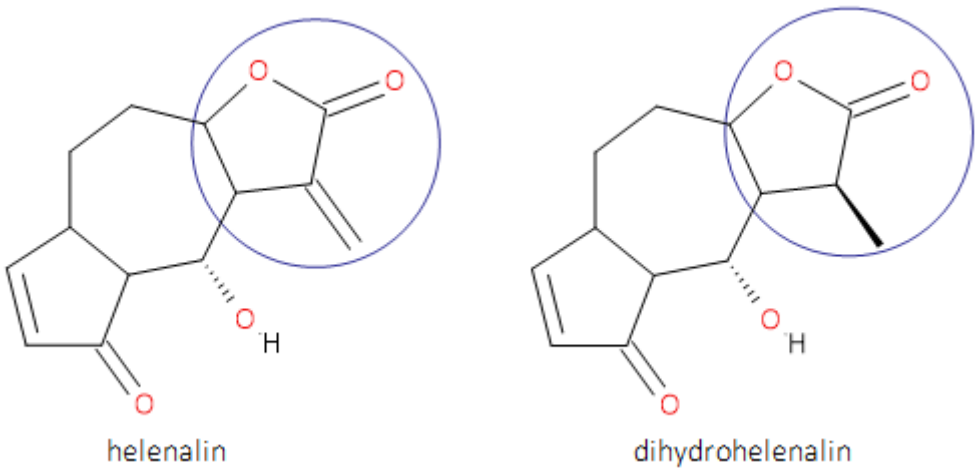


Figure 1. Sesquiterpene lactones in *Arnica montana*: helenalin and dihydrohelenalin. Helenalin and its derivatives are dominant in *A. montana* ssp. *montana* while dihydrohelenalin and its derivates are present in higher amount in *A. montana* ssp. *atlantica* variety. Lactone moiety of the SLs is marked with blue line. Figure created using BIOVIA Draw software by Dassault Systèmes (Vélizy-Villacoublay, France; https://www.3ds.com/products-services/biovia/products/scientific-informatics/biovia-draw/).

Considering their pharmacological importance, helenalin and dihydrohelenalin (DH) have been shown to suppress the activity of immune response transcription factors (e.g. NF-κB and NF-AT) in T- and B-lymphocytes and epithelial cells (Klaas et al., 2002; Kriplani et al., 2017; Petrova et al., 2012). In comparison with DH, helenalin has proved to be more efficient in its anti-inflammatory activity. It affects the immune response by decreasing cytokine release *in vitro* and activating the mitochondrial pathway of apoptosis in important cells of the adaptive immune system, namely CD4+ T cells (Matos et al., 2021). Helenalin possesses a stronger anti-inflammatory activity, but is also more probable to cause allergic reactions while dihydrohelenalin has less reactive methyl instead of helenalin's unsaturated methylene group and presumably causes fewer side effects (Klaas et al., 2002). Allergic reactions such as contact dermatitis are commonly induced by SLs with α-methylene-γ-lactone moieties, including helenalin, because they can penetrate the skin and form complexes with host proteins that promotes generation of allergen-specific cells of immune system (Drogosz & Janecka, 2018).

1.2.3 Biosynthesis pathway of sesquiterpene lactones

Being the largest and most diverse group of plant secondary metabolites, some terpenes can be found among a vast number of plant species while others are species-specific (Zhou & Pichersky, 2020). All terpenes are synthesised from two basic precursors isopentenyl diphosphate (IPP) and dimethylallyl diphosphate (DMAPP) which can origin from both plastidic 2-C-methyl-D-erythritol-4-phosphate, and the cytosolic mevalonate (MVA) pathway (Figure 2)

(Bröker et al., 2020). Farnesyl diphosphate (FPP), formed by fusion of two IPP molecules with DMAPP, is a precursor of germacrene A in sesquiterpenoid synthesis, but also a starting point for sterols synthesis, including squalene, and farnesol synthesis (Bröker et al., 2020).

Genes encoding enzymes required to produce germacrene A from FPP are conserved in Asteraceae species and recently biosynthesis pathways for some sesquiterpene lactones from Asteraceae have been partly elucidated (Frey et al., 2020; Gou et al., 2018; Ramirez et al., 2013). Initial part of sesquiterpene lactones biosynthesis pathway includes enzyme germacrene A synthase (GAS) which catalases synthesis of germacrene A from FPP (Figure 2). Germacrene A is subsequently oxidized by germacrene A oxidase (GAO) which is a heme binding enzyme from cytochrome P450 superfamily with monooxygenase activity yielding germacrene A acid (GAA). The next step predominately includes another heme binding enzyme from cytochrome P450 superfamily which catalyses stereoselective or non-stereoselective hydroxylation of germacrene A acid. Some isomers of hydroxylated GAA can spontaneously lactonize while others require enzymatic catalysation of the reaction (Frey et al., 2018; Gou et al., 2018; Matos et al., 2021; Ro et al., 2006).

Frey et al. (2020) confirmed a cytochrome P450 from sunflower, namely HaCOS/CYP71BL9, which is responsible for hydroxylation of germacrene A acid at 6α position, yielding a precursor for costunolide. Costunolide is an important precursor of several SLs, and also displays bioactivities such as anti-inflammatory, anti-carcinogenic and anti-viral activities (Ikezawa et al., 2011). A similar compound to costunolide is inunolide, a key intermediate of C12,8α-sesquiterpene lactones such as helenalin.

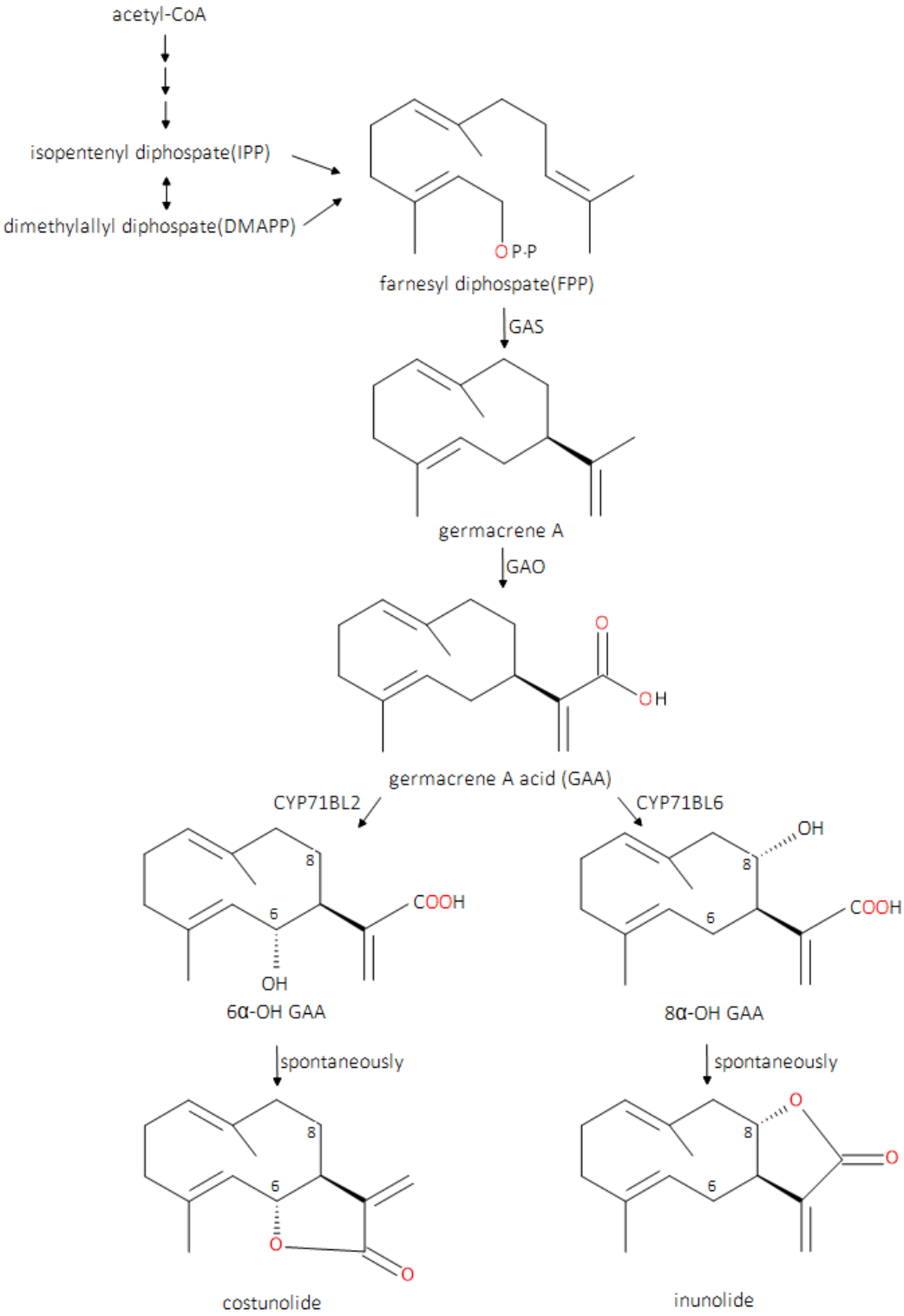


Figure 2. Putative biosynthesis pathway of sesquiterpene lactones precursors costunolide and inunolide in Asteraceae. The mevalonate (MVA) pathway, followed by synthesis of germacrene A from farnesyl diphosphate (FPP) catalysed by germacrene A synthase (GAS), subsequent oxidation to germacrene A acid (GAA) by germacrene A oxidase (GAO). Further steps depicted include CYP71BL2, *Lactuca sativa* GAA 12,6α-hydroxylase/costunolide synthase and CYP71BL6, *Inula hupehensis* GAA 12,8α/β- hydroxylase. Adapted from Gou et al. (2018). Figure created using BIOVIA Draw software by Dassault Systèmes (Vélizy-Villacoublay, France; https://www.3ds.com/products-services/biovia/products/scientific-informatics/biovia-draw/).

Gou et al. (2018) investigated the biosynthesis pathway of sesquiterpene lactones from an Asteraceae family medicinal plant, *Inula hupehensis.* They have found that *Inula hupehensis* cytochrome P450 monooxygenase, called CYP71BL6, can catalyse hydroxylation of germacrene A acid in non-stereoselective manner, producing both 8α- and 8β-hydroxyl germacrene A acid stereoisomers. Due to its favourable stereo-configuration 8α-hydroxyl germacrene A acid spontaneously converts into inunolide (Figure 2). Additionally, they confirmed that the expression level of CYP71BL6 correlates with presence of *I. hupehensis*-specific sesquiterpene lactones in different plant organs.

For Arnica, no SLs biosynthesis pathway genes have been described and characterized in literature so far and no genomic sequences of Arnica are available in public data bases making gene identification time-consuming and labour-intensive. Prior to the start of this thesis, putative candidate genes of sesquiterpene lactone biosynthesis pathway in *Arnica montana* were identified in the AG Prüfer working group using discriminating primers derived from other Asteraceae sequences and molecular methods such as 5´/3´RACE polymerase chain reactions. Their function remains elusive: *A. montana* germacrene A synthase (*AmGAS)*, *A. montana* germacrene A oxidase (*AmGAO)* and *A. montana* cytochrome P450 monooxygenase (*AmCYP)* genes.

1.3 Yeast as production platform

Synthetic biology and metabolic engineering provide an alternative approach to produce plant natural products, such as sesquiterpenes, in sufficient quantities using microbial fermentation (Liu et al., 2021). *Saccharomyces cerevisiae* is a well-known model organism of molecular biology and a promising host for metabolic engineering with established methods for gene editing and heterologous gene expression. Commercialized biopharmaceutical products synthesized by recombinant yeast include hormones like insulin and glucagon, as well as industrial scale production of progesterone, hydrocortisone, artemisinic acid (anti-malarial drug) and resveratrol (potent antioxidant) (Kulagina et al., 2021).

Native producers of sesquiterpenoids include plants, fungi, microorganisms, and insects, however, a limit for production in the organisms are small quantities and poor purities of the compounds. Yeast have an internal MVA pathway which makes them a good host for potential production of sesquiterpene lactones through modification of the pathway (Liu et al., 2021).

Paddon and Keasling (2014) reviewed production of sesquiterpenoid artemisinic acid in *S. cerevisiae* and *Escherichia coli* with altered MVA pathway and heterologously expressed MVA pathway genes, respectively. Artemisinic acid is a precursor for *Artemisia annua* sesquiterpene lactone artemisinin which is considered as a standard treatment for malaria. Artemisinic acid that can be further chemically converted to artemisinin was synthesized using a modified yeast strain, and optimized fermentation and extraction methods providing a solution to produce more reasonably priced medicine that is functionally equivalent to the plant-derived artemisinin. They achieved production as high as 25 g of artemisinic acid per one litre of *S. cerevisiae* culture. *E. coli* expressing MVA pathway proved to be unsuitable for the expression of eukaryotic P450 enzymes needed for biosynthesis of the sesquiterpenoids mostly due to nature of P450 membrane-bound proteins and lack of inner organelles, such as endoplasmic reticulum in bacteria (Hausjell et al., 2018; Paddon & Keasling, 2014).

There are multiple strategies used to increase the flux of FPP towards sesquiterpene synthesis in *S. cerevisiae*. The strategies follow different approaches such as mutating enzymes of competing pathways, introducing metabolic switches, upregulating key gene expression or gene copies, creating protein fusions and several others (Liu et al., 2021).

One strategy includes modification of MVA pathway by overexpressing the enzymes preceding FPP synthesis and suppression of genes in competing metabolic pathways such as triterpene pathway starting with squalene or synthesis of farnesol (Figure 3)(Bröker et al., 2020). More precise, modifications comprise overexpression of rate limiting enzymes of MVA pathway, 3-hydroxy-3-methylglutaryl-CoA (HMG-CoA) synthase (ERG13) and truncated 3-hydroxy-3-methylglutaryl-coenzyme reductase (tHMGR). The overexpression, combined with deletion of negative MVA pathway regulator *ROX1*, significantly increases MVA flux towards FPP. To tackle following challenges, this strategy includes replacement of squalene synthase promoter with copper-repressible promoter that can be easily regulated with addition or exclusion of copper salt in the cultivation media, and knockout of phosphatases, such as lipid phosphate phosphatase 1 (LPP1), which would move FPP pool towards farnesol synthesis. For heterologous expression of cytochrome P450 enzymes, it is common to include its electron donor, NADPH-cytochrome P450 oxidoreductase, essential for its functionality (Hamann & Møller, 2007; Ikezawa et al., 2011; Ro et al., 2006). These modifications of *Saccharomyces cerevisiae* sesquiterpene producing strains are considered to show high potential with

improvements in the future. Additionally, *S. cerevisiae* do not accumulate sesquiterpenes inside the cells, which contributes to their non-toxicity to yeast, and are characterized by high-cell density cultures and fermentation process control (Liu et al., 2021).

This metabolically modified yeast strain lays the basis for the characterization of the potential SLs biosynthesis genes derived from Arnica.

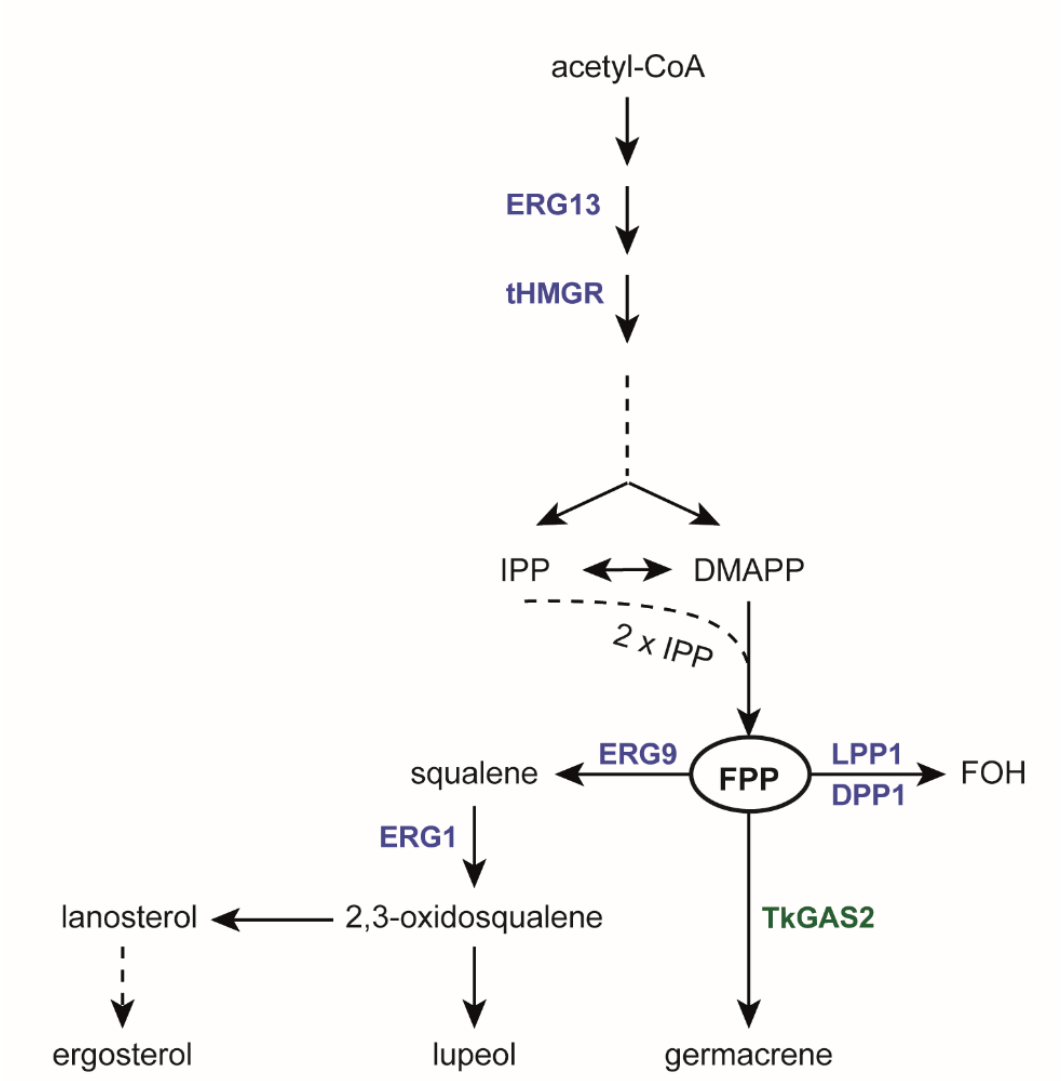


Figure 3. Modified mevalonate (MVA) pathway in *Saccharomyces cerevisiae* and redirected farnesyl diphosphate (FPP) flux towards sesquiterpene lactones as in Bröker et al. (2020). ERG13-3-hydroxy-3-methylglutaryl-CoA (HMG-CoA) synthase; tHMGR- truncated 3-hydroxy-3-methylglutaryl-coenzyme A (HMG-CoA) reductase; IPP- isopentenyl diphosphate; DMAPP- dimethylallyl diphosphate; FPP- farnesyl diphosphate; ERG9-squalene synthase; FOH- farnesol; LPP1- lipid phosphate phosphatase 1; DPP1- diacylglycerol pyrophosphate phosphatase; ERG1- squalene epoxidase; TkGAS2- *Taraxacum koksaghyz* germacrene-A synthase.

1.4 Thesis objectives

Plant material obtained by harvesting *A. montana* wild populations, also already threatened by habitat loss, is an inconvenient source of its secondary metabolites. Content and quality of bioactive sesquiterpene lactones greatly depends on environmental conditions the plants grow in. On the other hand, specific growth conditions and soil requirements make commercial cultivation of Arnica difficult and unviable, especially when plant's naturally low abundance of bioactive substances is considered.

To examine potentially more efficient and sustainable alternatives for acquiring bioactive compounds from *A. montana,* insight into their biosynthesis pathway is needed. The biosynthesis pathway of helenalin and dihydrohelenalin remains unknown up to date. Basic knowledge about potential candidate genes for biosynthesis and information about localisation of SLs in different plants tissues could offer the basis for future research and development.

Considering above stated, the aims of my master's thesis are:

- Functional characterization of potential candidate genes, that might be involved in SL biosynthesis, in the heterologous expression system yeast *Saccharomyces cerevisiae*
- Analysis of potential candidate gene expression patterns by qPCR and sesquiterpene lactones derivates content distribution in different plant tissues analysed by GC-MS after extraction

3. Materials and Methods

3.1 Materials

3.1.1 Equipment

All technical devices used in work for this thesis are listed below (Table 1).

Table 1. Used technical devices and respective providers.

Device	Provider
Autoclave CV-EL 12L/18L	CertoClav (Traun, Austria)
Biostep Dark Hood DH-50	Biostep GmbH (Burkhardtsdorf, Germany)
Centrifuge 5424	Eppendorf AG (Hamburg, Germany)
Centrifuge Allegra® X-15R (Rotor: SX4750)	Beckman Coulter (Krefeld, Germany)
CFX96 Touch™ Real-Time Detection System	Bio-Rad Laboratories GmbH (Munich, Germany)
Cleanbench Herasafe 2020	Heraeus Instruments (Hanau, Germany)
Concentrator plus	Eppendorf AG (Hamburg, Germany)
DNA/RNA UV Cleaner UVC/T-M-AR LTF	Labortechnik GmbH & Co. KG (Wasserburg, Germany)
Eppendorf BioSpectrometer®	Eppendorf AG (Hamburg, Germany)
Eppendorf Eporator	Eppendorf AG (Hamburg, Germany)
Freeze dryer alpha 4-2 LSCplus	Martin Christ Gefriertrocknungsanlagen GmbH (Osterode am Harz, Germany)
GC-MS-QP 2010 Ultra High-End	Shimadzu Corporation (Kyoto, Japan)
Incubation Shaker Multitron Pro	Infors AG (Bottmingen/Basel, Switzerland)
pH meter inoLab® pH720	WTW (Weilheim, Germany)
Rocket™ Evaporator System	Thermo Fisher Scientific (Waltham, USA)
Safe 2020 Class II Biological Safety Cabinet	Thermo Fisher Scientific (Waltham, USA)
Universal Oven UN110	Memmert GmbH & Co. KG (Schwabach, Germany)
UV transilluminator	Biostep GmbH (Burkhardtsdorf, Germany)
Vortex-Genie 2	Scientific Industries (Bohemia, USA)
Wide Mini-Sub® Cell GT system	Bio-Rad Laboratories GmbH (Munich, Germany)

3.1.2 Software

All software used for processing the data is listed below (Table 2).

Table 2. Software and respective providers.

Software	Provider
Argus-X1® documentation software	Biostep GmbH (Burkhardtsdorf, Germany)
BioRad CFX Manager	BioRad (Munich, Germany)
BIOVIA Draw	Dassault Systèmes (Vélizy-Villacoublay, France)
GCMSsolution Software	Shimadzu Corporation (Kyoto, Japan)
SeqBulider Pro™	DNASTAR Inc. (Madison, USA)
SeqMan Pro	DNASTAR Inc. (Madison, USA)

3.1.3 Organisms

Organisms used in the scope of this thesis' experiments are listed in Table 3.

Table 3. Organisms and their properties descriptions.

Organism	Strain/Variety	Description
Escherichia coli	NEB10β strain	Δ(ara-leu) 7697 araD139 fhuA ΔlacX74 galK16 galE15 e14- Φ80dlacZΔM15 recA1 relA1 endA1 nupG rpsL (StrR) rph spoT1 Δ(mrr-hsdRMSmcrBC) (New England Biolabs, Ipswich, UK)
Saccharomyces cerevisiae (JNB strains were kindly provided by Jan Niklas Bröker, AG Prüfer; NM strains are part of this thesis)	JNB586	rox1::tHMGR/erg13 dAUX Perg9::PCTR3 dAUX LPP1d ncp1::AtR1
	JNB621	rox1::tHMGR/erg13 dAUX Perg9::PCTR3 dAUX LPP1d ncp1::AtR1 1A1_AmGAS (NheI)
	JNB622	rox1::tHMGR/erg13 dAUX Perg9::PCTR3 dAUX LPP1d ncp1::AtR1 1A1_AmGAS (NheI) 1C1_AmGAO1 (PmeI)

Table 3. continued:

	NM13	rox1::tHMGR/erg13 dAUX Perg9::PCTR3 dAUX LPP1d ncp1::AtR1 1A1_AmGAS (NheI) 1E1_AmCYP7 (AflII)
	NM14	rox1::tHMGR/erg13 dAUX Perg9::PCTR3 dAUX LPP1d ncp1::AtR1 1A1_AmGAS (NheI) 1C1_AmGAO1(PmeI) 1E1_AmCYP7 (AflII)
	NM15	rox1::tHMGR/erg13 dAUX Perg9::PCTR3 dAUX LPP1d ncp1::AtR1 1A1_AmGAS (NheI) 1E1_AmCYP1 (BstEII)
	NM16	rox1::tHMGR/erg13 dAUX Perg9::PCTR3 dAUX LPP1d ncp1::AtR1 1A1_AmGAS (NheI) 1C1_AmGAO1(PmeI) 1E1_AmCYP1 (BstEII)
	NM17	rox1::tHMGR/erg13 dAUX Perg9::PCTR3 dAUX LPP1d ncp1::AtR1 1A1_AmGAS (NheI) 1E1_AmCYP2 (AflII)
	NM18	rox1::tHMGR/erg13 dAUX Perg9::PCTR3 dAUX LPP1d ncp1::AtR1 1A1_AmGAS (NheI) 1C1_AmGAO1(PmeI) 1E1_AmCYP2 (AflII)
Arnica montana	ssp. *montana*	Arbo cultivar (Eskusa GmbH, Germany)
	ssp. *atlantica*	WT, Spanish (Eskusa GmbH, Germany)

3.1.4 Cultivation media

All cultivation media were sterilized by autoclaving for 15-20 minutes at 120 °C and previously sterile filtered antibiotics were added in cooled media when needed. Yeast extract peptone dextrose adenine (YPDA) medium contained either glucose (YPDA, pH 7) or galactose (YPDA-Gal, pH 7.2). The media were prepared as follows (Table 4).

Table 4. Composition of lysogeny broth (LB), yeast extract peptone dextrose adenine (YPDA) and supplement dropout (SD) media.

Lysogeny broth (LB) medium, pH 7	
Tryptone	10 g/L
NaCl	10 g/L
Yeast extract	5 g/L
Agar (for solid medium)	16 g/L
Ampicillin (Amp)	100 mg/L
Kanamycin (Kan)	50 mg/L
YPDA medium, pH 7/pH 7.2	
Tryptone	20 g/L
Yeast extract	10 g/L
Glucose/galactose	20 g/L
Adenine hemisulfate salt	0.03 g/L
MOPS buffer (pH 7.2)	50 mM
Supplement dropout (SD) medium	
Glucose	20 g/L
Yeast nitrogen base without amino acids	6.7 g/L
Amino acid dropout	According to providers instructions
Agar (for solid medium)	16 g/L

3.1.5 Vectors

The vectors used are listed below (Table 5).

Table 5. Vectors and their properties description.

Vector	Selectable marker	Properties and provider
pCR®II-TOPO	Ampicillin, Kanamycin	LacZ reporter gene, T7 and Sp6 promoter, f1 and pUC ori, MCS (Thermo Fisher Scientific, Waltham, USA)
Gateway™ pENTR™ 3C	Chloramphenicol, Kanamycin	attL sites, rrnB terminator, pUC ori, ccdB (lethal gene), MCS, Gateway® entry vector (Thermo Fisher Scientific, Waltham, USA)
pAG305GAL-ccdB (1E1)	Ampicillin, Chloramphenicol	attR sites, GAL1 promotor, integrating plasmid, LEU2 auxotrophic marker Gateway® destination vector (a gift from Susan Lindquist , Addgene kit #1000000011 (Addgene, Watertown, USA))

3.1.6 Oligonucleotides

Oligonucleotide sequences (primers) used for polymerase chain reactions are listed in Table 6.

Table 6. Primers used in this work and the respective sequences. Restriction sites are bolded.

Primer	Sequence (5´- 3´)
5´RACE AmCYP6_GSP2	CGGAATCAGTTCAAAATGCAACCCTTG
5´RACE AmCYP6_GSP4	GTACCAACACTTTCTCCAAGCACATTGAGC
AmCYP6 for_int	CCCGAGAGATCATGAAGAC
AmCYP5_491bp_for	CATAAGTGAAATAAACCAGAGTTT
AmCYP4_457bp_for	CTAAACTGTCCAATCATGTGTT
AmCYP2_BamHI_for_Sc	AAA**GGATCC**AAAAAAATGTTCCCATGGCTTGTTAC
AmCYP5_BamHI_for_Sc	AAA**GGATCC**AAAAAAATGTTGAGTTTGTTGGAGC
AmCYP5_XhoI_rev	AAA**CTCGAG**TCAATAACCATCAATAGCCA
AmCYP1_427bp_for	CGTATTGGCGTAACCTTAG

Table 6. continued:

AmCYP2_858bp_rev	CAACTCAAGTAAAGCATGAACA
AmCYP2_SalI_rev	AAAGTCGACTTAGATCTCCCCAATTATATTTAG
AmCYP6 5´UTR for	ACTGAATCAAACAATCATGG
AmCYP6 3´UTR rev	AGTGCTTATGTAGTTCTCAC
3´RACE_AmGAO	CTCGTGGTCCGAGAAACATT
5´RACE_AmGAS1_GSP2	GTTGTGTGTAGATCAAGTTCAACA
5´RACE AmCYP5_GSP1	GTTCACCATAAGGAGAAATTGTGG
LPP1_STOP-Frag_for	CTTGATGACCCCAGTATAT
LPP1_STOP-Frag_rev	CTATCAACAAAGTCAGGCC
M13_for	GTAAAACGACGGCCAG
M13_rev	CAGGAAACAGCTATGAC
pAGD_bw	CAGGTTGTCTAACTCCTTCC
pAGD_fw	CTAGAACTAGTGGATCCCCCATC
pENTR4_for	CCTGTTAGTTAGTTACTTAAGCTC
pENTR4_rev	GTAACATCAGAGATTTTGAGACAC
AmCYP6 BamHI for	AAAGGATCCAAAAAAATGGCATTTTATTCATCCCTAC
AmCYP6 XhoI rev	AAACTCGAGTTAGAACCGAGGACTCACTGTAAC
5´LEU for yeast	CCCTTTTATGGATTCCTAAATCCTCGAG
3´LEU rev yeast	GTATTGGTAGATTTAGTACTGAAGAGG
3´LEU rev II yeast	CTGGGTAAGGATGATGCATTAGC
pAG 3´LEU	CATCAGAGCAGATTGTACTGAGAGTGC
pAG 5´LEU	CGCATCTGTGCGGTATTTCACACC
qRT_AmEF1a_1 for	TGGTCGAGACCTTTGCTGAG
qRT_AmEF1a_1 rev	ACTCCATTCGACCAAAACGAA
qRT_AmETIF5a_1 for	AGAGGGAAAGGACCTGGTGG
qRT_AmETIF5a_1 rev	AAGCTTTTGCTGCTGTTGATGAT
qRT AmGAS 1 for	ACGAGATCCGCCTTCTAACCGATGC
qRT AmGAS 1 rev	GCTAGTTGCTCCTCAAGTTCGGCGT
qRT AmGAO 1 for	TCGTCCCACATCGGCCACCAAAAAC
qRT AmGAO 1 rev	AGTCTCGGGTCTGTTGGCGAAGGTT
qRT AmCYP7 1 for	AGGGTGCAAAAGGGTGGTGATCTTG
qRT AmCYP7 1 rev	GCTCTCACTTCCGCTTGCGCTTTCT
qRT AmCYP1 1 for	TGCGGTGGATCACGACGGTGATTTC
qRT AmCYP1 1 rev	AGGCAGGTTTGGAATGTCCGCATC
qRT AmCYP2 1 for	AGCCCAGGAGATGTTGGGTGGATCG
qRT AmCYP2 1 rev	GTCGTATCAACTCCCCCGGTAAGCAC
qRT AmCYP5 1 for	TGGATGAGTTCTTGACGGGTGTGGT
qRT AmCYP5 1 rev	TGCCGTTGTATCCGTAGCAGCAACA
qRT AmCYP8/9 1 for	TGCTTCTGGCTATCGGCTAAATCATGC

Table 6. continued:

qRT AmCYP8/9 1 rev	CCCGTGCAGTTTCGGATTTCACTTTGC
qRT AmCYP6 1 for	GTAAACAAACGAAAGGGGGAATTAGG
qRT AmCYP6 1 rev	CGCTTTCAGATACCGCATGTTCTTC

3.2 Methods

3.2.1 Molecular methods

3.2.1.1 Polymerase chain reaction

The polymerase chain reaction (PCR) is a validated molecular biology method for DNA amplification. Thanks to properties of thermostable DNA polymerases, it is possible to obtain newly synthetized DNA in short time and from small amount of template DNA. The reaction consists of multiple steps, starting with denaturation when the strands of double stranded DNA are denatured that allows the next step, annealing of oligonucleotides complementary to template DNA, called primers, to hybridize to both ends of fragment which is to be amplified. Next step is elongation, catalysed by a thermostable DNA polymerase by adding deoxyribonucleoside 5′-triphosphates (dNTPs) from primers' 3´end in 5´-3´direction yielding a new DNA strand. One repetition of these steps is called thermal cycle and one PCR usually consists of 20-40 cycles. Initial denaturation and final elongation steps are not repeated, they ensure complete denaturation of template DNA and complete elongation of amplicons, respectively. Annealing temperature depends on primer sequence and elongation time depends on polymerase catalysis rate and amplicon size. All PCRs were performed on Biometra TProfessional trio thermocycler (Analytik-Jena, Jena, Germany)

Phusion® PCR

Phusion® PCR was used for subsequent cloning when it was crucial to ensure the correct sequence is amplified in PCR. This reaction was catalysed by Phusion® High-Fidelity DNA Polymerase (Thermo Fisher Scientific, Waltham, USA) because of its proofreading 3' - 5' exonuclease activity and low error rate. Conditions of reaction described in Table 7.

Table 7. PCR mixture and conditions for Phusion® HF polymerase. *- annealing temperatures are primer pair specific.

Components	V / µL	Step	T / °C	Duration / s	
5x Phusion® HF buffer green	10	Initial denaturation	98	300	
10 mM dNTPs	1	Denaturation	98	20	35 cycles
10 mM forward primer	1	Annealing	*	30	
10 mM reverse primer	1	Elongation	72	15-30 s/kb	
Template DNA	1	Final elongation	72	300	
Phusion® HF DNA polymerase	0.5	Conservation	4	∞	
ddH_2O	to 50				

Colony PCR

Colony PCR was used to confirm that *E. coli* or *S. cerevisiae* colonies contained plasmids with gene of interest. For yeast colony PCR, some colony material was picked into 1,5 mL Eppendorf tube, resuspended in 20 µL 0.2%(w/v) SDS and cooked at 95 °C for 10 minutes to lyse the cells. The lysate was pelleted by centrifugation at 11000 g for 30 s and 0.3 µL of supernatant was used as template for PCR. Another method for lysing yeast cells was used: a colony was touched with a pipette tip which was then dissolved in 50 µL 20 mM NaOH, vortexed, cooked at 95 °C for 10 minutes, centrifuged at 11000 g for 5 min and 1 µL of supernatant was used as template for PCR. For *E. coli* colony PCR, a colony was touched with a pipette tip which was then put in a PCR tube containing PCR mixture. Conditions of reaction described in Table 8.

Table 8. Polymerase chain reaction mixture and conditions for colony PCR using MangoTaq™ polymerase. *- annealing temperatures are primer pair specific.

Components	V / µL	Step	T / °C	Duration / s	
5x MangoTaq™ Buffer	6	Initial denaturation	95	600	
50 mM $MgCl_2$	1.5	Denaturation	95	30	32 cycles
10 mM dNTPs	0.6	Annealing	*	30	
10 mM forward primer	0.6	Elongation	72	30-60 s/kb	
10 mM reverse primer	0.6	Final elongation	72	300	
MangoTaq™ DNA polymerase	0.06	Conservation	4	∞	
ddH_2O	to 30				

3.2.1.2 Agarose gel electrophoresis

Agarose gel electrophoresis is a method that enables separation of nucleic acids by size. Nucleic acids samples are loaded on a gel and when electric field is applied, they move through the gel. Since DNA and RNA molecules are negatively charged thanks to their phosphate groups, they move towards the positive pole of the electric field. Agarose gel is a porous material when solidified, with pore size depending on the percentage of agarose. Smaller molecules move faster through the pores of the gel than bigger ones, which allows separation of nucleic acids by size.

The DNA and RNA samples were separated on 1% (w/v) agarose gels (Bio-Budget Technologies GmbH, Krefeld, Germany) prepared using 1x Tris-acetate-EDTA (TAE) buffer with addition of 20 µL/L Midori Green Advance (Biozym Scientific GmbH, Hessisch Oldendorf, Germany) prior to polymerization. The intercalating dye emits light at 530 nm upon excitation with UV-light at 290 nm. Gels of different sizes were prepared using Wide Mini-Sub® Cell GT gel trays (Bio-Rad Laboratories GmbH, Munich, Germany) and a comb to create wells in the gel. Gel electrophoresis was performed in gel electrophoresis cells of different sizes (Bio-Rad Laboratories GmbH, Munich, Germany) filled with 1x TAE buffer. If not included in PCR reaction reagents, 10x DNA loading dye was added to the samples before loading them. As a reference for size of separated nucleic acid fragments, 5 µL of Thermo Scientific™ GeneRuler™ 1 kb Plus DNA Ladder (Thermo Fisher Scientific, Waltham, USA) was used (Figure 4). Agarose gel electrophoresis was performed applying 100-120 V for 18-25 min. To visualise and document the nucleic acids fragments separated on the gel, Biostep Dark Hood DH-50 along with a digital camera and the Argus-X1® documentation software (Biostep GmbH, Burkhardtsdorf, Germany) were used.

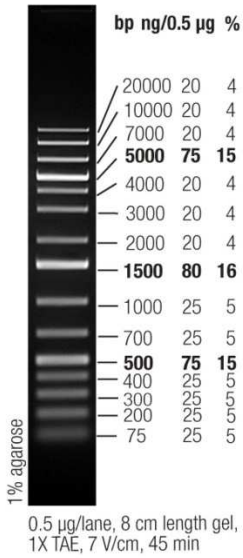


Figure 4. Thermo Scientific™ GeneRuler™ 1 kb Plus DNA Ladder (Thermo Fisher Scientific, Waltham, USA). Source: https://www.fishersci.de/shop/products/fermentas-generuler-1kb-plus-dna-ladder/p-4529751

3.2.1.3 DNA purification and plasmid isolation

For obtaining purified DNA for further use, after a PCR or restriction reaction and agarose gel electrophoresis, NucleoSpin® Gel and PCR Clean-up kit (Macherey-Nagel GmbH & Co. KG, Düren, Germany) was used according to the manufacturer's instructions. To purify DNA from agarose gel, the fragment of appropriate size was cut out using a scalpel and UV transilluminator (Biostep GmbH, Burkhardtsdorf, Germany) and the agarose was dissolved in a supplied buffer. The use of NucleoSpin® Gel and PCR Clean-up kit allows DNA fragments to be specifically eluted by buffer combinations of different chaotropic agents and ionic strength, that allow DNA to bind to or elute from the column.

An *E. coli* colony of interest was grown overnight at 37 °C and 140 rpm in 4 mL LB medium with appropriate antibiotics. For isolation of plasmid DNA from *E. coli* NucleoSpin Plasmid kit (Macherey-Nagel GmbH & Co. KG, Düren, Germany), NucleoSpin Plasmid/Plasmid (NoLid): Isolation of high-copy plasmid DNA from *E. coli* protocol was used according to the manufacturer's instructions. The process includes cell lysis, precipitation of proteins, genomic DNA, and cells organelles residue. The final step includes precipitation, washing and elution of plasmid DNA using a column supplied in the kit. The plasmid DNA was stored in elution buffer at -20 °C until further use.

3.2.1.4 DNA restriction and ligation

Combining of restriction and ligation reactions allows the generation of recombinant DNA constructs.

Restriction endonucleases are enzymes able to cleave double stranded DNA at specific palindromic sequences called restriction sites. The reaction results in either blunt or sticky ends of DNA fragments suitable for subsequent ligation. DNA restriction was used to create specific single strand overhangs at DNA fragments or vector backbones for subsequent DNA ligation or to verify correct integration in a plasmid followed by visualisation using agarose gel electrophoresis. Restriction endonucleases required for described purposes were supplied from New England BioLabs (Ipswich, Massachusetts, United States). Different restriction enzymes required different buffers, incubation temperature and time, and were used according to manufacturer's instructions.

Digested DNA fragments and vector backbones with corresponding single strand overhangs were ligated using T4 DNA ligase (Promega, Fitchburg, USA). This ligase catalyses the formation of a phosphodiester bond between 5'-phosphate and 3'-hydroxyl groups of nucleotides located in proximity. For estimation vector and insert concentrations, 2 µL of DNA were visualized using gel electrophoresis. DNA ligation reaction mixture was prepared as described in Table 9 and incubated for 2 hours at room temperature or at 16 °C overnight.

Table 9. DNA ligation reaction mixture. Vector and insert DNA are mixed at 1:3 molar ratio.

DNA ligation reaction mixture	
Vector DNA : Insert DNA	1:3
T4 DNA Ligase 10x Buffer	2 µL
T4 DNA Ligase	1 µL
ddH_2O	to 20 µL

3.2.1.5 cDNA synthesis

Following the manufacturer´s instructions, cDNA was synthesized using PrimeScript RT reagent kit (Takara Bio, Japan). The contents of the kit include RNase inhibitor, reverse transcriptase, dNTPs, dT oligonucleotides, random 6-mers and reaction buffer. Due to the 3' poly-A tail specific for mRNA, the kit enables mRNA transcription to cDNA. For 500 ng of total

RNA, 2 µL 5x PrimeScript RT -Mix 1 and RNase free water were mixed and put to thermal cycler at 37 °C for 15 min and 85 °C for 15 s.

3.2.1.6 TOPO® TA Cloning®

TOPO® TA Cloning® enables cloning of a DNA fragment into the pCR®II-TOPO vector (TOPO™ TA Cloning™ Kit, Thermo Fisher Scientific, Waltham, USA) without requirement for restriction enzymes and ligase. The linearized vector has 3´ thymine overhangs and covalently bound topoisomerase I which allows ligation of DNA fragments with 3´adenine overhangs. Polymerases with non-template dependent terminal transferase activity produce 3´ adenine overhangs at both ends of double stranded DNA, generating suitable product for ligation. Purified PCR products (10-100 ng, according to manufacturer's instructions) were A-tailed at 72 °C for 10 min using MangoTaq™ DNA Polymerase (Bioline GmbH, Luckenwalde, Germany). TOPO® TA Cloning® ligation reaction was incubated at room temperature for at least 5 minutes. Reaction mixtures for A-tailing and TOPO® TA Cloning® were prepared as follows in Table 10.

Table 10. Reaction mixtures for A-tailing and TOPO® TA Cloning® reactions. For both reactions 10-100 ng of PCR product was used.

A-tailing		TOPO® TA Cloning®	
Purified PCR product	14 µL	A-tailed PCR product	1.4 µL
10 mM dNTPs	1 µL	1:4 diluted salt solution	0.3 µL
50 mM $MgCl_2$	1 µL	pCR®II-TOPO vector	0.3 µL
5x MangoTaq™ Buffer	4 µL		
MangoTaq™ DNA polymerase	0.1 µL		

3.2.1.7 Gateway® Cloning

Gateway recombinational cloning is based on the specific integration and excision of bacteriophage λ into and out of the *E. coli* genome (Reece-Hoyes & Walhout, 2018). An entry vector carrying gene of interest in between attL1 and attL2 sites that allow site specific recombination with attR1 and attR2 sites in destination vector catalysed by LR clonase enzyme mix. In this recombination reaction, a cassette holding chloramphenicol-resistance gene and the lethal ccdB gene is replaced by the gene of interest, allowing efficient and easy selection of positive clones.

Prior to my work on this thesis, pENTR™ 3C (Thermo Fisher Scientific, Waltham, USA) entry clones carrying *AmCYP1, AmCYP2, AmCYP8* and *AmCYP9* genes were prepared, as well pAG305 clone for expression in yeast carrying *AmCYP7* gene (no GeneBank entry numbers available, internally comunicated results – Lena Grundmann, AG Prüfer). In my work, to generate entry clones, enzymatic restriction/ligation reactions described in section *3.2.1.4 DNA restriction and* ligation, Gateway™ pENTR™ 3C entry vector and coding sequence of gene of interest were used. Subsequent LR recombination reaction into pAG305GAL-ccdB destination vector (Addgene, Watertown, USA) for expression in yeast, Gateway® LR Clonase® II enzyme mix (Invitrogen, Carlsbad, USA) was used. The recombination reaction mixture was prepared as follows (Table 11).

Table 11. Gateway® LR Cloning reaction mixture.

Gateway® LR Cloning reaction mixture	
pENTR™ 3C-CYP entry clone	150 ng
pAG305GAL-ccdB destination vector	150 ng
1x TE buffer pH 8	to 7 µL
Gateway® LR Clonase® II	1 µL

3.2.1.8 Quantitative real-time PCR

Quantitative real-time PCR (qPCR) is a method that allows to quantify the amount of amplified DNA present in the PCR sample in real-time. When an intercalating fluorescent dye complexes with double stranded DNA, it emits fluorescence at a specific wavelength, and along with fluorescence detector, it allows the reaction to be followed in real-time. SYBR-Green dye that was used for this experiment absorbs at 494 nm and emits at 521 nm when in complex with double stranded DNA. The amplification of DNA can be quantified because the fluorescence increases proportionally to the amount of double-stranded DNA. Quantification cycle value (Cq), also known as cycle threshold (Ct) is the cycle at which the fluorescence rises above the background signal, and it correlates with the concentration of the gene of interest in the template. The lower the Cq value, the higher initial concentration of the template for the gene of interest. Along with Cq value, information about melting temperatures of the qPCR products is generated. It gives information about the purity of the PCR product as unspecific amplicons melting temperature differs from the target. Melting curve was generated by

continuous temperature increase of 0.5 °C every 5 s from 58 °C to 95 °C and simultaneous detection of the fluorescence signal.

Prior to my work on the thesis, primer pairs (Table 6) for each gene were tested and primer efficiencies determined along with internal control genes. In my work, qPCR was used to quantify the expression levels of a specific gene of interest. cDNA was synthesized from total RNA of a sample by reverse transcription, as described in *3.2.1.5 cDNA synthesis.* qPCR reaction mixtures were prepared under a UV clean bench (DNA/RNA UV Cleaner UVC/T-M-AR (LTF Labortechnik GmbH & Co. KG, Wasserburg, Germany)) using KAPA SYBR® FAST qPCR Master Mix and corresponding cDNA dilutions in RNase-free water for amplification of different genes (1:1, 1:2.5, 1:5, 1:10). All samples were analysed in three technical replicates using CFX9 Touch™ Real-Time PCR Detection System (Bio-Rad Laboratories GmbH, Munich, Germany). Reaction mixtures and cycling program for the qPCRs are shown in Table 12. As reference or internal control genes the expression of the genes encoding elongation factor 1 alpha (*AmEF1a*) and eukaryotic translation initiation factor 5 alpha (*AmETIF5a*) were analysed for all samples. Negative controls consisted of two technical replicates containing total RNA samples and one ddH_2O control. Data was analysed using BioRad CFX Manager software (BioRad, Munich, Germany). Delta Cq and 95% confidence interval values of three technical replicates were normalized to reference genes *AmEF1a* and *AmETIF5* and calculated considering cDNA samples dilutions and biological replicates when applicable.

Table 12. qPCR mixture and cycling program. *- annealing temperatures are primer pair specific.

Components	V / µL	Step	T / °C	Duration / s	
KAPA SYBR® FAST qPCR Master Mix	5	Initial denaturation	95	180	
2 mM primers mix	2.5	Denaturation	95	3	40 cycles
cDNA sample dilution (1:1, 1:2.5, 1:5, 1:10)	2.5	Annealing and elongation	*	30	
		Melt curve generation	58-95	+0.5 °C/5 s	

3.2.1.9 Sequencing

To determine the sequence of DNA fragments or plasmids, 150 to 300 ng were diluted in 5 mM Tris/HCl, pH 8.5 buffer (Macherey-Nagel, Düren, Germany) to total volume of 14 µL

and 2 µL of either forward primer or reverse primer were added. Samples were sent to the Sequencing and DNA Analysis Service at the Fraunhofer Institute of Molecular Biology and Applied Ecology in Aachen where the samples were sequenced using the chain-terminating inhibitor method developed by Sanger, Nicklen and Coulson (1977). Sequences were analysed using the programmes SeqMan Pro and SeqBuilder Pro (DNA Star, Madison, USA).

3.2.2 Work with *Saccharomyces cerevisiae*

3.2.2.1 Lithium acetate transformation

For transformation of *S. cerevisiae* strains with increased flux toward FPP and integrated *Arabidopsis thaliana* NADPH-cytochrome P450 reductase 1 (*AtR1*), *AmGAS* and *AmGAO* genes in their genome, lithium acetate (LiAc) method was used. For successful transformation exogenous DNA must pass through the cell wall and plasma membrane and be delivered in the cytosol to reach the nucleus. The presence of LiAc enhances transformation efficiency, especially when combined with polyethylene glycol (PEG), single stranded carrier DNA and heat shock. To have gene of interest, in my case *AmCYP*s integrated in the yeast genome at a specific site, plasmid carrying the gene was linearized at the *LEU2* auxotrophic marker sequence that is homologous to one on the yeast chromosome III. The homology region must be long enough for the integration to be successful and efficient.

Preculture was prepared in 5 mL YPDA media and incubated at 30 °C and 140 rpm overnight. Next day 50 mL YPDA main culture was inoculated with overnight culture of optical density at 600 nm (OD_{600}) set to 0,2 and incubated 30 °C and 140 rpm until OD_{600} reached 0,4. The culture was centrifugated at 1000 g for 10 minutes, cell pellet washed in 40 mL sterile 1x Tris-EDTA (TE; 10xTE, pH 8: 100 mM Tris/HCl, 10 mM EDTA) buffer, centrifuged again and resuspended in 1,5 mL sterile 1x LiAc solution (1 M LiAc : 10x TE buffer : ddH2O; 1:1:8) for 10 min at room temperature. pAG305 plasmid carrying *AmCYP* genes was linearised using restriction enzymes as described in *3.2.1.4 DNA restriction and ligation.* To 1 or 2 µg linearized plasmid DNA and 100 µg carrier DNA denatured (5 minutes at 95 °C), 100 µL aliquot of yeast suspension in LiAc was added, followed by 700 µL freshly prepared PEG solution (1 M LiAc : 10x TE buffer : 50% (w/v) PEG3350; 1:1:8). The mixture was resuspended and incubated at 30 °C and 160 rpm for 30 minutes. On that, 80 µL of dimethyl sulfoxide was added, resuspended, and incubated for 20 minutes at 42 °C followed by cooling on ice for 1 minute. The mix was then centrifugated for 30 s at 20000 g and the pellet was washed in 1x TE. The cells pelleted again,

resuspended in 1 mL YPDA, and incubated at 30 °C and 140 rpm overnight. Next day, the cells were pelleted, washed in 1x TE and the pellet was finally resuspended in 100 µL 1x TE. The suspension was plated on SD medium lacking appropriate amino acids undiluted and in 1:10 and 1:100 dilutions in 1x TE and the plate was incubated for 3 days at 30 °C.

3.2.2.2 Cultivation for galactose-induced expression

First, pre-cultures were grown in 5 mL YPDA media at 30 °C and 140 rpm overnight. Second pre-cultures were inoculated the next day in 50 mL YPDA media (pH 7.2) containing 150 µM $CuSO_4$ and 50 mM MOPS buffer (pH 7.2) and OD_{600} adjusted to 0.2. The cultures were incubated at 30 °C and 140 rpm until they reached OD_{600} of 0.4 when they were centrifuged at 1000 g for 10 minutes. For main culture expression induction, the cell pellet was resuspended in 50 mL or 500 mL YPDA media (pH 7) containing galactose instead of glucose, 150 µM $CuSO_4$, 50 mM MOPS buffer and appropriate volume of 50 times concentrated supplement mixture (Table 13). The main cultures in YPDA-galactose media were grown at 30 °C and 140 rpm until they reached OD_{600} of 2.

Table 13. Composition of supplement mixture for yeast cultivation.

50× supplement mixture	
5-Aminolevulinic acid	264.8 mg
Flavin adenine dinucleotide	138.2 mg
Flavin mononucleotide	215.35 mg
Iron(ii) sulfate heptahydrate	17.5 mg
Millipore ddH_2O	to 100 mL

3.2.2.3 Extraction of germacrene A acid

Main cultures of OD_{600} 2 were collected and extracted three times with 5 mL hexane (Carl Roth GmbH, Karlsruhe, Germany). The hexane layer was separated by centrifugation 5 min at 1000 g and pooled in large glass vials. Using Rocket™ Evaporator System (Thermo Fisher Scientific, Waltham, USA), the pooled hexane was evaporated, and the residue was resolved in 200 µL hexane and transferred in glass vials with inlet and membrane screw cap, ready for further GC-MS analysis.

3.2.2.4 Extraction of potential novel compounds

For extraction of potential novel compounds, the shaking of flasks in cultivation was set at 140 rpm, while other conditions were the same as is described in section *3.2.2.3 Extraction of germacrene A acid*. Variations in main culture volume and extraction methods were tested. Main cultures of either 50 mL or 500 mL were cultivated and equal volumes from each culture were extracted three times with 1/10 culture volume of hexane or 1/5 culture volume of ethyl acetate (EtOAc; Carl Roth GmbH, Karlsruhe, Germany). The pooled hexane extracts and pooled EtOAc extracts were then evaporated using Rocket evaporator and dissolved in 200 µL hexane or 200 µL methanol (Carl Roth GmbH, Karlsruhe, Germany) in glass vials with inlet and membrane screw cap. If there were particles present after dissolving, the samples were filtered and again evaporated and dried using vacuum concentrator (Concentrator plus, Eppendorf AG, Hamburg, Germany). The residue was then dissolved in 200 µL of hexane or methanol (MeOH) in glass vials with inlet and membrane screw cap ready for GC-MS analysis. As external standard for detection of novel compounds, costunolide was prepared in MeOH in concentration 0.1 mg/mL.

3.2.2.5 GC-MS analyses of yeast culture extracts

Gas chromatography–mass spectrometry (GC-MS) is an analytical method that allows separation of vaporized substances followed by detection of mass of the substances, and their identification and quantification. The samples are heated up to a temperature where they are in the gas state and carried through a column by an inert gas as mobile phase. Some compounds are prone to heat conversion at the high temperatures; hence their more stable isomers are detected. The inside of the column is covered by a stationary phase that ensures separation of substances according to their chemical and physical properties. The substances then enter MS system where they are ionized and subjected to an electric or magnetic field. The ions are separated according to their mass-to-charge (m/z) ratio and detected, resulting in a mass spectrum. The mass spectrum displays the ions signal intensity as a function of the m/z ratio. Using commercially available libraries of mass spectra, unknown compounds and target substances can be identified and quantified.

The GC-MS analyses of *S. cerevisiae* culture extracts were performed by GC-MS-QP 210 Ultra High-End (Shimadzu Corporation, Kyoto, Japan) with Rtx-5MS column (stationary phase: 5% diphenyl/95% dimethyl polysiloxane; length: 30 m, diameter: 0.25 mm, film thickness: 0.25

µm; Restek GmbH, Bad Homburg, Germany) and helium as mobile phase. The injection volume was 1 µL, column oven temperature 40 °C, injection temperature 250 °C and 1:5 split injection ratio. The gas chromatography temperature program: 40 °C 1 min, 25 °C/min until 180 °C, 180 °C 5 min, 15 °C/min until 300 °C, 300 °C 3 min, 75 °C/min until 330 °C, 330 °C 5 min. Mass spectrometry was performed using electron ionisation method with 70 eV and following temperature settings: ion source temperature was 200 °C and interface temperature 250 °C.

The mass spectra obtained were analysed in GCMSsolution Software (Shimadzu Corporation, Kyoto, Japan) and compared against the mass spectra library supplied with the software.

3.2.3 Work with *Escherichia coli*

3.2.3.1 Transformation of E. coli NEB10β

Electrocompetent *E. coli* NEB10β cells were used for plasmid amplification. When the bacteria are exposed to brief but strong impulses of electrical field, their cell wall is transiently destabilized which allows DNA to enter the cells. The method is called electroporation. In my work, to perform it, 1 µL of ligation reaction was mixed by pipetting with 50 µL competent cells on ice. The cells were then transferred to ice-cold electroporation cuvettes (Bio-Budget Technologies GmbH, Krefeld, Germany) and put in the electroporator (Eppendorf, Hamburg, Germany). An electric pulse of 1800 V was applied for 5 ms, the cells were resuspended in 800 µL LB medium and transferred to a 1.5 mL Eppendorf tube (Eppendorf, Hamburg, Germany). The cells were incubated for one hour at 37 °C and 140 rpm to regenerate. Afterwards, the cells were centrifuged for 1 minute at 11000 g, resuspended in 100 µL medium which was plated on LB media plates containing antibiotics required for selection of transformants and incubated at 37 °C overnight until visible colonies formed. In case of transformation with pCR®II-TOPO vector, the bacteria were plated on LB media plates containing isopropyl β-d-1-thiogalactopyranoside (IPTG, 23.8 µg/mL) and 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside (X-gal, 75 µg/mL) along with an antibiotic, that allow blue-white screening, with white colonies carrying a plasmid with inserted DNA fragment.

3.2.4 Plant material

Plant material analysed in the scope of this thesis included *Arnica montana* ssp. *montana* and ssp. *atlantica* plants cultivated in tissue culture and greenhouse conditions.

In tissue culture chamber, the plants were grown at 23 °C under long-day conditions (light: 16 h, dark: 8 h) with light intensity of 100 µmol photons/m^2s

In the greenhouse, the plants were cultivated in pots filled with soil (peat) under long-day conditions. The temperature in the greenhouse varied: 24 °C during the day and 20 °C at night. Artificial lights were switched on in case the ambient light intensity dropped below 700 µmol photons/m^2s.

Prior to the start of work for my thesis, flowerheads of two plants of each variety were collected at different time points when they were completely open. Around three leaves of the same plants were collected 4 months after flowering but only for one plant of each variety since for the two other plants the leaves had already decayed. The tissue was homogenized using cold mortar and pestle with addition of liquid nitrogen, collected, and stored at -80 °C until further use.

In my work, for RNA isolation and compounds extraction, young leaves of nine *A. montana* plants, (both AMM and AMA) grown in tissue culture, were harvested and shock-frozen in liquid nitrogen. The samples were then homogenized using cold mortar and pestle with addition of liquid nitrogen, collected in cold Eppendorf tubes, and stored at -80 °C until further use.

3.2.4.1 RNA isolation

RNA was isolated from the frozen plant tissue samples using innuPREP Plant RNA Kit (Analytik-Jena, Jena, Germany) following the kit protocol with DNase step included. The DNase step is crucial to ensure no genomic DNA is amplified in expression analysis. The protocol steps include cell lysis, discarding of column-bound gDNA, RNA precipitation and binding to a column, incubation with DNase to remove gDNA residues, and washing of purified RNA. The total RNA was finally eluted from the column in 50 µL RNase-free water. After checking RNA integrity on agarose gel electrophoresis, RNA samples were ready for cDNA synthesis as described in *3.2.1.5 cDNA synthesis*.

3.2.4.2 Helenalin and dihydrohelenalin extraction for GC-MS analysis

Homogenized flowerheads and leaves samples stored at -80 °C were lyophilized using Freeze dryer alpha 4-2 LSCplus (Martin Christ Gefriertrocknungsanlagen GmbH, Osterode am Harz, Germany). To 250 mg of tissue powder 15 mL EtOH/H_2O (1:1; p.A. grade; Carl Roth GmbH,

Karlsruhe, Germany) was added along with 37.,5 µL of internal α-Santonin standard (1% [w/v] in MeOH (>99%)) for shaking extraction at room temperature overnight. The samples were then centrifugated at 1000 g for 10 minutes and supernatant was transferred to a glass vial. The extracts were evaporated to approximately 50% of starting volume using Rocket evaporator (Thermo Fisher Scientific Waltham, USA). The evaporated extract was then three times extracted with 15 mL EtOAc and centrifuged at 4000 g for 5 minutes for phase separation. The upper EtOAc phases were pooled in a glass vial and completely evaporated in Rocket evaporator. The residue in glass vials were resolved 14 mL chloroform (HPLC grade; Carl Roth GmbH, Karlsruhe, Germany) while shaking 1 hour at room temperature and filtered using folded filter paper. Filtered samples were evaporated using Concentrator plus (Eppendorf AG, Hamburg, Germany) and the residues resolved in 500 µL MeOH while shaking at room temperature for one hour. The samples were centrifugated at 20000 g for 5 minutes and 250 µL of the extract was transferred into GC-MS vial with inlet.

3.2.4.3 GC-MS analysis of A. montana flowerheads and leaves extracts

The GC-MS analyses of *A. montana* tissue extracts were performed by GC-MS-QP 210 Ultra High-End (Shimadzu Corporation, Kyoto, Japan) with Rtx-5MS column (stationary phase: 5% diphenyl/95% dimethyl polysiloxane; length: 30 m, diameter: 0.25 mm, film thickness: 0.25 µm; Restek GmbH, Bad Homburg, Germany) and helium as mobile phase. The injection volume was 1 µL, column oven temperature 120 °C, injection temperature 290 °C and 1:10 split injection ratio. The gas chromatography temperature program: 120 °C 1 min, 10 °C/min until 270 °C, 270 °C 20 min, 15 °C/min until 330 °C, 330 °C 10 min. Mass spectrometry was performed using electron ionisation method with 70 eV and following settings: ion source temperature was 230 °C and interface temperature 290 °C.

The results were analysed in GCMSsolution Software (Shimadzu Corporation, Kyoto, Japan), and resulting spectra were compared with reference spectra obtained by literature search. Internal α-Santonin standard was used for quantification of detected compounds.

4. Results

Arnica montana is a medicinal plant harvested from wild populations for obtaining bioactive compounds helenalin and dihydrohelenalin from the flowerheads. These bioactive sesquiterpene lactones are present in the plants in low amount. Cultivation of Arnica requires specific conditions like the ones in nature and poses multiple difficulties. Alternatives to harvesting wild populations and cultivation such as production in microbial or plant heterologous systems, require more insight into bioactive SLs biosynthesis pathway genes and enzymes. Candidate genes encoding for *A. montana* SLs biosynthesis pathway enzymes: germacrene A synthase, germacrene A oxidase and another cytochrome P450 monooxygenase were tested.

Prior to this thesis, *Saccharomyces cerevisiae* JNB strains with modified MVA pathway, increased flux towards FPP and expressing *Arabidopsis thaliana* NADPH-cytochrome P450 oxidoreductase (*AtR1*) were transformed with *Arnica montana* genes coding for germacrene A synthase (*AmGAS*) and germacrene A oxidase (*AmGAO*). Furthermore, potential candidate gene sequences coding for enzymes in following steps of the pathway, putative cytochrome P450 monooxygenases from *Arnica montana* (*AmCYP*s), in total seven sequences: *AmCYP1*, *AmCYP2*, *AmCYP5*, *AmCYP6*, *AmCYP7, AmCYP8* and *AmCYP9*, were previously identified bioinformatically and isolated. In this thesis, *AmGAS, AmGAO and AmCYP* genes were functionally analysed in *Saccharomyces cerevisiae* heterologous expression system. In addition, expression patterns of *AmGAS, AmGAO and AmCYP* genes in different plant tissues including flowerheads and leaves of greenhouse grown, as well as leaves of plants grown in tissue culture, were assessed. In the following chapters, results of the experiments are described.

4.1. Identification and verification of *AmCYP6* gene

For identification and verification of *AmCYP6* genes three RNA samples from leaves of three plants each, from both AMM and AMA variety, were isolated prior to my thesis. These RNAs were used to verify sequences of the other *AmCYP* genes prior to this thesis while sequence of *AmCYP6* was not successfully proven yet. So, cDNA was synthesized from the pooled RNA samples and used as template for proofreading PCR with primers complement to the gene 5´ and 3´ untranslated regions (see *3.2.1.5 cDNA synthesis* and *3.1.6 Oligonucleotides*). Resulting amplicons were cloned into TOPO® TA vector and *AmCYP6* sequence was sequenced. *AmCYP6* sequences from AMM and AMA differed one from another by a few single nucleotide

polymorphisms which did not affect amino acid sequence (I. Multiple sequence alignment, *8. Supplementary data*). The correct sequence was amplified from the TOPO vector (TOPO-CYP6 Sp.5) and cloned in Gateway entry vector pENTR™ 3C. The transformants were confirmed by colony PCR, test restriction digest and sequencing (Figure 5).

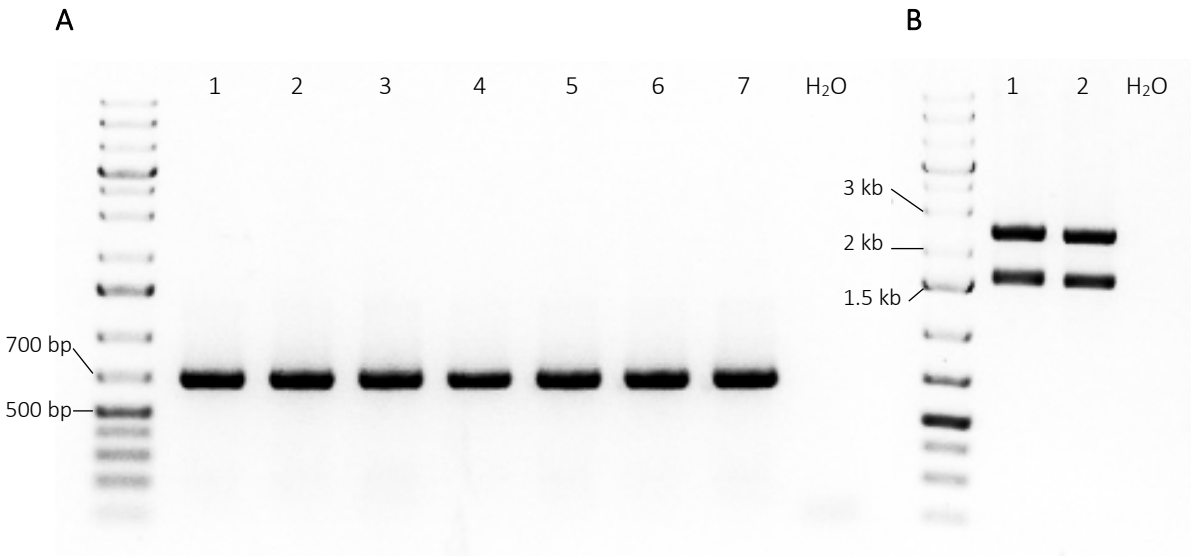


Figure 5. Confirmation of pENTR3C-AmCYP6 clones by colony PCR.
(A) Colony PCR of *Escherichia coli* NEB10β colonies carrying pENTR3C-AmCYP6 plasmids (1-7). Expected amplicon: ~600 bp (primers: pENTR4_for, 5´RACE AmCYP6_GSP4; Ta=60 °C). (B) Test digest of pENTR3C-CYP6 plasmids with EcoRI, colonies no. 1 and 2. Expected fragments: ~1.8kb and ~2.3 kb. Marker: Thermo Scientific™ GeneRuler™ 1 kb Plus DNA Ladder (Thermo Fisher Scientific, Waltham, USA) (See Figure 4 for size reference). H_2O- non-template control.

4.2. Cloning of *AmCYP* genes into pAG305 vector

To transform *S. cerevisiae*, ensure integration in their genome and expression of genes of interest in the heterologous system, *AmCYP* genes had to be cloned into a suitable expression vector. In this case Gateway cloning strategy was used. cDNA sequences of *AmCYP1, AmCYP2, AmCYP5, AmCYP8, AmCYP9* and here constructed *AmCYP6* in entry vector pENTR™ 3C were transferred via LR-reactions to pAG305GAL-ccdB vector (described in *3.2.1.7 Gateway® Cloning*). The vector has LEU2 auxotrophic marker, sequence homologous to yeast genome that facilitates homologous recombination and selection of transformants. Expression of transgenes inserted in the pAG305 vector is under control of GAL1 galactose inducible promotor. Correct *E. coli* clones carrying pAG305-CYP constructs were confirmed by colony PCR using corresponding primers, test restriction digest and sequencing (Figure 6). Work with *AmCYP5*

was not continued because of repetitive difficulties encountered when cloning it into pENTR3C vector: deletion, frame shift and early stop codon in the sequence.

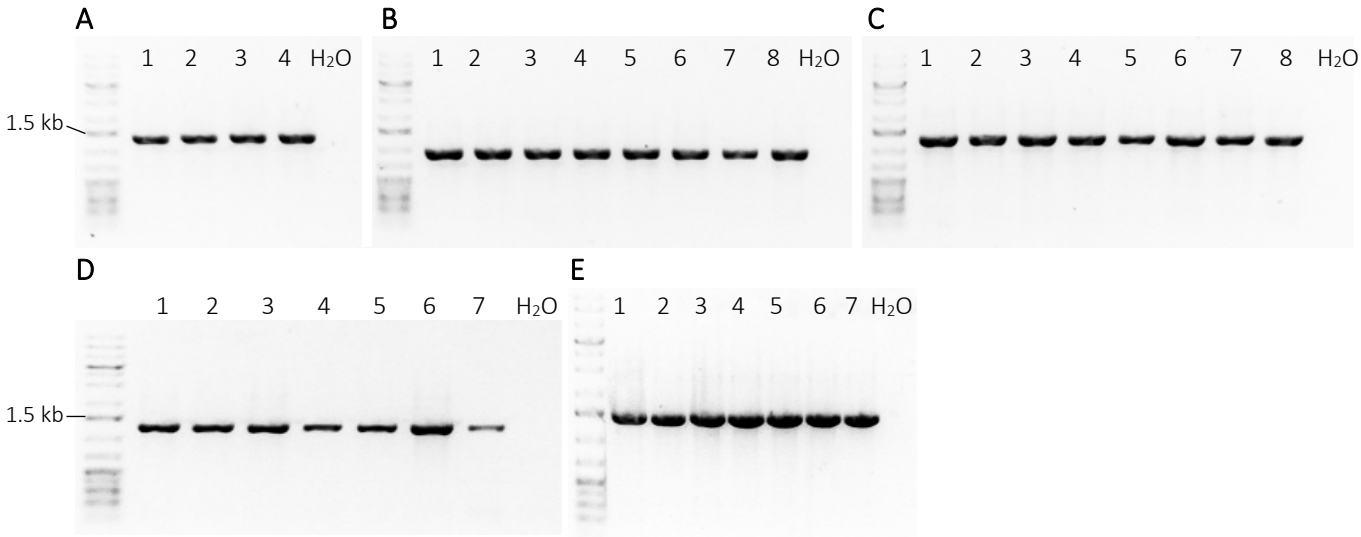


Figure 6. Verification of pAG305-AmCYP clones by colony PCR.

Colony PCR of *Escherichia coli* NEB10β colonies carrying pAG305-AmCYP plasmids (number of colonies tested: 4-8). pAG305-AmCYP plasmids and respective expected amplicons: (A) AmCYP1: ~1.3 kb (primers: pAGD_bw, AmCYP1_427bp_for, Ta= 52 °C); (B) AmCYP2: ~950 bp (primers: pAGD_fw, AmCYP2_858bp_rev, Ta= 55 °C); (C) AmCYP8, (D) AmCYP9 and (E) AmCYP6: ~1.2 kb (primers: pAGD_bw, AmCYP4_457bp_for ,Ta= 55 °C (C,D); AmCYP6 for_int, pAGD_bw, Ta= 54 °C (E)). Marker: Thermo Scientific™ GeneRuler™ 1 kb Plus DNA Ladder (Thermo Fisher Scientific, Waltham, USA) (See Figure 4 for size reference). H_2O- non-template control.

4.3. Analyses of heterologous expressed genes in yeast via GC-MS

To assess expression and functionality of heterologous *A. montana* genes in *S. cerevisiae*, the transformed yeast, along with control strains were cultivated. The expression of genes of interest, namely *AmGAS, AmGAO* and *AmCYP*s, inserted in yeast genome under control of the inducible GAL1 promotor was activated by addition of galactose to cultivation medium (see *3.2.2.2 Cultivation for galactose-induced expression*). The liquid cultures were extracted using different solvents to obtain fractions of interest and to test their efficiency when extracting potential novel compounds.

In GC-MS analysis, compounds of interest were identified by specific ion mass/charge ratio and similarity search to compounds data collected in mass spectra library. Due to heat-induced rearrangement compounds detected were β-elemene as germacrene A isomer, a product of AmGAS expression, and patchoulane as germacrene A acid isomer, a product of

AmGAO expression (Bröker et al., 2020; personal communication - Boje Müller, AG Prüfer) (*3.2.2.5 GC-MS analyses of yeast culture extracts*) .

4.3.1. Verification of *AmGAS* and *AmGAO* transformants

S. cerevisiae JNB strains were first tested by colony PCR to ensure they carry the genes of interest, *AmGAS* and *AmGAO,* in their genomes. Colonies with confirmed insertion of the genes by PCR were then cultivated in 50 mL shaking flask batch cultures to assess the expression and function of these genes in *S. cerevisiae* expression system. Fractions extracted from the yeast cultures were analysed by GC-MS at ion mass/charge ratio corresponding to germacrene A as β-elemene at m/z 93 and germacrene A acid, as its isomer patchoulane at m/z 81, specific retention times and mass spectra. Farnesol content in the cultures was also measured at m/z 69, to assess efficiency of strategies used to redirect FPP flux towards SLs synthesis (data not shown). Colony expressing both *AmGAS* and *AmGAO* with highest content of compounds of interest was further used for transformation with *AmCYP* genes.

Strain JNB586 is a control strain of modified yeast that does not contain transgenes from sesquiterpene lactone biosynthesis pathway but expresses a NADPH-reductase from *Arabidopsis thaliana*, a redox partner for cytochrome P450 enzymes. As expected, no β-elemene was detected in JNB586 cultures' extracts (Figure 7 A). Relatively high content of β-elemene, that greatly varies amongst different colonies, was detected in JNB621, which is based on JNB586 additionally expressing AmGAS. Furthermore, JNB622 strains (based on JNB621 strain additionally expressing AmGAO) produced β-elemene in relatively high amount, with highest content in culture extracts of colony no.1. Colonies in which extracts β-elemene was detected expressed at least *AmGAS* gene from *A. montana* coding for germacrene A synthase which successfully catalyses germacrene A synthesis from farnesyl diphosphate.

Germacrene A acid (GAA) isomer, patchoulane, was detected in JNB622 colonies no. 1 and 2 (Figure 7 B). This strain additionally expresses *AmGAO* gene showing that *A. montana* GAO catalyses oxidation of germacrene A to GAA. Among the JNB622 colonies, colony no. 1 produced GAA detected as its isomer patchoulane in highest amounts and exhibited good growth rate (assessed by OD_{600} measurement, data not shown) in cultivation conditions tested. Thus, JNB 622-1 proved to be an optimal host for heterologous expression of *AmCYPs* among the ones tested and was used for further experiments.

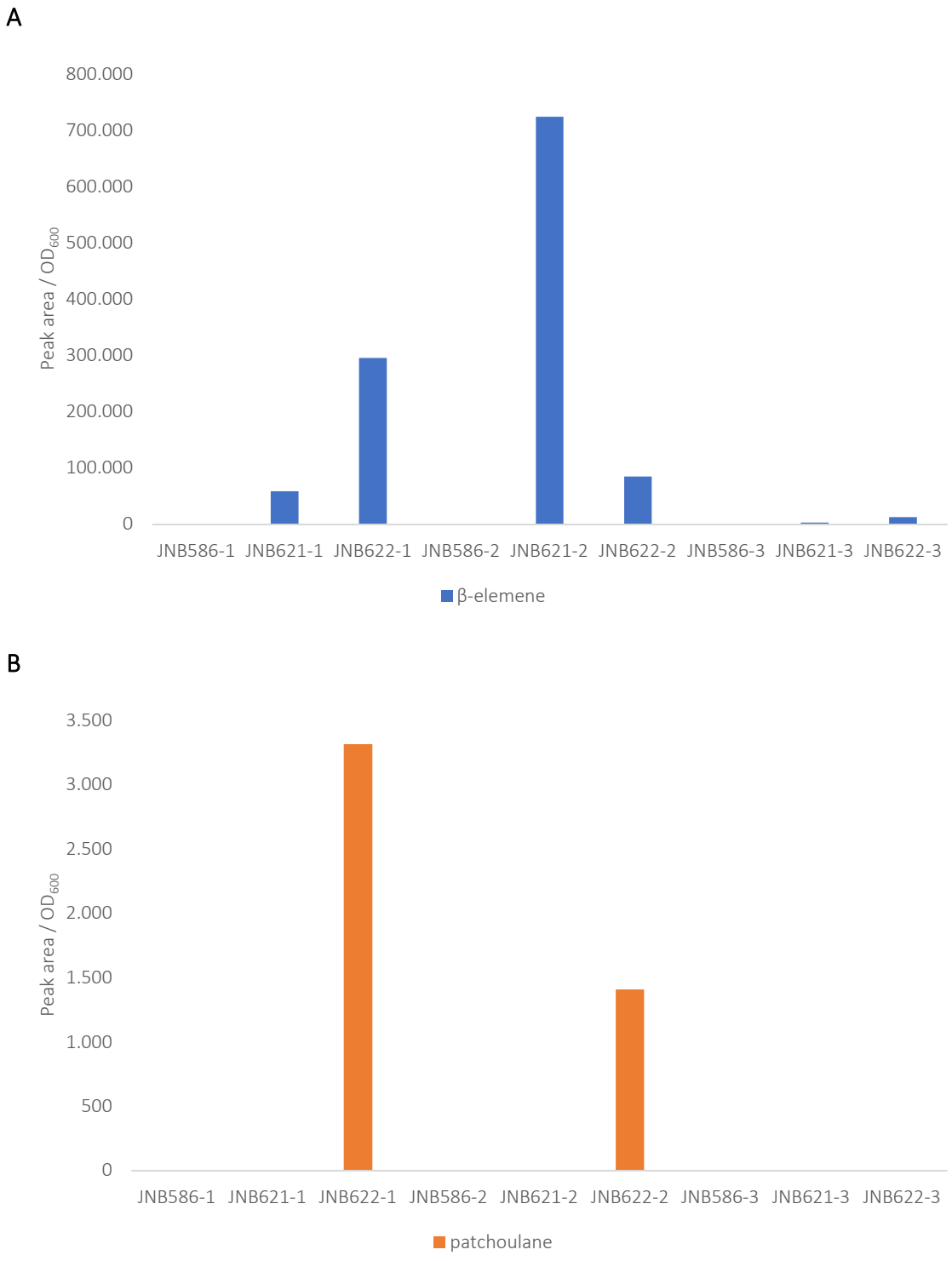


Figure 7. β-elemene and patchoulane content obtained by GC-MS analysis of *Saccharomyces cerevisiae* liquid culture extract after cultivation with induced expression of transgenes from *Arnica montana*.
Depicted are peak areas representing detected germacrene A isomer, β-elemene (A) and germacrene A acid isomer, patchoulane (B) corrected by the respective extracted culture OD_{600} values for three colonies of each tested strain. JNB586: control strain; JNB621: expressing *AmGAS* transgene; JNB622: expressing *AmGAS* and *AmGAO* transgenes. Numbers next to the strain names correspond to respective tested colony.

4.3.2. Verification of *AmCYP* transformants

Saccharomyces cerevisiae strains JNB621-1 and JNB622-1, proven to produce germacrene A and GAA, respectively, were transformed with linearized pAG305 expression vector carrying *AmCYP1*, *AmCYP2*, *AmCYP5*, *AmCYP6*, *AmCYP7, AmCYP8* and *AmCYP9* genes using lithium acetate transformation method. Two rounds of transformation did not result in transgenic yeasts with integrated pAG305-CYP vector. After testing of multiple restriction enzymes for vector linearisation and mock transformation with closed plasmid the protocol was optimized. Use of convenient restriction enzymes and prolonged incubation time from 7 to 20 minutes eliminated transformation difficulties (*3.2.2.1 Lithium acetate transformation;* Table 3). The transformants were first tested by colony PCR for integration of the pAG305-AmCYPs expression vector in the yeast genome (Figure 8).

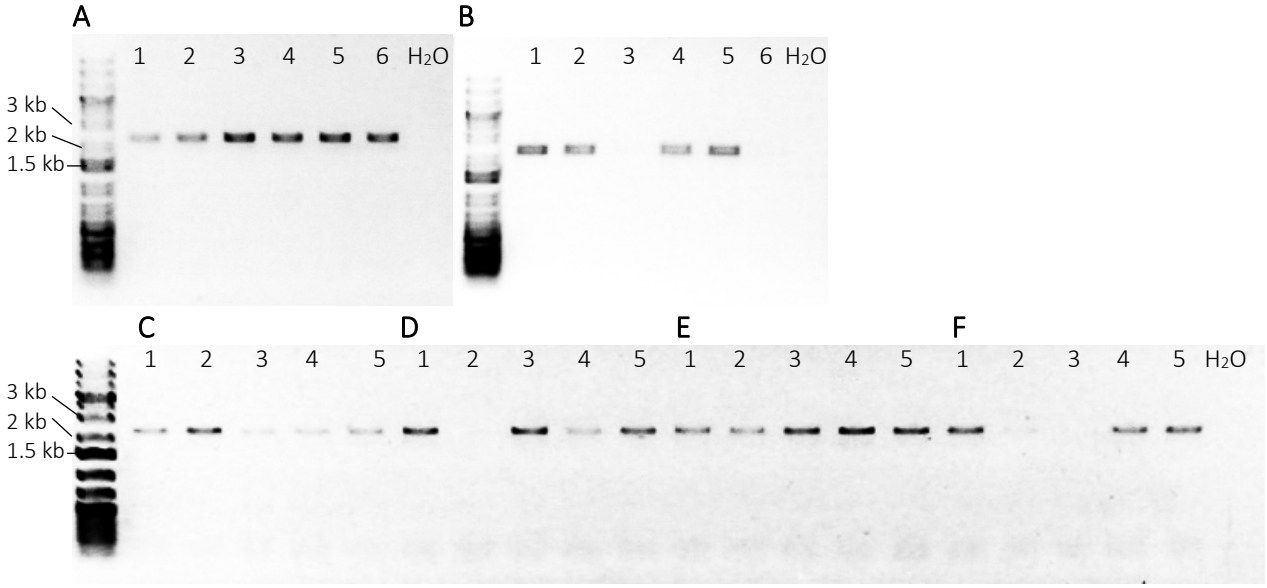


Figure 8. Verification of integration of pAG305-AmCYPs in *Saccharomyces cerevisiae* genome by colony PCR. Colony PCR of NM yeast colonies based on JNB621-1 and JNB622-1 for confirming integration pAG305-AmCYP plasmids (number of colonies tested: 5-6). Expected amplicons size: ~2.3 kb (primers: 5´LEU for yeast, pAG 3´LEU, Ta=61 °C). NM strains carrying pAG305-AmCYP plasmids and respective base strains: (A) NM13: AmCYP7 in JNB621-1); (B) NM14: AmCYP7 in JNB622-1; (C) NM15: AmCYP1 in JNB621-1, (D) NM16: AmCYP1 in JNB622-1; (E) NM17: AmCYP2 in JNB621-1; (F) NM18: AmCYP2 in JNB622-1. Marker: Thermo Scientific™ GeneRuler™ 1 kb Plus DNA Ladder (Thermo Fisher Scientific, Waltham, USA) (See Figure 4 for size reference). H_2O- non-template control.

Confirmed yeast strains carrying *AmGAS* and/or *AmGAO* and *AmCYP* genes are named NM followed by respective number. To detect potential conversion of GAA to inunolide by *AmCYP*s, the strains were cultivated using different culture volumes and extraction solvents, and extracts were analysed via GC-MS. For GC-MS analysis of potential novel compounds, fractions extracted from the yeast cultures were analysed at ion mass/charge ratio corresponding to inunolide and costunolide at m/z 232 and hydroxyl GAAs at m/z 250. External costunolide standard was used, as the compound is of similar structure and properties to expected inunolide. Costunolide standard dissolved in methanol but did not dissolve in hexane and was detected at retention time around 12.6 minutes when analysed in parallel with samples that were resolved in methanol.

The tested strains with corresponding transgenes were NM13 (*AmGAS, AmCYP7*), NM14 (*AmGAS, AmGAO, AmCYP7*), NM15 (*AmGAS, AmCYP1*), NM16 (*AmGAS, AmGAO, AmCYP1*), NM17 (*AmGAS, AmCYP2*), and NM18 (*AmGAS, AmGAO, AmCYP2*). Other *AmCYP* transformants were not cultivated and tested for compounds detection in the scope of this thesis due to time limitations. The tested strains were first cultivated in 50 mL shaking flask batch cultures, extracted using ethyl acetate and finally dissolved in methanol (for LC-MS analysis) as described in literature. However, due to malfunction of the LC-MS equipment and lack of time, LC-MS analyses were not performed in the scope of my master's thesis. The samples dissolved in methanol were analysed by GC-MS, although we were aware that this method is not optimal for separation and detection of expected novel compounds.

Extraction with ethyl acetate and dissolving in methanol resulted in detection of smaller peak height and area for β-elemene in all tested NM strains and patchoulane in NM strains expressing GAO, GAS and CYPs when compared to hexane mediated extraction of the same cultures using GC-MS (data not shown). However, since only ß-elemene was detected in strains NM13, NM15, NM17 (expressing GAS and CYP) these experiments indicated that no oxidation of germacrene A to GAA is catalysed by the enzyme encoded by *AmCYP* genes. Detected patchoulane in strains NM14, NM16 and NM18 (expressing GAS, GAO and CYP) confirmed synthesis of GAA is catalysed by the enzyme encoded by *AmGAO* in presence of *AmCYPs.* Additionally, novel compounds, including expected inunolide, were not detected in strains expressing *AmCYP* genes, when compared to detected costunolide external standard peak.

The experiments were proceeded with cultivation of smaller number of strains in larger culture volumes and combinations of hexane or ethyl acetate as extraction and hexane and methanol as second solvents suitable for GC-MS and LC-MS analysis. To acquire higher compound content opposed to the low contents detected via GC-MS when using ethyl acetate and methanol solvent combination, larger yeast culture volumes were tested. Different solvents combinations were tested to compare chromatograms and compounds detected, in aim for searching a solution for more efficient detection of potential novel compounds via GC-MS (described in *3.2.2.4 Extraction of potential novel compounds*). In continuation, a fewer number of strains with higher potential were tested due to complexity of cultivation of larger volume cultures.

Amongst the *AmCYP* candidate genes, *AmCYP7* gene has the most similar sequence to genes coding for germacrene A acid converting cytochrome p450 enzyme in *Inula hupehensis* and is potentially the most promising candidate for encoding an enzyme that catalyses conversion of GAA to inunolide in Arnica (personal communication - Lena Grundmann, AG Prüfer). Hence, *S. cerevisiae* transformants carrying the *AmCYP7* gene, strain NM14, were chosen for cultivation and further analysis. Due to inconsistency of GC-MS data gathered by measuring samples with different combination of extraction and final solvents used and no detection of novel compounds compared to analysis of *AmGAS* and *AmGAO* transformants (data not shown), only data where hexane was used in both cases is shown (Figure 9 A and B).

Cultivation of JNB622-1 and NM14 (three colonies) 500 mL cultures, hexane extraction and GC-MS analysis of the samples extracted with and dissolved in hexane were performed in parallel. In the experiments, no further peaks corresponding to potential novel compounds were detected in extracts of NM14 strain compared to extracts of the control strain JNB622-1. Germacrene A isomer in form of β-elemene was detected in all the cultures extracts, but in lower amounts in NM14 cultures compared to JNB622-1 control strain expressing only *AmGAS* and *AmGAO* (Figure 9 A). The lowest β-elemene content was observed for NM14-1 extract, lower by half compared to JNB622-1 culture extract.

Germacrene A acid isomer, patchoulane was detected in both JNB622-1 and NM14 cultures extracts (Figure 9 B). Content of patchoulane was distinctively, up to three times lower in yeast cultures expressing *AmCYP7*, as shown for NM14-2, compared to control strain expressing only *AmGAS* and *AmGAO.*

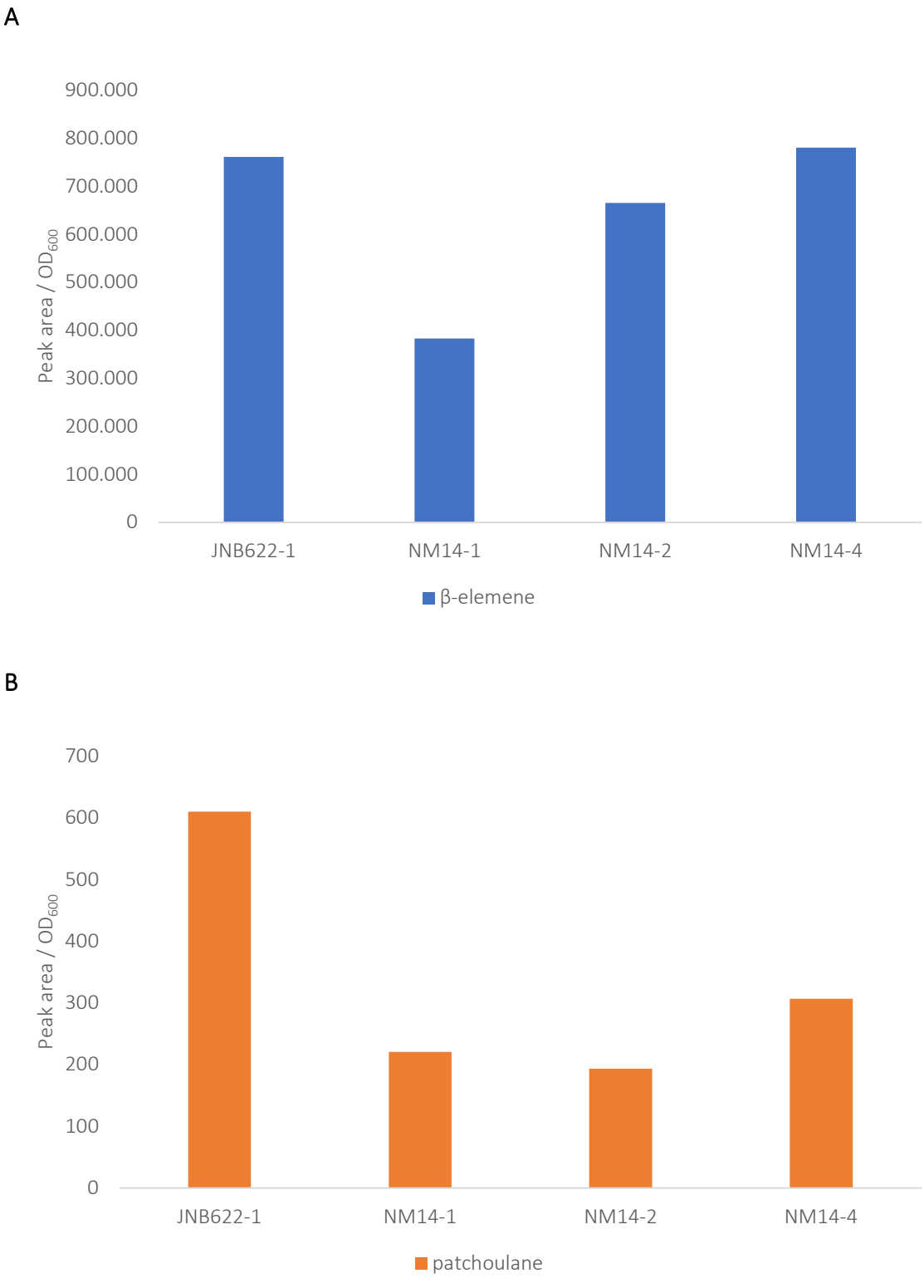


Figure 9. β-elemene (A) and patchoulane (B) content obtained by GC-MS analysis of *Saccharomyces cerevisiae* liquid culture extract after cultivation with induced expression of transgenes from *Arnica montana*. Depicted are peak areas representing detected β-elemene at m/z 93 (A) and patchoulane at m/z 81 (B) corrected by the corresponding culture OD_{600}. JNB622: expressing *AmGAS* and *AmGAO* transgenes; NM14: expressing *AmGAS*, *AmGAO* and *AmCYP7* transgenes.

4.4. Analyses of compound content and gene expression in *Arnica montana*

To further characterise *AmGAO, AmGAS* and *AmCYP* genes and compare their expression patterns with bioactive compounds content in *A. montana* tissues, quantitative PCR was performed, and bioactive SLs extracted and analysed by GC-MS. Preliminary expression analysis results of greenhouse grown AMA variety flowerheads and leaves collected at the same timepoint revealed *AmGAS, AmGAO* and *AmCYP7* were exclusively expressed in flowerheads (internally communicated results – Lena Grundmann, AG Prüfer). Further task was to examine if the expression patterns in AMM variety tissues correspond with findings in AMA and perform the analysis in parallel.

For that purpose, plant tissue was obtained from *A. montana* of both ssp. *montana* and ssp. *atlantica* cultivated in greenhouse and tissue culture conditions. Flowerheads showing optimal level of maturity for obtaining highest amount of bioactive substances of two greenhouse grown plants were collected. Hence, the number of flowerheads differed: three flowerheads from each plant of AMM variety and four and one flowerheads from plants of AMA variety were harvested. To check expression in leaves after flowering during vegetative growth, their leaves were collected four months after flowering, but only from one plant of each variety since the other two plants had already reached the stage of winter dormancy with no vital leaves. Plants grown in tissue culture (TC) conditions are non-flowering plants, hence only their young, developing leaves were collected. The tissue was processed as described in *3.2.4 Plant material*. In the following chapters genes *AmCYP8* and *AmCYP9* are referred to as *AmCYP8/9* due to high similarity in their sequence and amplification using the same primer pair in PCR. Expression of *AmGAS, AmGAO, AmCYP1, AmCYP2, AmCYP5, AmCYP6, AmCYP7, AmCYP8/9* genes was normalized to the reference genes *AmEF1a* and *AmETIF5* (see *3.2.1.8 Quantitative real-time PCR*).

4.4.1. Expression analysis of *AmGAS*, *AmGAO* and *AmCYPs* in *A. montana*

First, relative expression of *AmGAS, AmGAO, AmCYP1, AmCYP2, AmCYP5, AmCYP6, AmCYP7, AmCYP8/9* genes in *A. montana* flowerheads was analysed (Figure 10 A). In both AMM and AMA varieties' flowerheads, *AmCYP7* showed the highest relative expression and *AmCYP2* the lowest along with *AmCYP6* gene in AMM. Variations in expression patterns included higher expression levels of *AmGAS*, *AmGAO, AmCYP5* and *AmCYP6* genes in AMA variety, with the most pronounced difference observed in the expression levels of *AmCYP6* gene. *AmCYP8/9* had

a higher expression level in AMM flowerheads, while *AmCYP1, AmCYP2* and *AmCYP7* expression levels did not differ in the varieties. Relatively low calculated confidence interval (95%) values represented by error bars indicate the Cq value or expression levels among three technical replicates and two biological replicates is consistent.

In contrast to preliminary expression analysis of AMA leaves collected at the timepoint of flowering, leaves of greenhouse grown AMM and AMA four months after flowering were tested. Expression patterns of genes of interest in leaves of greenhouse grown AMM and AMA plants exhibited a larger dissonance between the two varieties (Figure 10 B). No expression of *AmGAS* and *AmCYP7* genes was detected in the samples. Expression level of *AmGAO* is very low, at least three hundred times lower, in AMM compared to flowerheads and not detectable in AMA leaves. These results correspond to preliminary results where no expression of *AmGAS, AmGAO* and *AmCYP7* was detected in AMA leaves collected at the timepoint of flowering. Relative expression levels of *AmCYP1* and *AmCYP2* are higher in AMM samples while expression levels of *AmCYP8/9*, *AmCYP5* and *AmCYP6* genes are higher in AMA variety. Error bars size representing 95% confidence interval indicate that Cq values measured in three technical replicates are very much alike.

Tested leaf tissue collected from plants cultivated in greenhouse conditions were mature at point of harvesting and no freshly developing leaves were available for analysis. To assess expression pattern in young, developing leaves, they were harvested from plants cultivated in tissue culture conditions. Tissue culture grown plants were readily available and more sustainable choice for acquiring the young leaf tissue we wanted to test. For this purpose, young leaves of nine plants of each variety were harvested and processed.

Relative expression of *AmGAS, AmGAO, AmCYP1, AmCYP2, AmCYP8/9, AmCYP5, AmCYP6* and *AmCYP7* genes in leaves of AMM and AMA plants cultivated in tissue culture is depicted in Figure 10 C. It is important to note that the expression analysis was performed using cDNA samples generated from pooled RNA from nine individual plants. Generally, variation of expression levels amongst the varieties was lower than in tissue from plants cultivated in greenhouse conditions. Expression levels of *AmCYP6* and *AmCYP7* genes did not differ amongst the varieties. Expression levels of *AmGAO* and *AmCYP1* were higher in AMM, while *AmCYP2*, *AmCYP8/9* and *AmCYP5* genes expression levels were distinctively higher in

AMA variety TC plant leaves. Consistency in pipetting of three technical replicates is indicated by low values of 95% confidence interval represented by error bars.

Figure 10. Gene expression patterns in different tissues of *Arnica montana* grown under different conditions. Relative expression levels of *AmGAS, AmGAO, AmCYP1, AmCYP2, AmCYP8/9, AmCYP5, AmCYP6* and *AmCYP7* genes in AMM and AMA flowerheads (A), leaves from greenhouse grown plants four months after flowering (B) and leaves from tissue culture (TC) grown plants (C) are shown (ΔCq values normalized to reference genes *AmEF1a* and *AmETIF5*) as log10-transformed mean values of three technical replicates (consisting of cDNA from one plant flowerheads and calculated with two biological replicates (A), from one plant leaves (B) and from leaves of 9 individual plants (C)). Error bars represent the 95% confidence interval.

In summary, when comparing gene expression patterns in *A. montana* flowerheads, leaves from greenhouse or tissue culture grown plants (Figure 10), there are multiple distinctions worth pointing out. Expression of *AmGAS* and *AmCYP7* genes differed greatly in different tissues. They were expressed at high levels in flowerheads and in TC leaves while no *AmGAS* or *AmCYP7* expression was detected in leaves of Arnica plants grown in greenhouse conditions. This is similar to the expression of *AmGAO*, but it was expressed at very low levels in greenhouse grown AMM leaves. Expression levels of other *AmCYP*s in the compared tissues did not differ distinctly.

4.4.2. Bioactive compounds extraction and analysis

A. montana ssp. *montana* and ssp. *atlantica* tissue from plants growing in greenhouse and tissue culture conditions was used for extraction of bioactive SLs fractions and GC-MS analysis (*3.2.4.2 Helenalin and dihydrohelenalin extraction for GC-MS analysis*). Lyophilized tissue of greenhouse grown plants' flowerheads and tissue culture plants' leaves (same as for gene expression analysis) was used. Compound extraction and GC-MS analysis of AMM and AMA leaves from flowering plants grown under greenhouse conditions were kindly provided as reference data (Figure 13).

Fractions extracted from the plant tissue were analysed by GC-MS at ion mass/charge ratio corresponding to helenalin and its derivatives at m/z 244 and dihydrohelenalin and its derivatives at m/z 246, specific retention times and mass spectra (*3.2.4.3 GC-MS analysis of A. montana flowerheads and leaves extracts*).

Ion abundance chromatograms at the sesquiterpene lactones' specific mass-to-charge ratios obtained by GC-MS analysis of AMM and AMA flowerheads were distinctively different. In AMM flowerheads, overall lower ion count was measured. Looking at the SLs content, helenalin derivatives were dominant and present in larger amounts represented by the peak heights, than dihydrohelenalin derivatives (Figure 11 A).

In AMA flowerheads extract GC-MS analysis ion abundance peaks representing helenalin derivatives were low and unsharp indicating low amounts of the compounds (Figure 11 B). Dominant peaks corresponded to dihydrohelenalin derivatives and were present in high amounts in the AMA variety flowerheads judging by the peaks' height and area.

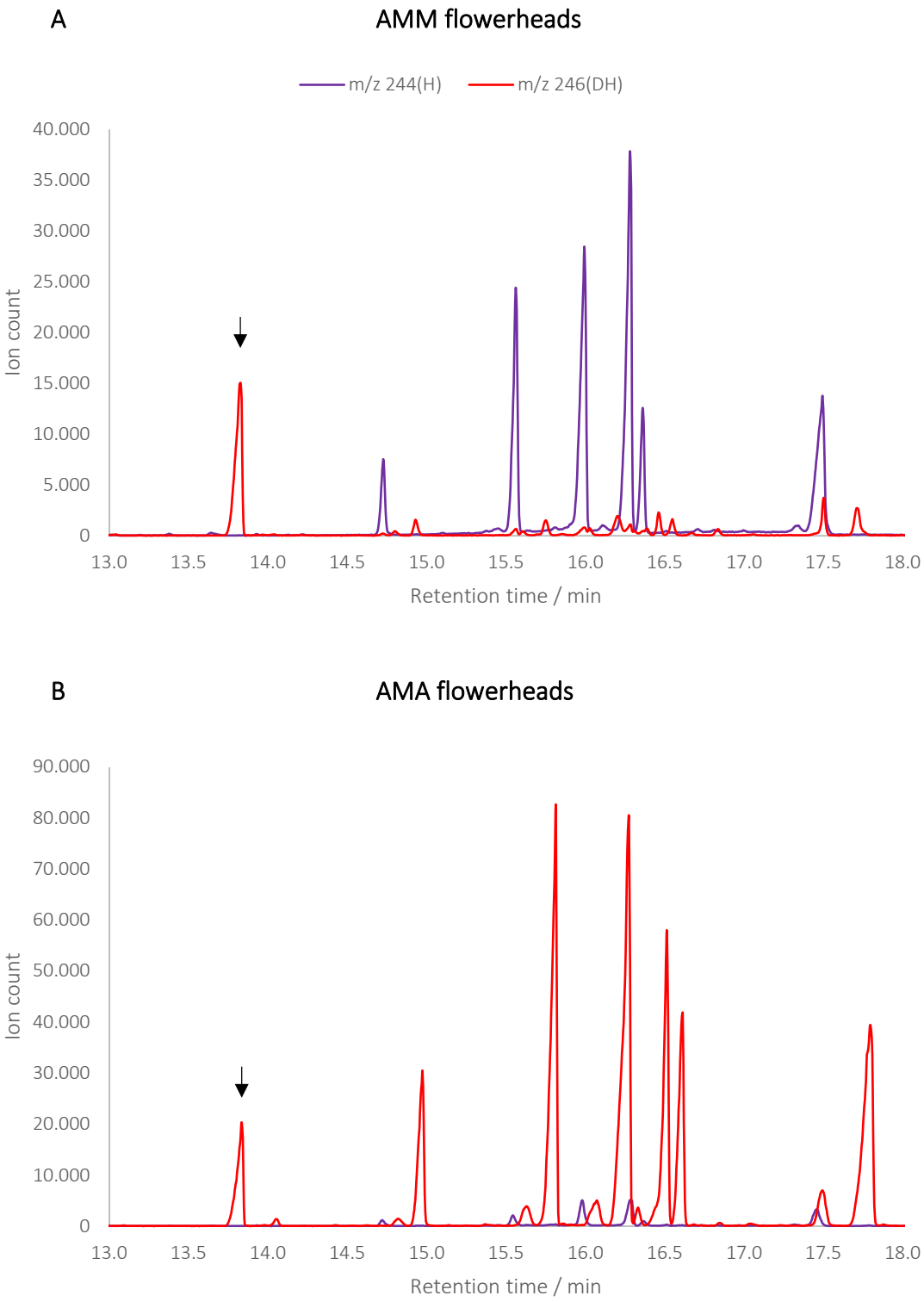


Figure 11. Ion abundance chromatogram from GC–MS analysis of sesquiterpene lactones helenalin (H, m/z 244) and dihydrohelenalin (DH, m/z 246) from *Arnica montana* ssp. *montana* (A) and *atlantica* (B) flowerheads from two plants grown in greenhouse conditions. Internal standard α-santonin peak (m/z 246) is marked by an arrow.

Ion abundance chromatograms acquired by GC-MS analysis of *A. montana* SLs in leaves of both varieties of tissue culture grown plants are represented by Figure 12. In contrast to distinct differences of compounds detected in AMM and AMA flowerheads, dihydrohelenalin derivatives were dominant and present in large amount in both AMM and AMA TC leaves. In AMM TC leaves extract chromatogram, two sharp peaks of helenalin were present at retention time of 14.75 and 16.00 minutes, while in the AMA chromatograms, they were present in much lower amount. Overall, dihydrohelenalin derivatives were dominantly detected in both varieties' extracts and were detected in AMA TC plants in high amounts presented by height and area of the peaks at ions m/z 246.

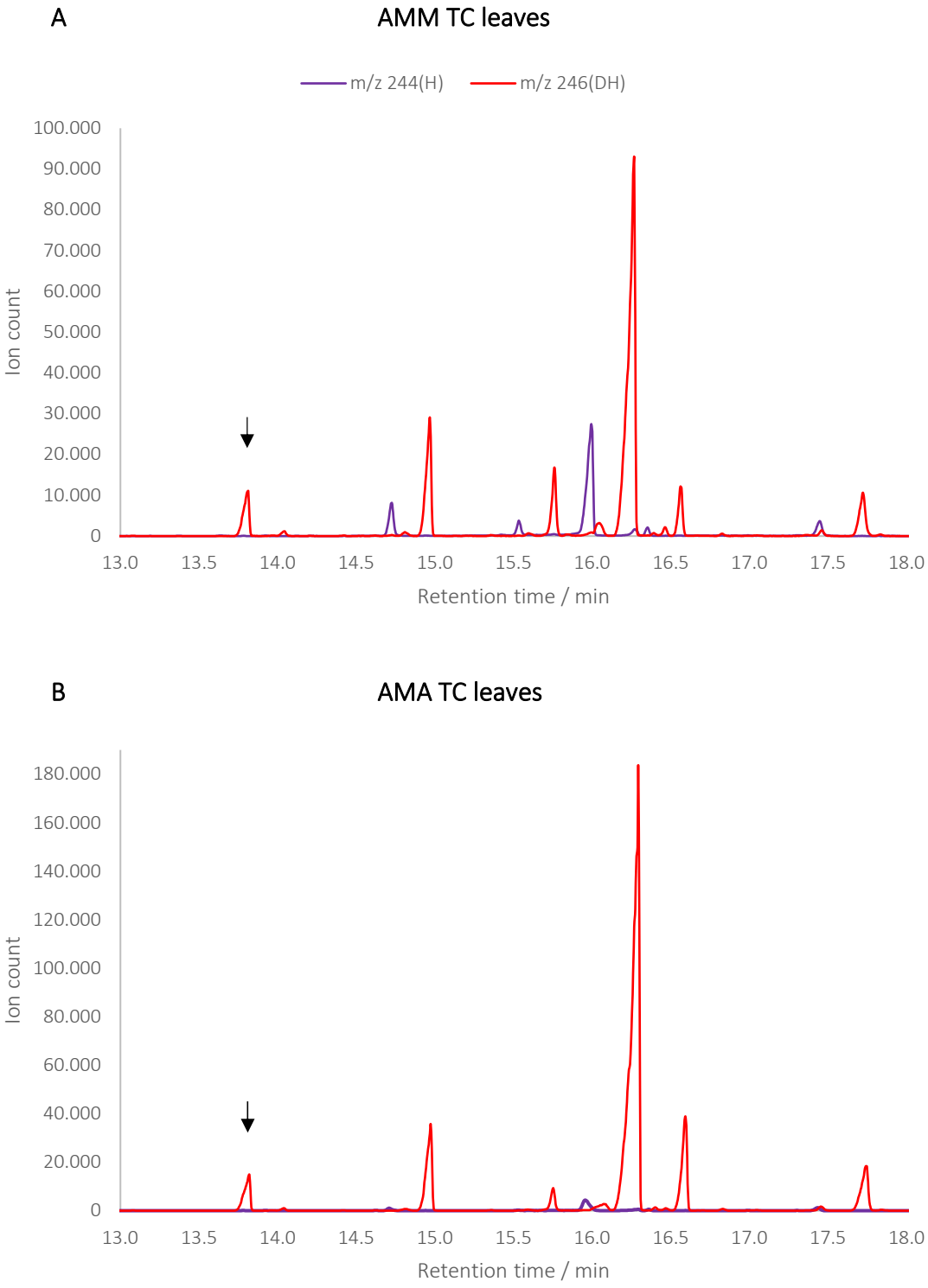


Figure 12. Ion abundance chromatogram from GC–MS analysis of sesquiterpene lactones helenalin (H, m/z 244) and dihydrohelenalin (DH, m/z 246) from *Arnica montana* ssp. *montana* (A) and *atlantica* (B) leaf extracts from nine plants grown in tissue culture conditions. Internal standard α-santonin peak (m/z 246) is marked by an arrow.

Based on the previously described ion abundance chromatograms (Figures 11 and 12), Helenalin and dihydrohelenalin derivatives content was calculated by mass of the compounds per gram of dry tissue using known concentration of internal α-santonin standard for quantification (Figure 13). Helenalin and its derivatives were present in AMM flowerheads in the highest amount of 9.86 mg/g and in AMM TC leaves at 6.08 mg/g dry tissue weight. Content of dihydrohelenalin and its derivates peaked in AMA TC leaves at 36.64 mg/g, followed by 26.70 mg/g in AMA flowerheads. In AMM and AMA flowerheads, dihydrohelenalin derivatives and helenalin derivatives content was as low as 2.68 and 2.42 mg/g, respectively. The content of both SLs is relatively low in AMM and AMA greenhouse grown leaves compared to other tissue samples with higher DH derivates levels than H derivates levels in both varieties.

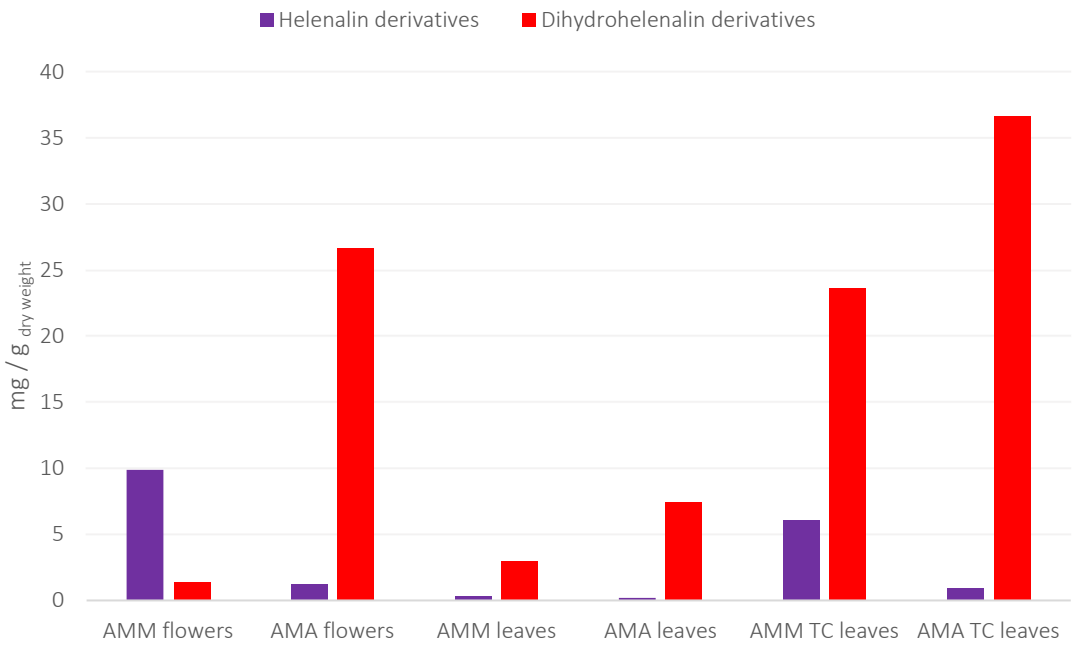


Figure 13. Helenalin derivates and dihydrohelenalin derivatives content in different tissue and cultivation conditions. *A. montana* ssp. *montana* (AMM) and ssp. *atlantica* (AMA) grown in greenhouse and in tissue culture (TC) conditions.
Depicted is content of the compounds in µg per g of dry tissue calculated in relation to α-santonin internal standard.

5. Discussion

In recent years, alternative production of plant natural product as pharmaceutically interesting substances has become focus of many researchers. High yield production of SL artemisinin from *A. annua* in yeast heterologous system described by Paddon & Keasling (2014) is a promising example. However, to develop such a SLs production platform, various factors, including identification and characterisation of biosynthesis pathway genes, need to be elucidated and optimised. *A. montana* is an endangered species and alternative, sustainable production methods of its bioactive SLs is a good step to take towards its wild populations' preservation.

To my best knowledge, enzymes involved in sesquiterpene lactone biosynthesis pathway, as well as localisation of the pathway in *Arnica montana* tissue have not been yet described.

In the first part of this research, the focus was on the functional analysis of *AmGAS, AmGAO* and *AmCYP* candidate genes for biosynthesis of sesquiterpene lactones in *Arnica montana.* The genes were tested in heterologous expression system optimised for production of sesquiterpenes: metabolically engineered *Saccharomyces cerevisiae*. Different methods of extraction of compounds produced by enzymes encoded by transgenes were assessed.

In the second part, expression patterns of genes potentially involved in SLs biosynthesis were correlated with SLs accumulation in different *Arnica montana* tissues.

5.1 Heterologous expression of *Arnica montana* genes in *Saccharomyces cerevisiae*

The function of *AmGAS* and *AmGAO* genes was verified in metabolically engineered *S. cerevisiae* heterologous system. When cultivated in previously described conditions in modified yeast strains, germacrene A synthase from *A. montana* encoded by *AmGAS* used farnesyl diphosphate as substrate and produced germacrene A which was detected via GC-MS as its isomer β-elemene (Figure 7 A). Subsequently, in strains expressing both *AmGAS* and *AmGAO* genes, germacrene A acid isomer, patchoulane, was detected (Figure 7 B). This indicates *AmGAO* encodes a functional enzyme that oxidises germacrene A to GAA in yeast. Thus, this is the first functional analysis of *GAS* and *GAO* genes from *Arnica montana* providing a good basis for expression in heterologous yeast system and further characterisation of candidate genes of SLs biosynthesis pathway in Arnica. The reactions and genes encoding the enzymes are conserved in Asteraceae, and were previously described in a number species

including Helianthus *annuus*, *Lactuca sativa* and *Barnadesia spinosa* (Ikezawa et al., 2011; Nguyen et al., 2010).

Cytochrome P450 enzymes (CYPs) that convert GAA as the next step in sesquiterpene lactone synthesis producing costunolide and inunolide in Asteraceae species, have been in focus in the past years. Frey et al. (2020) recently identified and characterised costunolide synthase from *Helianthus annuus* (HaCOS). HaCOS catalyses hydroxylation of GAA to costunolide. Gou et al. (2018) described cytochrome P450 (CYP71BL6) that catalyses hydroxylation of GAA in *Inula hupehensis* into 8α-hydroxy GAA that spontaneously converts to inunolide. In this work, the analysis of a total of seven AmCYPs was prepared by cloning the corresponding coding sequences of *AmCYPs* into yeast transformation vectors and successfully transforming seven CYP constructs into GAS/GAO expressing yeast strain JNB622. However, due to technical problems with LC-MS instrument, the analyses could not be performed as planned and it was necessary to switch to GC-MS. This was accompanied by various problems. For LC-MS, samples should be extracted with ethyl acetate and redissolved in methanol. However, measurement of the samples by GC-MS did not yield any additional peaks, furthermore the already confirmed presence of germacrene A or GAA could not be verified. Reason for this probably lies in nature of the solvents. Used solvents, arranged in order from non-polar to very polar are hexane, ethyl acetate (acetic acid ethyl ester) and methanol. Germacrene A and GAA and have low polarity making them soluble in non-polar hexane and non-soluble in more polar methanol. Expected novel compounds, inunolide and/or 8α-hydroxy GAA have hydroxyl and lactone (ester) functional groups which makes them more polar, hence more soluble in polar solvents such as ethyl acetate and methanol. Therefore, for comparison, the samples were extracted and redissolved with/in hexane and analysed by GC-MS. This showed that again no GAA oxidation was detected but a lower germacrene A and GAA isomers content in NM14 yeast expressing *AmGAS, AmGAO* and *AmCYP7* transgenes compared to JNB622-1 expressing only GAS and GAO (Figure 9). These lower contents might indicate that in NM14 strain GAA is further converted. However, this can only be speculated, but *AmCYP7* is probably a good candidate gene considering its sequence similarity to *I. hupehensis CYP71BL6*, proven to convert GAA into either 8α- or 8β-hydroxy GAA*,* and positive correlation of the transcript abundance and SLs accumulation in *A. montana* tissues (Figures 10 and 13). However, lower content of germacrene A and GAA isomers could also imply lower efficiency of

GAS and GAO enzymes in AmCYP7 expressing strain. Since no novel products were detected in strains expressing *AmCYP* genes, the approach for detection of compounds to verify gene function should be modified for future analysis as discussed in the following.

The modified *S. cerevisiae* JNB622-1 host strain expresses *A. thaliana* NADPH-cytochrome P450 oxidoreductase gene (*AtR1*) to support enzymatic activity of Arnica's cytochromes P450 monooxygenases introduced into the yeast genome. It is possible that expression of natural redox partner, a NADPH-cytochrome P450 oxidoreductase from *A. montana,* along with the *AmGAS* and *AmCYP* genes could improve the enzymes activity in heterologous system (Frey et al., 2018)

Heterologous expression of *A. montana* genes in *Saccharomyces cerevisiae* could be optimized using various methods. One of them is codon optimisation for expression in yeast as one example that could lead to more efficient expression and folding of plant proteins in yeast cells (Frey et al., 2018). Another option is optimization of cultivation conditions for expression. Next, as showcased with original JNB yeast strains, employment of native yeast enzymes and pathways, and their modification for more efficient precursor accumulation and substrate redirection makes great base for functional analysis of heterologous genes, as well as for latter potential production of plant natural products.

Gas chromatography-mass spectrometry has limitations. To be able detect compounds by GC-MS, they must be volatile and thermally stable. Otherwise, the compounds might convert to thermally stable isomers at high temperatures applied in the process or simply not detected. Sesquiterpene lactones are mostly thermolabile and less volatile compounds which makes them harder to separate using gas chromatography. Possible solutions to this limitation is derivatization of the potential product of AmCYP7 activity thereby transforming it into a stable compound (Ivanescu et al., 2015). In Arnica plants the final compounds of SL biosynthesis, helenalin and dihydrohelenalin, are mostly present in form of their ester derivatives which makes them more volatile and detectable using GC-MS (Perry et al., 2009). Other solution might be to use method for separation and detection of less and non-volatile compounds such as liquid chromatography paired with mass spectrometry (LC-MS). Examples of successful detection of sesquiterpene lactones via LC-MS are costunolide, inunolide and many others (Frey et al., 2020; Gou et al., 2018). Physical properties of SLs might have interfered with separation and detection of novel compounds, possible precursors of helenalin and

dihydrohelenalin, such as inunolide via GC-MS in yeast culture performed in this work. LC-MS analysis could not be performed due to previously stated reasons.

An alternative expression host for sesquiterpene lactone biosynthesis genes from Arnica could be *Nicotiana benthamiana.* This model plant is an efficient platform for transient expression and *in vivo* characterisation of plant biosynthetic enzymes (Hansen et al., 2020). The molecular environment in plants in contrast to microbial expression systems such as *S. cerevisiae* may be better for expression of plant transgenes. Lipid composition of plant endoplasmic reticulum (ER) membrane differs from yeast ER composition which could affect function of membrane bound proteins such as cytochromes P450 enzymes. In addition, CYPs redox partner, NADPH-reductase are naturally expressed in plants while in yeast they are not. *N. benthamiana* expression system possesses plant chaperones that could improve folding of transgenes which could subsequently increase expression efficiency and protein function. Reconstruction of costunolide biosynthesis pathways and transient expression of the Asteraceae *Cichorium intybus* GAO, GAS and costunolide synthase in *N. benthamiana* leaves resulted in costunolide production of up to 60 ng/g (Liu et al., 2011).

5.2 Sesquiterpene lactones biosynthesis pathway candidate genes expression patterns and bioactive compounds content in *Arnica montana*

To study metabolic function of enzymes encoded by *AmGAS, AmGAO* and *AmCYP* genes *in planta,* bioactive sesquiterpene lactones content was analysed by GC-MS along with the gene expression patterns in different *A. montana* tissues. To assess differences, both AMM and AMA variety plants were used, grown either in greenhouse conditions or tissue culture conditions. Tissue of plants grown in greenhouse conditions was harvested at later developmental stages (flowerheads and leaves) while young leaves were harvested from plants grown in TC conditions.

SLs, helenalin and dihydrohelenalin content was the highest in tissue culture grown young leaves and flowerheads of greenhouse grown plants (Figure 13). This finding indicates that the compounds are possibly biosynthesized in these tissues. SLs content measured by GC-MS was low in leaves of greenhouse grown plants. Looking at the tested genes expression in the tissues, *AmGAS, AmGAO* and *AmCYP7* were poorly or not expressed in greenhouse grown plants' leaves samples (Figure 10 B). On the other hand, expression levels of *AmGAS, AmGAO*

and *AmCYP7* were high in flowerheads and TC leaves samples (Figure 10 A and C). This implicates involvement of these genes in biosynthesis of sesquiterpene lactones from Arnica. Similar correlation of SLs accumulation and gene candidate gene expression are reported by Gou et al. (2018) for *CYP71BL6* gene from *I. hupehensis*. The highest levels of *CYP71BL6* transcript were observed in leaves and flowers that corresponded *I. hupehensis* SLs, burrodin and 2α-acetoxy-inuviscolide, content detected by LC-MS/MS. Furthermore, they strengthened the correlation by treating *I. hupehensis* seedlings with plant signalling molecule, methyl jasmonate, which after 48h treatment showed significant increase of the SLs content as well as *CYP71BL6* transcript levels. This kind of experiment could also be conducted in future testing of SLs biosynthesis candidate genes from Arnica.

In Asteraceae, SLs biosynthesis takes place in various plant tissues and their abundance in one tissue does not necessarily mean their synthesis occurs in this tissue. For example, *Cichorium intybus* accumulates costunolide in the roots and *Helianthus annuus* SLs are highly concentrated in glandular trichomes but not all are produced in the trichomes (Liu et al., 2011; Spring et al., 2020). The expression patterns of tested genes along with results of SLs analysis in different tissues imply SLs biosynthesis pathway is active in leaves in early developmental stages and in Arnica flowerheads, but this should be further investigated.

Although there is a clear genetic distinction of AMM and AMA plants (Schmiderer et al., 2018), *AmCYP* sequences isolated from them are very similar (personal communication – Lena Grundmann, AG Prüfer). These subspecies, however, do exhibit differences in SLs content (Perry et al., 2009). Helenalin is dominant in AMM, and dihydrohelenalin in AMA variety, as confirmed by results previously described (Figure 13). This difference in SLs content is not so distinct in young, developing TC leaves. Furthermore, in young TC leaves, dihydrohelenalin is dominantly accumulated in both varieties. These results might indicate conversion of helenalin to dihydrohelenalin, hence reduction of methylene to methyl functional group, happens in early developmental stages in *A. montana* leaves. To speculate, this reduction reaction is probably catalysed by an enzyme which is possibly expressed in greater abundance in AMA compared to AMM plants, and in developing tissues compared to differentiated plant tissue. Moreover, methyl group makes dihydrohelenalin less reactive when compared to helenalin's methylene functional group, making dihydrohelenalin more stable. Hence, dihydrohelenalin might have a role as SLs storage form accumulated in leaves that can be converted to helenalin

in plants by an oxidation reaction when needed. These speculations would be interesting to examine in future research on *A. montana* SLs biosynthesis genes.

6. Conclusion

This study provided basis for further research of sesquiterpene lactones biosynthesis genes in *Arnica montana*.

Candidate genes functional analysis in *Saccharomyces cerevisiae* expression system was successful for *AmGAS* and *AmGAO* genes. These genes encode respective germacrene A synthase and germacrene A acid oxidase, and their enzymatic activity was confirmed in modified yeast expression host. Candidate genes *AmCYP*s were only functionally analysed to a certain extent, but candidate *AmCYP7* shows great potential for further investigation of its role in SLs biosynthesis pathway due to high homology with characterized CYPs such as *CYP71BL6*. *Saccharomyces cerevisiae* is a good host for heterologous expression of *A. montana* genes and elucidation of biosynthesis pathway of helenalin and dihydrohelenalin.

Analysis of potential candidate gene expression patterns by qPCR and sesquiterpene lactones derivatives content distribution in different plant tissues analysed by GC-MS were conducted. *AmCYP7* transcript abundance was positively correlated to bioactive SLs accumulation in *A. montana* flowerheads and young leaves. This correlation further strengthened *AmCYP7* role as a good candidate gene for *A. montana* SLs biosynthesis pathway. Helenalin is dominantly present in *A. montana* ssp. *montana* flowerheads, and dihydrohelenalin content is dominant in both varieties young, developing leaves and *A. montana* ssp. *atlantica* flowerheads providing preliminary evidence for their biosynthesis at different developmental stages during the plant´s life cycle.

7. References

Bröker, J. N., Müller, B., Prüfer, D., & Gronover, C. S. (2020). Combinatorial Metabolic Engineering in *Saccharomyces cerevisiae* for the Enhanced Production of the FPP-Derived Sesquiterpene Germacrene. *Bioengineering*, *7*(4), 1–13. https://doi.org/10.3390/bioengineering7040135

Douglas Kinghorn, A., & Seo, E. K. (1996). Plants as Sources of Drugs. *ACS Symposium Series*, 647, 179–193. https://doi.org/10.1021/bk-1996-0647.ch012

Drogosz, J., & Janecka, A. (2018). Helenalin - A Sesquiterpene Lactone with Multidirectional Activity. *Current Drug Targets*, *20*(4), 444–452. https://doi.org/10.2174/1389450119666181012125230

Frey, M. (2020). Traps and Pitfalls—Unspecific Reactions in Metabolic Engineering of Sesquiterpenoid Pathways. *Molecules*, *25*(8). https://doi.org/10.3390/MOLECULES25081935

Frey, M., Klaiber, I., Conrad, J., & Spring, O. (2020). CYP71BL9, the Missing Link in Costunolide Synthesis of Sunflower. *Phytochemistry*, *177*(May), 112430. https://doi.org/10.1016/j.phytochem.2020.112430

Frey, M., Schmauder, K., Pateraki, I., & Spring, O. (2018). Biosynthesis of Eupatolide - A Metabolic Route for Sesquiterpene Lactone Formation Involving the P450 Enzyme CYP71DD6. *ACS Chemical Biology*, *13*(6), 1536–1543. https://doi.org/10.1021/acschembio.8b00126

Gou, J., Hao, F., Huang, C., Kwon, M., Chen, F., Li, C., Liu, C., Ro, D. K., Tang, H., & Zhang, Y. (2018). Discovery of a Non-stereoselective Cytochrome P450 Catalyzing either 8α- or 8β-hydroxylation of Germacrene A Acid from the Chinese Medicinal Plant, *Inula hupehensis*. *Plant Journal*, *93*(1), 92–106. https://doi.org/10.1111/tpj.13760

Hamann, T., & Møller, B. L. (2007). Improved Cloning and Expression of Cytochrome P450s and Cytochrome P450 Reductase in Yeast. *Protein Expression and Purification*, *56*(1), 121–127. https://doi.org/10.1016/j.pep.2007.06.007

Hansen, N. L., Miettinen, K., Zhao, Y., Ignea, C., Andreadelli, A., Raadam, M. H., Makris, A. M., Møller, B. L., Stærk, D., Bak, S., & Kampranis, S. C. (2020). Integrating Pathway Elucidation with Yeast Engineering to Produce Polpunonic Acid the Precursor of the Anti-obesity Agent Celastrol. *Microbial Cell Factories*, *19*(1), 1–17. https://doi.org/10.1186/s12934-020-1284-9

Hausjell, J., Halbwirth, H., & Spadiut, O. (2018). Recombinant Production of Eukaryotic Cytochrome P450s in Microbial Cell Factories. *Bioscience Reports*, *38*(2), 20171290. https://doi.org/10.1042/BSR20171290

Hollmann, V., Donath, T. W., Grammel, F., Himmighofen, T., Zerahn, U., & Leyer, I. (2020). From Nutrients to Competition Processes: Habitat Specific Threats to *Arnica montana* L. Populations in Hesse, Germany. *PLoS ONE*, *15*(5). https://doi.org/10.1371/journal.pone.0233709

Ikezawa, N., Göpfert, J. C., Nguyen, D. T., Kim, S. U., O'Maille, P. E., Spring, O., & Ro, D. K. (2011). Lettuce Costunolide Synthase (CYP71BL2) and its Homolog (CYP71BL1) from Sunflower Catalyze Distinct Regio- and Stereoselective Hydroxylations in Sesquiterpene Lactone Metabolism. *Journal of Biological Chemistry*, *286*(24), 21601–21611. https://doi.org/10.1074/jbc.M110.216804

Ivanescu, B., Miron, A., & Corciova, A. (2015). Sesquiterpene Lactones from Artemisia Genus: Biological Activities and Methods of Analysis. *Journal of Analytical Methods in Chemistry*, *2015*. https://doi.org/10.1155/2015/247685

Klaas, C. A., Wagner, G., Laufer, S., Sosa, S., Della Loggia, R., Bomme, U., Pahl, H. L., & Merfort, I. (2002). Studies on the Anti-inflammatory Activity of Phytopharmaceuticals Prepared from Arnica Flowers. *Planta Medica*, *68*(5), 385–391. https://doi.org/10.1055/s-2002-32067

Kriplani, P., Guarve, K., & Baghael, U. S. (2017). *Arnica montana* L. – A Plant of Healing: Review. *Journal of Pharmacy and Pharmacology*, *69*(8), 925–945. Blackwell Publishing Ltd. https://doi.org/10.1111/jphp.12724

Kulagina, N., Besseau, S., Godon, C., Goldman, G. H., Papon, N., & Courdavault, V. (2021). Yeasts as Biopharmaceutical Production Platforms. *Frontiers in Fungal Biology*, *2*, 50. https://doi.org/10.3389/ffunb.2021.733492

Kumar Srivastava, A. (2018). Significance of Medicinal Plants in Human Life. *Synthesis of Medicinal Agents from Plants* (pp. 1–24). Elsevier Ltd. https://doi.org/10.1016/b978-0-08-102071-5.00001-5

Leoni, V., Borgonovo, G., Giupponi, L., Bassoli, A., Pedrali, D., Zuccolo, M., Rodari, A., & Giorgi, A. (2021). Comparing Wild and Cultivated *Arnica montana* L. from the Italian Alps to Explore the Possibility of Sustainable Production using Local Seeds. *Sustainability (Switzerland)*, *13*(6), 3382. https://doi.org/10.3390/su13063382

Liu, C. L., Xue, K., Yang, Y., Liu, X., Li, Y., Lee, T. S., Bai, Z., & Tan, T. (2021). Metabolic Engineering Strategies for Sesquiterpene Production in Microorganism. *Critical Reviews in Biotechnology*. Taylor & Francis. https://doi.org/10.1080/07388551.2021.1924112

Liu, Q., Majdi, M., Cankar, K., Goedbloed, M., Charnikhova, T., Verstappen, F. W. A., de Vos, R. C. H., Beekwilder, J., van der Krol, S., & Bouwmeester, H. J. (2011). Reconstitution of the Costunolide Biosynthetic Pathway in Yeast and *Nicotiana benthamiana*. *PLoS ONE*, *6*(8). https://doi.org/10.1371/journal.pone.0023255

Matos, M. S., Anastácio, J. D., & Dos Santos, C. N. (2021). Sesquiterpene Lactones: Promising Natural Compounds to Fight Inflammation. *Pharmaceutics 13*(7). Pharmaceutics. https://doi.org/10.3390/pharmaceutics13070991

Nguyen, D. T., Göpfert, J. C., Ikezawa, N., MacNevin, G., Kathiresan, M., Conrad, J., Spring, O., & Ro, D.-K. K. (2010). Biochemical Conservation and Evolution of Germacrene A Oxidase in Asteraceae. *Journal of Biological Chemistry*, *285*(22), 16588–16598. https://doi.org/10.1074/jbc.M110.111757

Paddon, C. J., & Keasling, J. D. (2014). Semi-synthetic Artemisinin: A Model for the use of Synthetic Biology in Pharmaceutical Development. *Nature Reviews Microbiology 12*(5), 355–367. Nature Publishing Group. https://doi.org/10.1038/nrmicro3240

Padilla-Gonzalez, G. F., dos Santos, F. A., & Da Costa, F. B. (2016). Sesquiterpene Lactones: More Than Protective Plant Compounds With High Toxicity. *Critical Reviews in Plant Sciences*, *35*(1), 18–37. https://doi.org/10.1080/07352689.2016.1145956

Perry, N. B., Burgess, E. J., Rodríguez Guitián, M. A., Franco, R. R., Mosquera, E. L., Smallfield, B. M., Joyce, N. I., & Littlejohn, R. P. (2009). Sesquiterpene Lactones in *Arnica montana*: Helenalin and Dihydrohelenalin Chemotypes in Spain. *Planta Medica*, *75*(6), 660–666. https://doi.org/10.1055/s-0029-1185362

Petrova, Maria, Zayova, E., Geneva, M., Dimitrova, L., Vitkova, A., & Stanilova, M. (2021). Multiplication and Conservation of Threatened Medicinal Plant *Arnica montana* L. by i*n vitro* Techniques. *Agriculturae Conspectus Scientificus*, *86*(1), 57–65.

Petrova, Mariya, Zayova, E., Vassilevska-Ivanova, R., & Vlahova, M. (2012). Biotechnological Approaches for Cultivation and Enhancement of Secondary Metabolites in A*rnica montana* L. *Acta Physiologiae Plantarum*, *34*(5), 1597–1606. https://doi.org/10.1007/s11738-012-0987-x

Petrova, Mariya, Zayova, E., & Vlahova, M. (2013). Induction of Hairy Roots in *Arnica montana* L. by *Agrobacterium rhizogenes*. *Central European Journal of Biology*, *8*(5), 470–479. https://doi.org/10.2478/s11535-013-0157-6

Petrovska, B. B. (2012). Historical Review of Medicinal Plants' Usage. *Pharmacognosy Reviews*, *6*(11), 1. https://doi.org/10.4103/0973-7847.95849

Ramirez, A. M., Saillard, N., Yang, T., Franssen, M. C. R., Bouwmeester, H. J., & Jongsma, M. A. (2013). Biosynthesis of Sesquiterpene Lactones in Pyrethrum (*Tanacetum cinerariifolium*). *PLoS ONE*, *8*(5), e65030. https://doi.org/10.1371/journal.pone.0065030

Reece-Hoyes, J. S., & Walhout, A. J. M. (2018). Gateway Recombinational Cloning. *Cold Spring Harbor Protocols*, *2018*(1), pdb.top094912. https://doi.org/10.1101/PDB.TOP094912

Ro, D. K., Paradise, E. M., Quellet, M., Fisher, K. J., Newman, K. L., Ndungu, J. M., Ho, K. A., Eachus, R. A., Ham, T. S., Kirby, J., Chang, M. C. Y., Withers, S. T., Shiba, Y., Sarpong, R., & Keasling, J. D. (2006). Production of the Antimalarial Drug Precursor Artemisinic Acid in Engineered Yeast. *Nature*, *440*(7086), 940–943. https://doi.org/10.1038/nature04640

Salmerón-Manzano, E., Garrido-Cardenas, J. A., & Manzano-Agugliaro, F. (2020). Worldwide Research Trends on Medicinal Plants. *International Journal of Environmental Research and Public Health*, *17*(10). https://doi.org/10.3390/IJERPH17103376

Sanger, F., Nicklen, S., & Coulson, A. R. (1977). DNA Sequencing with Chain-terminating Inhibitors. *Proceedings of the National Academy of Sciences of the United States of America*, *74*(12), 5463–5467. https://doi.org/10.1073/PNAS.74.12.5463

Schmiderer, C., Torres-Londono, P., Lutz-Röder, A., Duwe, V. K., & Novak, J. (2018). *Arnica montana* subsp. *atlantica*: Really a Subspecies? *Genetic Resources and Crop Evolution 2018 65:7*, *65*(7), 1829–1838. https://doi.org/10.1007/S10722-018-0653-2

Seemann, A., Wallner, T., Poschlod, P., & Heilmann, J. (2010). Variation of Sesquiterpene Lactone Contents in Different *Arnica montana* Populations: Influence of Ecological Parameters. *Planta Medica*, *76*(08), 837–842. https://doi.org/10.1055/S-0029-1240797

Shafi, A., & Zahoor, I. (2021). Metabolomics of Medicinal and Aromatic Plants: Goldmines of Secondary Metabolites for Herbal Medicine Research. *Medicinal and Aromatic Plants* (pp. 261–287). Elsevier Inc. https://doi.org/10.1016/b978-0-12-819590-1.00012-4

Spring, O., Schmauder, K., Lackus, N. D., Schreiner, J., Meier, C., Wellhausen, J., Smith, L. V., & Frey, M. (2020). Spatial and Developmental Synthesis of Endogenous Sesquiterpene Lactones Supports Function in Growth Regulation of Sunflower. *Planta*, *252*(1). https://doi.org/10.1007/s00425-020-03409-y

Sugier, D., Sugier, P., & Gawlik-Dziki, U. (2013). Propagation and Introduction of *Arnica montana* L. into Cultivation: A Step to Reduce the Pressure on Endangered and High-Valued Medicinal Plant Species. *The Scientific World Journal*, *2013*, 11. https://doi.org/10.1155/2013/414363

Surmacz-Magdziak, A., & Sugier, D. (2012). In vitro Propagation of *Arnica montana* L.: An Endangered Herbal Species of Great Importance to Medicine. *Hortorum Cultus*, *11*(2), 127–140.

Valkute, T. R., Aratikatla, E. K., Gupta, N. A., Ganga, S., Santra, M. K., & Bhattacharya, A. K. (2018). Synthesis and Anticancer Studies of Michael Adducts and Heck Arylation Products of Sesquiterpene Lactones, Zaluzanin D and Zaluzanin C from *Vernonia arborea*. *RSC Advances*, *8*(67), 38289–38304. https://doi.org/10.1039/C8RA06238B

Wiesner, J., & Kairies, H. (2014). Assessment report on *Arnica montana* L., *flos*. www.ema.europa.eu

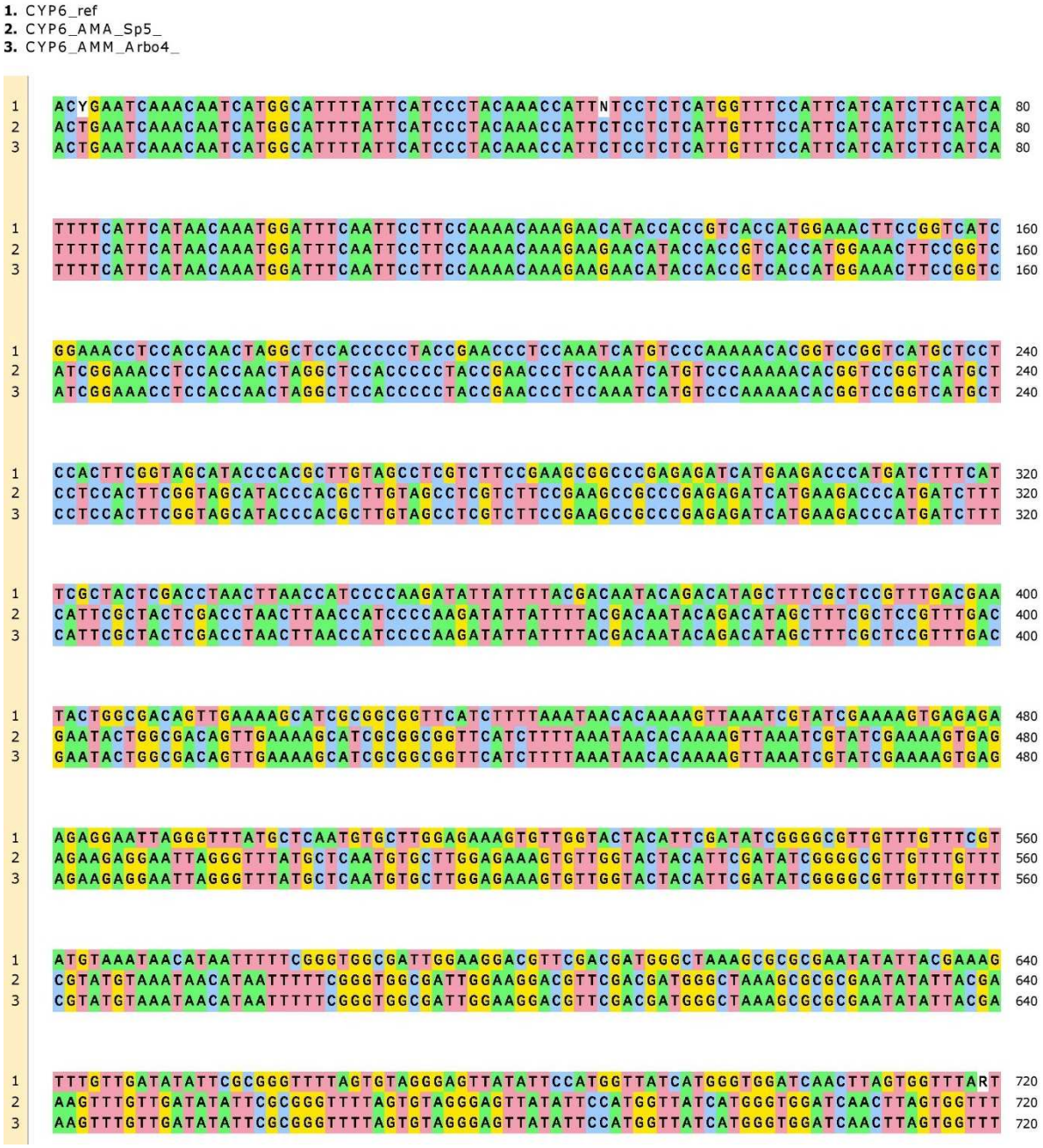
Zhou, F., & Pichersky, E. (2020). More is Better: The Diversity of Terpene Metabolism in Plants. *Current Opinion in Plant Biology* 55, 1–10. Elsevier Ltd. https://doi.org/10.1016/j.pbi.2020.01.005

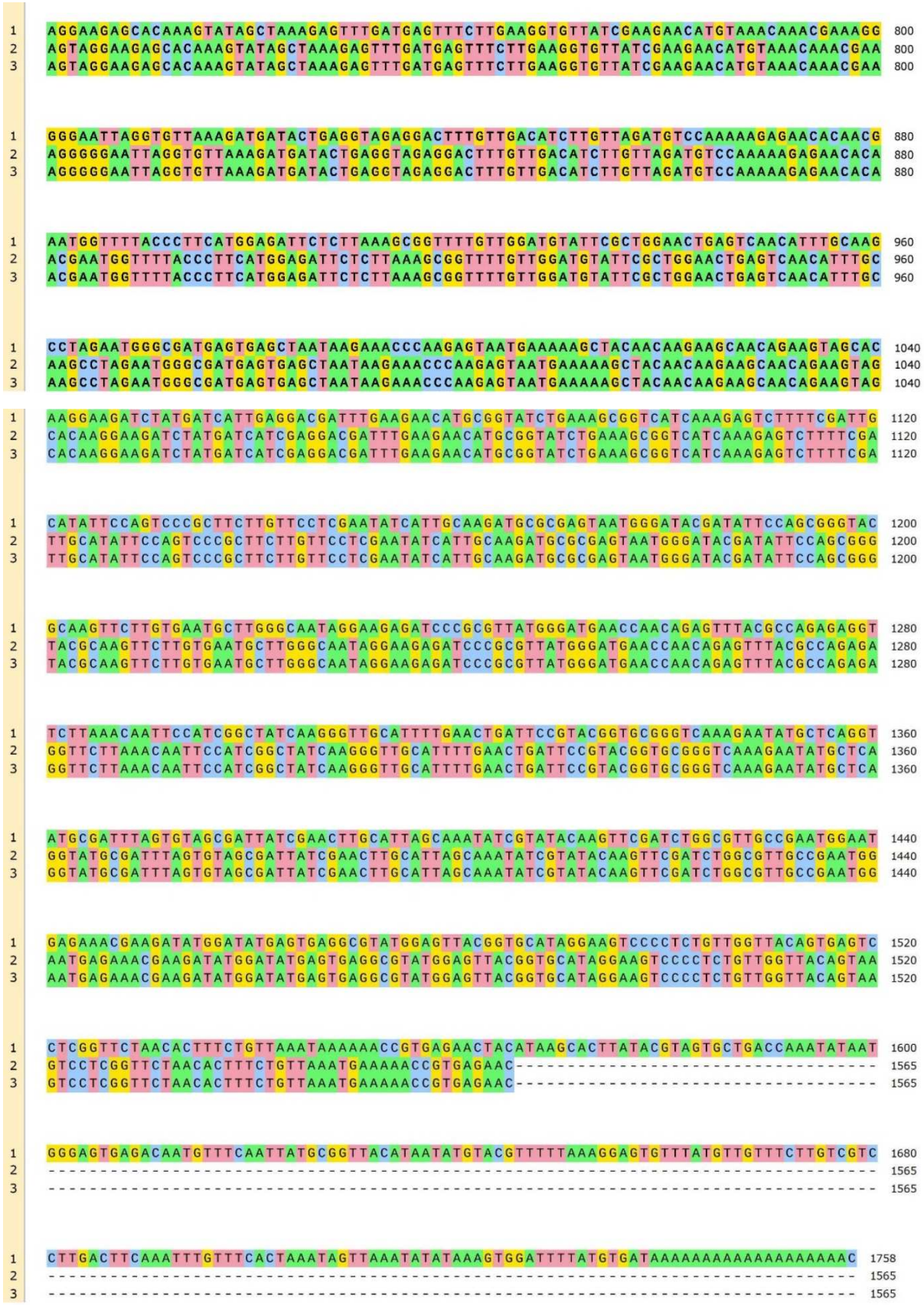
8. Supplementary data

I. Multiple sequence alignment

Multiple sequence alignment was performed using MUSCLE online tool (version 3.8; https://www.ebi.ac.uk/Tools/msa/muscle/) and visualised in SnapGene® software (version 5.1; https://www.snapgene.com/snapgene-viewer/).

AmCYP6 gene sequences from *A. montana* ssp. *montana* (CYP6_AMM_Arbo4) and ssp. *atlantica* (CYP6_AMA_Sp5) were aligned to reference *AmCYP6* sequence identified prior to my thesis via 5' and 3' RACE using non-proofreading polymerase (CYP6_ref).





Curriculum vitae

I was born in Rijeka and grew up in Punat on island of Krk. After finishing gymnasium programme at High school Hrvatski kralj Zvonimir, I enrolled in bachelor studies in 2015 and master studies in 2018, both in Molecular biology at Faculty of Science, University of Zagreb. During my master studies, I spent one semester on Erasmus+ exchange studies with focus on biotechnology on Technical University of Madrid, Spain. In July 2021, I started Erasmus+ traineeship program at the Institute of Plant Biology and Biotechnology in Prof. Dr. Dirk Prüfer's group where I also did my master's thesis research.