

Master's thesis in Biotechnology

Expression of Terpene Synthases from Edible Fungi using Yeast Chassis

Helena Buhrgard

Master student

Biotechnology, Faculty of Engineering, Lund University

-Supervisors-

Sotirios Kampranis

Professor

*Section for Plant Biochemistry,
Department of Plant and Environmental Science,
Faculty of Science, University of Copenhagen*

Ed van Niel

Senior lecturer

*Applied Microbiology, Department of Chemistry,
Faculty of Engineering, Lund University*

-Co-supervisor-

Chi Zhang

Postdoc

*Section for Plant Biochemistry,
Department of Plant and Environmental Science,
Faculty of Science, University of Copenhagen*



Submitted: 2024-06-17

LUND
UNIVERSITY

Abstract

Secondary metabolites produced from organisms are compounds that are not directly connected to the growth of organisms but that serve many important purposes for them in nature. Terpenes and terpenoids form a section of the secondary metabolites, produced from terpene synthase (TPS) enzymes. Fungal species are highly dependent on secondary metabolites and especially the terpenoid compounds for numerous adaptive tasks, such as defense and formation of symbiotic relationships. Despite the importance of terpenes and terpenoids in fungi and the large number of fungal species compared to plant species, the corresponding TPS genes in fungal genomes are much less studied than those from plants. In this work, as part of a larger study at UCPH, unexplored TPSs from edible fungal species were studied, to contribute to the enzyme characterizations and product exploration. 31 TPSs enzymes from fungal genomes of shiitake mushroom *Lentinula edodes*, oyster mushroom *Pleurotus ostreatus*, porcini mushroom *Boletus edulis*, jelly fungus *Auricularia subglabra* and cheese fungi *Penicillium roqueforti*, *Penicillium biforme*, and *Penicillium camemberti* were expressed. The genes were introduced in yeast through multicopy plasmids, using the uracil-specific excision reagent (USER) cloning technique to construct the plasmids together with an inducible GAL1 promoter. The products were analyzed *in vivo* with headspace solid phase microextraction (HS-SPME) as well as *in vitro* with hexane extraction, using gas chromatography mass spectrometry (GC-MS). From the results it could be concluded that three TPSs produced mainly monoterpenes, nine TPSs mainly sesquiterpenes, and one TPS mainly diterpenes. One putative sesquiterpene without a provided name was detected as well as sesquiterpenes sinularene and myltayl-4(12)-ene which could not be found in fungal species. The monoterpene synthase (mono-TPSs) belonged in majority to the Ascomycota phylum and the sesquiterpene synthase (sesqui-TPSs) in majority to the Basidiomycota phylum. The catalytic activity of the TPS genes was traced back to the phylogenetic tree, where especially TPSs producing monoterpenes located in one cluster and TPSs producing sesquiterpenes of cadalane structures in another. An additional experiment with expression of Erg20p(N127W), a gene described to accumulate GPP in yeast cells, led to an unexpected increase in sesquiterpenes. Furthermore, a third of the transformants were induced into buffered medium (pH 6.5) to analyze the correlation between pH and enzymatic activity. The buffered induction led to an overall greater sesquiterpene production for all except the three tested inactive transformants that still showed no terpene peaks.

Popular science summary

Fungal species exist everywhere on earth, highly important in various types of ecosystems. There are many sorts of fungi, including the toxic molds growing on bread, the yeast used for beer production, and mushrooms that can be picked in the forest. Fungal species can be utilized for many purposes, where a large aspect is their role in the food sector. They are included in dishes or used in food industries, such as for various meat substitutes and cheese production. Fungi produce a great array of small molecules, providing the species with specific bioactivities. One group of the small molecules are the terpene hydrocarbons and their modified structures termed terpenoids. The compounds have been less studied in fungi compared to plants, and the understanding of product formations is low. However, as terpenes and terpenoids can carry a variety of properties, such as antimicrobial, anti-inflammatory, and anticancer, and fungal species are highly consumed by humans, the interactions between the produced terpenes and terpenoids with the human microbiota is important.

This thesis project was part of a larger study at University of Copenhagen, aiming to provide knowledge on the terpene production from edible fungi. Yeast cells were used to express the terpene-producing enzymes TPSs. GC-MS was used to analyze the formed products. TPSs from 4 edible mushrooms, specifically shiitake mushroom, oyster mushroom, porcini mushroom, and jelly fungi, as well as 3 cheese fungi were studied. Not all TPSs demonstrated activity, but the active presented either one or a few terpene products, or they were able to produce a greater range of terpenes. A large portion of the detected terpenes had been discovered in fungi before, some in edible species and some in the same species, but sesquiterpenes not described in fungi were also produced. The work studied how pH affected the activity of the TPSs hence the terpene production. This thesis work provided insights into terpenes formed from edible fungi, allowing continuous studies on potential terpenoid productions.

Table of Contents

Abbreviations	6
1. Introduction	7
1.1 Project background	7
1.2 Aim and scope of the thesis work	7
2. Theory	8
2.1 Fungi	8
2.2 Terpenes and terpenoids	8
2.2.1 Monoterpenes	10
2.2.2 Sesquiterpenes	10
2.2.3 Diterpenes	10
2.2.4 Triterpenes	10
2.3 Terpenoid biosynthetic pathway in fungi	11
2.3.1 Production of IPP and DMAPP	11
2.3.2 Production of direct precursors and products	11
2.3.2.1 Prenyltransferases	12
2.3.2.2 Terpene synthases	13
2.3.2.2.1 Class I terpene synthases	13
2.3.2.2.2 Class II terpene synthases	13
2.3.2.2.3 Monoterpene synthases	14
2.3.2.2.4 Sesquiterpene synthases	15
2.3.2.2.5 Diterpene synthases	16
2.3.2.2.6 Triterpene synthases	17
2.3.2.3 Tailoring enzymes	17
2.4 The unexplored fungal TPSs	17
2.5 Host systems for terpene synthase expression	18
2.5.1 <i>Escherichia coli</i>	18
2.5.2 <i>Saccharomyces cerevisiae</i>	18
2.6 Methods for terpene synthase study	19
2.6.1 Gene cloning	19
2.6.1.1 USER cloning	19
2.6.2 Transformation	20
2.6.3 Enzyme production	20

2.6.4 Compound detection	20
3. Preparatory work	21
3.1 Tree assembly	21
3.2 Selection of TPS genes	21
4. Material and Methods	23
4.1 Material	23
4.1.1 Gene sequences	23
4.1.2 Vector system	24
4.1.3 Strains	26
4.1.4 Media	27
4.2 Gene and primer preparation	27
4.3 Plasmid construction	27
4.4 Verification of plasmid construct	29
4.5 Preparation and transformation of <i>S. cerevisiae</i>	30
4.6 Induction of transformed yeast cells	30
4.7 <i>In vitro</i> enzyme assay	30
4.8 GC-MS analysis	31
5. Results and Discussion	32
5.1 <i>In vivo</i> analysis	32
5.1.1 Single transformations	32
5.1.1.1 Monoterpene production	32
5.1.1.2 Sesquiterpene production	33
5.1.1.3 Diterpene production	36
5.1.2 Co-transformations	37
5.2 <i>In vitro</i> analysis	38
5.2.1 Cell hydrolysis reactions	39
5.2.2 Mono-TPSs	39
5.2.2.1 Comparison <i>in vivo</i> and <i>in vitro</i> mono-TPSs	40
5.2.3 Ses-TPSs	41
5.2.3.1 Comparison <i>in vivo</i> and <i>in vitro</i> sesqui-TPSs	46
5.2.4 Di-TPSs	48
5.2.4.1 Class I and class II di-TPSs	48
5.3 pH analysis	49
5.4 Inactive genes and further work	52

5.5 TPS catalytic activity in phylogenetic tree	53
5.6 Terpene documentation in fungal species	58
5.6.1 Documentation of monoterpenes	58
5.6.2 Documentation of sesqui- and diterpenes	59
6. Conclusion	59
7. Future work	60
8. References	61
Appendix	68

Abbreviations

AI	Arithmetic Index
Di-TPS	Diterpene synthase
DMAPP	Dimethylallyl pyrophosphate
FPP	Farnesyl diphosphate
GC-MS	Gas chromatography mass spectrometry
GFPP	Geranylfarnesyl diphosphate
GGPP	Geranylgeranyl diphosphate
GPP	Geranyl diphosphate
HS-SPME	Headspace solid phase microextraction
IDI	Isopentenyl pyrophosphate isomerase
IPP	Isopentenyl pyrophosphate
MEP	Methylerythritol-phosphate
MEV	Mevalonate
Mono-TPS	Monoterpene synthase
NerPP	Neryl diphosphate
NPP	Nerolidyl diphosphate
OSC	Oxidosqualene synthase
PCR	Polymerase chain reaction
PTTS	Prenyltransferase terpene synthase
RT	Retention time
Ses-TPS	Sesquiterpene synthase
TIC	Total ion chromatogram
TPS	Terpene synthase
Tri-TPS	Triterpene synthase
USER	Uracil-specific excision reagent

1. Introduction

1.1 Project background

Fungal species produce a large mixture of secondary metabolites, consisting of compounds with various functions. One type of secondary metabolites are the terpenes and terpenoids. Terpenes are hydrocarbons of a general $(C_5H_8)_n$ structure, with non-polar and usually rather volatile character, whereas terpenoids arise from modifications of the terpenes, with functional groups added to the terpene scaffolds (1). Terpenes are produced as TPSs cyclize linear isoprenyl diphosphate chain substrates. These hydrocarbon-based compounds present a great variety of bioactivities, such as antimicrobial and cytotoxic activities, allowing fungi to adapt and survive in different environments (2). The fungal terpenome is unexplored, compared to the much more investigated plant terpenome, and terpenoids in terms of their bioactivities and biosynthesis pathways existing in fungal species need further exploration. Since a great part of fungal species are used for edible purposes, in foods, and for industrial purposes, and fungi further reside naturally in humans, the interactions of the bioactive terpenes and terpenoids with the human microbiome are of great interest. Fungi have previously been linked to several diseases due to the contribution to an imbalanced gut microbiome, known as dysbiosis, but the mechanisms of exactly how fungi affect the human microbiome and induce dysbiosis are not fully understood.

A study performed at the Section for Plant Biochemistry at University of Copenhagen hypothesizes that the secondary metabolites from fungi, such as terpenoids, play a major role in affecting the microbiome. However, as the knowledge on the terpenoids formed by fungal species is limited, the TPSs and tailoring enzymes need to be characterized to be able to fully grip the terpenoids present in fungi. The study hence aims to screen for TPS genes and the clusters in which TPS usually locate in the fungal genome, to elucidate the terpenoid biosynthesis pathways and characterize the products related to the genes. Especially edible mushrooms, fungi used in food processes, and fungal species residing in the human gut will be in focus, where the terpenoid products will in a later stage in the study be tested on the human microbiome for an understanding on how these fungal metabolites influence this system. The knowledge how the metabolites affect will be highly valuable in preventions of dysbiosis related diseases.

1.2 Aim and scope of the thesis work

The work in this project is a part of the study at University of Copenhagen. The aim of the thesis work is to contribute to the exploration of the fungal terpenome, with a focus on edible fungal species. The exploration is reached by expressing selected TPS genes in yeast and analyzing the products with GC-MS. The focus lies on the expression of putative TPS enzymes solely and not on genes coding tailoring enzymes. The TPS enzyme characterization will allow insights into the initial terpene scaffolds produced from the fungal species, and this knowledge can later be used for elucidating terpenoid pathways.

The scope of the work included TPS gene cloning, expression, and analysis in *S. cerevisiae*. Plasmid constructs carrying an inducible promoter and selected TPS genes were created with USER cloning. The plasmids were firstly verified by colony polymerase chain reaction (PCR) and sequencing, and correct plasmids were transformed into a yeast host system. The yeast cells were either kept intact for product analysis *in vivo* or lysed for release of TPSs, which were tested *in vitro* for product formation. GC-MS

was used for both *in vivo* and *in vitro* analysis. Furthermore, production patterns were studied as the TPSs were co-expressed with a monoterpene-boosting gene as well as induced in buffered medium.

2. Theory

2.1 Fungi

Fungi are eukaryotes that exist and grow everywhere, in the air, earth, and all animals (3). The number of fungal species on Earth varies greatly between sources, and has been described in ranges of 500,000 up to around 19 million species existing, but the formal number is much lower, with documentation of around 155,000 (4). The fungi kingdom is classified on several levels, from a higher order of nine different phyla, such as Ascomycota and Basidiomycota that both come from the Dikarya subkingdom, with lineages down to a vast amount of specific species (5). The fungi are divided into groups, of the common mold, yeast, and mushroom. Molds are filamentous fungi composed of thin threads, known as hyphae, that grow into a mycelium. Enzymes are released from the mycelium that will allow digestion of compounds into food sources for the mold. Yeast is a unicellular fungus that usually grows in number through budding. A common yeast species is the baker's yeast. Lastly, mushrooms are multicellular filamentous organisms that grow over and under ground, and have in general terms been defined as macro fungi (6). Recognized mushroom species are in the span of 16,000, but according to estimations this is only around 10 % of all mushroom species that exist. Out of this 16,000 described, 7,000 mushroom species are considered edible to various degrees and 3,000 are considered purely edible (7).

The uses of fungi, both uni- and multicellular, for edible causes are plenty. In addition to using mushrooms in dishes, the fermentation industry to produce beverages and food rely greatly on the use of fungi. This includes the making of bread, beer, and wine by yeast, as well as the growth of mold species on cheese (8). Tempeh is made from inoculation with mold species that are allowed to ferment (9). Furthermore, the fungal protein mycoprotein is produced by fermenting fungi and it can be used to make high-protein food such as Quorn (10).

All fungi species produce numerous natural products, also called secondary metabolites, to survive and to adapt to their environment. Unlike the primary metabolites, secondary metabolites are not directly connected to the normal growth of fungi but they serve great importance by allowing the fungi species to communicate and compete with neighbors for surrounding nutrients as well as survive various predators (11). They come from the numerous biosynthetic pathways that exist in fungi and can be organized into classes of compounds, such as polyketides, non-ribosomal peptides, hybrid non-ribosomal polyketides/peptides, and the terpenoids (12). The last class, also known as isoprenoids, constitutes a major part of the fungal secondary metabolites. Today more than 110,000 terpenoids have been discovered, mostly isolated from plants (11, 13). Terpenoids are nonetheless produced from organisms other than fungi and plants, including bacteria, insects, and animals (14).

2.2 Terpenes and terpenoids

Terpenes are compounds of a commonly cyclic structure with the general molecular formula $(C_5H_8)_n$, which gives them a lipophilic character. Terpenoids are modified terpenes, with additional enzymes, such

as atom additions or altered number of methyl groups. All terpenes and further terpenoids are built from the 5-carbon molecule isoprene (C_5H_8) (Figure 1) and are classified based on the number of isoprene units (hemiterpenes) that make up the structure. Some terpene classes are, in an increasing number of isoprene units, monoterpenes (C_{10}), sesquiterpenes (C_{15}), diterpenes (C_{20}), sesterterpenes (C_{25}), triterpenes (C_{30}), and tetraterpenes (C_{40}) (Table 1). Sesquiterpenoids, diterpenoids, and triterpenoids are considered the three major classes of terpenoids in fungi. Furthermore, terpenoids can originate from other biosynthetic routes, such as meroterpenoids and indole terpenoids (2, 11).

This report will focus on the common fungal terpenes as well as the monoterpenes, including insights into their derivatives.

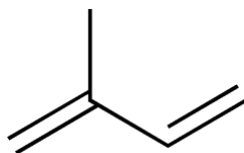


Figure 1. The structure of an isoprene molecule (15).

Table 1. List of terpene carbon skeletons, with increasing number of carbon and isoprene units (2). From the listed terpene scaffolds, further chemically altered terpenoids are created.

Class	# carbon atoms	# isoprene units
Monoterpenes	10	2
Sesquiterpenes	15	3
Diterpenes	20	4
Sesterterpenes	25	5
Triterpenes	30	6
Tetraterpenes	40	8
Polyterpenes	>40	>8

Terpenes and derivatives help with electron transport in the fungus, as well as contribute to formation of cell walls and membranes. Furthermore, they act as chemical defense and promote formation of symbiotic relationships. Terpenoids are of interest for humans in many areas due to the various biological activities, e.g. the pharmaceutical departments have great interest in terpenoids as the compounds display antimicrobial, antiinflammatory and antiviral properties (2). The anticancer drug Taxol and antimalarial drug Artemisinin are two known terpenoids (16). The compounds are also highly valued as fragrance and flavor additives (17).

The common fungal terpenes as well as the monoterpenes will now be discussed more in depth.

2.2.1 Monoterpenes

Monoterpene scaffolds are composed of 10 carbon atoms, with an acyclic, monocyclic, or bicyclic structure. Monoterpenes and their derivatives are, due to their smaller size, usually volatile, making them well-suited for air signaling purposes. They have great potential for perfumery fragrance as well as pharmaceutical uses (18). 1,000 monoterpenoids have been described from 38 monoterpene scaffolds (19).

Relatively few monoterpenes and derivatives have been identified from fungal species, compared to the 15 to 30 carbon terpenes (11). Although production of monoterpenoids is less common than sesqui- and diterpenoids in fungi, research has revealed monoterpenes produced in species, such as linalool and terpineol from a variety of strains as well as limonene from e.g. the edible fungi species *Lentinula edodes*, known as shiitake mushroom (20).

2.2.2 Sesquiterpenes

The sesquiterpenes, despite being larger than monoterpenes with a 15-carbon skeleton, present volatile properties found in both Ascomycota and Basidiomycota (21). 300 sesquiterpene scaffolds have been described and more than 7,000 sesquiterpenoid compounds (19). These 15 carbon compounds can, similarly to monoterpenes, be a-, mono-, and bicyclic, even though the acyclic structure is not as common for sesquiterpenes. Furthermore, additional tricyclic formations can be found in sesquiterpene structures (20).

As has been mentioned, sesquiterpenes and derivatives are a major part of the natural product formation in fungi, where cytotoxicity has been presented as the major activity by sesquiterpenoids (22). Mycotoxins, some being sesquiterpenoids, are produced from a variety of filamentous fungi carrying plant pathogenic effects, meaning they can infect plants, such as those used for feed and foods. These filamentous fungi are for instance *Penicillium* and *Aspergillus* (23).

2.2.3 Diterpenes

The 20 carbon diterpenes present even larger macrocyclic arrangements. Despite most diterpenoids being described from plants, in the span of the years 2010 to 2020 around 400 diterpenoids were identified in fungi. These are mainly tri and tetra ring structures. Diterpenes described from fungal species have a great variety of biological activities, but as for sesquiterpenoids, cytotoxicity has been described as the major activity. Diterpenes identified from fungal species are for instance from species of genus *Penicillium* that have been characterized to produce tetracyclic cyclopiane diterpenoids (24).

2.2.4 Triterpenes

Triterpenes are 30 carbon structures, mostly of multicyclic arrangements which can be built into triterpenoids as well as steroid and sterol compounds. A vital compound for fungi is ergosterol, a fungal cell membrane sterol (23). Ergosterol comes from the triterpene precursor lanosterol. Fungi can contain varying amounts of ergosterol, and for instance the edible fungi *Pleurotus ostreatus* and *Lentinula edodes* have been reported with high levels (25). Lanosterol is also a precursor to a defined group of tricyclic triterpenoids, called lanostanes (26). Triterpenoids are found especially in medicinal fungi, mostly of the lanostane skeleton (27).

2.3 Terpenoid biosynthetic pathway in fungi

Terpenoids are produced through an overall four step process, including in a very broad manner the following reactions: 1) the two initial precursors for terpenoids are produced, the 5 carbon monomers isopentenyl pyrophosphate (IPP) and dimethylallyl pyrophosphate (DMAPP). These compounds are of the same isoprene structure but in activated forms. 2) IPP and DMAPP polymerize to various degrees, forming the linear isoprenyl diphosphate precursors geranyl diphosphate (GPP), farnesyl diphosphate (FPP), geranylgeranyl diphosphate (GGPP), and geranylgeranyl farnesyl diphosphate (GFPP). 3) rearrangement and cyclization of the precursors produce various terpene scaffolds. 4) the terpene scaffolds are often modified to obtain the specific final terpenoids (17, 28). The steps will be discussed more thoroughly.

2.3.1 Production of IPP and DMAPP

IPP and DMAPP production in fungal species is through the mevalonate (MEV) pathway, which differs from e.g. plants where the methylerythritol-phosphate (MEP) pathway may also be utilized (29). The fungal MEV pathway starts in the cytoplasm with the primary metabolite acetyl-CoA (Figure 2). The last step produces IPP from mevalonate 5-pyrophosphate. With isopentenyl pyrophosphate isomerase (IDI) enzyme, IPP can isomerize into DMAPP (2).

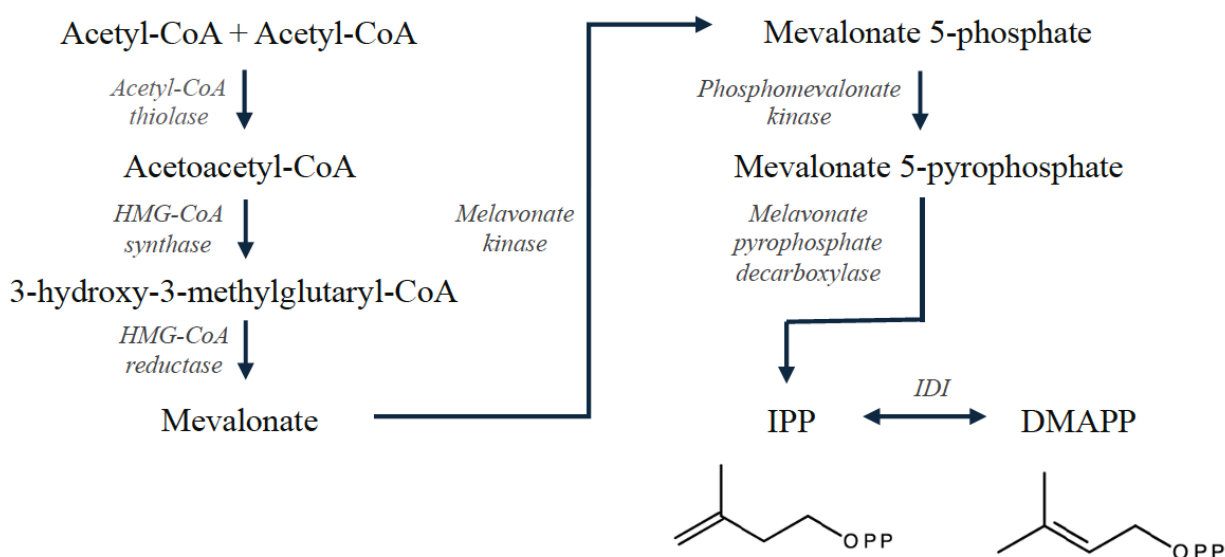


Figure 2. The mevalonate pathway is used for production of terpenoid precursors in fungi. The IPP precursor is isomerized into DMAPP, catalyzed by IDI. OPP - diphosphate group. The structures and figure are made by the author (2).

2.3.2 Production of direct precursors and products

Direct precursors GPP, FPP, GGPP, and GFPP that are used for terpenoid production are formed from IPP and DMAPP (Figure 3). The 10-carbon precursor GPP is produced through the fusion of DMAPP and IPP, further cyclized into a monoterpene scaffold. The sesquiterpene precursor FPP is produced by fusing GPP to IPP, and further addition of IPP to FPP forms GGPP, the precursor for diterpene scaffolds. Two FPP will fuse together to a squalene, then a 2,3-oxidosqualene, and start the pathway to triterpenes and similarly will two GGPP form phytoene, allowing formation of tetraterpenes. GFPP of 25 carbon atoms is formed by addition of IPP to GGPP and the compound produces sesterterpenes. The terpene scaffolds can

lastly be further modified into various terpenoids (12). The reaction steps are catalyzed with different types of enzymes, further discussed.

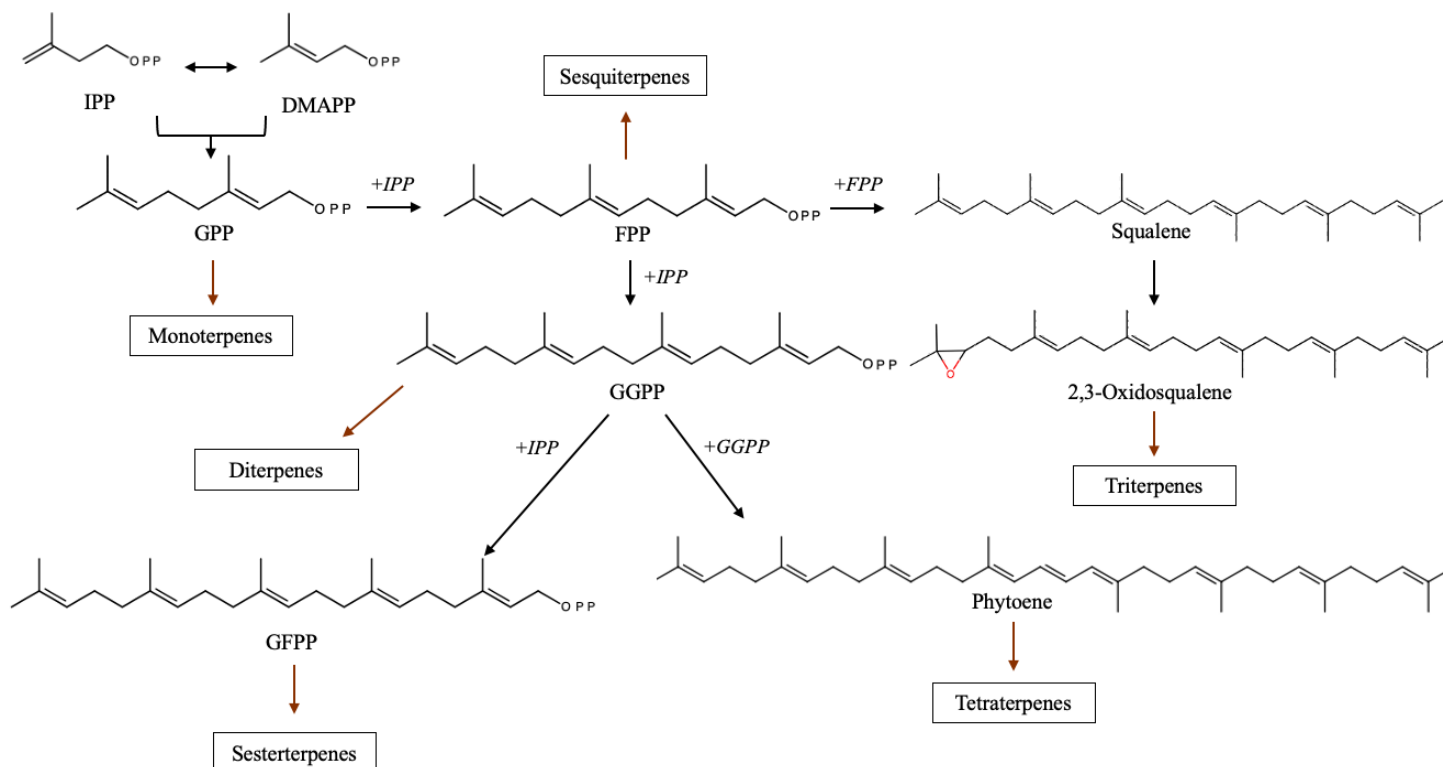


Figure 3. Terpene scaffolds are formed from various direct precursors (red arrows). Monoterpenes are produced from GPP, sesquiterpenes from FPP, diterpenes from GGPP, and sesterterpenes from GFPP. Triterpenes are made from fusion of two FPP molecules and tetraterpenes are made from fusion of two GGPP molecules. OPP - diphosphate group. The figure and structures are made by the author (2).

2.3.2.1 Prenyltransferases

Prenyltransferases will catalyze the biosynthesis of direct precursors, both of solely DMAPP/IPP origin and a mixed origin where other pathway products enter the terpenoid biosynthesis. For the precursors of DMAPP/IPP origin, chains with varying numbers of isoprene units are produced. The required enzymes for production of GPP, FPP, GGPP, and GFPP are geranyl diphosphate synthases, farnesyl diphosphate synthases, geranylgeranyl diphosphate synthases, and geranylgeranyl farnesyl diphosphate synthases, respectively (30).

The synthases catalyze the fusion of IPP with DMAPP or an existing direct precursor, through a condensation of head-to-tail form, where the diphosphate group is positioned at the tail. The head-to-tail fusion allows one of the diphosphate groups to stay attached to the molecule (30). More specifically, two conserved DDxxD motifs in the enzymes enable the removal of the diphosphate groups from IPP and DMAPP. This generates a carbon cation in the molecules, allowing the chain elongations. Naturally produced terpenoids identified in nature usually present to originate from *E* - precursors, meaning an all-through trans formation of the isoprenyl diphosphate chains, however cis-prenyltransferases exist, forming *Z* - double bonds in the precursor (31).

The formation of squalene, initial precursor for triterpenoids, and phytoene, precursor for tetraterpenoids, is catalyzed with the prenyltransferases squalene synthase and phytoene synthase, respectively. Instead of the head-to-tail fusion, the fusion occurs tail-to-tail of two FPPs or two GGPPs. This results in loss of both diphosphate groups (Figure 3) (28).

2.3.2.2 Terpene synthases

The linear precursors will be cyclized by TPSs, or terpene cyclases, into terpene scaffolds. The specificity of the TPS reactions varies greatly, making some TPSs highly specific whereas other TPSs will be able to produce multiple products. Especially sesquiterpene synthases have shown to be catalytically promiscuous (2). Mono- and sesquiterpene synthases, perhaps due to their similar structure, have presented bifunctionality with both mono- and sesquiterpene productions (19).

Based on their mechanism for the initial formation of a carbocation in the substrate, the TPSs are divided into enzymes of class I and class II (11). The two classes will be described, and the mono-TPS, ses-TPS, diterpene synthases (di-TPS), and triterpene synthases (tri-TPS) will be discussed in terms of activity and class belonging.

2.3.2.2.1 Class I terpene synthases

The class I TPSs are related to the prenyltransferases and promote an ionization-dependent reaction for the cyclization of the substrate (11). A hydrophobic active site is created in the enzyme by an α -helix structure, or the so called α -domain. In this site, two conserved parts, DDxxD and NSE/DTE, are located. These sequences bind the substrate at the diphosphate group. The reaction requires the presence of a magnesium ion. Diphosphate removal creates a carbocation from which numerous reactions can take place (Figure 4). A double bond in the molecule can react with the cation to form a cyclic structure (2). Further possible pathways are water quenching into a terpene alcohol, hydride shifts, or deprotonation (11).

2.3.2.2.2 Class II terpene synthases

Different from the strategy of class I TPSs, class II TPSs will perform a protonation-dependent reaction. The enzyme carries, similarly to class I, a hydrophobic active center, but instead this active site is located between two α -helix barrel folds, typically referred to as $\gamma\beta$ -domains. Class II catalyzes reactions by providing a proton from an aspartate motif, DxDD. This proton addition generates a carbocation (Figure 4). The carbocation can create cyclization as it reacts with another double bond in the substrate structure. The reactions will not remove the diphosphate group, as is done in class I enzymes (23).

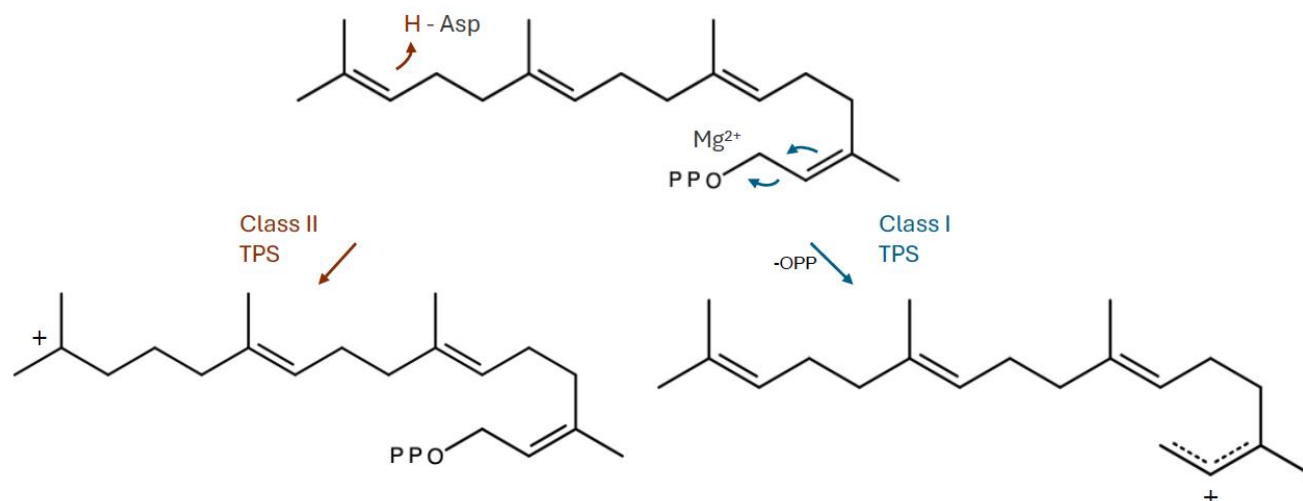


Figure 4. Example of how class I vs class II TPSs create an intermediate cation from GGPP. OPP - diphosphate group. The structures and figure are made by the author (32).

2.3.2.2.3 Monoterpene synthases

Mono-TPS uses GPP as a substrate for monoterpene formation. A limited number of mono-TPS have been detected in fungal species, some of these discovered in recent years. The mono-TPS enzymes are of class I, as the diphosphate group is removed (28). As previously mentioned, acyclic or cyclic structures may be formed, depending on the pathway taken. The monoterpene formation explained is that GPP is converted into a geranyl cation and either formed into scaffolds or further produced into a linalyl cation. Acyclic structures are formed from the geranyl cation, whereas mono- and bicyclic are formed as the linalyl cation cyclize (Figure 5). Hydride shifts, deprotonation, and further modifications create the final scaffold (33).

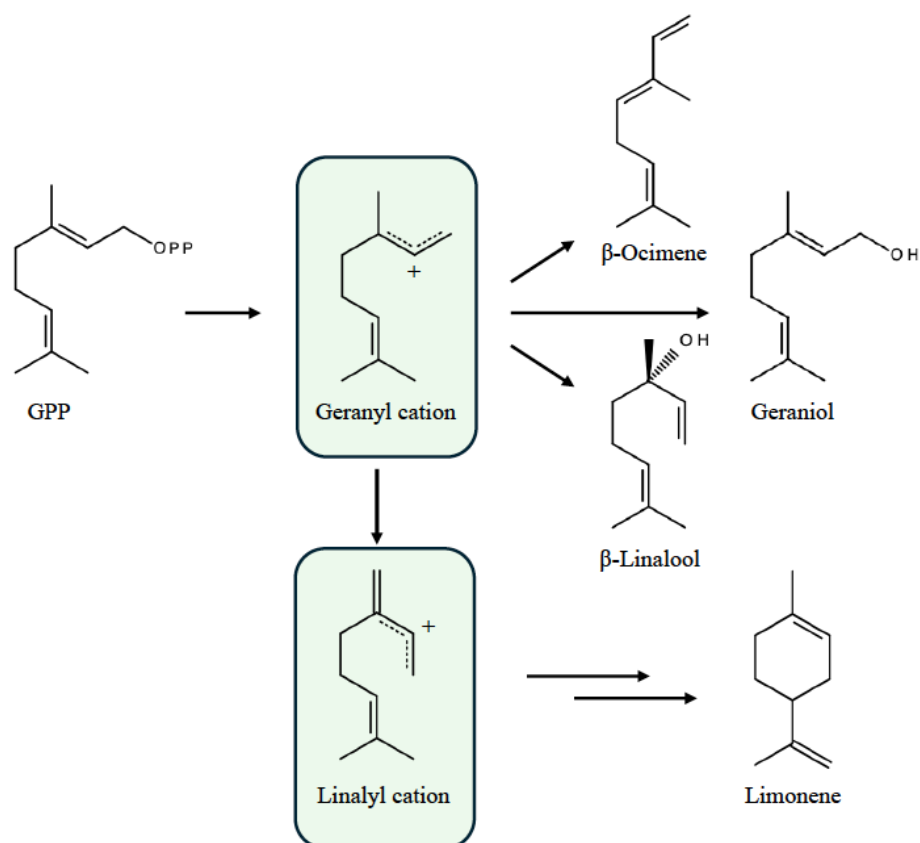


Figure 5. Example on formation of acyclic and monocyclic monoterpenes from GPP. OPP - disphosphate group. Compounds and the figure are made by the author (28).

2.3.2.2.4 Sesquiterpene synthases

Due to the active site being in a α -helix bundle in the enzyme, ses-TPSs are like mono-TPSs classified as class I TPSs. However, as has been mentioned, due to the longer FPP substrate with one more double bond in the structure, the diversity of sesquiterpenes is much greater than monoterpenes. Ses-TPS start the terpene scaffold production by ionizing FPP into a farnesyl cation, or FPP is isomerized into nerolidyl diphosphate (NPP) and further into nerolidyl cation as intermediate. The two cations have been described with two common cyclization pathways each to specialize into two forms of the germacradienyl cation, humulyl cation, or β -bisabolyll cation (farnesyl C1,6 and C1,10; nerolidyl C1,10 and C1,11) (Figure 6). The enzyme will adjust the ring formations into final scaffolds (deprotonation, hydride shift, etc.) or continue cyclizing into tricyclic sesquiterpenes (28, 34). Acyclic sesquiterpenes may also be formed, such as protonation into β -farnesene (Figure 6) or through water quenching into farnesol or nerolidol.

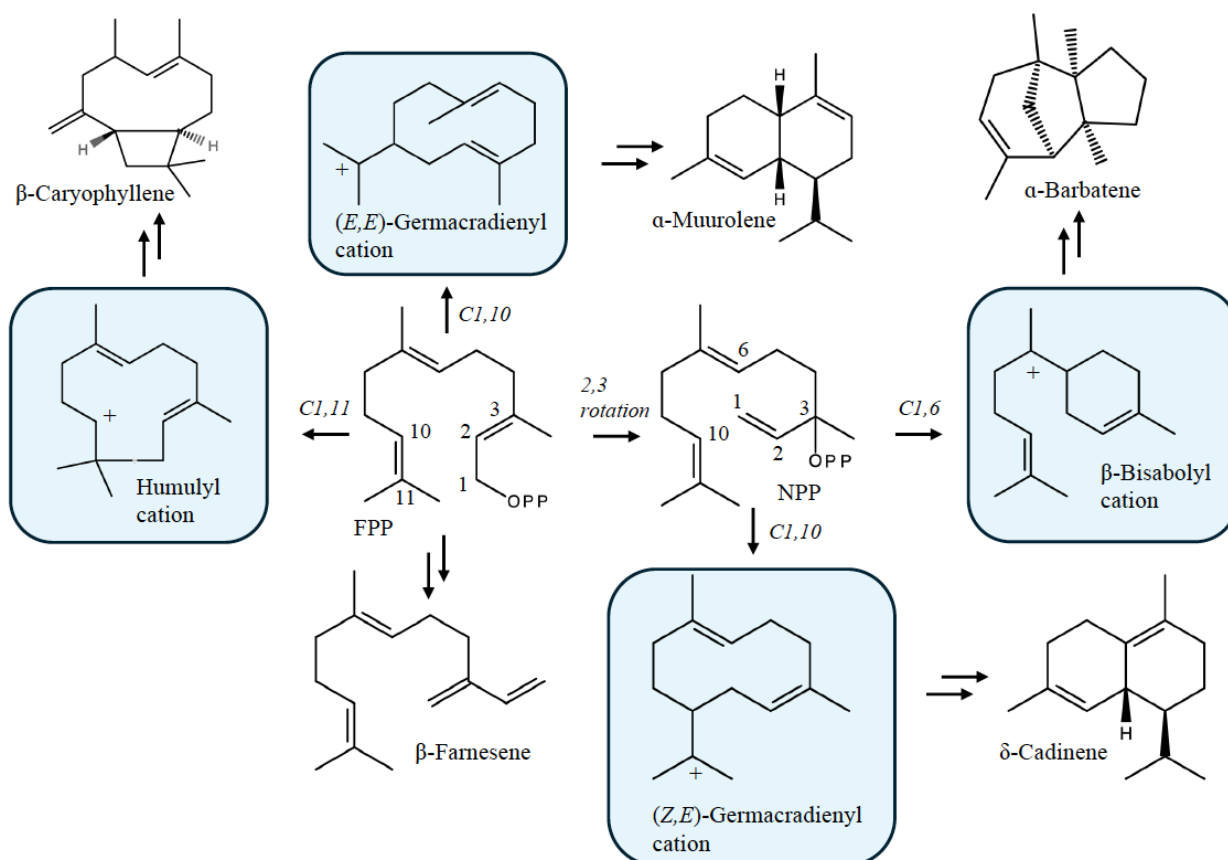


Figure 6. Four cyclizations that are possible for either FPP or its isomer NPP as well as acyclic formations, with β -farnesene as example. The four cations are further formed into the final terpenes. Four compounds are shown, described in sources as possible final products from the specific cyclization pathway. OPP - diphosphate group. Compounds and the figure are made by the author (28).

2.3.2.2.5 Diterpene synthases

Diterpenes can be produced in different ways, depending on if di-TPS class I or class II is involved (Figure 4). Class I enzymes serve as monofunctional enzymes and will perform a one-step reaction, like the route of sesquiterpene synthases. However, a two-step reaction may also occur when class II cyclizes GGPP, followed by class I that cleave the diphosphate group and further cyclize the compound. Fungal di-TPS can carry the class I α -domain (N terminal) and the class II $\gamma\beta$ -domains (C terminal) in the same so-called bifunctional di-TPS enzyme (23).

Chimeric enzymes exist furthermore in another form in fungal species, as enzymes that carry either PT+class I TPS or PT+class II TPS functions, termed prenyltransferase terpene synthases (PTTSs). The PTTSs allows the polymerization of the direct precursor and further cyclization to form the terpene scaffolds. For di-TPS, this entails that activity from GGPPS is present to form GGPP, along with activity from TPS, either class I or II, to form diterpenes. The PT may play an important role in substrate specificity as the TPS may accept various linear prenyl chains, but the PT will provide the exact correct ones. This is important for organisms to avoid producing unnecessary products (31).

2.3.2.2.6 Triterpene synthases

The biosynthesis of triterpenes is not very understood in fungal species, as the research has mostly focused on plant triterpenes (23). Tri-TPS, more commonly known in eukaryotes as oxidosqualene cyclases (OSC), will synthesize 2,3-oxidosqualene substrates into triterpene scaffolds, where 2,3-oxidosqualene is formed from squalene by an enzyme called squalene monooxygenase (Figure 3). 2,3-oxidosqualene is in fungi mainly further transformed into lanosterol or the related molecule protostadienol through the lanosteryl cation and protosteryl cation. The OSC for lanosterol formation is called lanosterol synthase, and it is the main OSC enzyme in fungi. Lanosterol is the precursor to several triterpenoids as well as to previously mentioned ergosterol that is essential for membrane formation. The tri-TPS carry class II enzyme function meaning they present a protonation-dependent reaction (2, 23).

2.3.2.3 Tailoring enzymes

The tailoring enzymes have the final job of modifying some terpenes to terpenoids. These modifications can be complicated, and if highly complex terpenoids are produced, several reactions by a variety of tailoring enzymes might be needed. The tailoring enzymes are a large group with various activities, such as cytochrome P450, transferases, and isomerases. Oxidation of the terpene skeleton is the most common method, where cytochrome P450 are the major tailoring enzymes for this. It has been estimated that 97 % of all terpenoids are modified by cytochrome P450 enzymes (35). However, dioxygenases or dehydrogenases can also perform the oxidation to alcohols, acids, or aldehydes. Furthermore, transferases can create esters and glycosides (2).

Fungi have been understood to display a physical closeness of the enzymes required for production of final terpenoids, meaning that genes coding the enzymes can be found in clusters in the genome. This is highly important for the identification of terpenoid biosynthetic pathways, as the surrounding genome of an explored TPS can be studied to detect possible tailoring enzymes for further modifications of the produced terpene scaffold, such as the cytochrome P450s enzymes (11).

2.4 The unexplored fungal TPSs

Fungal TPS genes and terpenoid clusters are rather unexplored compared to plant TPS, even though fungi species have been estimated to reach up to several million while plants have been estimated to exist in numbers of 400,000. This opens great opportunities and needs for exploring fungal species to understand the biosynthesis pathways they exhibit for various terpene products and modified structures. Especially the terpenoid pathways and products from species in the mushroom-forming phylum Basidiomycota are unexplored compared to Ascomycota, despite the fact Basidiomycota constitutes 30 % of the documented fungal species and genome studies of Basidiomycota and Ascomycota suggest that Basidiomycota fungal strains carry an average of 12 ses-TPS genes in the genome, whereas Ascomycota presents 3.5 ses-TPS genes on average (23, 36).

As of 2023, a bit more than 150 ses-TPSs from mushroom-forming fungal species had been characterized (37). Mono-TPSs in fungal species are less studied than sesqui-TPSs. Sesqui-TPSs have been shown to be able to produce monoterpenes from the GPP substrate, but few true mono-TPSs have been present in studies (23).

The fungal genome is becoming more available due to development in DNA sequencing methods and various databases such as JGI offer a large library of sequenced fungal genomes. Once interesting genes have been identified, such as TPS genes, they can be further studied by using heterologous expressions. There are plenty of hosts available for this.

2.5 Host systems for terpene synthase expression

The expression of terpenes and further terpenoids from fungi organisms in host systems is advantageous as it allows the isolation of the fungal genes of interest to study them one by one. This provides a better understanding of the specific biosynthetic pathway. Furthermore, as the production of secondary metabolites usually are triggered by environmental changes, it may be very difficult to map the complete terpene product formation by solely analyzing fungal samples in the laboratory and genome studies are therefore necessary (38). The process of implementing the pathways of interest includes first obtaining the single TPS gene or cluster of enzymes. Then, the gene(s) is inserted into the host genome or introduced through a non-integrative plasmid vector system. There are various factors to consider when choosing a host. Two host systems are highly used today for terpenoid production, namely the bacterium *Escherichia coli* or the yeast *Saccharomyces cerevisiae*.

2.5.1 *Escherichia coli*

E. coli is advantageous to use as the cells grow very fast, allowing fast paced laboratory work. Furthermore, this organism is well-known and a diversity of tools for engineering exist (39). The bacteria produce small amounts of terpenoids, allowing a low background (40). *E. coli*, being a prokaryote, uses the MEP pathway to produce terpenoid precursors. The MEP pathway starts with the two substrates pyruvate and glyceraldehyde-3-phosphate and produces IPP and DMAPP through several steps. Despite the higher theoretical yield compared to the MEV pathway, *E. coli* has been shown to suppress this pathway. Therefore, plenty of research has been put into engineering the MEV pathway into *E. coli*, to enable greater precursor production (41). Lastly, *E. coli* will as host system display production issues of eukaryotic cytochrome P450 due to folding issues, as *E. coli* lacks an endomembrane system for P450 attachment (42).

2.5.2 *Saccharomyces cerevisiae*

S. cerevisiae is a well-established and commonly used host system, with a well-known genome, high robustness, and a rather fast growth. A reason to use yeast in production of final terpenoid structures, instead of for instance *E. coli*, is the organism's ability to perform modifications on the terpene scaffolds due to its good production of the eukaryotic cytochrome P450 (42). Yeast has no natural production of terpenes and therefore provides low background signals in constructed production (43).

Even though mono-, sesqui-, and diterpene products are not produced naturally in yeast, the possibility exists as yeast produce the precursors IPP, DMAPP, but also the direct ones such as GPP and FPP. The biosynthesis of precursors in *S. cerevisiae* is through the MEV pathway, in the cytosol (43). Previous works show that yeast strains expressing TPSs excrete terpenes into the culture medium where the volatile compounds (mainly mono- and sesquiterpenes) can be absorbed in the head-space or where overlay methods with for instance hexane can retain the volatile and lipophilic compounds (44, 45). However, larger terpenes have greater tendency to accumulate in the yeast cell than the smaller terpene

compounds (45). Excretion by microorganisms is especially important for cells when the metabolic pathways are unbalanced and the cell homeostasis compromised, as happens when modified strains are induced (46).

IPP and DMAPP will naturally in yeast cells be mostly directed to production of FPP and further to the membrane compound ergosterol. FPP is therefore an important precursor in yeast. However, unlike FPP, GPP is not abundant in yeast, as no other metabolites arise from GPP for yeast function. The enzyme Erg20p in yeast carries bifunctionality of GPPS and FPPS activity and hence catalyzes the two steps, DMAPP and IPP to GPP and further GPP to FPP (Figure 7). The high activity of Erg20p seems to be the explanation for why GPP concentrations in yeast are low. This further seems to be the reason why yeast have presented low yields of monoterpenoids (47). Looking at other precursors, GGPP may also be produced but as GPP in lower concentrations. Based on this, yeast systems have been successful at producing various sesquiterpene compounds but less successful in mono- and diterpene production (48).

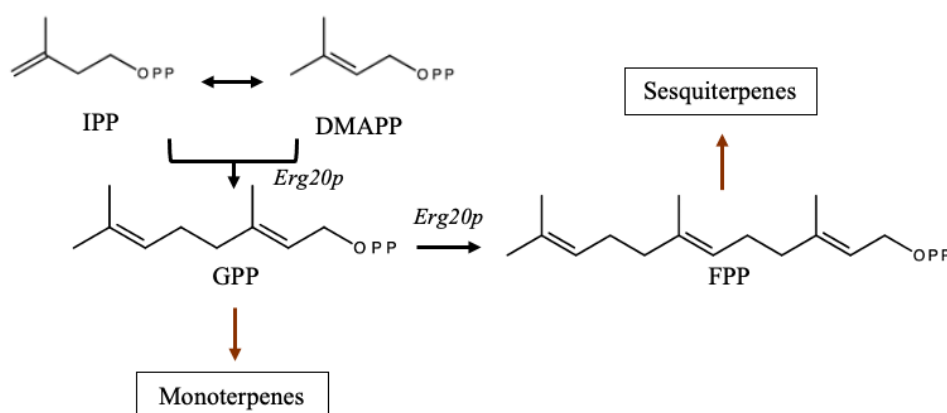


Figure 7. The production of IPP and DMAPP to GPP and further to FPP is in yeast produced by the bifunctional enzyme Erg20p. Figure is created by the author (49).

2.6 Methods for terpene synthase study

Several factors can be considered when choosing a setup for the TPS expression in the chosen host system. Gene cloning method, transformation of the yeast cells, enzyme production, and product detection are four essential steps.

2.6.1 Gene cloning

The TPS genes can be introduced into the host system through chromosomal integration or episomal plasmids, specifically multicopy or centromeric plasmids. Chromosomal integration is a more controlled and uniform method as the final vector numbers may vary cell to cell. Integration is also a better method for creating more stable strains. However, gene introduction with episomal plasmids is advantageous over integration when considering its simple procedure (50). USER cloning is one way to construct plasmids.

2.6.1.1 USER cloning

Uracil-specific excision reagent cloning allows insertion of multiple gene sequences into a vector. The vector contains a total of three recognition sites, one for a restriction enzyme and two for nicking-enzymes. Treatment with the enzymes will produce a vector with 3' overhangs in each strand. The gene

sequences act as template for PCR, using specifically designed primers carrying an overhang at the 5' end. The overhang is complementary to the vector overhang and has one thymine (T) replaced with an uracil (U). DNA polymerase will build the strands and the final product displays a slightly longer DNA sequence, with overhangs. Enzymes, such as uracil DNA glycosylase and DNA glycosylase-lyase, perform uracil excision and backbone breakage, allowing uracil overhang removal. The resulting 3' overhang will allow the ligation of sequences to vectors. If multiple sequences, such as promoter and gene of interest, are to be inserted, the same method is applied where the overhangs in the middle complement each other (51).

2.6.2 Transformation

In the transformation step, the gene of interest is introduced into the host. The yeast often needs to be prepared into competent cells for the transformation. Methods used for plasmid transformation are for instance the electroporation method where voltage is applied to the cell solution, which creates pores for the DNA to enter. Sorbitol and other chemicals such as lithium acetate are included to create electrocompetent cells. The lithium method includes creating competent cells in lithium, such as lithium acetate, and polyethylene glycol (PEG), adding the DNA, and exposing the cells to a heat shock treatment. A third method can be using glass beads, however a rather low transformation efficiency has been obtained. After the transformation, the cells may be plated and grown for selection (52).

2.6.3 Enzyme production

Utilizing a constitutive or inducible promoter are two ways to produce the chosen enzymes in yeast. Constitutive promoters are not controlled by stimuli, but instead are continuously active in the cells. By including an inducible promoter in a vector system, the enzyme production may be controlled in the host. Common inducible promoters in yeast are the galactose-inducible promoters, such as GAL1 or GAL2, which are proven to be strong promoters with high expression levels (53).

2.6.4 Compound detection

When the host system has produced terpenes, an extraction step collects the products to run them in GC-MS. As has been mentioned in Section 2.5.2, the formed terpenes can be extracted with different methods. Two common ones are the solid phase microextraction (SPME) or using an organic overlay, such as hexane. Overlay is a solvent-based approach, where the products are extracted into the top layer and the extracted phase can be injected into the GC column. As terpenes are rather non-polar, an organic phase can allow them to partition into this phase well. SPME instead uses a coated fiber to solvent-free extract the compounds, either from the liquid of a sample with direct-immersion SPME or from above the sample with headspace SPME (HS-SPME) (54).

The extraction in SPME is either through adsorption or absorption whereas the desorption is thermal or solvent based. The fiber coating will affect the analysis, as how well the analyte is retained will depend on its interaction with the coating. The interaction depends for instance on polarity and size of the analyte (55). SPME is advantageous due to the great isolation and enrichment of the analytes. Furthermore, the method is solvent-free and especially HS-SPME is non-destructive to samples as the extraction occurs without physical contact (54). HS-SPME extracts analytes of good or semi-good volatility well.

Furthermore, less interference from extraction of unwanted cell compounds can be accomplished in HS-SPME compared to direct-immersion SPME, where the fiber is in direct contact with the sample (56).

3. Preparatory work

3.1 Tree assembly

Preparatory work for the master's thesis included selection of species and construction of a phylogenetic tree that presented the TPS genes from selected species. Furthermore, TPS genes to be studied were selected from the tree. The preparatory work was performed by supervisor Chi Zhang.

The phylogenetic tree was constructed by firstly identifying TPSs from genomes of edible mushrooms and fungi. This was done with the Hidden Markov Model (HMM), using several TPS domain profiles from InterPro (57, 58). TPS protein sequences in the fungal genomes were found, with an E-value of 10^{-5} used as threshold for significance. TPS candidates were chosen based on a few requirements, E-value and sufficient protein sequence length being two. These candidates together with several characterized TPSs from previous works were aligned and FastTree was applied to build the phylogenetic tree (58).

3.2 Selection of TPS genes

TPS genes in the phylogenetic tree were selected in batches. Figure 8 presents the created phylogenetic tree as well as the selected TPS genes for this work. The TPS genes originated from the shiitake mushroom *Lentinula edodes* (Lenafn1, Lened1, LedB17, Lenedo1, Lentinedodes1, LedVB361, and LedTMI1633), the oyster mushroom *Pleurotus ostreatus* (PleosDSM11191 and PleosPC9), the porcini mushroom *Boletus edulis* (Boledp1 and Boled5), and the jelly fungi *Auricularia subglabra* (Aurde3), all from the phylum Basidiomycota. Additionally, genes were chosen from cheese fungi of phylum Ascomycota, *Penicillium roqueforti* (Proq16744 and Proq16728), *Penicillium biforme* (Pcas15300 and Pbif18567), and *Penicillium camemberti* (Pencam1).

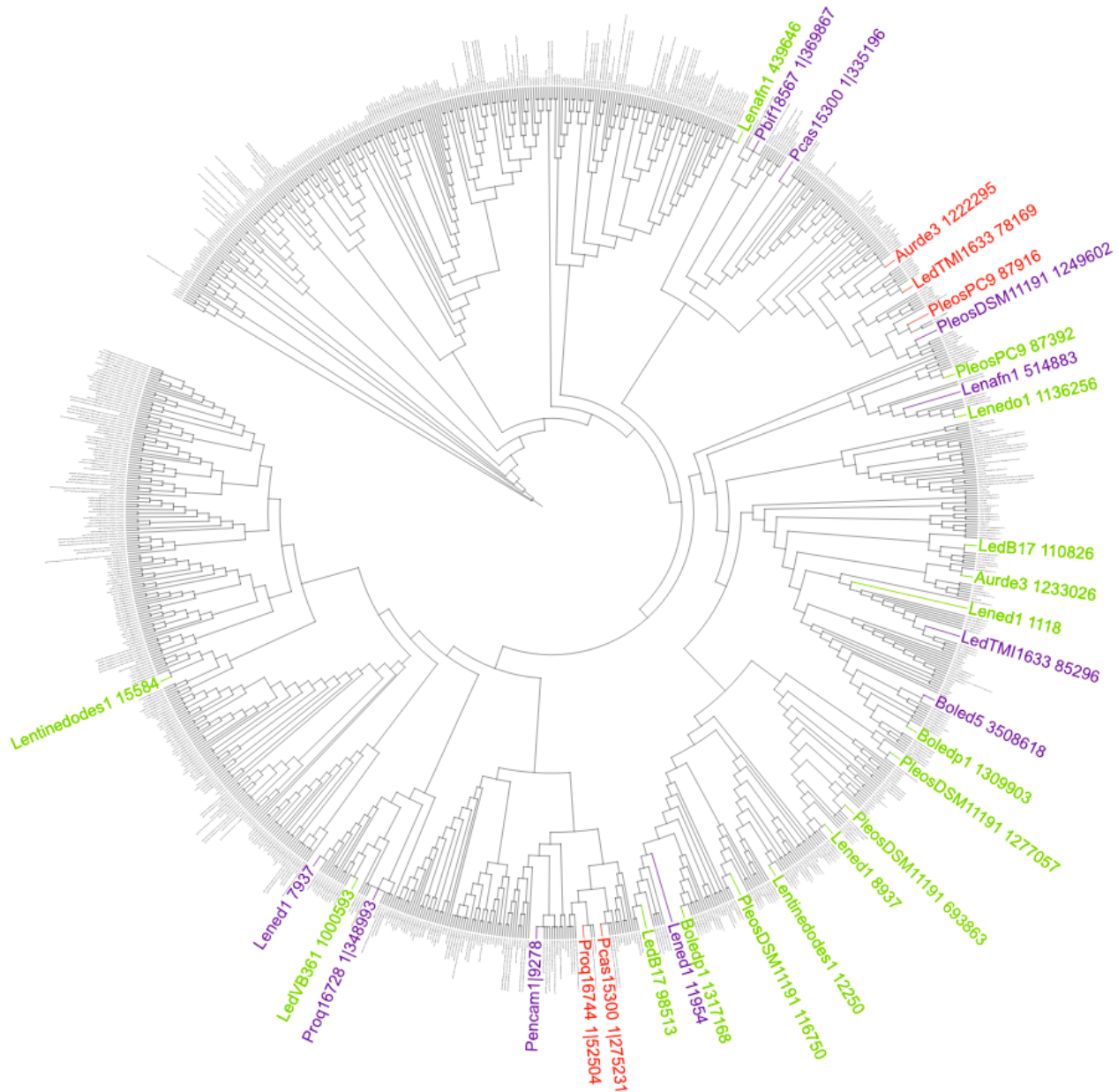


Figure 8. Phylogenetic tree used in this work, created by Chi Zhang and modified by the author. Genes are labeled with fungi origin and code. The genes were selected at three times (batches) where green is from batch 1, red is from batch 2, and purple is from batch 3. See link attached for visualizing a high-resolution tree. All genes starting with CHA are characterized TPSs. The phylogenetic tree was obtained in UGENE where the 31 selected TPS genes were modified with larger font and color (58).

4. Material and Methods

4.1 Material

4.1.1 Gene sequences

Throughout the project, a total of 31 TPS genes were expressed in the host system for product analysis (Table 2). The genes are divided into 3 batches, according to their time of selection. The gene and primer sequences were ordered from TAG Copenhagen A/S. The GAL1 promoter prepared for USER cloning was obtained from PBio lab. Gene sequences for TPS and primers are listed in Table S4 and S5. Primers for ADH1 and Cyc1 were obtained from the PBio lab.

Table 2. Code for the selected gene as labeled in the phylogenetic tree, ID, and species (name, strain, and version) that it was obtained from. All fungal strains can be found at JGI (59).

Gene code	ID	Origin Species strain version (v)
Batch 1		
439646	A1	<i>Lentinula edodes</i> sp2 HRB7682 ss15 v1.0
693863	B1	<i>Pleurotus ostreatus</i> DSM 11191 v1.0
1118	C1	<i>Lentinula edodes</i> NBRC 111202
87392	D1	<i>Pleurotus ostreatus</i> PC9 v1.0
98513	E1	<i>Lentinula edodes</i> B17 v1.3
8937	F1	<i>Lentinula edodes</i> NBRC 111202
1309903	G1	<i>Boletus edulis</i> Prilba v1.0
1136256	H1	<i>Lentinula edodes</i> Le(Bin) 0899 ss11 v1.0
1317168	I1	<i>Boletus edulis</i> Prilba v1.0
12250	J1	<i>Lentinula edodes</i> W1-26 v1.0
1233026	K1	<i>Auricularia subglabra</i> v2.0
15584	L1	<i>Lentinula edodes</i> W1-26 v1.0
110826	M1	<i>Lentinula edodes</i> B17 v1.3
116750	N1	<i>Pleurotus ostreatus</i> DSM 11191 v1.0
1000593	O1	<i>Lentinula edodes</i> VB361 v1.0
1277057	P1	<i>Pleurotus ostreatus</i> DSM 11191 v1.0

Gene code	ID	Origin Species strain version (v)
Batch 2		
87916	A2	<i>Pleurotus ostreatus</i> PC9 v1.0
52504	B2	<i>Penicillium roqueforti</i> ITEM 16744 v1.0
275231	C2	<i>Penicillium biforme</i> ITEM 15300 v1.0
78169	D2	<i>Lentinula edodes</i> TMI1633 v1.0
1222295	E2	<i>Auricularia subglabra</i> v2.0
Batch 3		
369867	A3	<i>Penicillium biforme</i> ITEM 18567 v1.0
1249602	B3	<i>Pleurotus ostreatus</i> DSM 11191 v1.0
7937	C3	<i>Lentinula edodes</i> NBRC 111202
335196	D3	<i>Penicillium biforme</i> ITEM 15300 v1.0
85296	E3	<i>Lentinula edodes</i> TMI1633 v1.0
3508618	F3	<i>Boletus edulis</i> BED1 v4.0
348993	G3	<i>Penicillium roqueforti</i> ITEM 16728 v1.0
11954	H3	<i>Lentinula edodes</i> NBRC 111202
514883	I3	<i>Lentinula edodes</i> sp2 HRB7682 ss15 v1.0
9278	J3	<i>Penicillium camemberti</i> FM 013

4.1.2 Vector system

High-copy plasmid pWUS was used for all experiments, carrying an ampicillin resistance gene (AmpR) for bacterial selection as well as a tryptophan (Trp, W) marker to be able to select for transformed yeast cells. The vector consisted of 5,736 bp and was provided by PBio lab. In Figure 9, pWUS and its specific genes are shown. ADH1 and Cyc1 are backbone terminators in the pWUS vector. The vector was cut prior to the thesis work with AsiSI as well as Nb.BsmI at the restriction sites to create the correct destination for gene ligation, see the USER cassette in Figure 10.

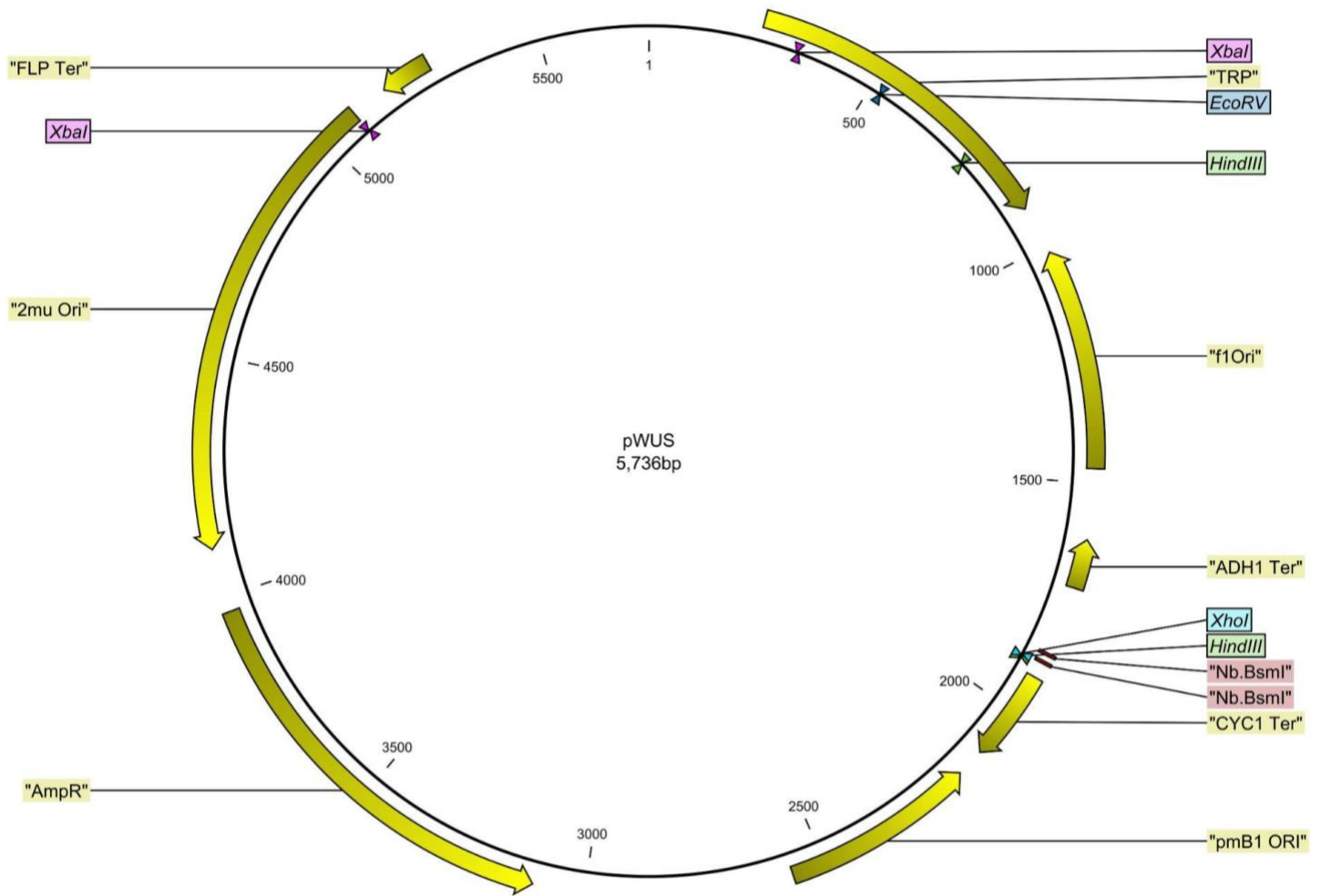


Figure 9. The high-copy plasmid pWUS of 5,736 base pairs was utilized for plasmid construction, carrying ADH1 and Cyc1 backbone terminators and Nb.BsmI restriction sites for gene insertion. The vector carried an ampicillin resistance gene (AmpR) for bacterial selection as well as a tryptophan gene (Trp) for yeast selection. The figure of pWUS was obtained from CLC Main Workbench 22 (QIAGEN), made prior to the thesis work.

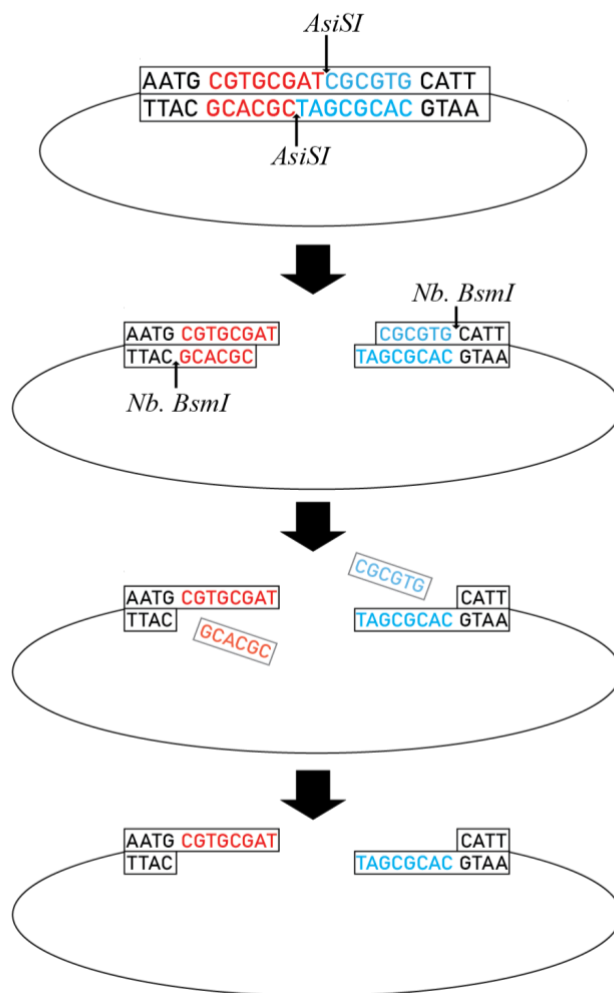


Figure 10. The pWUS vector carried a USER cassette, consisting of 14 bp. The USER cassette was cut previously to the thesis work at *AsiSI* and *Nb.BsmI* to create the final cut pWUS vector, ready for gene ligation. The figure is made by the author.

A pLUS vector with a gene for Erg20p(N127W) inserted was used for co-transformation. This vector consisted of 6,970 bp and carried instead a selection marker for leucine (Leu, L). The vector was provided by PBio lab.

4.1.3 Strains

For sub-cloning of the pWUS vector carrying promoter and TPS gene, the bacterium *E. coli* was utilized. For expression of terpene synthases and terpene production, yeast cells were transformed with the purified pWUS plasmid that had been subcloned in *E. coli*. For terpene synthase expression, *S. cerevisiae* AM158 was used, constructed for greater FPP and GGPP production, with deletion of non-essential genes for increased terpene production (60). AM158 was tryptophan (W), uracil (U), leucine (L), and histidine (H) auxotroph.

4.1.4 Media

W, U, L, and H were essential for AM158 cell growth. W and L were provided through the pWUS vector and the pLUS vector, respectively. *S. cerevisiae* AM158 was cultivated in Yeast extract Peptone Dextrose (YPD) medium supplemented with 40 % glucose. *E. coli* was cultivated in Lysogeny broth (LB) medium. For transformation, Solution 1 and Solution 2 (lithium acetate and PEG3350) were both provided by PBio lab. The Glu-UHWL medium for growth of transformed *S. cerevisiae* cells contained H₂O and equal parts glucose, OD, and YNB, prepared by PBio labs. The GalRaf-UHWL medium for induction of TPS genes was obtained from PBio lab and contained H₂O and equal parts galactose+raffinose, DO, and YNB, where galactose induces the gene transcription. For verification on gel agarose, GelRed was used, provided by PBio labs. The 5X Assay buffer for *in vitro* assay consisted of 100 mM MgCl₂, 2.5 mM MnCl₂, 50 % (w/v) glycerol, and 250 mM HEPES-KOH (pH 8).

4.2 Gene and primer preparation

The genes and primers obtained from TAG Copenhagen A/S were obtained as powder and as concentrated solutions respectively. The genes were dissolved in 50 µL H₂O and the primers diluted with H₂O to final concentrations of 100 pmol/µL. The primers were diluted an additional 10X in H₂O before further usage.

4.3 Plasmid construction

The selected genes with GAL1 promoter were introduced into the pWUS vector with USER cloning (51). All genes from batch 1, 2, and 3 were introduced with the same technique. For sequence preparation, Q5U PCR was performed by mixing 20 µL H₂O with 2 µL each of gene specific forward and reverse USER primers that carried flanking uracil parts. 1 µL of the gene template was added followed by 25 of µL NEBNext® Q5U® Master Mix (NEB). The products were verified on gel, where 2 µL GelRed was mixed with 3 µL of the PCR product and 3 µL loaded. The gel ran for approximately 10 minutes. The PCR products were purified with MicroElute Cycle-Pure Kit (Omega Bio-Tek), according to given instructions. For ligation, 0.5 µL of cut vector was mixed with 0.5 µL purified gene and 0.5 µL USER prepared GAL1 promoter, followed by addition of 1 µL cutsmart buffer (NEB) and 2 µL water. Lastly, 0.5 µL USER® Enzyme (NEB) was added and the solution ran in PCR. Figures 11 and 12 present the ligation procedure and final plasmid construct. Plasmid construction for batch 1 was performed prior to this thesis work.

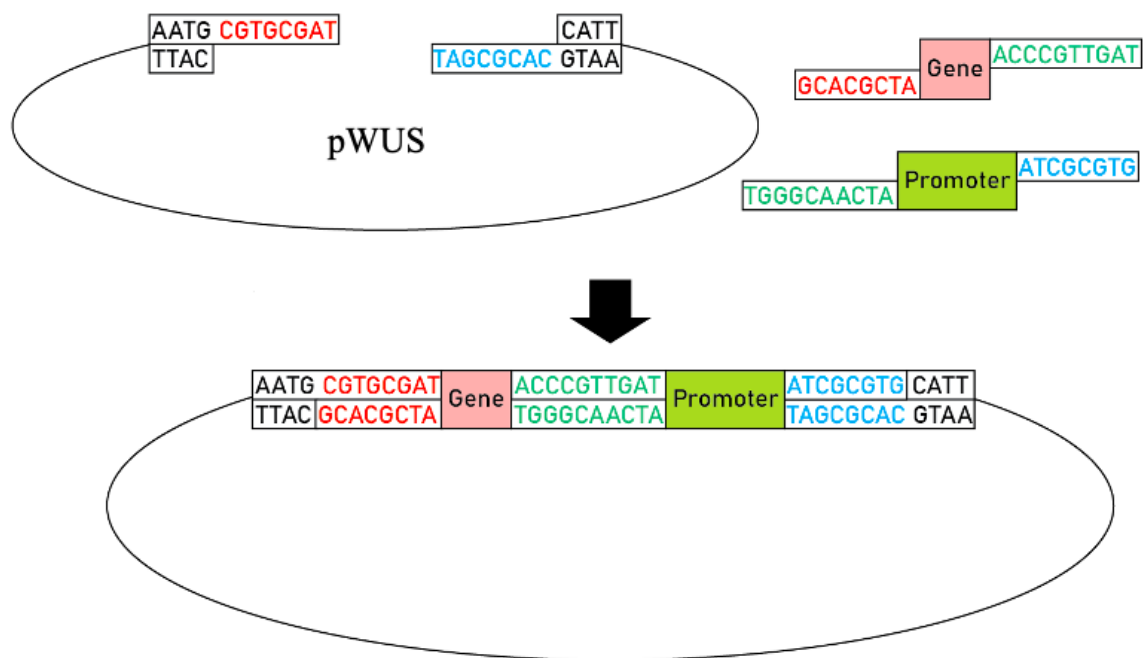


Figure 11. Promoter and TPS gene were prepared through the USER cloning technique by PCR with uracil containing primers followed by treatment with USER® Enzyme, allowing the presented promoter and gene to ligate into the cut pWUs vector (Figure 10). The figure is made by the author.

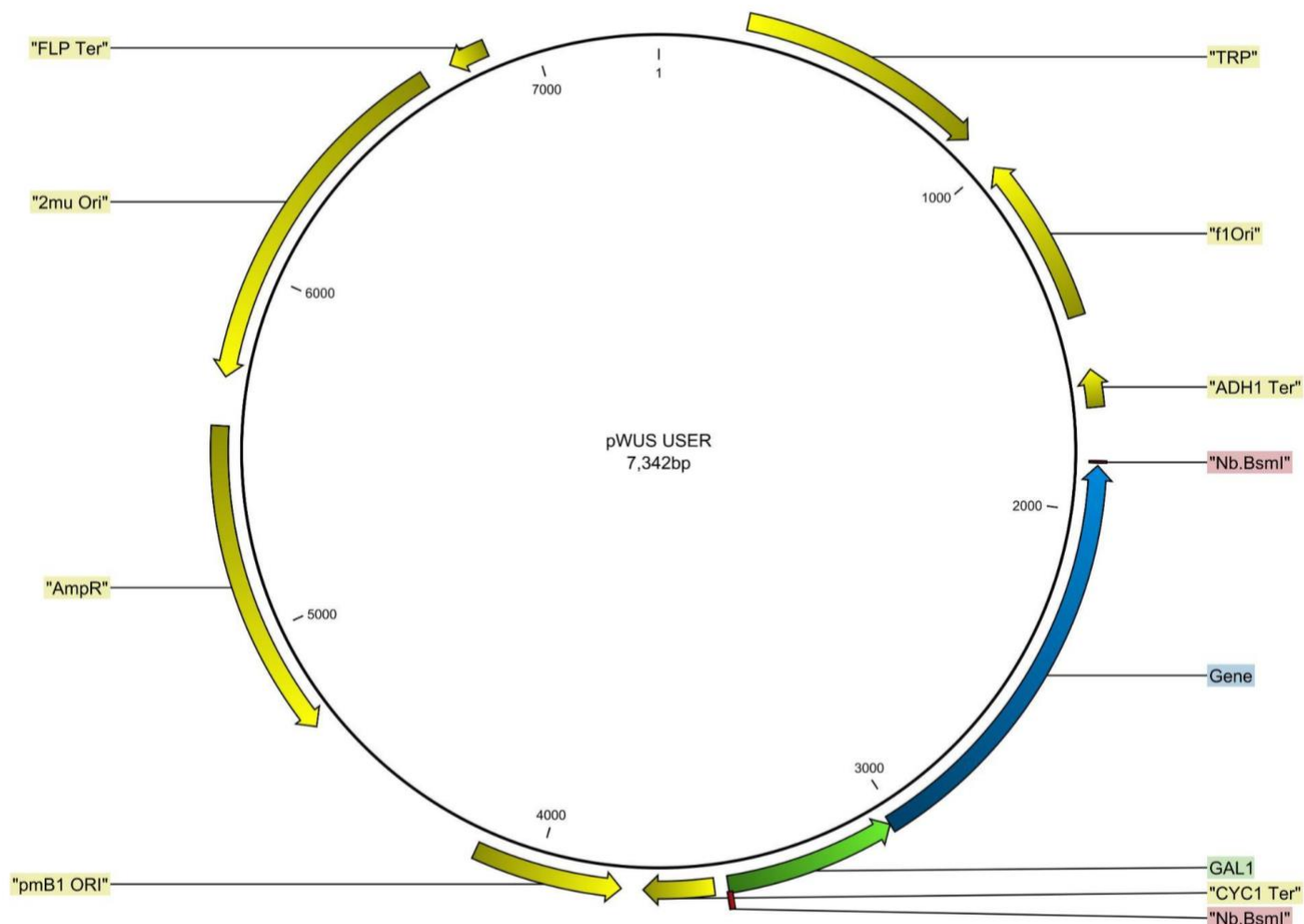


Figure 12. The high-copy plasmid pWUS was utilized for plasmid construction, carrying ADH1 and Cyc1 backbone terminators and Nb.BsmI restriction sites for gene insertion. The vector carried an ampicillin resistance gene (AmpR) for bacterial selection as well as a tryptophan gene (Trp) for yeast selection. The GAL1 promoter and the studied TPS gene (blue labeled fragment) were inserted between Nb.BsmI restriction sites with USER cloning. The figure is modified from CLC Main Workbench 22 (QIAGEN).

4.4 Verification of plasmid construct

E. coli was transformed to culture the plasmid constructs. Competent *E. coli* cells were chemically prepared with Mix & Go *E. coli* Transformation Kit (T3001) (Zymo Research). The prepared pWUS plasmids carrying the gene of interest were added to frozen competent *E. coli* cells and the mixture spread on LB plates with carbenicillin selection. The plates were incubated in 37 °C for 1 day.

Colony PCR was performed by mixing 5 µL 2x Master, 3 µL H₂O, and 1 µL of each primer. The primer combination was either the forward with ADH1 or the reverse with Cyc1. 8 colonies from the incubated plates were picked individually and added to each tube. The PCR products were run in a gel electrophoresis, with 3 µL of GelRed added as ladder and to each PCR product. The gel was run for approximately 10-15 minutes.

For plasmid sequencing, 2-3 colonies that were correct in the gel analysis were picked and transferred to a LB + carbenicillin medium for cultivation. The liquid culture was then used for plasmid purification. The plasmid purification was performed with E.Z.N.A. Plasmid DNA Mini Kit I (Omega Bio-Tek), according to given instructions. The purified plasmids were mixed in two separate tubes with either the forward or the reverse primer, alternatively the backbone ADH1 or Cyc1 primers, and sent to Azenta for sequencing (Sanger sequencing). The obtained sequences were verified by assembling the gene sequence with the obtained sequences from forward and reverse primers to ensure that no mutations occurred and that an overlap between the two sequences existed. Plasmid construct verification for batch 1 was performed prior to this thesis work.

4.5 Preparation and transformation of *S. cerevisiae*

Yeast strain AM158 was inoculated overnight in YPD medium in a 30 °C shake incubator, changed into fresh medium with an inoculate concentration of 5 %, and finally incubated at 30 °C for an additional 3-4 hours. The OD was measured when the cell concentration visually seemed to be too high. 5 mL yeast cell extract was spun down at 4200 rpm. The supernatant was removed, and pellet washed with 250 µL Solution 1. Solution 1 was spun down, and supernatant removed. Pellet was resuspended in 25 µL Solution 1 + 10 µL ssDNA (D9156, Sigma Aldrich) + 250 µL Solution 2. 5 µL DNA was added, mixed well, and the solution was heat shocked at 42 °C for 30 minutes, at 700 rpm. Solution was spun down at 6,000 rpm, supernatant removed, and pellet resuspended in 1 mL water. The solution was spun down at the same speed, supernatant removed, and pellets resuspended in 50 µL H₂O. The solution was spread on glucose agar plates, supplemented with U, L, and H (50X), and incubated at 30 °C for 2-3 days. For co-transformations, the same protocol was used, with an additional 5 µL of pLUS+Erg20p(N127W) included, and solution spread on agar supplemented with U and H (50X).

4.6 Induction of transformed yeast cells

5 colonies were picked from the incubated plates and added to 5 mL Glu-W or Glu-LW medium (50X diluted amino acids). The solution was incubated overnight at 30 °C shaking. Either the cells were induced or changed into fresh glucose medium until induction could be performed. Induction was performed by transferring a volume of the culture to a new tube, corresponding to OD 0.5. The cells were spun down at 1500 rpm for 3 minutes and washed twice with 3 mL H₂O. The clean pellet was resuspended in 2 mL of GalRaf-W or GalRaf-LW medium (50X diluted amino acids) and transferred to 20 mL SPME vials.

4.7 *In vitro* enzyme assay

The induction for the *in vitro* assay was performed according to Section 4.6, however 25 mL GalRaf-W or GalRaf-LW was used. The induced yeast cells were spun down and washed with 5 mL H₂O, then mixed with 500 µL 50 mM HEPES-KOH (pH 8) and 500 µL 0.5 mm glass beads. The solution was vortexed and cooled on ice 1 minute each in 10 cycles, followed by centrifugation at 9000 rpm for 5 minutes. 95 µL supernatant was extracted and mixed with 50 µL 5X Assay buffer, 75 µL 100 mM DTT, 20 µL H₂O, and 10 µL isoprenyl diphosphate substrate in room temperature, followed by an immediate addition of 250 µL hexane layered on top. The solutions were added in vials and stored at 30 °C

overnight. The vials were shaken and centrifuged, followed by extraction of 80 μ L hexane phase. To test class II activity and prenyltransferases, the GGPP vials were mixed with 250 μ L 2M HCl/83% EtOH, incubated 10 minutes at 37 °C, shaken, and 175 μ L 10% NaOH added. The vials were shaken, centrifuged, and another 80 μ L hexane phase was extracted from the vials. GC-MS program B was used for analysis as described in Section 4.8.

4.8 GC-MS analysis

All terpene products were analyzed with GC-MS, a Nexis GC-2030 combined with a single quadrupole GCMS-QP2020 NX GC-MS spectrometer (Shimadzu), with an EI ionization method. Nitrogen was used as carrier gas. The terpenes produced *in vivo* were extracted from the headspace above the sample using a HS-SPME fiber, allowing the terpenes to accumulate in the coating on the HS-SPME fiber. The terpenes produced in the *in vitro* enzyme assay (Section 4.7) had been extracted to the hexane phase and were injected through a liquid injection tool. The coating on the HS-SPME fiber was 50/30 μ m Divinylbenzene/Carboxen/Polydimethylsiloxane, which according to the provider was compatible with volatiles and semi volatiles of sizes MW 40-275 and carbon number between 3-20 (57348-U, Sigma Aldrich). The HS-SPME fiber and the liquid injection tool could both be mounted to the GC-MS. The column in GC (HP-5MS) had a length of 30.0 m, 0.25 mm inner diameter, and film thickness of 0.25 μ m. The maximum temperature of the column was 345.0 °C.

GC-MS programs are labeled A and B, both slightly adjusted throughout the project. Program A was used for the HS-SPME injection method. The program had an injection temperature of 250 °C and the analysis was performed with a starting column temperature of 50 or 60 °C. 3 °C/min rate was kept to 170 °C, followed by 40 °C/min up to 260 °C, with a holding time of 3 min. The ion source temperature was 250 °C and the scan range was 40 *m/z* - 300 *m/z*. Program B was used for samples in hexane solvent, with an injection temperature of 250 °C and inlet column temperature of 60 °C. 10 °C/min was applied to reach 300 °C, with a holding time of 6 min. The ion source temperature was 250 °C and the scan range was either 40 - 350 *m/z* or 50-350 *m/z*. Terpenes were verified by extracting *m/z* 93, a common ion in fragmentation of terpenes (61). Peak integration was performed using TIC chromatograms and selecting all likely terpene peaks. The abundance in percentage of each compound produced by the TPS was calculated.

Two methods were used to study fragmentation patterns, the cutoff at 90% similarity to library patterns was used for a simple approach in combination with a more advanced approach where the fragmentation patterns were visually studied to conclude similarity, especially verifying the presence of the main ion peaks and somewhat the ratio between them. The GC-MS libraries NIST08 and Wiley Registry 8th were used for terpene compound identification, as well as the external library *Identification of Essential Oil Components by Gas Chromatography/Mass Spectrometry*, ed 4.1, providing fragmentation spectra (62-64). Arithmetic Index (AI) with retention time (RT) were provided for compounds, where AI was calculated with an equation, listed in the book, as a function of the C4-34 paraffin RT values. To be able to use the external library a linear regression was made from the AIs and RTs of a sesquiterpene standard, containing longifolene, trans-caryophyllene, α -humulene, allo-aromadendrene, valencene, and farnesol, that was run with the selected TPSs. AIs of the uncertain peaks were then calculated from the linear regression using their RTs, and fragmentation patterns with a variance of around $AI \pm 4$ were visually

studied to identify probable compounds. The focus of the comparison lied on identifying the main ion peaks in the provided spectra that existed in the unknown compound.

5. Results and Discussion

The results from the *in vivo* analysis, using HS-SPME as extraction method, as well as the *in vitro* analysis, using hexane extraction, will be presented and discussed. All analyses were performed with a negative control (yeast carrying empty pWUS plasmid) to verify TPS product formation. The products were identified as described in Section 4.8 where the uncertain products (similarity < 90 % to GC-MS library fragmentation pattern) were manually studied to understand whether the suggested product seemed correct (further discussed in the report). In the case of several possible products with similar fragmentation similarity scores, the terpene product with the highest similarity score was chosen.

5.1 *In vivo* analysis

All 31 TPS genes, inserted individually in pWUS vector with the GAL1 promoter using USER cloning, were induced in GalRaf medium in closed 20 mL vials and the terpenes formed from the transformant were analyzed by running the samples in the 20 mL vials in GC-MS. A HS-SPME fiber was utilized for terpene extraction in the headspace above the sample. A negative control was run simultaneously in the GC-MS program. The negative control was yeast transformed with an empty pWUS vector.

All 31 TPS genes in the pWUS vector were also transformed into yeast cells together with a pLUS-Erg20p(N127W). This pLUS plasmid had been ligated with a GAL1 promoter and a mutated Erg20p gene (Erg20p(N127W)) prior to thesis work. The transformants carrying both pWUS and pLUS were induced in GalRaf medium in 20 mL vials and the terpenes formed in the headspace extracted with HS-SPME fiber. For transformants containing pWUS-TPS and pLUS-Erg20p(N127W), a negative control was run simultaneously in the GC-MS program. The negative control was yeast transformed with an empty pWUS vector and the pLUS-Erg20p(N127W) plasmid.

The GC-MS analysis for transformants carrying pWUS-TPS (single transformation) or carrying both pWUS-TPS and pLUS-Erg20p(N127W) (co-transformation) were performed with Program A and identified as described in Section 4.8.

5.1.1 Single transformations

The genes listed in Table 2 were cloned individually into the pWUS vector with a GAL1 promoter and expressed in yeast cells. The products were analyzed with HS-SPME in GC-MS. Out of 31 transformants in total, 13 produced terpenes. Three strains showed mainly monoterpene production, nine showed mainly sesquiterpene production, and one showed diterpene production. The spectrum peaks are summarized in tables, divided based on the type of terpenes mainly produced. Products with at least 10% abundance out of total terpene production are included.

5.1.1.1 Monoterpene production

Products detected from TPS genes B2, C2, and J3 are presented in Table 3 (abundance out of total terpene production $\geq 10\%$). These three TPSs in Table 3 were analyzed for product formation once. B2 produced

the acyclic monoterpenes β -linalool (25.4 %), β -citronellol (36.6 %), and geraniol (14.0 %), eluting in accordance with their boiling points (65-67). Two other rather prominent peaks for B2 were geranyl acetate and citronellyl acetate (Figure S1). C2 produced ocimene (67.4 %) and the sesquiterpene farnesene (22.9 %), both acyclic, where farnesene is expected to elute later than ocimene as it carries 5 additional carbons. J3 produced two ocimene peaks, albeit of lower intensities than the other two TPSs. B2 is slightly more promiscuous with 13 terpene peaks in the chromatogram (Figure S1).

Table 3. Three TPSs, B2, C2, and J3, were identified to produce monoterpenes as major compounds. The products and their corresponding peak number labeled in Figure S1 are presented, with RT, abundance out of total terpene formations, and intensity. A combined percentage from remaining terpenes < 10% is included.

Gene ID	Total number of terpene peaks	Products	RT [min]	Abundance [%]	Intensity [x100,000]
B2	13	1 β -Linalool	14.2	25.4	8.2
		2 β -Citronellol	19.7	36.6	16.7
		3 Geraniol	20.9	14.0	7.1
		Remaining		24.0	
C2	9	1 β -Ocimene	12.1	67.4	32.2
		2 α -Farnesene (C15)	31.6	22.9	28.1
		Remaining		9.7	
J3	2	1 β -Ocimene	11.7	66.4	0.083
		2 Alloocimene	15.5	33.6	0.045
		Remaining		0	

5.1.1.2 Sesquiterpene production

Almost all TPSs in Table 4 produced a great variety of sesquiterpenes, with some uncertain and smaller peaks, providing the remaining unlisted peaks with rather high summed percentages. Especially B1 and I1 presented a total of 19 and 24 terpene peaks respectively (Figure S3). The idea that especially ses-TPSs produce many products has been mentioned in Section 2.3.2.2. From batch 1, G1 and O1 were analyzed once for product formation with GC-MS program listed in Table S6. B1, I1, N1, and P1 were analyzed twice, aiming to obtain a clearer chromatogram where the second GC-MS was run with Program A described in Section 4.8. The results obtained from analysis with Program A are presented.

The products are provided with RT, abundance, and intensity (Table 4). B1 presented three major products, β -cadinene, isodene, and δ -cadinene. G1 produced bicyclic valencene as its major product (50.2 %) as well as an improperly separated peak next to valencene, not included in the results. This gives G1 a rather high remaining percentage. I1 and N1 produced a variety of peaks, but with very similar chromatograms and products, with α -muurolene and δ -cadinene from both. I1 also generated tricyclic α -copaene at RT $\sim$ 26.3 min but <10% (Figure S3). O1 produced the acyclic β -farnesene (20.4 %) and nerolidol (46.8 %). P1 and D2 created peaks (>10 %) that presented poor similarity to the GC-MS libraries with main ion peaks not present (labeled *Unknown*). A2 showed β -himachalene (70.8 %) and cuparene (21.4 %) production. Lastly, A3 produced the bicyclic α -bergamotene as sole terpene product.

Chromatograms for G1 and O1 were produced by Chi Zhang with a previous program (Table S6), hence G1 and O1 have in general a shorter RT.

Table 4. The products with RT and peak percentage (abundance out of total terpene formations) are listed, as well as a combined percentage from remaining terpenes < 10%. The peak intensities are given for perspective. Peaks for G1 and O1 are produced in GC-MS by Chi Zhang (Table S6).

Gene ID	Total number of terpene peaks	Products	RT [min]	Abundance [%]	Intensity [x100,000]
B1	19	1 β -Cadinene	30.4	12.6	41.0
		2 Isoledene	31.4	10.3	24.2
		3 δ -Cadinene	32.5	50.1	104.3
		Remaining		27.0	
G1	10	1 Valencene	12.8	50.2	20.2
		Remaining		49.8	
I1	24	1 α -Amorphene	30.4	18.4	6.8
		2 α -Muurolene	31.4	19.1	8.7
		3 δ -Cadinene	32.3	26.7	11.7
		Remaining		35.8	
N1	11	1 α -Copaene	26.3	17.1	2.1
		2 Sinularene	30.4	26.7	0.68
		3 α -Muurolene	31.4	18.2	0.71
		4 δ -Cadinene	32.3	25.4	0.78
		Remaining		12.6	
O1	14	1 β -Farnesene	12.3	20.4	16.9
		2 Nerolidol	13.7	46.8	27.3
		Remaining		32.8	
P1	6	1 α -Cubebene	26.3	30.9	1.4
		2 Unknown	28.3	39.4	1.9
		3 δ -Cadinene	32.3	14.8	0.61
		Remaining		14.9	
A2	10	1 β -Himachalene	31.5	70.8	36.6
		2 Cuparene	31.7	21.4	11.9
		Remaining		7.8	
D2	14	1 Unknown A	27.9	26.3	84.3
		2 Unknown B	29.4	50.8	113.5
		Remaining		22.9	
A3	1	1 α -Bergamotene	32.0	100	0.55
		Remaining		0	

A2 has two peaks that are slightly present in the negative control, especially the second peak where cuparene appears to be (Figure S4). The fragmentation patterns are quite similar for both peaks to the respective control peak and the 93 ion peaks typical for terpenes are visible in the control (data not

shown). The negative control should not produce the terpene compounds at all, and therefore there is an uncertainty whether these are true product peaks. It is necessary to investigate the *in vitro* analysis further.

As for P1 and D2, the fragmentation patterns need to be further inspected. P1 has a low similarity score for peak 2 (Figure 13). However, by closely comparing the fragmentation of peak 2 to the provided library fragmentation pattern, the two patterns are rather similar. m/z 94 is a bit high compared to m/z 79 and 107, and the ion peaks of higher m/z seem to not correlate very well. However, the higher m/z ion peaks are all very small and hard to make out, perhaps they are not the deciding factor. The compound is not provided with a common name but is potentially a terpene of bicyclic structure (Figure 14) with the chemical name 1,2,5,5,8A-pentamethyl-1,2,3,5,6,7,8A-octahydro-naphthalen-1-ol.

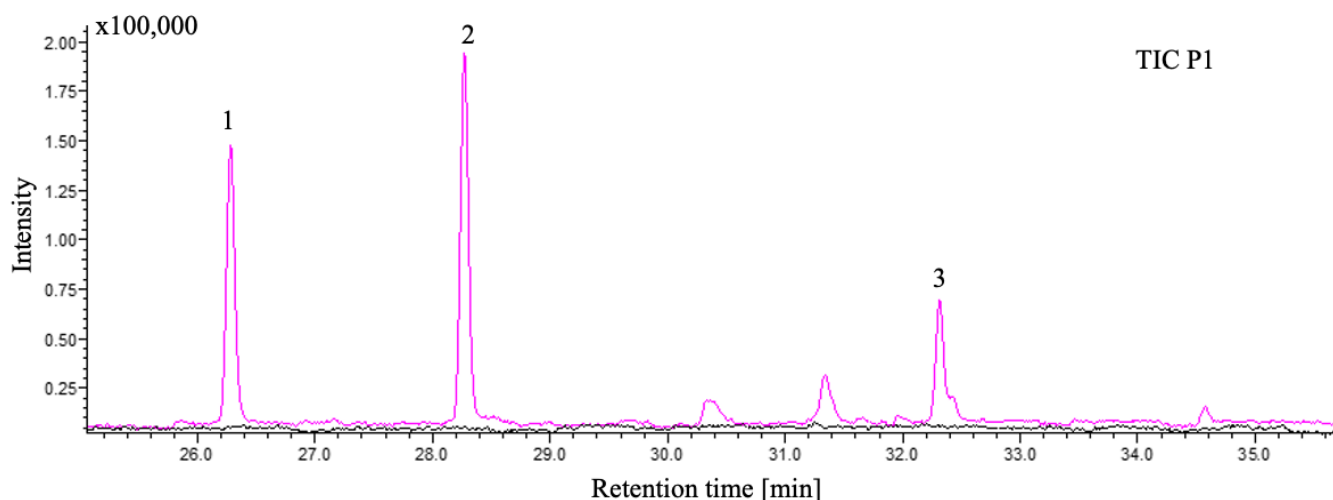


Figure 13. The chromatogram of P1, with peaks listed in Table 4 labeled in the figure. The negative control (yeast transformed with empty pWUS plasmid) is added (black line).

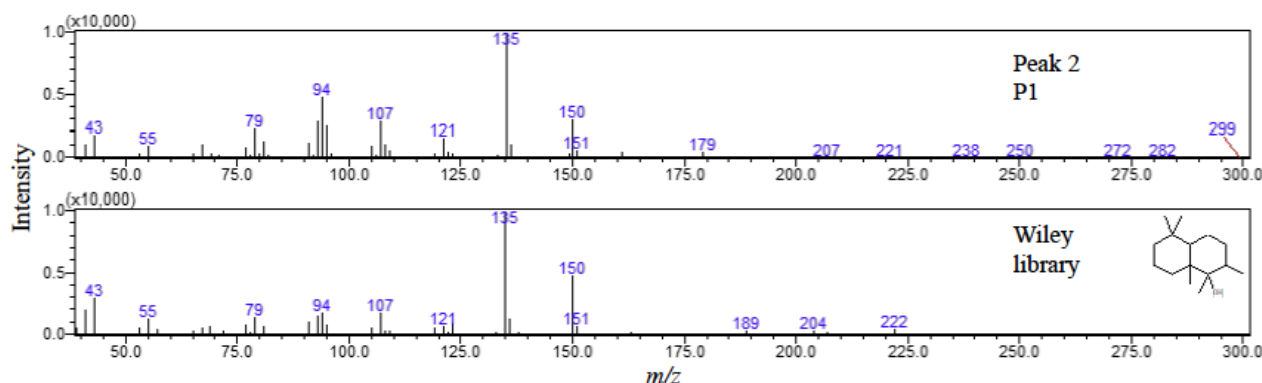


Figure 14. The fragmentation pattern of P1 peak 2, as shown in Figure 13, compared to the pattern for a structure provided by Wiley library. m/z 135 is the highest ion peak in both patterns.

Peaks 1 and 2 were the major peaks of D2 (>10% each) (Figure 15). The GC-MS library proposed them both as α -santalene, but of 81 and 82 % similarity, respectively. They are both missing the close appearance of ion peaks m/z 93 and 94, as well as m/z 79, 107, and 121 as main ion peaks (Figure 16). Hence, neither appear to be α -santalene. This can to be further investigated in the *in vitro* analysis for an additional compound production and detection.

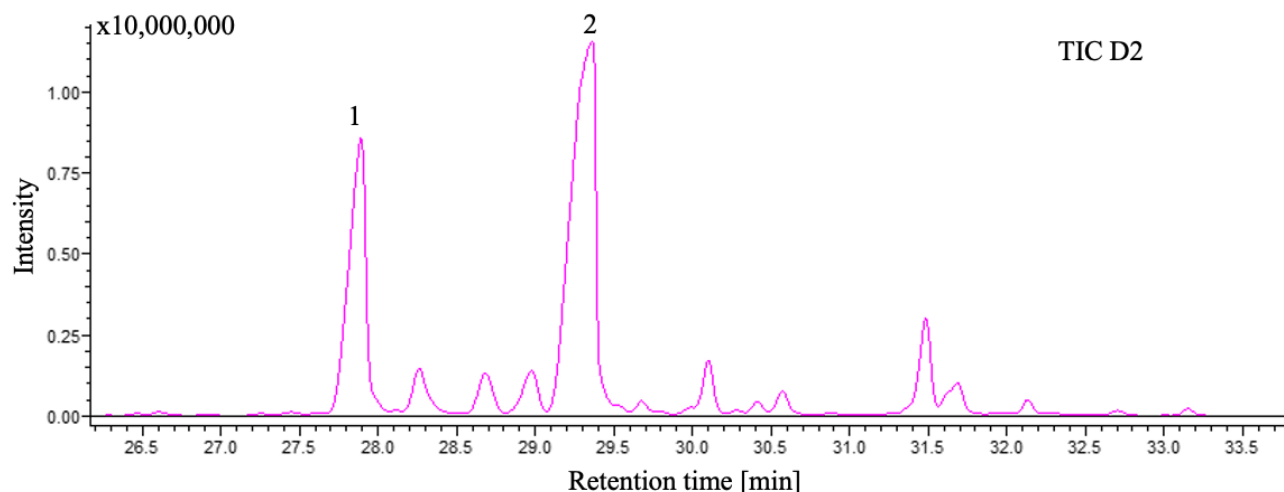


Figure 15. The chromatogram for D2, peaks listed in Table 4 (>10%) are labeled. The negative control (yeast transformed with empty pWUS plasmid) is added (black line).

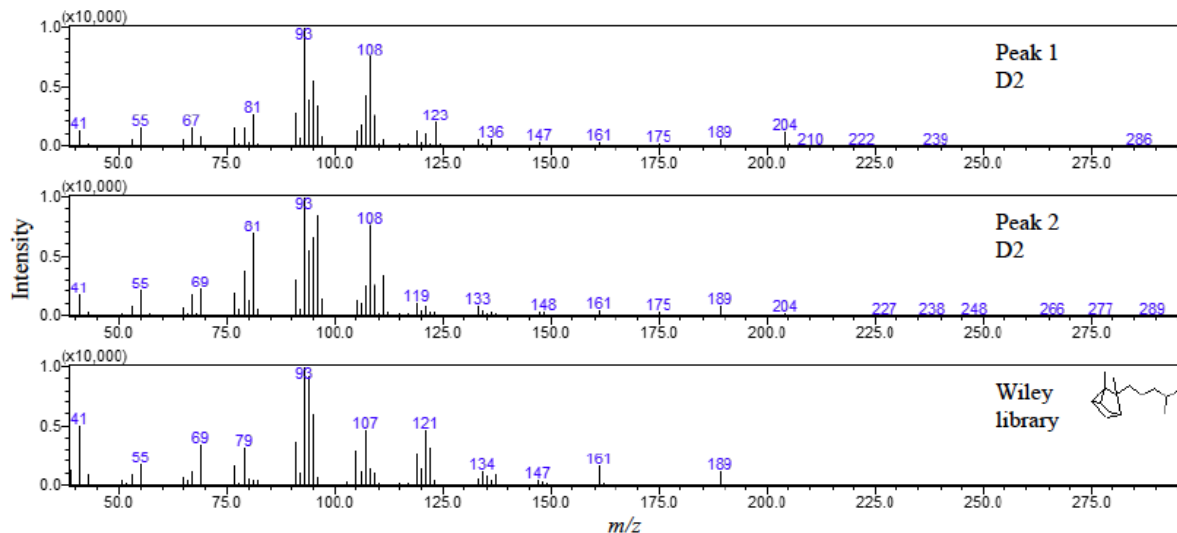


Figure 16. The fragmentation patterns for peak 1 and 2 for D2 as shown in Figure 15, as well as the provided fragmentation pattern by Wiley library. m/z 93 is highest in all patterns.

5.1.1.3 Diterpene production

D1 produced cembrene as the only product (Table 5). D1 product formation was analyzed twice, both times presenting cembrene as sole product. Cembrene has a higher RT than the sesquiterpenes run with the same GC-MS program A, which is reasonable as diterpenes are larger compounds than sesquiterpenes and should be retained by the column to a greater extent. The fragmentation pattern of this putative cembrene product will be discussed later in the *in vitro* analysis.

Table 5. One TPS showed to produce the diterpene cembrene. The product with RT and abundance in percentage are listed, as well as a combined percentage from the remaining terpene < 10%.

Gene ID	Total number of terpene peaks	Product	RT [min]	Abundance [%]	Intensity [x100,000]
D1	1	Cembrene Remaining	39.4	100 0	2.0

5.1.2 Co-transformations

pWUS-TPS was also co-transformed with pLUS-Erg20p(N127W). Based on the knowledge that the precursor for monoterpenes, GPP, has low abundance in yeast, an approach for increasing the GPP pool was performed. Ignea et al. introduced Erg20p(N127W), a mutated variant of Erg20p, that acted as a negative dominant component in the Erg20p dimer, inhibiting synthesis of FPP in the yeast cell. The idea was that by replacing N127 in the Erg20p dimer with a bulkier amino acid W, the activity of Erg20p towards FPP would decrease as the active site in the dimer complex is blocked. However, activity towards GPP would not decrease (47). Hence, in this project, TPSs from batch 1 and batch 2 were, besides from being singly expressed in yeast, co-introduced into yeast with pLUS-Erg20p(N127W).

The correlation between the results from co-transformation and the expected outcomes from the co-transformation varied between expressed TPSs. The expectations were that the GPP pool would increase, and monoterpenes produced to a greater extent, if the expressed TPS carried monoterpene activity. None of the TPSs co-transformed with pLUS-Erg20p(N127W) produced additional monoterpenes. Instead, the monoterpene production appeared to have decreased and the sesquiterpene production increased.

In Figure 17, the chromatograms for single and co-transformation of B2 and C2 are shown. B2 initially produced monoterpenes (Table 3), but the co-transformation presents two sesquiterpene peaks, which was unexpected. At the same time, the monoterpene peaks are greater for the single transformations. The same is visible for C2, also unexpected.

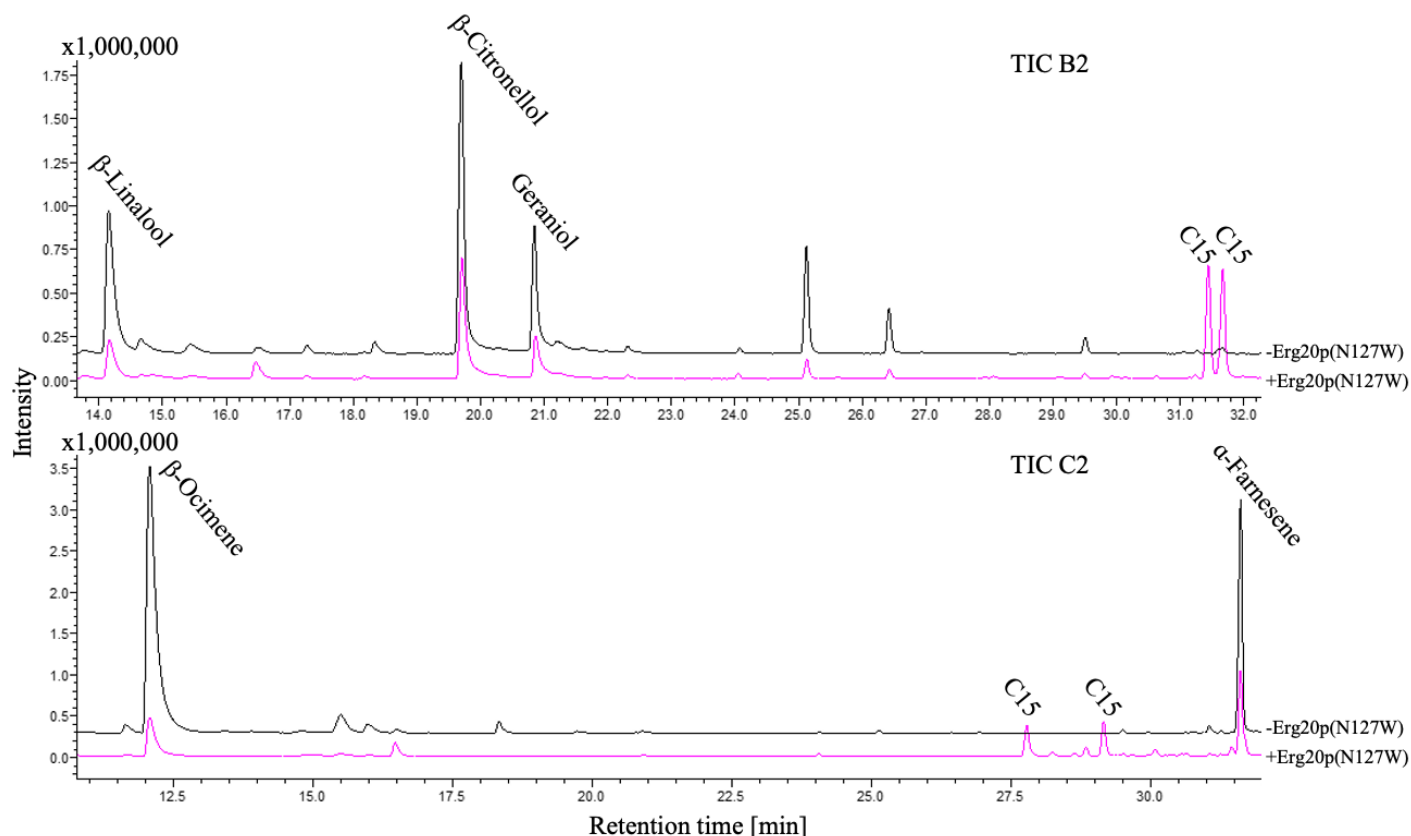


Figure 17. Chromatograms for B2 and C2, presenting peaks from transformation without *Erg20p(N127W)* and with *Erg20p(N127W)*. Note that the baselines are shifted to visualize each sample. Single transformation without *Erg20p(N127W)* suggests a greater monoterpene production as all monoterpene peaks presented in Table 3 are more intense for B2 and C2. Meanwhile, co-expression with *Erg20p(N127W)* shows two sesquiterpene peaks (labeled C15) for both B2 and C2, visible at RT 27.5 to 32 min.

Due to time limitations as well as the fact that the co-transformations led to unpredicted and unexplained behaviors in the cells, TPSs from batch 3 were not co-transformed with pLUS-*Erg20p(N127W)*. However, it is interesting to note that compounds are not just shifted in amount of production, but new compounds seem to have been produced. In addition to these TPSs that produced differently in single and co-transformation, several of the inactive genes in the single transformation produced sesquiterpene compounds in the co-transformation, of various intensities but especially present for E1, K1, and M1, however of some uncertainties (Figure S6). This suggests that the TPSs are active and functioning, but issues arise in the transformants. Retesting the co-transformation would be necessary to verify the occurrences. The products formed in co-transformations are not included as true products as the understanding of what occurred is low.

5.2 *In vitro* analysis

The *in vitro* analysis was performed by adding either GPP (monoterpene substrate), FPP (sesquiterpene substrate), or GGPP (diterpene substrate) to the cell extract after the cells had been induced and lysed. GC-MS program B was used for analysis. The aim of running *in vitro* as an additional analysis method

was to provide an extra analysis and hopefully more certainty regarding the product formations by the TPSs.

In the *in vitro* analysis, fewer transformants were active compared to HS-SPME. Furthermore, less as well as smaller product peaks were in general detected from the *in vitro* analysis. This was probably because a small amount of hexane containing unknown, and perhaps low, amounts of extracted terpenes was injected into the GC column, whereas HS-SPME analysis allowed terpene accumulation in the fiber coating and is considered a sensitive method. However, some TPSs produced several peaks still, so peak areas of at least 10% are presented also for the *in vitro* analysis. The results from the *in vitro* correlated somewhat with the *in vivo* HS-SPME results. Note that in Table 6 and 7, products from GGPP addition are not presented.

5.2.1 Cell hydrolysis reactions

All transformants produced to various degrees the monoterpene geraniol from GPP, even though the product was not necessarily detected in HS-SPME. The control also produced geraniol, albeit generally of lower amounts compared to the TPS-producing strains but suggesting that geraniol was produced without TPSs present. Hence, geraniol is not included as a product for transformants but B2, discussed later. Geraniol may be hydrolyzed from GPP by phosphatases located in the yeast cells, when the GPP substrate is in excess (47). GPP is as has been discussed, a substrate of low concentrations in yeast, therefore it is believable that the cell wants to quickly eliminate the compound when detected in higher concentrations. Hence, transformants produced geraniol, but this was most likely cell stress and not a product formed by the enzyme. Geraniol was produced in very uneven amounts, some in the million magnitude and some barely detected. It is possible, as the mixture with substrate were made in larger batches, uneven distribution of the substrate to the various transformants occurred and some samples were given less GPP, meaning those cells did not exhibit stress.

5.2.2 Mono-TPSs

The three TPSs presented from HS-SPME analysis as potential mono-TPSs are listed in Table 6. Products from the *in vivo* HS-SPME analysis are listed in the last column for comparison. Note that geraniol is for B2 included as the only product. For C2, β -ocimene was 100 % of the monoterpene production from GPP, excluding geraniol, whereas FPP addition led to sesquiterpene production of 100 % farnesene. J3 did not present any terpene peaks from GPP or FPP addition.

Table 6. Mono- and sesquiterpenes from GPP and FPP addition, respectively, in order of RT in the chromatogram. The terpenes present in Table 3 (in vivo HS-SPME results) are listed in the last column. All products from in vivo as well as in vitro are given abundance in rounded percentage for comparison.

Gene ID	<i>In vitro</i> products, [%]		<i>In vivo</i> products, [%]
	Monoterpenes from GPP	Sesquiterpenes from FPP	
B2	Geraniol, ~99		β -Linalool, 25 β -Citronellol, 37 Geraniol, 14
C2	β -Ocimene, 100	α -Farnesene, 100	β -Ocimene, 67 α -Farnesene, 23
J3			β -Ocimene, 66 Alloocimene, 34

5.2.2.1 Comparison *in vivo* and *in vitro* mono-TPSs

Geraniol is listed for B2 because it was detected with the HS-SPME method. β -Citronellol and β -linalool were also produced from B2 with HS-SPME, but not present in the *in vitro* analysis. The same handling of GPP by the cell as for geraniol production can occur for linalool production, meaning linalool might only be produced from the cell as it hydrolyses GPP (47). Linalool has also been shown to be produced from geraniol, by Steyer et al. (68). The same authors explained citronellol formation with cell function, where citronellol can be produced from a reduction step of geraniol. Further formations caused by cell activity are geranyl acetate and citronellyl acetate, both detected as smaller peaks in the HS-SPME analysis of B2 (Figure S1) (68). B2 also produced a nerol peak eluting slightly before geraniol, geraniol appeared to co-elute with other compounds and the abundances were hard to estimate. Nerol is typically produced from neryl diphosphate (NerPP), the *cis* form of GPP, but as NerPP is not present in yeast cells, it is most likely yeast cell reactions from geraniol, a described reaction (68, 69). Based on all information, it is possible that B2 produced geraniol as its sole product and is a geraniol synthase.

The β -ocimene production was 3 times greater (~60%) than α -farnesene production (~20%) in the HS-SPME analysis for C2. Hence, C2 appeared to be a bifunctional ocimene synthase with potential to produce farnesene. In the *in vitro*, the α -farnesene production was even lower, see Figure 18 for comparison between substrate additions. Note that GPP also produced farnesene. The idea behind the *in vitro* analysis is that the added substrate will be utilized by the enzyme, hence different substrate addition will allow production of the specific types of terpenes. As yeast cells contain endogenous FPP and since the method used in this work did not isolate the released enzyme, the enzyme may use endogenous FPP substrate as well. However, the GPP addition (no extra FPP added) allowed a sesquiterpene peak production of similar size as the FPP addition (extra FPP added), meaning either FPP was unevenly distributed in the larger pool for the *in vitro* analysis or not utilized by the enzyme.

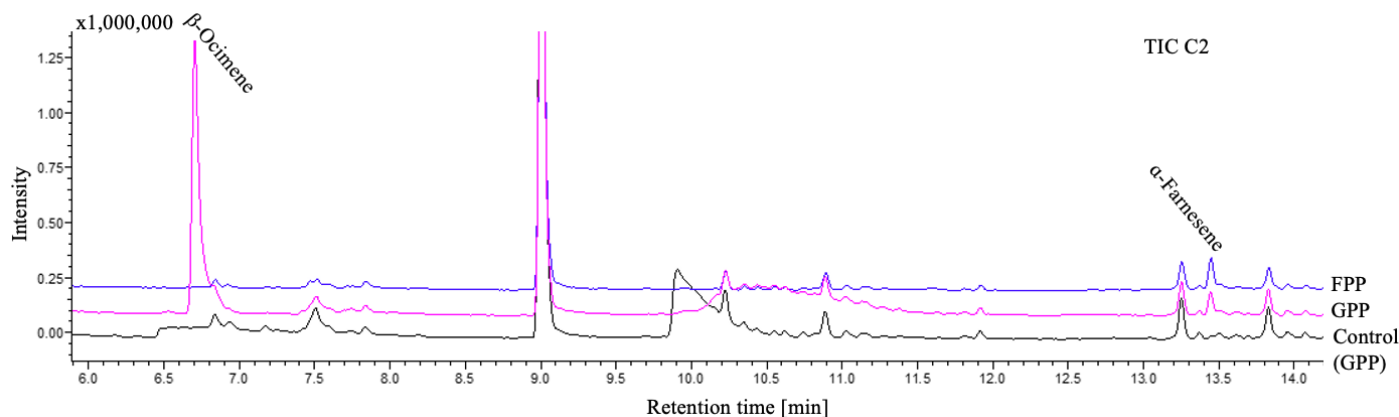


Figure 18. The chromatogram for C2 (mono- and sesquiterpene peaks). GPP and FPP addition are presented, as well as the negative control with GPP addition due to lack of combined controls. Note that the baselines are shifted to visualize each sample. GPP addition produces a significant β -ocimene peak. FPP addition produces a slightly higher α -farnesene peak than GPP addition.

J3 produced two ocimene isomers in HS-SPME. These are not detected in the *in vitro* analysis, most likely because the peaks were rather small in HS-SPME (Figure S1), and a lower concentration of terpenes may have entered the GC column at hexane injection. It is therefore still possible that J3 is a mono-TPS that produces two ocimene isomers.

5.2.3 Ses-TPSs

As for C2, GPP addition allowed sesquiterpene production for B1, I1, N1, O1, A2, and D2. Again, probably due to endogenous FPP that was released as the yeast cells were lysed. The sesquiterpene peaks were for almost all TPSs greater at FPP addition than GPP addition, discussed later.

B1, I1, and D2 produced peaks with the HS-SPME as well as in the *in vitro* analysis that presented under 90% similarity to the GC-MS libraries. B1 in the *in vitro* presented a similar elution pattern as when using the HS-SPME method (Table 4), and due to the fragmentation pattern being most like the compounds identified in that analysis, they were assumed to be the same, i.e. germacrene D, isolatedene, and δ -cadinene (Table 7). However, I1 produced an uncertain peak not shown in the *in vivo* analysis and the fragmentation patterns provided by the GC-MS libraries for D2 were still not convincing. *Identification of Essential Oil Components by Gas Chromatography/Mass Spectrometry, ed 4.1* was used as an external library to identify peaks (62). D2 and I1 were retested in GC-MS with program B together with a standard sample with six sesquiterpenes. The standard sample was labeled according to given instructions (Figure 19). The sixth peak (farnesol) was not detected.

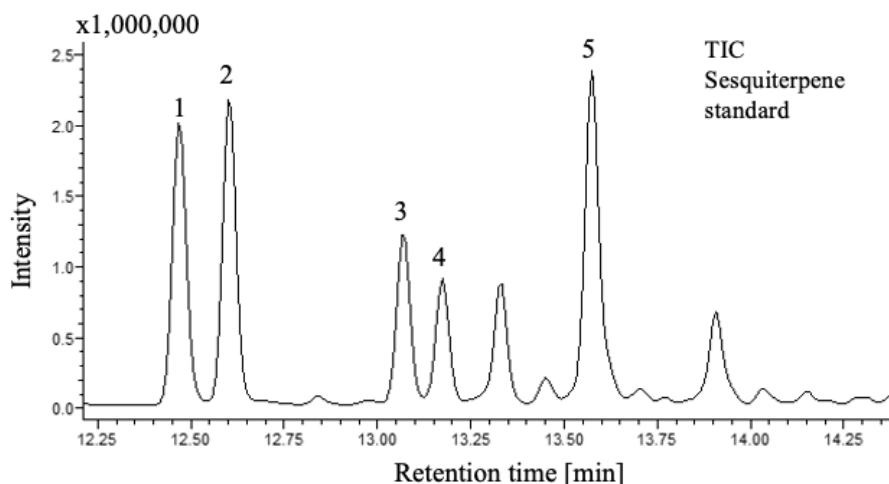


Figure 19. Sesquiterpene standard produced with GC-MS program B. Labels: 1 longifolene, 2 trans-caryophyllene, 3 α -humulene, 4 allo-aromadendrene, 5 valencene.

With RTs in Figure 19, the method provided in Section 4.8 was used for product identification. The AIs of the compounds were taken from the book and plotted to each other. An equation could be obtained from the graph, used to calculate AIs of the unknown peaks (Figure 20).

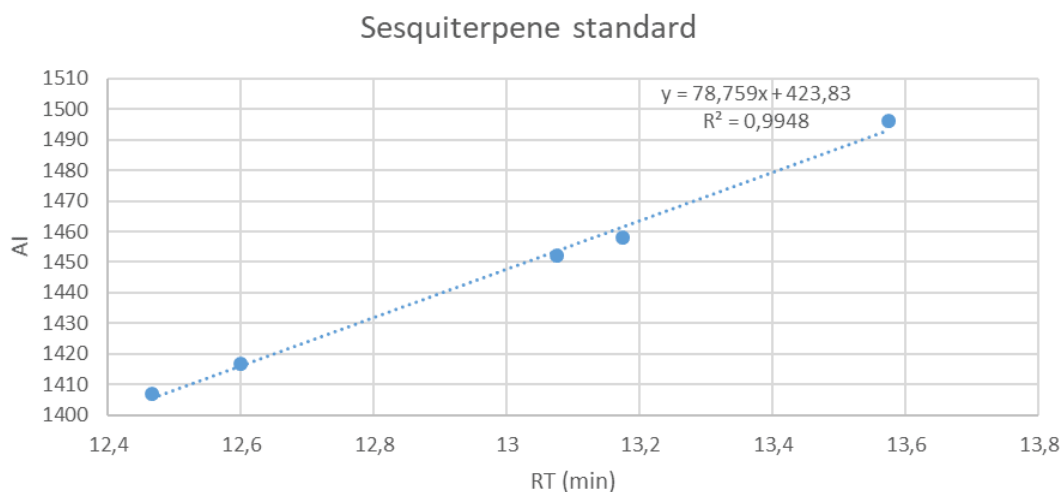


Figure 20. The products from Figure 19 are plotted RT to AI (shown as blue circles). AIs are given by the book. The equation for the linear regression is provided in the graph, with corresponding R^2 value.

The two uncertain peaks of D2 (Table 7) with RTs 12.45 and 13.03 generated with the equation in Figure 20 AIs of 1404 and 1450 that could be utilized to search in the external library. Peak 1 in the *in vitro* analysis, that corresponded to *Unknown A* in Table 4 and 7, was less than 10 % as well as slightly outside the linear regression range (Figure 20), but still studied to identify compound *Unknown A*. Fragmentation patterns for peak 1 presented several small ion peaks in the higher m/z range but overall showed a high similarity to the fragmentation pattern of α -barbatene with AI 1407 (Figure 21).

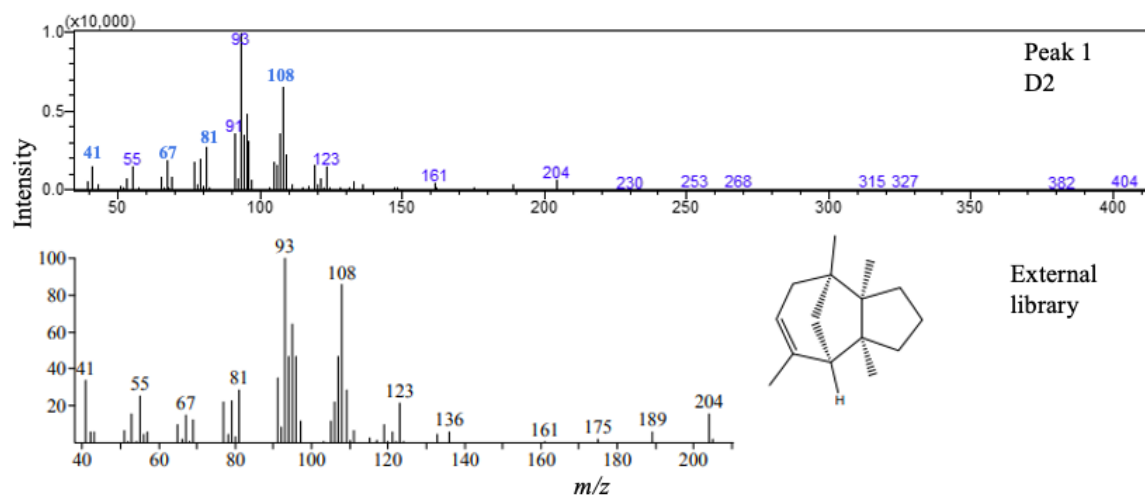


Figure 21. The most similar compound from the utilized book, referred to as α -barbatene. m/z 41 and 51 in peak 1 appear to be slightly lower than the library. Note that ion peak labels have been manually added to the unknown pattern for better comparison. m/z 93 is highest in both patterns.

For Unknown B of D2 (peak 2), the visually most similar compound was myltayl-4(12)-ene with AI 1445 (Figure 22). The ratio of m/z 93 and 108 is skewed as 93 appears too high. Furthermore, m/z 67 as well as 81 is higher in the unknown compound, and 41 lower, but otherwise there is a possibility that the compound is myltayl-4(12)-ene.

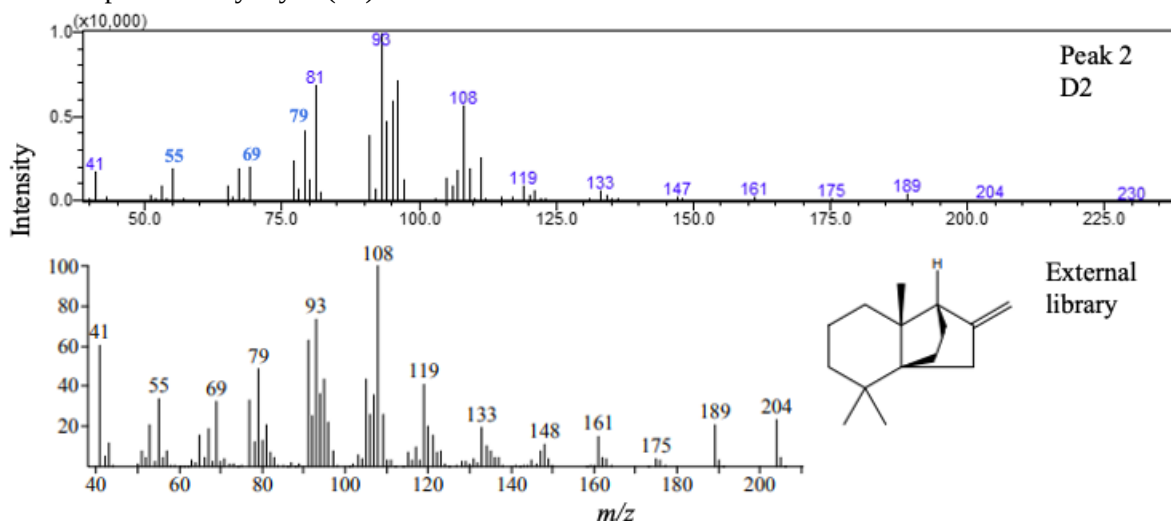


Figure 22. The most similar compound found from the utilized book, referred to as myltayl-4(12)-ene. m/z 67, 81, and 93 are higher in peak 2, and m/z 41 is lower, than the provided library. Note that ion peak labels have been manually added to the unknown pattern for better comparison.

The retesting of I1 and D2 produced much smaller or not even visible peaks compared to the first *in vitro* analysis, most likely due to old colonies on the plates, and the uncertain peak of I1 was not detected hence could not be verified with the external library. The fragmentation pattern of I1 was further studied for peak 2 (Table 7) as it showed under 90 % similarity to the library compound pattern. The fragmentation pattern from peak 2 presented as α -cadinol by the NIST library but had a skewed ratio of m/z 43 and 95,

as well as m/z 161 and 204. Otherwise, all main ion peaks are present (Figure 23), allowing the assumption that the peak is α -cadinol.

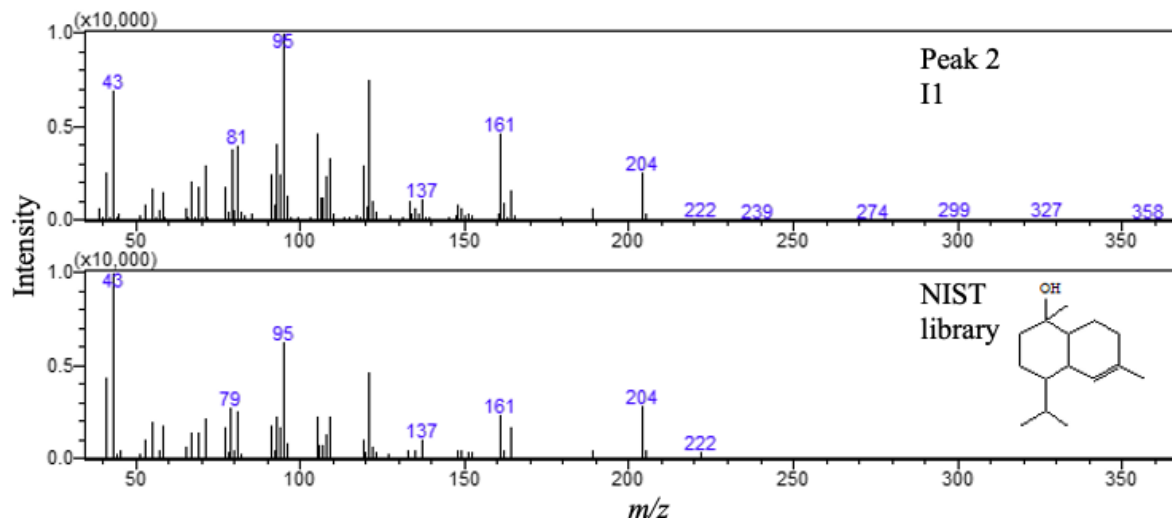


Figure 23. Fragmentation pattern for I1 from *in vitro* (peak 2) and the NIST GC-MS compound library for α -cadinol fragmentation.

The results from the *in vitro* analysis are summarized in Table 7 together with the nine TPSs producing sesquiterpenes as main products using the HS-SPME extraction. Several products identified in the *in vitro* were below 10 % in the *in vivo* testing using HS-SPME. B1 presented a rather high remaining percentage due to multiple unlisted products. D2 also had some remaining percentages, filled by peaks <10% such as α -barbatene and a rather uncertain sesquichamene peak. B1 and O1 presented new monoterpene production other than geraniol, their chromatograms are included later for discussion.

Table 7. The product peaks from GPP and FPP addition, in order of RT in the chromatogram, for all TPS genes active in the HS-SPME analysis. The terpenes present in Table 4 (in vivo HS-SPME results) are listed in the last column. All products from in vivo and in vitro are given abundance in rounded percentages for comparison, as well as remaining percentages for the in vitro analysis.

Gene ID	In vitro products		In vivo products
	Monoterpenes from GPP	Sesquiterpenes from FPP	
B1	1 Myrcene, 67 2 Nerol, 24 Remaining, 9	1 Germacrene D, 19 2 Isoledene, 12 3 δ -Cadinene, 24 4 δ -Cadinol 10 Remaining, 35	Germacrene D, <10 β -Cadinene, 13 Isoledene, 10 δ -Cadinene, 50 δ -Cadinol, <10
G1		1 Farnesol, 100	Valencene, 50
I1		1 δ -Cadinol, 70 2 α -Cadinol, 12 3 Farnesol, 17 Remaining, 1	α -Amorphene, 18 α -Muurolene, 19 δ -Cadinene, 27 δ -Cadinol, <10 α -Cadinol, <10
N1		1 δ -Cadinol, 100	α -Copaene, 17 Sinularene, 27 α -Muurolene, 18 δ -Cadinene, 25 δ -Cadinol, <10
O1	1 β -Linalool, 100	1 Nerolidol, 100	β -Farnesene, 20 Nerolidol, 47
P1		1 Unknown, 53 2 δ -Cadinene, 47	α -Cubebene, 31 Unknown, 39 δ -Cadinene, 15
A2		1 β -Himachalene, 100	β -Himachalene, 71 Cuparene, 21
D2		1 * α -Barbatene 2 Myltayl-4(12)-ene, 68 3 β -Himachalene, 15 Remaining, 17	Unknown A, 26 Unknown B, 51 β -Himachalene, <10
A3			α -Bergamotene, 100

* in vitro product <10%

5.2.3.1 Comparison *in vivo* and *in vitro* sesqui-TPSs

O1 and A2 correlated rather well in the two experiments. From the *in vivo* HS-SPME results, it was unclear whether A2 produced β -himachalene and cuparene as they were both present in the negative control, β -himachalene very minor but cuparene more present. From the *in vitro* analysis, cuparene was either not detected or not produced. Due to the presence in the control for the *in vivo* analysis as well as no production for *in vitro*, perhaps cuparene was not a true product, making A2 a potential β -himachalene synthase.

O1 produces nerolidol that, like geraniol, may be produced through cell hydrolysis when FPP is in excess (47). However, nerolidol is produced in HS-SPME when FPP is not in excess. Furthermore, according to Figure 24, the nerolidol peak is greater when FPP is added as substrate compared to addition of GPP, with no visible nerolidol peak in the control. This means that the enzyme most likely uses both endogenous and added FPP to produce nerolidol. The β -linalool peak, formed at GPP addition, has uncertain origin. The linalool peak might be produced like the assumed production pathway for B2, by the cell handling excess GPP. However, the ratio of geraniol to linalool from the control and O1 differs greatly (Figure 24). O1 is the only TPS strain producing a great amount of linalool instead of just geraniol, like the other strains, such as I1 that produced geraniol with intensity of almost 10^7 but negligible amount of linalool. Lastly, β -farnesene was detected with HS-SPME, produced in a one-step reaction from FPP as nerolidol (34). It is possible that O1 is a nerolidol - linalool synthase that allows production of β -farnesene.

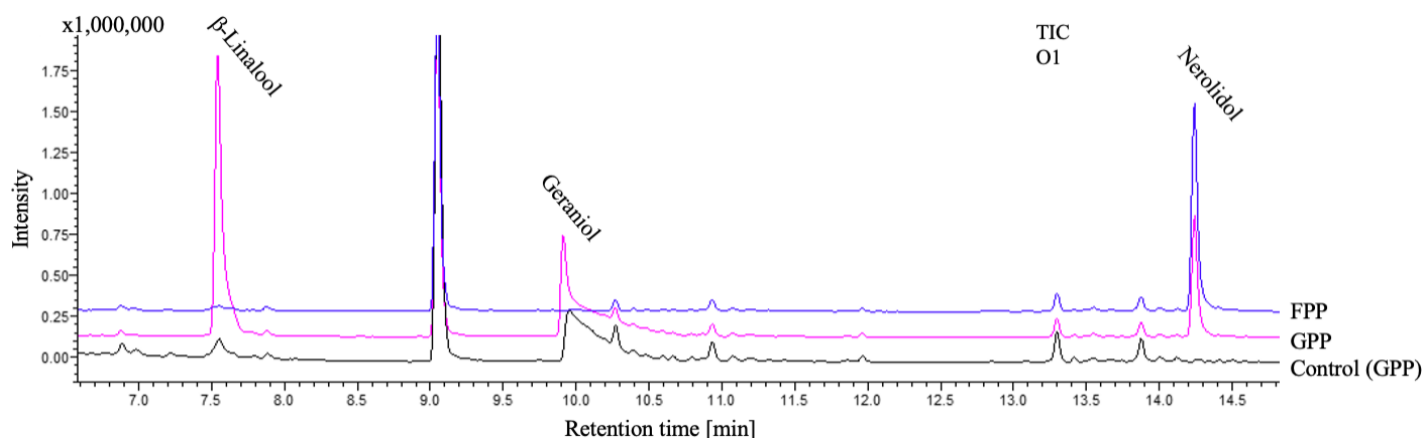


Figure 24. The TIC chromatogram for O1, with products listed in Table 7 labeled. Note that the baselines are shifted to visualize each sample. The geraniol peak most likely co-elutes with something else, assumed as an extra peak is detected in the dragged-out peak. The negative control with GPP addition is included, due to the lack of a combined GPP and FPP negative control.

For B1, I1, N1, P1, and D2 it was more difficult to conclude what the actual product formations were. B1, as for O1, allowed monoterpene production other than geraniol (Table 7) allowing the assumption that the B1 is also a bifunctional mono- and sesqui-TPS. *Agrocybe aegerita* has been presented with a very similar bifunctional and promiscuous TPS, producing, among other products, myrcene and δ -cadinene (70). Nerol is most likely yeast cell reactions from geraniol, as suggested for B2. The control produced a peak when δ -cadinene elutes (Figure 25), verified with fragmentation pattern analysis and extraction of m/z 93 as a non-terpene. However, the peak most likely makes δ -cadinene too high in TIC and therefore of lower abundance. Two rather large peaks close to δ -cadinol are an uncertain α -longipinene peak as well as α -

cadinol. As for C2, sesquiterpenes can be produced from endogenous FPP released from the cell. However, endogenous peaks are low for FPP addition compared to the control and GPP addition, so perhaps a less concentrated part of the released yeast solution, containing endogenous cell products and TPSs, was taken. As a conclusion, TPS B1 presents low catalytic specificity, allowing production of a great variety of both mono- and sesquiterpenes, but is most likely some sort of bifunctional δ -cadinene/germacrene D/myrcene synthase.

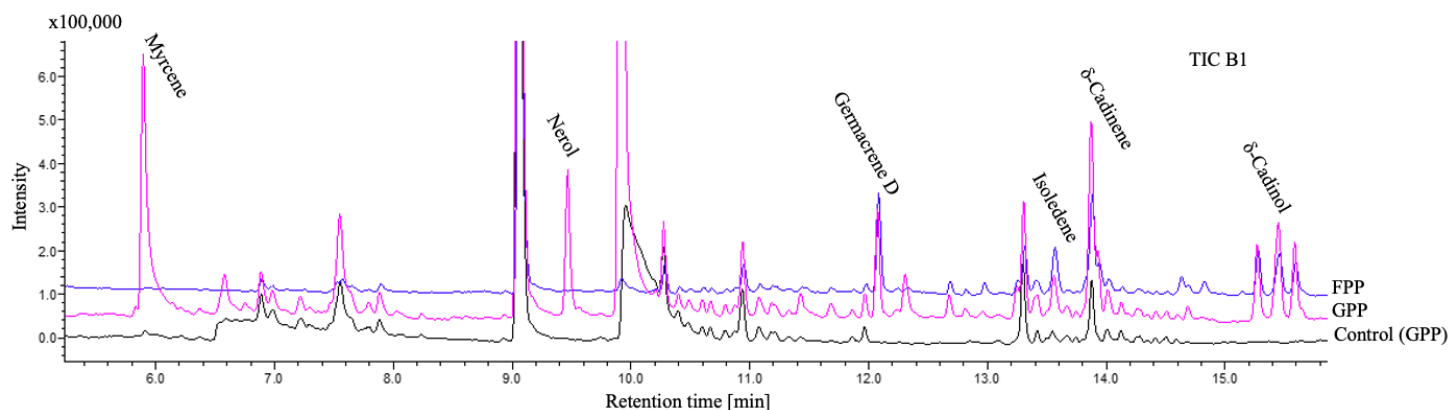


Figure 25. Chromatogram for B1. Note that the baselines are shifted to visualize each sample. The terpenes listed in Table 7 are labeled in the chromatogram. The negative control with GPP addition is included, due to the lack of a combined GPP and FPP negative control.

The *in vivo* in comparison to the *in vitro* skewed abundance ratios for I1 and N1. With HS-SPME extraction, δ -cadinol is less than 10 % of the terpene production for both, but it is the major one/only one produced in the *in vitro* analysis. As different extraction methods were used, it is believable that the abundances would vary a bit between the methods. The interaction of the analyte to the coating on the SPME fiber has been described in Section 2.6.4 to be related to analyte polarity and size, perhaps the coating interacted and retained the less polar compounds better, which skewed the peak ratios. This would entail that the compounds other than δ -cadinol detected in HS-SPME for I1 and N1 were in fact of lower concentrations than the analysis indicated. This seems to have occurred also for B1, where the ratio of δ -cadinol to the cadinene compounds is skewed. Another explanation is that the terpenes from *in vivo* analysis were not produced in the *in vitro* analysis, discussed later. For I1, as farnesol was not detected in HS-SPME for any of the transformants, the same hydrolysis reaction by the cell is likely the explanation for the production of the sesquiterpene farnesol by I1, as well as G1, and some transformants shown inactive in the *in vivo* analysis (data not shown) (47). In conclusion, I1 and N1 may both be δ -cadinol synthases of lower catalytic specificity.

P1 is assumed to produce the same putative but nameless sesquiterpene in the *in vitro* analysis as was detected with the HS-SPME, as a similar fragmentation pattern and the same elution order to δ -cadinene is found. The peak for δ -cadinene seemed to co-elute with something else (data not shown) which is probably the reason for a higher percentage in the *in vitro* vs using the HS-SPME method *in vivo*. The unknown product could not be analyzed towards the external library as no production was detected when running with the standard solution. D2 produced two unknown peaks *in vivo*, now distinguished as possible myltayl-4(22)-ene as the main peak, and α -barbatene. Whether β -himachalene or α -barbatene is second most produced compound is hard to determine, perhaps the coating accumulated β -himachalene better.

G1 and A3 produced nothing in the *in vitro*. It appears as if the enzyme was either not active, low activity, or the products did not enter the organic overlay phase.

As a conclusion from the FPP addition to analyze sesquiterpene production, more substrate needs to be added, or the compounds need to somehow be concentrated in the hexane phase as several sesqui-TPSs produced small sesquiterpene peaks, especially B1, N1, and P1.

5.2.4 Di-TPSs

5.2.4.1 Class I and class II di-TPSs

The GGPP addition tested di-TPS activity, class I and potentially bifunctional, able to remove the diphosphate group as well as class II di-TPSs that are unable to remove the diphosphate group. Class II verification was performed with acid (Section 4.7) as the more polar diphosphate compound would not necessarily position in the organic phase hence the diphosphate group was first removed. As larger terpene products have a greater tendency to be trapped in the cells, previously discussed, the *in vitro* analysis was important to verify that no products were missed *in vivo*. However, no diterpenes were detected, except D1 that produced cembrene, albeit very small, in the acidified GGPP mixture (Table 8). The experiments allowed some sesquiterpene production like GPP addition.

Table 8. The product peaks from GPP and FPP addition, for D1. The terpene present in Table 5 (*in vivo* HS-SPME results) is listed in the last column. All products from *in vivo* and *in vitro* are given abundance in percentage for comparison.

Gene ID	In vitro products				In vivo products
	Monoterpenes from GPP	Sesquiterpenes from FPP	Diterpenes from GGPP	Diterpenes from GGPP (acid)	
D1		Farnesol, 100		Cembrene, 100	Cembrene, 100

The fragmentation patterns for the cembrene peak were compared to verify if the peaks produced in the *in vivo* and *in vitro* correspond to the library. One observation is that the molecular ion peak m/z 272 is undeniably smaller in both *in vitro* and *in vivo* than the library fragmentation pattern. All other peaks are of similar intensity between produced peaks and library, but perhaps the EI source led to an unusually small molecular ion (Figure 26).

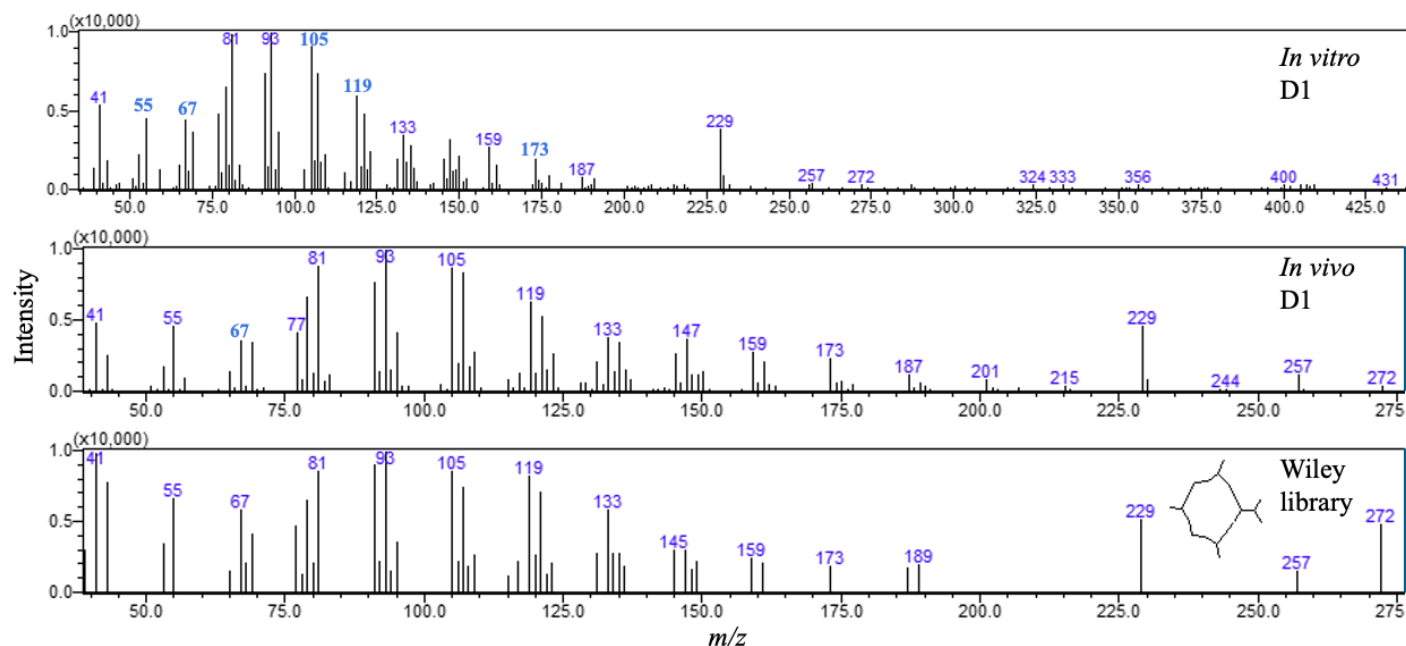


Figure 26. The *in vitro* fragmentation presents m/z 81, 93, and 105 as the highest peaks. Note that ion peak labels have been manually added to the unknown pattern for better comparison. The fragmentation pattern for the *in vivo* peak as well as the cembrene fragmentation pattern and structure provided by Wiley library is shown.

5.3 pH analysis

Yeast metabolites secreted into the cultivation medium can cause pH drops, especially if the cells go into fermentation mode which occurs when oxygen runs out. Yeast cells are rather impermeable to protons and keep in general a constant pH, but a very low external pH can affect the intracellular pH as protons flux in when nutrients and such are taken up (71). Hence, it is possible that the accumulated external pH changes the intracellular yeast pH. Especially enzymes that are produced in the yeast cells can be affected.

The TPS enzymes are produced as the yeast cells are diluted in the GalRaf induction medium. A pH measurement was performed on recovered GalRaf medium from 3 days of induction for later *in vitro* analysis. Another volume of GalRaf had been used for induction than used for *in vivo* and the pH of GalRaf for *in vivo* might not have been the same, but this was never tested. However, as the *in vitro* induction experienced a low pH of 3.09 (Table 9), the *in vivo* probably experienced a somewhat low pH too. It should be noted that the pH of the unbuffered GalRaf was not measured prior to induction, but either way is the pH on the GalRaf medium after induction low. As the *in vivo* analysis is performed in this low pH GalRaf medium but the *in vitro* is performed in a 5X Assay Buffer (GalRaf removed and enzymes resuspended), the pH conditions are not the same throughout the experiments. This can affect the terpene production, as an acidic environment can alter the enzyme as well as the substrate, allowing a changed interaction between them (72).

Table 9. pH of recovered GalRaf medium.

Medium	pH	Temperature [°C]
Recovered GalRaf	3.09	22

To verify how the pH affected the TPS catalytic activity, some TPSs that were re-tested due to uncertain peak production for *in vivo* HS-SPME were induced in two vials, one using the normal GalRaf medium used in previous experiments, one using buffered GalRaf of pH 6.5. The exact same setup was performed otherwise. The pH of the buffered induction medium after 3 days induction was not tested for this specific experiment but in a later setup it was measured to pH 5.8.

The control was first studied to verify that the endogenous cell production remained constant unbuffered vs buffered, which it did. This also suggested that the yeast grew with similar pace in the buffered vs unbuffered medium. The negative control in buffered medium is included for all. Figure 27 and 28 presents four out of six tested sesquiterpene producing strains. The TPSs appeared affected to various degrees, although all sesquiterpene producing TPSs that were tested in buffered medium produced greater peaks in the buffered induction medium. D2 and A3 produced slightly higher peaks of the same ratio (data not shown), whereas B1 produced 5 either new or much greater peaks in the buffered medium (Figure 27). Note that peak 2 appears to be germacrene D, also possibly present in the peak before peak 3, in both buffered and unbuffered. Germacrene D was produced as a major compound in the *in vitro* analysis, perhaps the pH adjustment allowed the enzyme to go back to the right formation and produce more germacrene D than in lower pH. However, δ -cadinol, with 10% production in the *in vitro*, is close to undetected in the buffered medium. Hence, the varying ratio between δ -cadinol and the other terpene peaks between the two analysis methods is most likely not related to the pH but to the different extraction methods used, as discussed in Section 5.2.3.1.

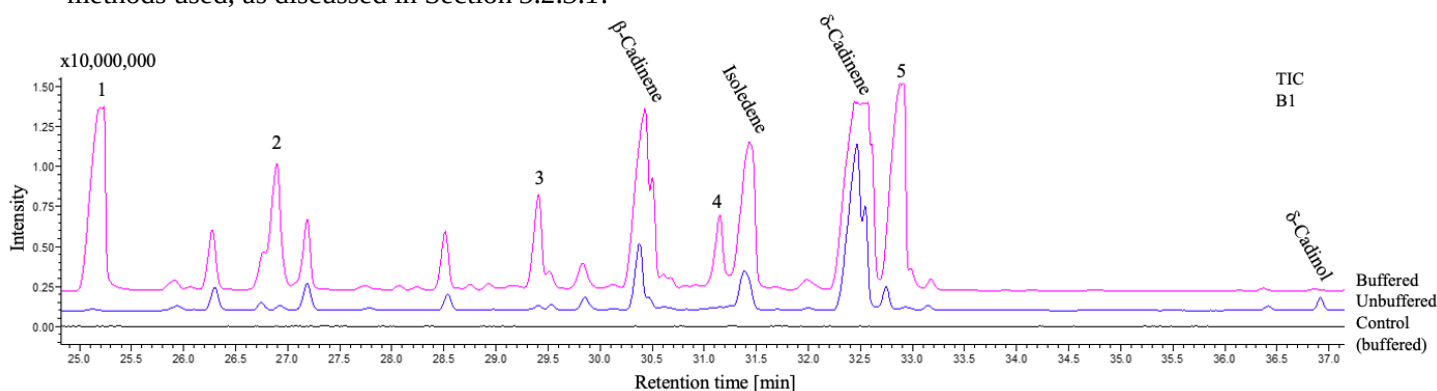


Figure 27. Chromatogram for buffered vs unbuffered B1. Note that the baselines are shifted to visualize each sample. B1 is labeled at peaks present in unbuffered HS-SPME as well as peaks highly present in the buffered medium, but almost or completely undetectable in the unbuffered. 1 - α -cubebene, 2 - germacrene D, 3 - uncertain sesquiterpene peak, 4 - β -cubebene, 5 - uncertain sesquiterpene peak.

It can be noted that the three TPSs in Figure 28 produced β -elemene, although β -elemene was not present in the unbuffered HS-SPME nor the *in vitro* analysis. Note that the buffered medium allowed generally more production of the products under 10 % in the *in vitro*, especially for I1, but not greater δ -cadinol

production (labeled), the main product in the *in vitro*. This suggests, as for B1, that the extraction methods gave different ratios between δ -cadinol and the rest.

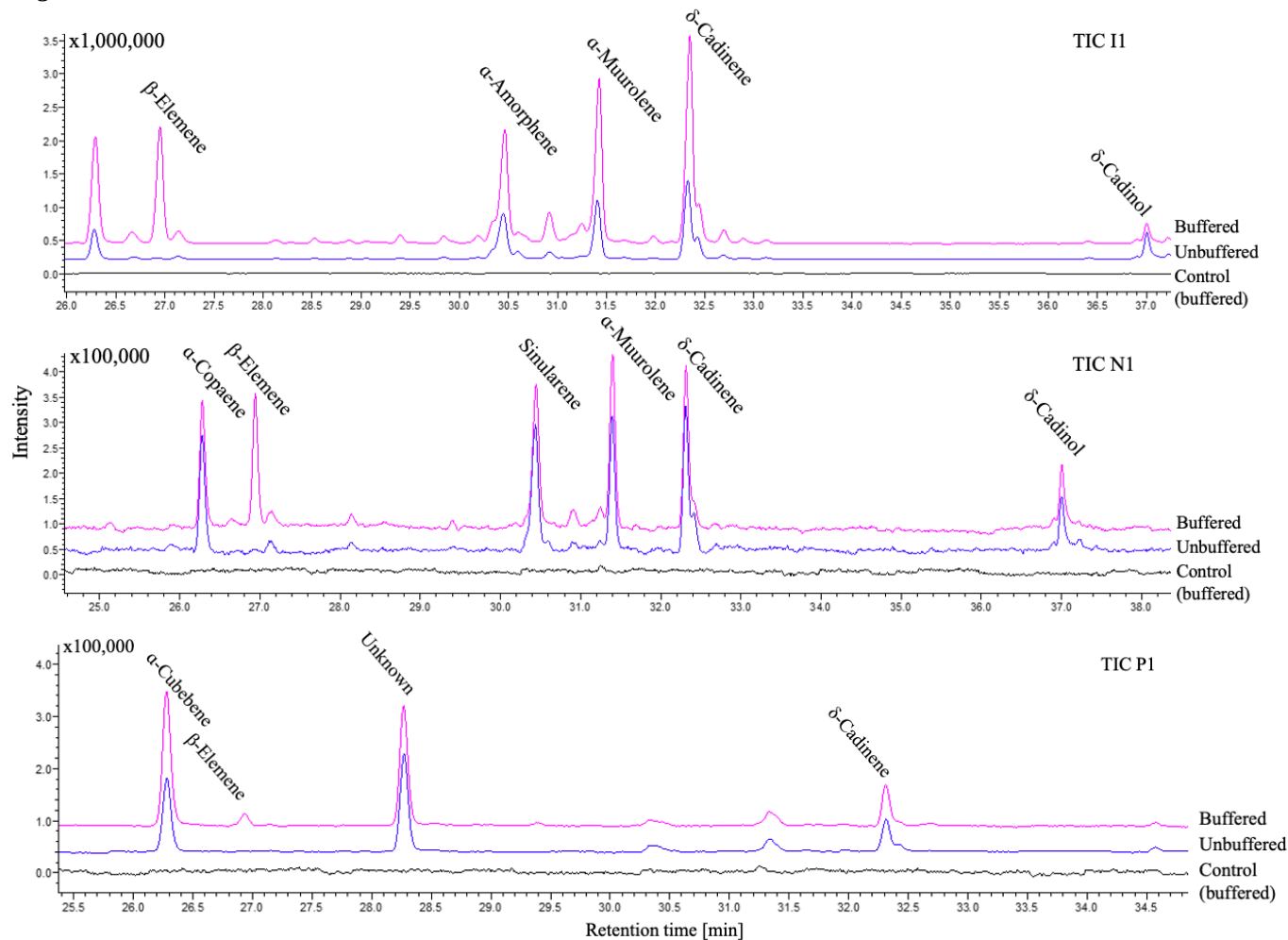


Figure 28. Chromatograms for buffered vs unbuffered I1, N1, and P1. The buffered medium produced higher peaks especially for I1 as well as a new β -elemene peak for the three TPSs. Note that the baselines are shifted to visualize each sample.

Figure 29 presents D1 that produced diterpene. This is the only tested one that produced less product in the buffered induction. It is very difficult to say if this is certain, and it is not evidence that di-TPSs prefer more acidic environments.

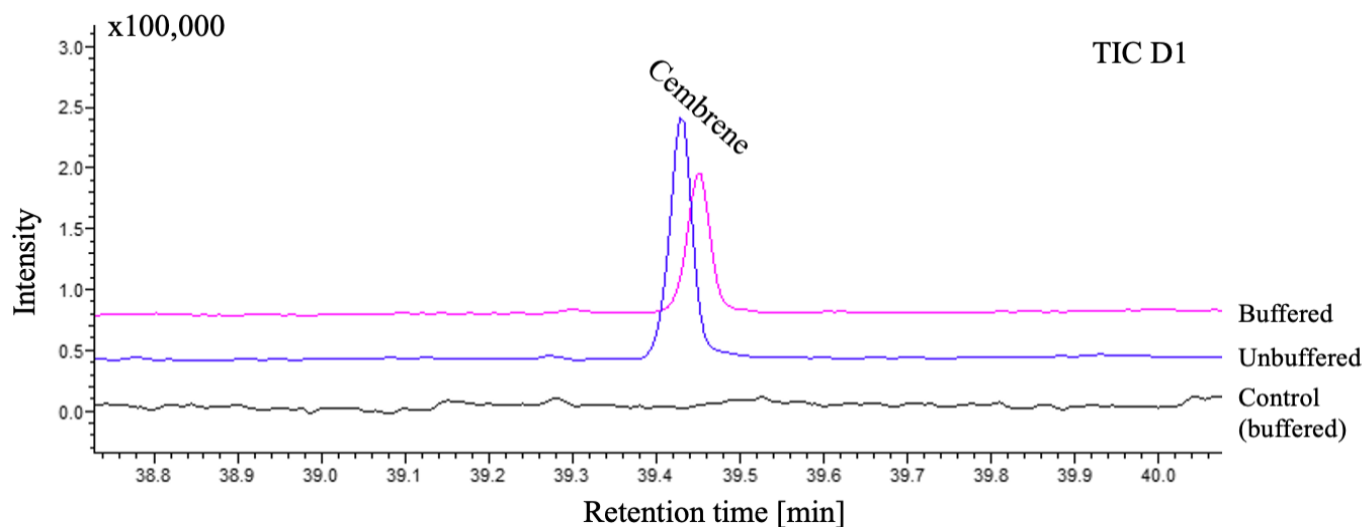


Figure 29. The cembrene peaks from D1, unbuffered vs buffered, are slightly horizontally and vertically shifted to visualize them better.

Adjusting the pH by buffering the induction medium did not appear to make the *in vivo* and *in vitro* more similar, except perhaps the germacrene D production by B1. However, even though the environments were different for *in vivo* and *in vitro* at production, they were both subjected to a potentially low pH at induction and enzymes or substrates may have changed structure, affecting their catalytic activity. Hence, the fact that the production changed in the buffered induction, such as β -elemene for three TPSs, might be reasonable. Enzymes may also denature if the pH changes much from the optimal one (72). Two different filamentous fungal species have been described with a intracellular pH ~ 7.6 of the hyphae cells meaning the buffered medium is most likely closer to the true pH of fungal cells (73). Three inactive transformants were buffered in the induction but still showed no production. However, all inactive transformants would need to be tested with buffered induction medium to verify if low pH caused enzyme inactivation. Other reasons for inactivity are possible, further discussed.

5.4 Inactive genes and further work

The TPSs presented inactive in the *in vivo* single transformation analysis as well as in the *in vitro* analysis can be inactive due to misfolded enzymes, such as from described low pH or other reasons, or the gene sequences might be wrong. The TPS genes in the fungal genomes were predicted with developed algorithms, where this tool allowed gene annotations to avoid manual annotations. However, previous work has expressed issues regarding identifying TPS genes with algorithms, especially clusters from Basidiomycota, partly because fungal species from Basidiomycota are not greatly studied (11). Alternative splicing is common in eukaryotes, but it is not very well understood in fungi. The idea is that pre-mRNA formed from one gene can be spliced in numerous ways to create an array of final protein sequences, controlled by gene introns (74). A study by Grützmann et al presented that Basidiomycota 8.6% of the genes undergo alternative splicing with a slightly lower number of 7.2% in Ascomycota (75). The genomes with the prepared annotations were used when ordering the genes and hence the obtained sequences might not be the correct ones. A collaborator of this study ran the TPS gene sequences in a trained algorithm to determine the possibility that the genes were in fact terpene synthases (76). Seven

potential TPS genes were not tested, but all tested genes were in the span around 90 to >99%, except C1 and J1, with values of 76% and 83 % respectively. They were inactive in the co-transformation, where also A1, H1, and L1 were inactive, two first ones not tested in the algorithm, however the third of 93 % score.

Approaches to see alternative substrate utilizations or activities of the genes were performed in this work. As has been stated in Section 2.3.2.1, prenyltransferases can produce *cis* forms of the isoprenyl diphosphates which may be used as a substrate by TPSs. Tomato glands have been proven to produce and utilize NerPP, the *cis* form of GPP, for monoterpene production, verifying that monoterpenes can be produced this way (77). The *in vitro* method in Section 4.7 was hence run by adding NerPP as substrate. However, nothing indicated that it was utilized as a substrate. With the same *in vitro* method, prenyltransferase activity of the inactive transformants was tested, due to the similar DDxxD domain between prenyltransferases and class I TPSs. IPP was mixed with either DMAPP, GPP, NerPP, or FPP. The active enzymes were not tested as it is assumed that their gene sequences are too short to be able to include all domains required to be a chimeric PTTS enzyme. However, the results presented no new activity.

Other possibilities are that the inactive genes are class I di-TPSs that need modification of the GGPP substrate by a class II di-TPS, as described in Section 2.3.2.2.5. Furthermore, it is possible that *cis* forms of the other isoprenyl diphosphate chains (FPP, GGPP) are produced by *cis* prenyltransferases (31).

Note that the co-transformation with pLUS-Erg20p(N127W) has allowed sesquiterpene production, leading to the assumption that in fact three TPSs, E1, K1, and M1, are most likely active but need further investigations to find reasons to why they do not produce in the single transformation or when tested *in vitro*. Sesquiterpene productions in the co-transformation is interesting but not considered true products until how this occurred can be elucidated.

5.5 TPS catalytic activity in phylogenetic tree

A large portion of the active TPSs presented a variety of products, especially TPSs producing sesquiterpenes. This correlates with the previous discussion regarding the fact that TPSs often are promiscuous. The *in vivo* analysis in combination with the *in vitro* analysis could give some insights into what the main formations are by the TPSs. A convincing conclusion regarding the product formation for some TPSs was hard to form, as the *in vitro* most likely missed several terpenes due to lower sensitivity and the extraction methods presented with varying peak ratios, perhaps due to coating selectivity of the HS-SPME fiber. Both theories need verification, hence the products that have been detected as over 10 % for *in vivo* and/or *in vitro* are presented for TPSs producing mainly monoterpenes (Figure 30), sesquiterpenes (Figure 31, 32, and 33), and diterpene (Figure 34). Products formed by more than one TPS are highlighted.

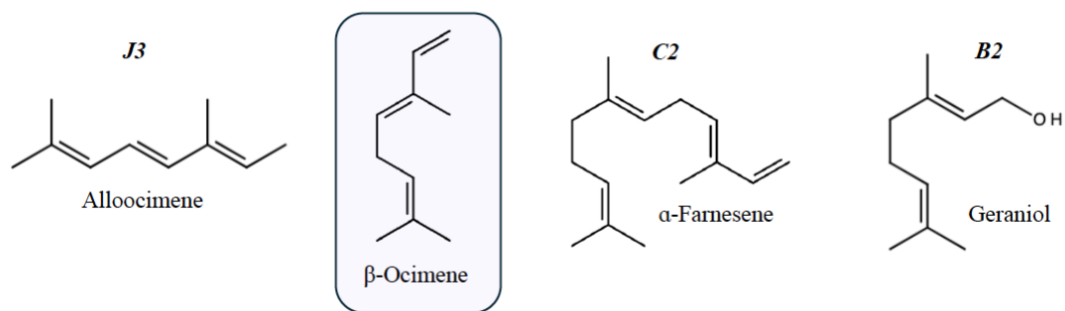


Figure 30. The products formed by TPSs B2, C2, and J3. β -ocimene is produced by C2 and J3. Structures are made by the author.

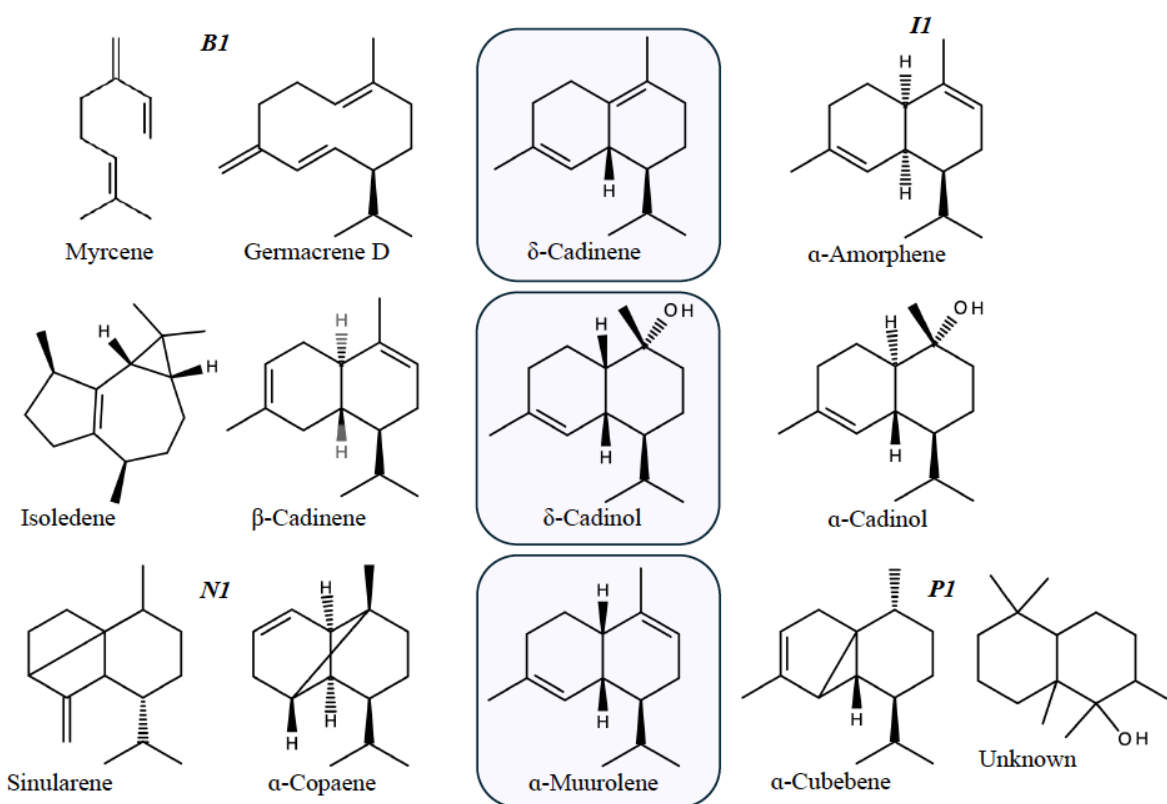


Figure 31. The products formed by TPSs B1, I1, N1, and P1. δ -Cadinene is produced from all four, δ -cadinol from B1, I1, and N1, and α -muulolene from I1 and N1. Structures are made by the author.

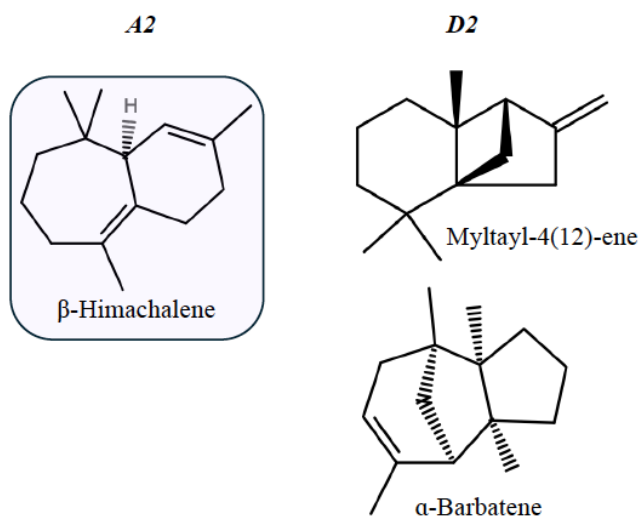


Figure 32. The products formed by TPSs A2 and D2. Structures are made by the author.

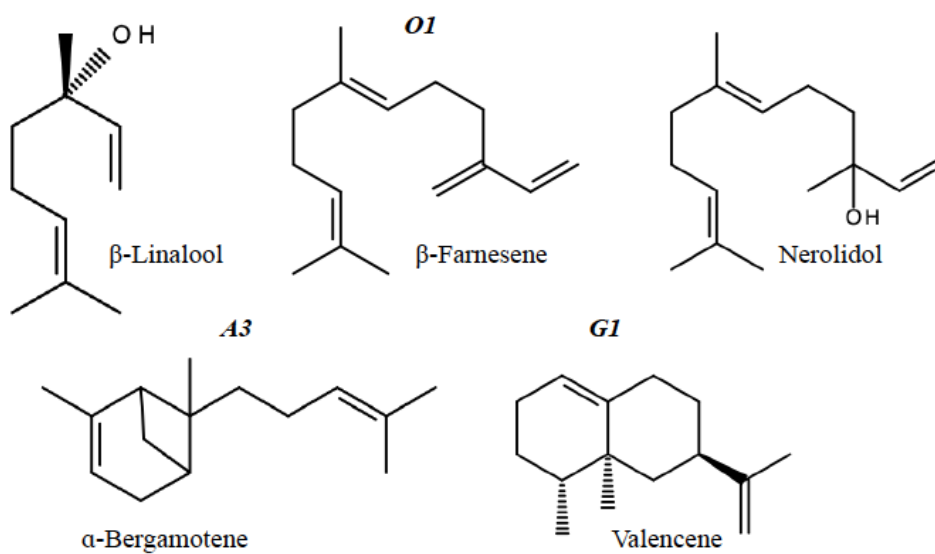


Figure 33. The products formed by sesqui-TPSs O1, G1, and A3. Structures are made by the author.

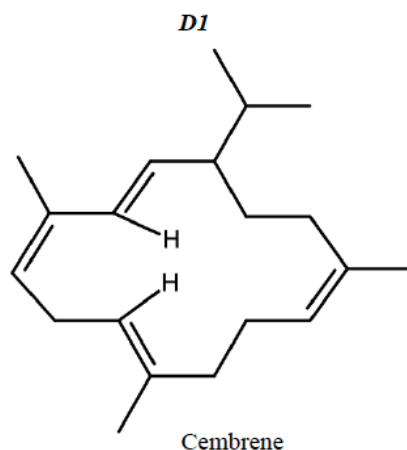


Figure 34. The structure of the diterpene cembrene, produced by D1. Structure is made by the author.

Functional work on characterized ses-TPS products from mushroom-forming fungi, performed by Wang et al, showed that the ses-TPSs divided into clades in the phylogenetic tree based on the substrate cyclization pathway taken, specifically the four cyclization pathways in Figure 6 were located in four different clades (37). The TPS clade divisions in the phylogenetic tree of this work can be studied the same way to find patterns.

Sesquiterpenes have been described to arise from either FPP or NPP (Figure 6). All four sesqui-TPSs in Figure 31 present, in various constellations, compounds of the common cadalene structure (Figure 35) that come from specifically C1,10 closure of FPP or NPP. Cadalane δ -cadinene from this work have been described in previous work arising from C1,10 closure of NPP and α -muurolene as well as δ -cadinol from FPP (Figure 6) (37). Germacrene D production is a bit more unclear with explained pathways both of C1,10 NPP cyclization as well as C1,10 or C1,11 FPP cyclization (22, 37).

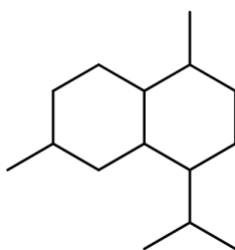


Figure 35. The cadalane structure, made by the author.

I1 and N1 present δ -cadinol as a likely main product (70% and 100% *in vitro*). They come from the same small clade in the phylogenetic tree, assuming this might represent the C1,10 ring closure from FPP. Furthermore, B1 and P1 produce germacrene D and/or δ -cadinene as products, although with an undiscovered pathway for the unknown product of P1. B1 and P1 are in a larger clade with I1 and N1, allowing the C1,10 from NPP or FPP but also rather promiscuous TPS enzymes to be in the lower right clade (pink in Figure 36).

The sesqui-TPSs A2 and D2 are found in the same small clade and are both relatively specific with β -himachalene as a product (Figure 32), a compound of the himachalane structures. β -himachalene has been described to arise from the C1,11 cyclization of NPP, into the humulyl cation as intermediate (78). Even though the α -barbatene structure resembles that of β -himachalene, α -barbatene has been described from the bisabolyll cation (C1,6 NPP closure), hence D2 appear to utilize several cyclization methods (22). The pathway for the main product of D2, myltayl-4(12)-ene, cannot be found. A3 is part of a larger clade with A2 and D2. A3 produce α -bergamotene that arise from the bisabolyll cation, same as α -barbatene (22). Hence, C1,6 NPP and C1,11 FPP closure seem to be in the upper right clade (green in Figure 36).

As for the three main mono-TPSs (Figure 30), B2, C2, and J3, all are in another clade, with specificity for GPP. O1 is included with the mono-TPSs in a larger clade, where O1 produce monoterpene β -linalool but also share production of the farnasane acyclic structure with C2 (nerolidol, β -farnesene, and α -farnesene) (Figure 30 and 33) (yellow in Figure 36).

The potentially inactive genes are located all throughout the tree. However, a group of 5 genes are in a clade to the middle right of the tree, perhaps they all are inactive due to the same issues that arose in evolutionary separation from other TPSs, such as other substrate preferences or being highly affected by pH (blue in Figure 36). Note that K1 and M1 in this clade showed production in the co-transformation. G1 is in this inactive clade and di-TPS D1 is in the larger clade, providing no found pattern for them.

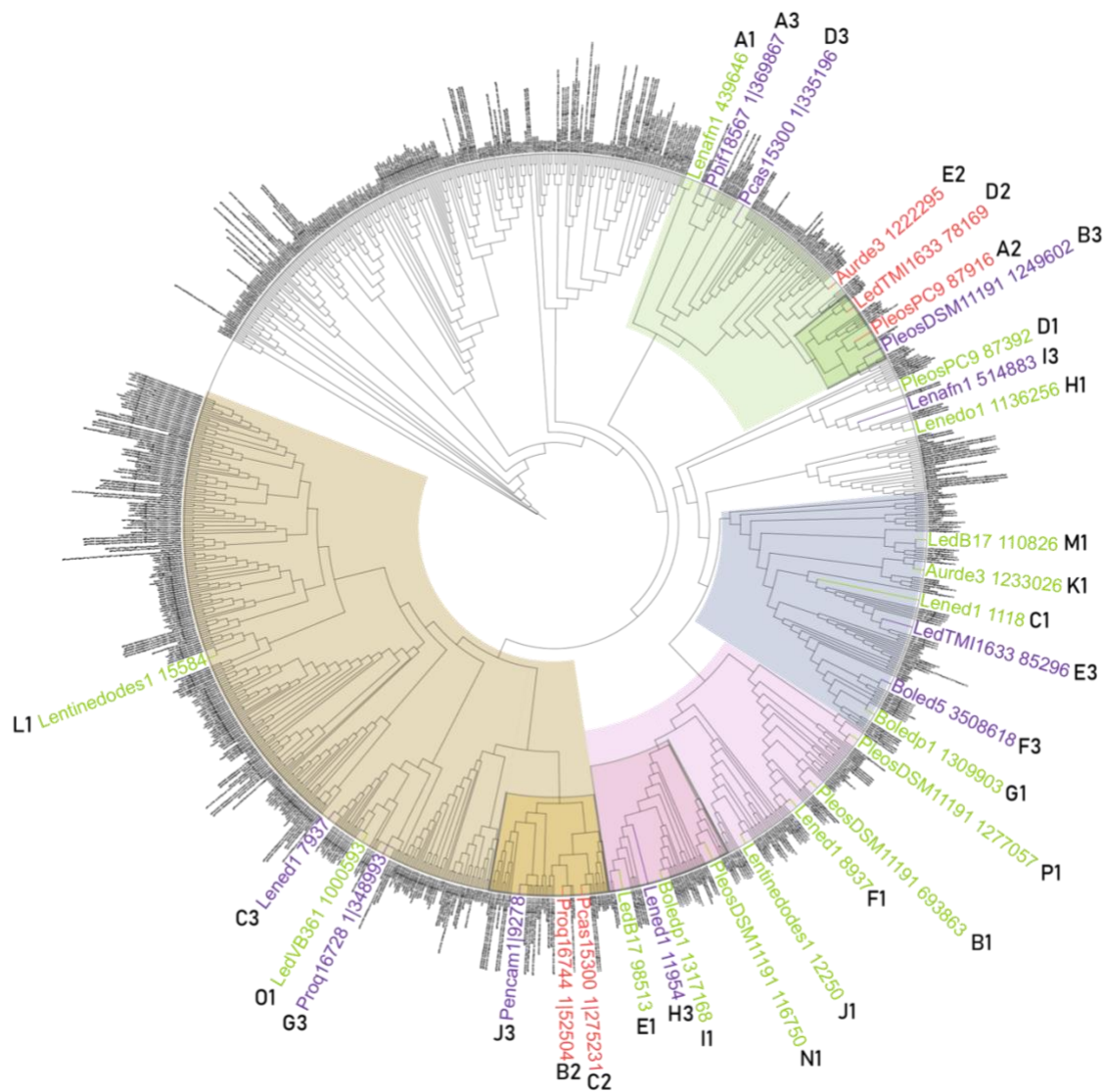


Figure 36. Phylogenetic tree, with highlighted parts according to clade division described in Section 5.5.

5.6 Terpene documentation in fungal species

The terpenes in this work have been found described in literature to various extents. Some have already been detected in the species used for this work. If that is the case, no new product for that species has been found, depending naturally on further modifications that can occur with clustered enzymes. The terpenes may also appear to be undescribed from the specific species but detected in other fungal species or e.g. plant species. Lastly, the terpene can be assumed as entirely new without any documentation in organisms. To get a final overview of the products and their occurrences, this will be discussed.

5.6.1 Documentation of monoterpenes

The main monoterpene producing TPSs B2, C2, and J3 were obtained from genomes of the *Penicillium* species, specifically used in cheese production. The detected monoterpenes in this work have to a much greater extent been documented from plants, with little to find about *Penicillium* production. The

monoterpene β -ocimene might be undescribed in the studied *Penicillium* species, but have been detected as products in fungi before (37). Geraniol also from *Penicillium*, famous for exhibiting several bioactivities including antifungal, have been detected as a minor product in endophytic fungi that reside in plant tissues (79, 80). Myrcene, produced as one of the products by B1 from *Pleurotus ostreatus*, is not a new compound for fungi as it has been described from several edible Basidiomycota species (37). β -Linalool, detected from *Lentinula edodes*, has been reported from the same mushroom species by analyzing freeze-dried samples (81).

5.6.2 Documentation of sesqui- and diterpenes

TPSs producing sesquiterpenes as well as diterpenes as products are obtained in this work mainly from the mushroom-forming species (*Pleurotus ostreatus*, *Boletus edulis*, and *Lentinula edodes*). Basidiomycota has, from genome analyses, been described as great producers of sesquiterpenes, as discussed in Section 2.4. Wang et al have performed a thorough study on mushrooms from Basidiomycota, where they gathered 174 ses-TPSs from 35 edible and non-edible species to which product formations have been described. The cadalane structures α -muurolene, δ -cadinene, and δ -cadinol have been considerably described, especially δ -cadinene and α -muurolene from *Pleurotus ostreatus*, from which they were produced also in this work. The cyclic terpenes germacrene D, α -copaene, α -cubebene, α -cadinol, and β -cadinene as well as the acyclic structures β -farnesene, α -farnesene, and nerolidol have with the same work been compiled from various Basidiomycota species, often edible ones (37). The sesquiterpenes α -bergamotene (A3, *Penicillium biforme*), isodene (B1, *Pleurotus ostreatus*), α -barbatene (D2, *Lentinula edodes*), valencene (G1, *Boletus edulis*), and β -himachalene (A2, *Pleurotus ostreatus* and D2, *Lentinula edodes*) as well as the diterpene cembrene (D1, *Pleurotus ostreatus*) cannot be found for their specific species but are described in various species of fungi by other works (11, 82-84).

Some less common sesquiterpenes have been produced from the edible mushrooms in this work. *Lentinula edodes* and *Pleurotus ostreatus* come second and third place respectively in world production volumes, hence obtaining a complete understanding of these edible mushroom product formations is of great interest (85). The minor product sinularene (N1, *Pleurotus ostreatus*) has been studied from a coral species and cannot be found in any fungal species (86). Furthermore, TPSs D2 and P1 obtained from *Lentinula edodes* produced myltayl-4(12)-ene as well as a putative sesquiterpene without a provided name. Myltayl-4(12)-ene is an uncommon structure, found only as a product in some oils such as oxidized olive oil, suggesting that it is a new product in fungi (87).

6. Conclusion

In this discovery-based work, uncharacterized TPSs from several edible fungal species were tested for product formation in yeast, by allowing *in vivo* production in combination with HS-SPME analysis as well as testing the enzymatic function *in vitro* and extracting the terpenes into a hexane phase. In summary, three TPSs were considered to produce mainly monoterpenes, nine mainly sesquiterpenes, and one produced a diterpene as sole product. However, the co-transformation with Erg20p(N127W) suggested that three more TPSs were active, as they presented sesquiterpene production. Three TPSs showed bifunctional activities, allowing both mono- and sesquiterpene production. The product pathways could to rather great extents be connected back to clades in the phylogenetic tree, one clade presenting

especially cyclization methods for cadalane production and another monoterpene production. Several sesquiterpene synthases formed multiple terpene products, while the monoterpene synthases and the diterpene synthase were rather specific. The sesquiterpene myltayl-4(12)-ene unknown to fungi were detected as a product, obtained from *Lentinula edodes*, as well as sinularene and one putative sesquiterpene without a provided name from *Pleurotus ostreatus*. The low pH of the induction medium seems to have affected the TPS catalytic activity, allowing the idea that the induction medium should be buffered for future work.

7. Future work

To fully understand the product formations, in terms of compounds and ratios, additional testing is required. To be more certain of the product formations, the analyses can be performed in duplicates or triplicates. Hexane extraction of the *in vivo* production in combination with a concentration step could handle issues regarding insensitivity and potentially ratio skewing. Furthermore, the clusters in which the studied TPSs can be located should be studied, to further elucidate the complete biosynthesis of terpenoid products from the unexplored terpenoid clusters. Several protein analysis experiments exist that can be performed to verify whether the inactive genes are expressed in the work. Possible reasons for inactivity or non-expression of putative TPSs have been described in this work, which may be utilized for future experiments.

8. References

1. Reyes BAS, Dufourt EC, Ross J, Warner MJ, Tanquilut NC, Leung AB. Chapter 4 - Selected Phyto and Marine Bioactive Compounds: Alternatives for the Treatment of Type 2 Diabetes. In: Atta ur R, editor. *Studies in Natural Products Chemistry*. 55: Elsevier; 2018. p. 111-43. <https://www.sciencedirect.com/science/article/pii/B9780444640680000048>
2. González-Hernández RA, Valdez-Cruz NA, Macías-Rubalcava ML, Trujillo-Roldán MA. Overview of fungal terpene synthases and their regulation. *World Journal of Microbiology and Biotechnology*. 2023;39(7):194. 10.1007/s11274-023-03635-y.
3. Ahmadjian V, Alexopoulos CJ, Moore D. Importance of fungi. In: Britannica TEoE, editor. *Encyclopaedia Britannica* 1999.
4. Antonelli A, Fry C. State of the World's Plants and Fungi 2023 2023 2024/03/15. Available from: <https://www.kew.org/sites/default/files/2023-10/State%20of%20the%20World%27s%20Plants%20and%20Fungi%202023.pdf>.
5. Naranjo-Ortiz MA, Gabaldón T. Fungal evolution: diversity, taxonomy and phylogeny of the Fungi. *Biol Rev Camb Philos Soc*. 2019;94(6):2101-37. 10.1111/brv.12550.
6. Milgroom MG. Fungi. In: Milgroom MG, editor. *Biology of Infectious Disease: From Molecules to Ecosystems*. Cham: Springer International Publishing; 2023. p. 109-22. https://doi.org/10.1007/978-3-031-38941-2_8
7. Chang ST, Wasser SP. *The Cultivation and Environmental Impact of Mushrooms*. Oxford University Press; 2017.
8. Dupont J, Dequin S, Giraud T, Le Tacon F, Marsit S, Ropars J, et al. Fungi as a Source of Food. *Microbiology Spectrum*. 2017;5(3):10.1128/microbiolspec.funk-0030-2016. 10.1128/microbiolspec.funk-0030-2016.
9. Hachmeister KA, Fung DY. Tempeh: A Mold-Modified Indigenous Fermented Food Made from Soybeans and/or Cereal Grains. *Critical Reviews in Microbiology*. 1993;19(3):137-88. 10.3109/10408419309113527.
10. Derbyshire EJ, Finnigan TJA. Chapter 16 - Mycoprotein: A futuristic portrayal. In: Bhat R, editor. *Future Foods*: Academic Press; 2022. p. 287-303. <https://www.sciencedirect.com/science/article/pii/B9780323910019000372>
11. Quin MB, Flynn CM, Schmidt-Dannert C. Traversing the fungal terpenome. *Nat Prod Rep*. 2014;31(10):1449-73. 10.1039/c4np00075g.
12. Avalos J, Limón MC. Fungal Secondary Metabolism. *Encyclopedia*. 2022;2(1):1-13. <https://www.mdpi.com/2673-8392/2/1/1>
13. Zeng T, Liu Z, Zhuang J, Jiang Y, He W, Diao H, et al. TeroKit: A Database-Driven Web Server for Terpenome Research. *Journal of Chemical Information and Modeling*. 2020;60(4):2082-90. 10.1021/acs.jcim.0c00141.

14. Vinha AF, Machado M, Oliveira MBPP. Terpenes from Fungi. In: Carochio M, Heleno SA, Barros L, editors. *Natural Secondary Metabolites: From Nature, Through Science, to Industry*. Cham: Springer International Publishing; 2023. p. 497-528. https://doi.org/10.1007/978-3-031-18587-8_15
15. NCBI. PubChem Compound Summary for CID 6557, Isoprene: National Institutes of Health; 2024 [Available from: <https://pubchem.ncbi.nlm.nih.gov/compound/Isoprene>.
16. Yang W, Chen X, Li Y, Guo S, Wang Z, Yu X. Advances in Pharmacological Activities of Terpenoids. *Natural Product Communications*. 2020;15(3):1934578X20903555. 10.1177/1934578x20903555.
17. Chang MCY, Keasling JD. Production of isoprenoid pharmaceuticals by engineered microbes. *Nature Chemical Biology*. 2006;2(12):674-81. 10.1038/nchembio836.
18. Liu Y, Ma X, Liang H, Stephanopoulos G, Zhou K. Monoterpenoid biosynthesis by engineered microbes. *J Ind Microbiol Biotechnol*. 2021;48(9-10). 10.1093/jimb/kuab065.
19. T R, Sharma D, Lin F, Choong YK, Lim C, Jobichen C, et al. Structural Understanding of Fungal Terpene Synthases for the Formation of Linear or Cyclic Terpene Products. *ACS Catalysis*. 2023;13(7):4949-59. 10.1021/acscatal.2c05598.
20. Trepá M, Sułkowska-Ziaja K, Kała K, Muszyńska B. Therapeutic Potential of Fungal Terpenes and Terpenoids: Application in Skin Diseases. *Molecules*. 2024;29(5):1183. <https://www.mdpi.com/1420-3049/29/5/1183>
21. Kramer R, Abraham W-R. Volatile sesquiterpenes from fungi: what are they good for? *Phytochemistry Reviews*. 2012;11(1):15-37. 10.1007/s11101-011-9216-2.
22. Dai Q, Zhang FL, Feng T. Sesquiterpenoids Specially Produced by Fungi: Structures, Biological Activities, Chemical and Biosynthesis (2015-2020). *J Fungi (Basel)*. 2021;7(12). 10.3390/jof7121026.
23. Schrader J. *Advances in Biochemical Engineering/Biotechnology*. Scheper T, editor: Springer International Publishing; 2015. 470 p. <https://link.springer.com/content/pdf/10.1007/978-3-319-20107-8.pdf>
24. Zhang FL, Feng T. Diterpenes Specially Produced by Fungi: Structures, Biological Activities, and Biosynthesis (2010-2020). *J Fungi (Basel)*. 2022;8(3). 10.3390/jof8030244.
25. Wang G, Wang Y, Wang B, Su M, Zhou S, Luo P, et al. Prevention and control effects of edible fungi and their active ingredients on obesity: An updated review of research and mechanism. *Journal of Functional Foods*. 2023;107:105621. <https://doi.org/10.1016/j.jff.2023.105621>.
26. Ríos J-L, Andújar I, Recio M-C, Giner R-M. Lanostanoids from Fungi: A Group of Potential Anticancer Compounds. *Journal of Natural Products*. 2012;75(11):2016-44. 10.1021/np300412h.
27. Zhao P, Guan M, Tang W, Walayat N, Ding Y, Liu J. Structural diversity, fermentation production, bioactivities and applications of triterpenoids from several common medicinal fungi: Recent advances and future perspectives. *Fitoterapia*. 2023;166:105470. <https://doi.org/10.1016/j.fitote.2023.105470>.

28. Brock NL, Dickschat JS. Biosynthesis of Terpenoids. In: Ramawat KG, Mérillon J-M, editors. *Natural Products: Phytochemistry, Botany and Metabolism of Alkaloids, Phenolics and Terpenes*. Berlin, Heidelberg: Springer Berlin Heidelberg; 2013. p. 2693-732. https://doi.org/10.1007/978-3-642-22144-6_121
29. Liang Z, Zhi H, Fang Z, Zhang P. Genetic engineering of yeast, filamentous fungi and bacteria for terpene production and applications in food industry. *Food Research International*. 2021;147:110487. <https://doi.org/10.1016/j.foodres.2021.110487>.
30. Abdallah II, Quax WJ. A Glimpse into the Biosynthesis of Terpenoids. *NRLS Conference Proceedings* 2017. <https://knepublishing.com/index.php/KnE-Life/article/view/981/2601>
31. Mitsuhashi T, Abe I. Chimeric Terpene Synthases Possessing both Terpene Cyclization and Prenyltransfer Activities. *ChemBioChem*. 2018;19:1106 – 14. 10.1002/cbic.201800120.
32. Raz K, Levi S, Gupta PK, Major DT. Enzymatic control of product distribution in terpene synthases: insights from multiscale simulations. *Current Opinion in Biotechnology*. 2020;65:248-58. <https://doi.org/10.1016/j.copbio.2020.06.002>.
33. Degenhardt J, Köllner TG, Gershenzon J. Monoterpene and sesquiterpene synthases and the origin of terpene skeletal diversity in plants. *Phytochemistry*. 2009;70(15):1621-37. <https://doi.org/10.1016/j.phytochem.2009.07.030>.
34. Nosenko T, Zimmer I, Ghirardo A, Köllner TG, Weber B, Polle A, et al. Predicting functions of putative fungal sesquiterpene synthase genes based on multiomics data analysis. *Fungal Genetics and Biology*. 2023;165:103779. <https://doi.org/10.1016/j.fgb.2023.103779>.
35. Wang Q, Cao R, Zhang Y, Qi P, Wang L, Fang S. Biosynthesis and regulation of terpenoids from basidiomycetes: exploration of new research. *AMB Express*. 2021;11(1):150. 10.1186/s13568-021-01304-7.
36. Wu J, Yang X, Duan Y, Wang P, Qi J, Gao JM, et al. Biosynthesis of Sesquiterpenes in Basidiomycetes: A Review. *J Fungi (Basel)*. 2022;8(9). 10.3390/jof8090913.
37. Wang S, Chen R, Yuan L, Zhang C, Liang D, Qiao J. Molecular and Functional Analyses of Characterized Sesquiterpene Synthases in Mushroom-Forming Fungi. *J Fungi (Basel)*. 2023;9(10). 10.3390/jof9101017.
38. Osbourn A. Secondary metabolic gene clusters: evolutionary toolkits for chemical innovation. *Trends in Genetics*. 2010;26(10):449-57. <https://doi.org/10.1016/j.tig.2010.07.001>.
39. Rinaldi MA, Ferraz CA, Scrutton NS. Alternative metabolic pathways and strategies to high-titre terpene production in *Escherichia coli*. *Nat Prod Rep*. 2022;39(1):90-118. 10.1039/d1np00025j.
40. Wang C, Liwei M, Park JB, Jeong SH, Wei G, Wang Y, et al. Microbial Platform for Terpene Production: *Escherichia coli* and Yeast. *Front Microbiol*. 2018;9:2460. 10.3389/fmicb.2018.02460.
41. Cao Y, Zhang R, Liu W, Zhao G, Niu W, Guo J, et al. Manipulation of the precursor supply for high-level production of longifolene by metabolically engineered *Escherichia coli*. *Scientific Reports*. 2019;9(1):95. 10.1038/s41598-018-36495-w.

42. Zhang C, Hong K. Production of Terpenoids by Synthetic Biology Approaches. *Frontiers in Bioengineering and Biotechnology*. 2020;8. 10.3389/fbioe.2020.00347.
43. Kampranis SC, Makris AM. Developing a yeast cell factory for the production of terpenoids. *Comput Struct Biotechnol J*. 2012;3:e201210006. 10.5936/csbj.201210006.
44. Herrero Ó, Ramón D, Orejas M. Engineering the *Saccharomyces cerevisiae* isoprenoid pathway for de novo production of aromatic monoterpenes in wine. *Metabolic Engineering*. 2008;10(2):78-86. <https://doi.org/10.1016/j.ymben.2007.11.001>.
45. Liu G-S, Li T, Zhou W, Jiang M, Tao X-Y, Liu M, et al. The yeast peroxisome: A dynamic storage depot and subcellular factory for squalene overproduction. *Metabolic Engineering*. 2020;57:151-61. <https://doi.org/10.1016/j.ymben.2019.11.001>.
46. Son S-H, Kim J-E, Park G, Ko Y-J, Sung BH, Seo J, et al. Metabolite trafficking enables membrane-impermeable-terpene secretion by yeast. *Nature Communications*. 2022;13(1):2605. 10.1038/s41467-022-30312-9.
47. Ignea C, Pontini M, Maffei ME, Makris AM, Kampranis SC. Engineering Monoterpene Production in Yeast Using a Synthetic Dominant Negative Geranyl Diphosphate Synthase. *ACS Synthetic Biology*. 2014;3(5):298-306. 10.1021/sb400115e.
48. Bureau J-A, Oliva ME, Dong Y, Ignea C. Engineering yeast for the production of plant terpenoids using synthetic biology approaches. *Natural Product Reports*. 2023;40(12):1822-48. 10.1039/D3NP00005B.
49. Wu T, Li S, Zhang B, Bi C, Zhang X. Engineering *Saccharomyces cerevisiae* for the production of the valuable monoterpene ester geranyl acetate. *Microb Cell Fact*. 2018;17(1):85. 10.1186/s12934-018-0930-y.
50. Amen T, Kaganovich D. Integrative modules for efficient genome engineering in yeast. *Microb Cell*. 2017;4(6):182-90. 10.15698/mic2017.06.576.
51. NEB. Applications of USER® and Thermolabile USER II Enzymes [Available from: <https://www.neb.com/en/applications/cloning-and-synthetic-biology/user-cloning/applications-of-user-and-thermolabile-user-ii-enzymes>].
52. Kawai S, Hashimoto W, Murata K. Transformation of *Saccharomyces cerevisiae* and other fungi: methods and possible underlying mechanism. *Bioeng Bugs*. 2010;1(6):395-403. 10.4161/bbug.1.6.13257.
53. Tang H, Wu Y, Deng J, Chen N, Zheng Z, Wei Y, et al. Promoter Architecture and Promoter Engineering in *Saccharomyces cerevisiae*. *Metabolites*. 2020;10(8). 10.3390/metabo10080320.
54. Sigma-Aldrich. Solid Phase Microextraction (SPME): Merck; [
55. Shirey R. Selecting the Appropriate SPME Fiber Coating –Effect of Analyte Molecular Weight and Polarity: Merck; [Available from: <https://www.sigmaaldrich.com/PT/en/technical-documents/technical-article/analytical-chemistry/solid-phase-microextraction/selecting-the-appropriate>].

56. Lancioni C, Castells C, Candal R, Tascon M. Headspace solid-phase microextraction: Fundamentals and recent advances. *Advances in Sample Preparation*. 2022;3:100035. <https://doi.org/10.1016/j.sampre.2022.100035>.
57. Classification of protein families [Internet]. Available from: <https://www.ebi.ac.uk/interpro/entry/InterPro/?search=terpene#table>.
58. Okonechnikov K, Golosova O, Fursov M. Unipro UGENE: a unified bioinformatics toolkit. 50.0 ed2024.
59. The Fungal Genomics Resource [Internet]. Available from: <https://mycocosm.jgi.doe.gov/mycocosm/home>.
60. Zhao Y, Liang F, Xie Y, Duan Y-T, Andeadelli A, Pateraki I, et al. Oxetane Ring Formation in Taxol Biosynthesis Is Catalyzed by a Bifunctional Cytochrome P450 Enzyme. *Journal of the American Chemical Society*. 2023. <https://doi.org/10.1021/jacs.3c10864>
61. Schilmiller A, Shi F, Kim J, Charbonneau AL, Holmes D, Daniel Jones A, et al. Mass spectrometry screening reveals widespread diversity in trichome specialized metabolites of tomato chromosomal substitution lines. *Plant J*. 2010;62(3):391-403. 10.1111/j.1365-313X.2010.04154.x.
62. Adams RP. Identification of essential oil components by gas chromatography/mass spectrometry. 4.1 ed: Allured publishing; 2017. https://diabloanalytical.com/ms/essential-oil-components-by-gcms/essential_oil_components_ebook.pdf
63. The NIST Mass Spectrometry Data Center [Internet].
64. Wiley Registry of Mass Spectral Data 8th Edition [Internet]. 2006.
65. NCBI. PubChem Compound Summary for CID 6549, Linalool, (+/-)- [Available from: <https://pubchem.ncbi.nlm.nih.gov/compound/Linalool#section=Boiling-Point>.
66. NCBI. PubChem Compound Summary for CID 7793, (-)-Citronellol [Available from: <https://pubchem.ncbi.nlm.nih.gov/compound/7793#section=Experimental-Properties>.
67. NCBI. PubChem Compound Summary for CID 637566, Geraniol [Available from: <https://pubchem.ncbi.nlm.nih.gov/compound/Geraniol#section=Boiling-Point>.
68. Steyer D, Erny C, Claudel P, Riveill G, Karst F, Legras J-L. Genetic analysis of geraniol metabolism during fermentation. *Food Microbiology*. 2013;33(2):228-34. <https://doi.org/10.1016/j.fm.2012.09.021>.
69. Ignea C, Raadam MH, Motawia MS, Makris AM, Vickers CE, Kampranis SC. Orthogonal monoterpene biosynthesis in yeast constructed on an isomeric substrate. *Nature Communications*. 2019;10(1):3799. 10.1038/s41467-019-11290-x.
70. Zhang C, Chen X, Orban A, Shukul S, Birk F, Too H-P, et al. *Agroclype aegerita* Serves As a Gateway for Identifying Sesquiterpene Biosynthetic Enzymes in Higher Fungi. *ACS Chemical Biology*. 2020;15(5):1268-77. 10.1021/acscchembio.0c00155.

71. Lucena RM, Dolz-Edo L, Brul S, de Moraes MA, Jr., Smits G. Extreme Low Cytosolic pH Is a Signal for Cell Survival in Acid Stressed Yeast. *Genes (Basel)*. 2020;11(6). 10.3390/genes11060656.
72. CreativeEnzymes. Effect of pH on Enzymatic Reaction [Available from: https://www.creative-enzymes.com/resource/effect-of-ph-on-enzymatic-reaction_51.html.
73. Parton RM, Fischer S, Malhó R, Papasouliotis O, Jelitto TC, Leonard T, et al. Pronounced cytoplasmic pH gradients are not required for tip growth in plant and fungal cells. *Journal of Cell Science*. 1997;110(10):1187-98. 10.1242/jcs.110.10.1187.
74. Muzafar S, Sharma RD, Chauhan N, Prasad R. Intron distribution and emerging role of alternative splicing in fungi. *FEMS Microbiol Lett*. 2021;368(19). 10.1093/femsle/fnab135.
75. Grützmann K, Szafranski K, Pohl M, Voigt K, Petzold A, Schuster S. Fungal alternative splicing is associated with multicellular complexity and virulence: a genome-wide multi-species study. *DNA Res*. 2014;21(1):27-39. 10.1093/dnares/dst038.
76. Samusevich R, Hebra T, Bushuiev R, Bushuiev A, Chatpatanasiri R, Kulhánek J, et al. Discovery and Characterization of Terpene Synthases Powered by Machine Learning. *bioRxiv*. 2024;2024.01.29.577750. 10.1101/2024.01.29.577750.
<https://www.biorxiv.org/content/biorxiv/early/2024/02/01/2024.01.29.577750.full.pdf>
77. Schillmiller AL, Schauvinhold I, Larson M, Xu R, Charbonneau AL, Schmidt A, et al. Monoterpenes in the glandular trichomes of tomato are synthesized from a neryl diphosphate precursor rather than geranyl diphosphate. *Proc Natl Acad Sci U S A*. 2009;106(26):10865-70. 10.1073/pnas.0904113106.
78. Vattekkatte A, Garms S, Brandt W, Boland W. Enhanced structural diversity in terpenoid biosynthesis: enzymes, substrates and cofactors. *Organic & Biomolecular Chemistry*. 2018;16:348-62. 10.1039/C7OB02040F.
79. Jayaram H, Marigowda V, Thara Saraswathi KJ. Secondary Metabolite Production and Terpenoid Biosynthesis in Endophytic Fungi *Cladosporium cladosporioides* Isolated from Wild *Cymbopogon martinii* (Roxb.) Wats. *Microbiology Research*. 2021;12(4):812-28.
80. Fajdek-Bieda A, Pawlińska J, Wróblewska A, Łuś A. Evaluation of the Antimicrobial Activity of Geraniol and Selected Geraniol Transformation Products against Gram-Positive Bacteria. *Molecules*. 2024;29(5):950. <https://www.mdpi.com/1420-3049/29/5/950>
81. Zhang L, Dong X, Feng X, Ibrahim SA, Huang W, Liu Y. Effects of Drying Process on the Volatile and Non-Volatile Flavor Compounds of *Lentinula edodes*. *Foods*. 2021;10(11). 10.3390/foods10112836.
82. Annaz H, Fakhouri KE, Bakrim WB, Mahdi I, Bouhssini ME, Sobeh M. Bergamotenes: A comprehensive compile of their natural occurrence, biosynthesis, toxicity, therapeutic merits and agricultural applications. *Critical Reviews in Food Science and Nutrition*. 2023.
https://www.researchgate.net/publication/369032502_Bergamotenes_A_comprehensive_compile_of_their_natural_occurrence_biosynthesis_toxicity_therapeutic_merits_and_agricultural_applications

83. Shahiri Tabarestani M, Rahnama K, Jahanshahi M, Nasrollahnejad S, Fatemi MH. Identification of volatile organic compounds of some *Trichoderma* species using static headspace gas chromatography-mass spectrometry. *Mycologia Iranica*. 2016;3(1):47-55. 10.22043/mi.2017.41532.1072.
84. Wu W, Tran W, Taatjes CA, Alonso-Gutierrez J, Lee TS, Gladden JM. Rapid Discovery and Functional Characterization of Terpene Synthases from Four Endophytic Xylariaceae. *PLoS One*. 2016;11(2):e0146983. 10.1371/journal.pone.0146983.
85. Tridge. 2020 Industry Report: Mushroom. 2021.
https://cdn.tridge.com/market_report_report/40/ea/8d/40ea8d79df9592c9434372a6555800b53d57dfc2/210203_Mushroom_Report.pdf
86. Kamel HN, Slattery M. Terpenoids of *Sinularia*.: Chemistry and Biomedical Applications. *Pharmaceutical Biology*. 2005;43(3):253-69. 10.1080/13880200590928852.
87. Al-Dabbas MM, Al-Jaloudi R, Abdullah MA, Abughoush M. Characterization of Olive Oil Volatile Compounds after Elution through Selected Bleaching Materials-Gas Chromatography-Mass Spectrometry Analysis. *Molecules*. 2023;28(18). 10.3390/molecules28186444.

Appendix

Supplementary tables

Table S1. Temperature program for Q5U PCR

Table S2. Temperature program for ligation PCR

Table S3. Temperature program for colony PCR

Table S4. Gene sequences

Table S5. Primer sequences

Table S6. Additional GC-MS program

Supplementary figures

Figure S1. Chromatogram for mono-TPSs

Figure S2. Chromatogram for sesqui-TPSs (prior analysis)

Figure S3. Chromatogram for sesqui-TPSs (1)

Figure S4. Chromatogram for sesqui-TPSs (2)

Figure S5. Chromatogram for di-TPS

Figure S6. Chromatogram for co-transformations

Table S1. Temperature program for Q5U PCR.

Step	Temperature (°C)	Time (hh:mm:ss)	Loop back to step	Passes
1	98	30	-	1
2	98	10	-	10
3	60	20	-	10
4	65	1:00	2	10
5	65	4:00	-	1
6	12	∞	-	1

Table S2. Temperature program for ligation PCR.

Step	Temperature (°C)	Time (hh:mm:ss)	Loop back to step	Passes
1	37	25:00	-	1
2	30	2:00	-	1
3	25	2:00	-	1
4	20	2:00	-	1
5	15	2:00	-	1
6	10	∞	-	1

Table S3. Temperature program for colony PCR. Step 4 used either 50 s or 1 min.

Step	Temperature (°C)	Time (hh:mm:ss)	Loop back to step	Passes
1	96	3:00	-	1
2	96	30	-	30
3	56	30	-	30
4	72	50/1:00	-	30
5	72	5:00	2	1
6	12	∞	-	1

Table S4. Gene sequences.

Gene code	ID	Sequence (5'-3')
439646	A1	ATGTCCTCTACCTCTCAGGATTCTGCACACCATTTCTTTGCTGAGGCACATCACCTAACATCTGAGGCTAAATTCACAATTACTGCTCTTCCTAATGCTGAAATTCGGGCAGTCGAAGCACTGGTCCACAAGCTCTATGCTGTTCGTCGCATACTAAGTGATTAGATGATCCTTATTCGGACCCAGTGGAGCTACAGTCTTCCCACAATGTCCGCAATTGAACGTCTTTGGCGTGATGTGCGTAAGGACAGCCTTGAGGCTACCGCAAGATATTTGAATACCTCGAACAGAATAATTTACTTGACATGTCAAACCTCTGTTTCAGCGCCTGTGCTTATTTGTGATTCCAGCCTCGCATCCAGGAATCCCTCAATCATACTGCTGATGCCTGGAATCATCACCGGCTCCGTACCGAGGGCCATAAATCACCTCTTGAATTTACGAGTTGAGCCGCGAGTATGCTAAAAACCCGAGGATATTGGACTGGAGATCCAGGGGATGATATTTCATACAGCACAACACCTTCTTATGGTTTGGAGCAGGAGGAGACTCCTGGTGATTACCAGGTGCCAGAAGAGGATAGTGATAACATCAGGGTCAATGAGGATGGTGATGTTACTTTGTACGTGAAGCCTTCGAGCACATGGATTTGACTTCAGACGTGATGATGGGAATTGGGGGATTGAGGTGTACTGCGAGGCTGTACAGCGTCTTAAGGTATTTGTAGAGAGCTTATAG
693863	B1	ATGTCCTCCCAGCCAACTCAAATCGTCTCTCCCTGACCTCCTGGCCATGTGCCCTCTGAAGGGATACACGAACCTCACTACAAAGAAGCAGCCGACAGATCTGTCCGCTGGATCGACAGCTACAACGTCTTCACTGACAGGAAGCGTGCCCTTCTTCGTCCAAGGGTGCAACGAGCTCCTTGTCTCCACACATTCCCATACGCACCTACGAGCAATTCAGAACATGCTGCGACTTGGTCAACCTACTATTTGTCATCGACGAAGTGAGCGACGACCAGAGCGGGAAGGATGCCCCGTGCAACAGGCCAGGTATTCTCAACGCTCTTTGGCGACCCAGAATGGAACGATGGTTCTTTACTGTCCAGAATCACAAAAGACTTTAGGGCCAGGTATTTCGCGCTCGCCGGACCAAAATCTTCTCGTCGCTTCTTGAAGCACTGCGAGGACTATATCAACGCAGTGTCCTGAGGCTGAGCTCCGTGAAAGGGCGGAAGTCTGGACATCGAGCCCTTCACTGCTCTTCGTCGCGGAAAACAGCGCGATTGCGCTCTGCTTCGGTCTCTTCGAATACGCACCTTGGCATCGATCTGCCCCGATGAAATATTTCAAGACAAGACATTCAGGATATGTACTTTGCTGCCGTCGACATGGTTAGCTGGAGCAACGATGTATACTCTATAACAATGGAGCAAGCCAAAGGACACAGCGGAAACAACATCGTCACCGTCTCATGAAATCCAAGGGAATGGGAATTCAGAAGCTGTGGACACGTTGGCGTTCACTTCCAACAGATTATGGACATTTATATGAATCAAAGTCGCGTCTACCATCATGGGGGCTGAAGTTGATGTTAATATCGCCCGCTACGTCGAGCACTTGGCCACTGGGCAAAAGGAAACCTTGACTGGAGTTTCGAAACTCAACGTTACTTCGGTGCTGAGCATCTCGAGGTGATGGCAACCAGAGTCGTGACTCTAAGGCCACGGCTCGCCCGAGGATTTTCGACGAATAG

Gene code	ID	Sequence (5'-3')
1118	C1	ATGGGTGCCCTCAACTCGAAGTCTTCAGTCTTCCAGATACGCTATGTCGCTGGGCATGGAGCCGGAAGCTGAATCCAAACTACAAGAAGGTCAAAG AAGAATCCAAGGCGTGGCTGGAAAGCTTTCACGCCCTTCCGCCCCAAAGCACAGAACGCCTTCAATCGTTGCGATTTCATTTTGCTAGCCTCGCTCGCA TACCAGATGCGTCTGCTGAGCATTGAGGACCGGTTGCGACCTTATGAACCTTTTCTTTGTGTTTATGATGAGTACACCGATGTCTGCGATGCCGTAGAA GCAAGAAAGCTAGCGGACATTGTTTATGGATGCACCTAGAAACCCGGACCTCCCTCGCCCCGAGAATGAATCTATAGTGGGAGAGATTACCAGGCAGT GA
87392	D1	ATGGTCTTTGTCCAATATCGCCTCCCTGACGTCTCTCTGTTATCTGTGGTCGGAATGTCCACGACCTGTTGCCCTCAATCCTCACTATGCCAACTCG ACGCTTATTTCTGTCGACTGGATCAATGCTCAGAAGCTGAGCCCTGCTACTCAGAAATCATGGATCGCAGTCCAAATGCCCTTGCTCGCCTCTCTCACA AATCCTCATATGACGCTTGACCAGCTTCGGGTCTGTCTGGAGTACATGGTTCTGTCTTGCGGTATGGAGGACTTAACAGACACTGCGTCAAGTGAAGA GTCCAAGAAGTGGGCTCACCTTCTTCATGGATCTACTCCGAAGCCCGGAAACACAGACGAAACGCGTGTGAAAGCTCACCAATGATCGGTAACGTT TGTGTCTGTAGTCTTCCGTGCGGGTAATGAAGGACCTTGGACCTTACCTACACCTTGCCGTTTCATCCAGAGCAATGATGAACATGTTTCGCGGAATGGT ACAAGAATCCTTGGACAGAGAGGCCGAAGAAACCGATGCCGATCAATCTACGCTTTCGCTCGAAGCTTACATCCATGCTAGAAGGACGACCATAGGC ATTCAACCTGTTTTGTTCTCGGCCGTTGGCCAGAGGTTTAGACCTACCAGACGAAGTATTGTGTTCTGAACAAGTAACGGGAATGGTCGATGCTAC ATTCGACATGGTCTTCTTGCAAACGATATATATTTCTACAAGAAGGAGCGCATCGCTGGTGCCGTTCAACACAACATTGTACCGTTGCAATGAATG ACGAATCCGCTCGAGTGGGAAGCTCCAACGTTTCAGGGCGCTGTGGACTACGCGCTACGAGTTGATCCACACGCGCTCATGATCGCTTCCAACGGCTATTA GCCGAGCTGTCTTCTACACCATGTGGAACACTTACCAAGACGCCCTTGAAAAATACGCAGAGAACATGCACGACTGGGCCGTCGGGAATGTTGTTT GGAGTTTGGCGAGCCCAAGATACTCCGTCTTCGACAACCCCTCTGCCAAGTCTACGATGATGATCGAATTTTATGCCCATTTGTGA
98513	E1	ATGCCTTCACTGCTCCTACAAAGTTTATTCTGCCAGATCTCGTATCTGACTGTCCCTTTCAACCACATGTTAGTCGTACAGAAAAACAGGTTACCACC GAAACGAAGAAGTGGCTTTTCCAAGGCGATAAATTTAGTGGCAAGATACCGGATAGATACCAACGGGCTGAATTTGTGGCTTGCTAGCCGCGATGGTCT ATCCTACTGACGCTTATCTCTCAATTAAGAGTCTGCAATGACTTCTCACCTATCTATTCCACTTGGACAACCTTTCTGACGATATGGATAACCAACGGAA CCGTAGCCACAGCAGATGTGGTGATGAATACCTTGTATCATCTGAAACCACAGCTCTCCTCGTATTTGCCAAAAATGACCAGAGATTATTGGAAGCGG CTCGTTAGCACTGCACTCTCCCGAGTCCAGCAACGCTTCATGGAGACGCTTGACTTCTTCCAGAGCGTAACGGAAACAGCCAACGATCGGCAAGGA GGGCGCAATACCTGATCTTGAATCATAACATTGCCCTTCGCAGAGATACGTCAGGATGCAAGACGCTCTTCTGATTAATCGAATACTCTAATAACCTCG ACATACCGGATGAAGTGATGGAACACCCACTCATTCGTAGCTTGGGCGAGGCTGCCAACGATTGGTTACGTGGTCTAACGACATCTTTTCTACAAAT GTTGAACAATCGAAAGGAGACACTCATAACATGATCCCGCTCATCATGAACCATCACTCCCTCGATCTACAGCCCGCAGTTGATCTGGTTGGTCAACT TTGCCAGCAGTCTATTGACCGCTTCATGAACGATCGAGCTCAGTTACCTTCTTGGGGTCCAGAGATAGATCGGCAGGTACAAATCTATGTCACTGGAC TGGAGGATTGGATGGTCGGTTCATGTCATTGGTTTTTGAGTCGGAAGGATTTTTGGCAAAGGGGGAGGGAAGTCAAGAAGTCGAGGGTGGTGA
8937	F1	ATGCGCAATTTCGAAAAGGTTTCGATCTTTGAGCTACCCCTTGAGGACTTGACGCGTTGCGTGGCTGGGGGCGAACTTTCTCACAGGATACTATCCCG ACCCGACGAGACCCTGACTGACCATATAAAAGCCTCAACATTTTACTCTTCCCGACCTATTGTGTCAGTCCGCTGATTGGATCAACAATTATAATGTCTT CACCGACCGAAAAAAGCGTTCTTTCATGCAAGGCTGCAACGAATTACTAGTGGCGCATACATACCTTTATGCTGGGTATGAGCAATTTAGAAGTTGCT GCGATTTTGTCAACCTGCTCTTCGTTGTTGATGAAGTTAGTGACGACCAAGATGGCAAAGATGCTCGCGCCACCGGACAGATCTATCTGAATACAATG CAAGACCCAGATTTTGACGACGGATCGACATTATGCAAAATCACAAAGATTTCAGAAAACGATACATGAGTTACGCTGTCTCTAATACGGCGAGAA GATTTTTAAAACACTGCGACGACTATATAGAATGCGTCTCAAGAGAAGCCGAGCTTCGTGAACAAGGGGAGGTCTCGACCTCGCCTCATTTTGTGCGAC CTTCGTCGAGAAAAACAGTGCCATTGCACTATGCTTTGGGCTTTTGAATATGTCTCGGACTCGATCTGCCACAAGAAAGTTTTCACGATCCCACTTTC ATGGAAGCATATTGGGCTCGCCCGACATGGTTTGTCTGGCAATGATGTGTACTCATACGACATGGAGCAAAGAAAAGGCCTAAGCGGCAACAAC ATTGTAACAGTACTAATGAAAGACAGAAATATCGAACTGCAAGCAGCTTCCGACTATATCGGATTGTATTTCAGGTTGCTCATGGATCGTTATCTGGA TGCTCGAAAGCGTATACCGTCGTGGGGCCGTGAGTTGATGCCGCCCTTGACAGCTATATGGATGGCCACAGCTCACTGGGTACGGGCAAGTCTTGTACT GGAGTTTCGAACGCAACGGTACTTTGGAGCTCGATTTTGGAAAGTCAAAGAAACTGGAATTGTCTATCTCAGCCCTCAGTCACCGGAAGAAGTATA AGCGACTCTGGTCTGAGAGCGACTAA
1309903	G1	ATGGAAACTCTTGCTTTTCATTTCACTCAGTGCACCTCATAGCCCTTCTCATCCGTTTCGCCCTTCTCTAACGCTCCTGGCGTGGTTTCCAGGAGGCGCAAG GGAATTACCAGCGAATCTATCCTTCGGGAAGAGATGCTCCTCAGGCAGAGCGTCAATAAATGTGAGACACCCACCAGAATCTTCATCCCGAACGCTGT TTGCCAGATGGCCATGGCCTCGTCGGATCAACCCATAATTACGCCGTGGTCAAGGAAGAAGCTGACGCATGGATGACGAGCTTCAAAGCCTTCAGTCC CAAAGCCCAAGAGTGTCTACAATCGCTGCAACTTCAACCTACTCGCATGTTTAGGGTGTCTTACAGCGAGTAAAGAGCATGTGTAGATCCGCTGTGATC TCATGCATCTGTTCTTCATGTTTCGACGAGCATCTGATAGGGCCACGCCCCTCCGAGGTTTGGAAAGCAAAGGAAGTATCATGGACGCGCTCGGGAAT CCTCACACGCCACGTCAGAGAGGGAATGGATTGGTGGGGAATAGCACGGCAATTTCTGGGAACGAGCCATACAAAATGCCACTGCGGTGCGCACAA AAGAGGTTCTGTCGGGTACCTGGATGAATATCTCGAGGCTGTGCTGAGGAATACATCGCAACAAAGAACATTCGATCTGTGATCTGAGTCTGAGTATA TCATCGCAGCAGCTCGTACAGCGGAGTGAAGCCTTCAATTACTGTCTGTTGAATTTGGCCTCGACATCCCTGATGAAGTTATGTACACCCCTGCCTTG CAGGAAATGATTCAAGCAGCTGTTGACATGATCGCTCTGACTAATGACATCTATTCTACAATCTAGAGCGGTTGCGTGGAGACGACTTCCACAATAT CATAACGTGCGTCATGAAAGAATACCAGACAGACGTGAACGGCGCCATGCTGTGGACGGAAGGCTTTCTGCTCAAAGCCGAGGAAAGGTTCAATGATG GCCATGGCGGCCCTTCCCCAATGGGAAGAGCCATTGAATTTTCAGGTCAAGGTGATTGCGATGGCGTGGGTCACTGGGTGCGGGCTAATGATGACT GGTGCTTCGAGGGTGAGAGGTATTTTGGTAAACAAGGGTCTGAGATCAAGGAAAATAGATGGATGTCGCTCCTGCCGAAACAAAAGGCAACGGAGA TTGGTCTGTGTCACATTGATAGTTCACCTGTTGTAG
1136256	H1	ATGCCACAATTGAGGGTGATTTTATGATTACTTTGTCTTGCTCTTTTGGTTGAAGTCAATGAGCGACGTTTCCAATGCCAGGCAGGCAATTTGATTTG GATGACATGATCAACCTCTTGAATACTTTGTCCGACAAATGAACCTTTCGACATCAATTAAGCAAGAAGACCCCTTGGTACATCTCGCTGCAAAAGTTTGC AAGGAAATGCTTCCCTCCTCTTTCTGCCTTATATCGTTATGATGCAACTCAAGGTACCATTTTGTACCTTCAAGCAGCTGCACAGGAAATCATGATCG CATGACACACATCAACCTTCACAGTAGACTCGTACCTTGTCTACCGCCGCAAAAGTAGTAGCCATACATCTGTATTGCTTTCACCCGGTGGATGTGCC AAGTTAAGCTTGAGCCCCGAGATCCTGCAAAACCGGCTTTTGAAGGCTACTGGAAGTCAACAATTGATCTGATTGCACTTCCAATGATCTGATTTC TACAAAAAAGAGAAATCACAAGGAGAAGACATCCACAATCTTGTATCGGTAGCTATGGTTGATCCTGCCATTCCAGTGAAGTCAGGAGACATTCAGG CTGCACTGGATTTTGCAGGTGTTAAAAATATCAGAGGCTTATCTTCGCTTTCAACAATGCAAGCTCGATGCAATTGCACAGACCGGCGCTTGGCGATTCT GAAACCGGTCAGAGATAATGCGCTATGTAGATATTTTACTGAAGACCATTTGATGGGAACATTGATGGCAAAATTGATCCTAGGACGGTACGCTACC AAGTCTTTGAGACCCCGAAGAGCGCGCGCAATGGTTGGTATCAATTTGCATTGTA
1317168	I1	ATGTCACCTATATCTTCCGCTCCCGCTCACTTACGACCATTCAACTTCTGACCTCGTCTCGCATTGCACGTTTCGACCTCCATACGGAACCGAAACCAC AAACCTCGTCAATCGAATCAAAGCGGTGGATTTCACCGGTGACCCCAACCTCGATGAAGTTGCTCGTCGCGCTTTTCATGGTCTCAAGGACGAAAA GCTCGCTCGGATGTGTTATGCCAATGCCGGTTACCACCAGCTCCGGGTCTGCAGCGATTTCCTCAACTGGCTTTTCCATTGGACAATCTCTCCGACGA CATGGACGATCGTGGTATAGACAACGTCGCCAACGTTGTGATGAACCTCATCCATCACCACACACGATACCGTTGTTCGCGCAGCTTAAACCGGATGA CAAGAGATTTTACAAAACGGATGATGCAAAACGGCTTCCCTGGTACTTGTGACAGCATTAAGAGACGATGGGGCTCTTCTTCCAAAGGGCATCCAACAG CAGGCCCTCGATCGTGCAACGGCGCCATACCTGATCTCGAGTCGTACATATCCCTTCGCCGCGACACGAGCGGGTGCAAAACCCAGCTGGGTACTCAT TGAATACGCACACAACCTCGATATCCCGGATGACGTCATGAATCATCTATCATCTGTAGCCTAGGTGAAGGCCACAAACGATCTCGTCACATGGTCCA

Gene code	ID	Sequence (5'-3')
		ATGACATCTTTTCGTACGACGTCGAACAGTCAAAAGGGCACACGCATAACATGATCGATGTTATCATGCAGGAGCAAGGGCTTGGTCTCCAAGATGC TGTC AACGTCGTCGGTGAATATTGCAAGGGCGCTATCAACCGGTTGTACGCGACCGAGCCAATTTGCCGTCCTGGGGACCCGCCATCGATCAGAAAG TTCGCGTGATGTCGACGGGCTTGCAATTGGATCGTGGGCAGTCTGCATTGGTCTGTTGACGCGAGAGATATTTGGGCAAGGCTGGGAAGCAGGT CAAGGCGACCCGCTCTCGTACGCTACGGCCCAAGACTGTTCAGCCGCCACTCCGAACGGCATAG
12250	J1	ATGTTCTCTACAATTGCCTAGTGCACCTACACTATTCGCTCATGATTATCAAGGAGCGCAATTGATGAAATATACAGATTATCTGAGGTATACTGAGAA CGAAGAAGTATCGTCCAGGATAAAAAAGGAGTAAGGGTTCTACATGTAGTCGCCCTGCACTCACGAACTCAGTCCTGACTCACGCTACGCTCCTTACCA CCAACTAAACGAAACGCCCTTCTTCTTTTGCACCCACATGATGCAAGCCAAGACCAAGACCTTTGCTAGTTCGCCGAAATCCCTCTCCAACACTGTC GAACAATTTTCATTCTCCGATTGATATCACACTGCCGTTTCCCTCATTTACCATCCCAAGGGGGATGTCGTGGCCGCGGAGTCTATCCGTTGGCT CGACACTAACTGTCTGATTGAGGGAAAAACGCAGGCGAGCTCTATATGGCTACGAGCAGGAGAACTGACTGCATTCTGCTACAACACCCGCTCC GATGAACGCCCTCAGAATAAATATCCGATTTCATGAATTATCTCTTCCACTTTGGGACAATATACGCCGATGGGATGATGACCAACGATACTGAGGCTCTATC CGACGCTGTAATGAATGCCTTGCAGTTCACCGACTCGAAGGTTTATCGTCTCACCCTGAGCTTAATTATGTGCAGCTCGAAGACGAACTGAACGCAG CGAAATTGGCCAGAGACTTCTGGGTCGTTGTATACGCGGAGCCGCTCTGGCTGCCAATCCCGGTTCAAAGATAATCTTGGACTTTTCTTCCAGGCT AGGTCGCGGTTGATTACGTCGGGAATCTCTGCTACCAGACCATGGAATATTTCAACGCGGAATAGGTCAAAGGTACCAATTCCTGGGAGCTGAAATCGA TGATATGGTGCAGCGCTACGTGCAAGGCTTGCAAGGACTGGATAATTGGATCGCTTCATTGGAGCTTTTGACTACTCGATACTTTGGAGAAGAGGGTCT AAGATGTCAAACAGCATAGATTTATTGAGCTTATGCCAATACGGGAGGAGAAACAAATGTTAG
1233026	K1	ATGGACTCGTGGCCATGGCCTCGCAAGCTGAACGTACACTACGAGGAAGTCAAACGGGAGTCGGACGAGTGGTTCTTCTCCTTCCACCCCTTCACGCC CAAGTCTTTGGCTGCGTTCGAGCGGTGCGACTTCGCAAAACTCATTGTCTTTGGAGGCAACTCGGCGTCGAAAGAGGAAATCCGCGTCGGCTGCGAC CTTATGAACCTCTTCTTCTCATCGACGAGTACACCGACCGCCAGCCGCTGGATACCGTCAAGACCACGTCGGAGATCGTCTTTCAGCCCTGAAACAA CCGGAGAGAGCTCGGCCGAGGAGAGATCCTTGTGCGTGTGTCGCAAAAGAGTTCTGGCAAAATGCAATCCAGCACTCGACTCCGAGGCTCGC CGCTTCTTTCGACGCTTCTCGTTGTACCTCGACTCGGTTATCGTCGAGGCGGACGATCGCGACAAGAAGTACATCCGACCGGTGCAAGGATACAT CGACATGCGCGCGCAACAACATCGGGCTCCGTCCTCGCTTTCATCGGCACCATGAAGCTCGACATGCCTGCTCCGGTCTGGGATCACAGGTACCTTC TCGACCTCGAGAACATCATCTCCGACTTCAACCGGTTTCGACAATGACATGGTTTCTTCAACAAGGAGCAGGCGTCGGATGATGCTGACTGCAACATC GTCAGGATTCTGATGCACAGCTCGAGACAGATTATTCGCGCGGATCGAGTGGATCGTCAAGGAGCACAAGAAGCTTCAGCGCGAGTTCTCTCGAGT TGGCCGCTTCTCTCCCGCAGCGTCTGCGCGACGACGGGTGTCGAGGTGACCCCAAGGTCGGGGAGCAGCTCTGGCTCTACATCGAGACATGCGGCA ATGGCCGCGGCCAACAAAGCAGTGGTCTGTTGAGAGCGGCGCTACTTTGGCAATAAGGGGTTGGAGATCCAGGAGACGCGTCTCGTGGCAAGTTG CCCAAGCGCGCCAACGATGACAACAAGCTCGGCCTGCCTCAGGCAATGTAATCATCAGATCCCGACTGTAGAAGTCTGCTAG
15584	L1	ATGCCGCCACCTAGCGCTGTTGGCAGCTATCAACTGCCGATCTCTTGTGCGATTTTCAAGACTATCATGAATTGAGCATCAACGGCCATTGTGCTGCC GCTAGTCTTGTCTCGGAGAAATGGGTTCTAGATATAGCTCAAGAAAATCAGGTCGCTGCTTGTCTTGGCTCCCTTCATCGAGAGACTCCCCAA CATGAAGCTCGGTTTACTTGGGGCATGCAGCTTCCCCAGCTCGGATTTGACCCAACCTCATTCTCATGATAGATTTTGTACTCTCGTTGTGCTCGATAC CGAGAGCCTGCTAAGATCAAATTCCTATCGCATCGCATTGGAGCCTTGGGAATGACACGTCCTCAACTCCGGCTACGTTGAAATATTGCAAAACACATC CTGTTTTACAGCGACTTCTGACCCGGCTCTCCAGGCTCCAACATAGTACAACCGAATCATGGCAGCAGCGTTTACGTCCTCTATAGTAGCCTTTTGGC ATTCACGGCAGAAAGCTATAGATTATCGACGCGAAGGTACTGCGCTAGCGCATCTAGAAATAGAGGAGTACATTGGTTTTTCGACGCGACTTGTAGCGG GATACCTTTTCATACTTGACTTGTCTGGAGGCTGTGCAAGGATTGGAACCTTGCTATCTCACCCTACAATGATGATTGCGTACGCTACGACGCTCGTGG GGGATATCATCGTGTGACCTGGGATGTATTTTCGTACAACATCGACCAAGTCTATTGATAATAAAATCAACATTGTGTCACTTTTACAATCCAAAAGA GAGATTGCATTGGTGCACCTCTATCAAAGAAGCGGCGATCTTATTTTCAAAGGATCGATGCTTTCAAATTTCTCGAAAAGTCTTTGCTTTTCATCTATG GCGCTTCGCTTCGGATCGTGCAAGTGATACTGATGCAACCGCCCTGAACCTCATTATGGAACAACCTTCGCTGGTCTATTITTCAGCAGCCCAAAAACC AAATAACCTGACCCGCTCTGTTACCGAGTTCCAAGCGCATGACATAGCACTCTATATCCGTGCATTGAAAGACTGCATGGTAGGGTCTTGTGATTGGG CGTAGCAGACAGACATATTTTCGGAACAAAAGGGGATACGGTCAAAGGTTTGGATGGGTGTTTTGTACCGAACCACAAATCGTCTCGCGCAGAAAC GACAGCCTAG
110826	M1	ATGGATGGACTGCGAAACCCTCACAAATGCTGTAATGTAAAGGGAGGCTCAATATTGAGAAAAATGGCAGCTGAATTCTGGGCAAAATACCATGTCAA TAGCTACCCATCTTCCAAAGCCCACTTTCTGCATGCATGTCATGAATGTATGTTCAGTTGTGGAACAAGCCGAGATCGATCCGGCAACAAAGTC CGTTGATCTCAAGATTACCTTGAAGTCCGGCAGCTGGGAGTTCACCGCTACCTGCTTCTTTTGAAGCTTATGACCTTTCTGATCATGTT ATATCCACCCGGCAATGAAAACCTTGACTCGATGTGCTGTGACATAAATTGCTTTGGGAAACGACATATACTCCTTTAAGCTCGAACAAGCTCGTGG AGATGGATCTCACAACATCATCACTGTGGTGATGATGGAGCTCAAGTTGGATTTACATGAGGCAATTGGAATGGGTAGGGCACTATCATCGCGAAAGA GAGCTGGAGTTCTTGAGAGCAGTAAAGGAACTACCATGTGGAGTCCAGAAATTTGACAGACAGTTGCAACATATGTCAATGGATTGTGAAACTGGG TTCGAGCTAACGATTGCTGGAGCTTTGAAAGCGGACGATATTTTGGCCAGGATGGTTTGAAGGTCCAGGAAACTCGGATGGTTACATTGGCACCGAA AGTTGCTTGA
116750	N1	ATGAAGATTCTCTCGATTCAAGTGCATATTTACCAGCGTTGGCCCCCACTTGAAGAGTGCAAAACTTGCCAGGAGTCAGCAAGAACGCTCTCTATG GCGTGTTACATACGCTGTCAACAGGGTCCCTCTCGAAGCACAGATCTCCGTACAGATCTGTGCGGGTTCAATGCAATCGACAGGTGGGCGGCTCGCG TCAGTATAAGAAGGCCAACGTCCTCCCTCAATCTGGAATCTCGAGCTTATCATCTTCACTTACCTGACCTTCCTGACCTTCCTGACCTATCCATCTGCGCG TCAACGACAACCTGTGAAGAAGTGGCGCGCGTCTCGGAACAGTGGCTACTCGACATGGCGAATCACAGCGAGCGCAACGTCGTGCTTTTCTTGGGCT CAAGGCGAGCGAGTTGACGGCAGCTTGTATCCCCGACGCCGACGCTTTCACCTCCGCGTATGCGTGGACTTCATGAACATTTTGTCAACCTCGATG ATTGGCTCGATTGATGATGTAGAGGGAACCCGTGGAATGCATGACTGCTGTATTGTGCGATGCGCGACCCGCTAAATTCGAGACGGATAAGCG CGCGGGTATCATGACGAAGTCATTCTTTCAGCCGATTTATCGAAACCGGCGGCGCTGGGTGTACAGAACGCTTCATTACACTATGGATCTCTTTTCA AGGCTGTTCGCCGTTCAAGCGGCTGATCGCGAGAAGGATGCCATTCTGACTCGATCTTACATCACCATTCTCGTCGAGATACGAGTGGCTGTAAACCT TGCTTCGCCCTTATTGAATACGCTCTCGAATCGACCTTCTGATGAGGTAGCTGAGCATCCGTTGATTTCGCTCCATCGGAAGAAGCGACGAATGACCT AGTGACATGGTCAAACGATTTGTTCTCTTACAACGTCGAGCAATCGAGAGGAGATACACACAACATGATTCCGGTCACTCATGCACCAACGCTGACCTG AATCTCAAAGAAGCAGTAGAATATGTGCGAGCTCTATGTAATCAAGCATCCAGAGGTTGCAAAACCGATCGCAAGAACCTCCCAACATGGGGTCCCG AAATTGACAGCTGGCAGTTTATGTGCAAGGACTCCAAAACATGGATCGTCGGTTCTCTGCATTGGTCTTTCGATTTCGAGCGCTACTTTGGTACCT CCGGCTTTGAAGTCAAGCAACAGCGCATAGTCAAACCTCTTCCAAAGGTTTCATCATGA
1000593	O1	ATGGCACCATTGCGATCTCACCTGTTGATACTACTAATGTTTCTTCGTACTTTCAGTTCTTTTCCCGGTCAAGAGCTGTGAGCTTGACGCTCTACCAGCC ATCCAACAAGGCTTGAACGATACAATTTACGGCTGTACGACTCCGGGTCAAAAAGAACGGAAAAAGCTGAATACCGGCATACCAACCCAGCAGGG AACATTTTGGCTTGAGCCTCGCGCTATGCGAAGCAGATCGGGTCGGTTATGTGGTAAAATTCATTGAGTTTCTCTGTATCGTCGATGATGTGATGGA GGATCTTCCCTTTGGAGAAGCATGCACAGAACATACCATTCTACGACAAGCACTACACGACAGCTACGACGAGCACCAATACGGCGGGAGAGCAGTC

Gene code	ID	Sequence (5'-3')
		AGTCTGATGAGGAACCTTTCTCTACGAGCTCCGGATAGAGCTCGTAGCATGGAAGACCCAAGCACTTCCTGTCTACTGAAGACTCTGGACACCACTCT CCGTGACAGAGACAGCAACGACTCAGAGTTTCTAACCTCGCAGGATATATTCCTATCGAAAAATGAATTTTGATTATGACTTTGTGTGCCAATCTAC TACGTTGGGCCATGTGCATCCCGCAACGATCCAAGATGATCCGCTAGCAAGGAGCTACGAGCACATCATTGGGGTCATCGTCGGCCTGACAAACGA TTATTTCTCTTGGGAGATGGAGCGCTAGAGCCTACAGATAGAATTAGGAATGCAGTTCTTGTCTTGATGAAGGAACATTCAGTCAATGAGATGCAAG CGAAGCCGATGCTGAAGGAAGTAATCATTGAGGAGGAGAGAAAGGCCAGGGAGTTGATGGCGAAATCATGGACCTACCAAGGAACATGAACCGT TGAAGCGCTATGTGGCTGGGATGGAGCTTTTTCAGCGGGATATAGCTTCTGGTGTGCAACTTGTCTAGATACCACCGACCACAGAACGAGGAAAG TTTTGTAGAGCAAGCTAAATTGTAA
1277057	P1	ATGCCTTCCAGTGGAATAGCAACACCAGCACTCGATATACGGCTACCTAATCTCATGGGCTACTGTCCATTGCCAATCTTTATTAACCAGCGTTATCGT TGGTCAAGGAGCGCTCTACCCTCTGGATACTAGAGCATGGGTTCACTACTGCGGGATCAGGAATCCTGGGACGGGAGCTACTGGCCAGTGTTC CACACCCACACGCCGGCAAAGAGAGCTCAAGATCGCTCGGATTTCTTAATCTCCTCATGGCCGCGGATGATTTAACGGATACGCAAGGTTGCGCG CAACGCGCAGATGACCTGTGATATATTTCTTGGTTCAATTGCGCGACCCCTCAATGCCTCCTCCTCGAATTTCTCCAGAGCGGTGGCTGACTTCGCCAA GCGGTACGAACCTACGGGCTCTTCCAATGCTCAGCAAGATTTATGAAGCATATGGAGTCGTAAGTTCGCAACAGTCGCGTCCCAAGCCAGCTGCGC GCAATATCTGTCACACCGGCGCTTGAGCCATTGCTCATCCAGCAGCAGAGAACTGCTGATGAAAGCGTAATACGACCTAATAGAGCTCGAAGCTCG GCATTGATTACCAACAGGATATACGGCGACCCAGAGTTTCTGGTAGCGTATGAAGCAGCGGTGCGACCTCGCATGTTGGGTAATGATATTTATTGCG TTCTTCAAAGAGATTGAGCAGGTCGAAGGGGTTACAATGTGATCAGGTTGATACATGGTGAATTTGGGCATGGATGCAAGCTGCATTTCGACTACG CCGGCGCAGAATTCGGAAGGATTTGGGATCGGATTCGAAGCTCGGAAGGCTACGATGCTTCCTGCTGGTGATGAACACGATGACGATGTCCGCGCTTC ATAGAAGGGCTAGAGATTTGGTGTATTGGGAATATATACTGTGCATTTGAGAGTCGACGTTATTTCCGGGAGCAGGGACGTGAGTTCATGGCAACAC AAATATGTTCCGTACAGACCAAGGCATCAATTATTGA
87916	A2	ATGCCTGTCAATATCCGTGACATTGTTGCCTTACATCTTCGCCGTTGCAACATACCTTATGAGGTACCTGTTCCCGATGTAAAGCTCCAGGCAGCATGC CTCAATTACGCCAAGAGCCACGGGTACGAAATTGGCGGCCAGAAGACTCTAGGGCCATGCATACCTGGAGGAGTCGTGATGGCCTCGAAGCGTTTCG GCCATATTGGCGACCTCCGACTCGCGTCATCATTTGGCATATACACGGCGTTCATGATTTACTTGGAGGATATCTCTCCAGCGACATCGAGGCTGTTG CACAGTTTAAACCAAGATTTTACCGTGCAGAGCCTCAGCTCGATAAAGTGCTTGACGACTTTTGGCGAGCTTCTTCGAGAGTTTCCCGACTATTTTTCGA CGGCTGGGTGCGATATGATCGTCACTCAACACTGAATTTTGTACGTCCTCTTTCATGGACGTGAGACAGAAGGCATGGACGTGATGGAACCGCG AAGCGATACCCAGGTTTGCACGGATTTTGAACGGGGCTCGCCCTCGCGTACACTCTCCTAGCGTTCCCGAAGTCGCTGCCAATATCAAGTTATATTC AGCGGTCCCGGAGATAAACGTATATGTTGAGAACACTAACGACGTTCTCTCATTTTACAAAGAACAGATGGCGGGAGACGATATCAACTATGCCAGC TTGCTTTCTATTCTTACAGTGTAAACGAAGTACGAAGCACTCATCCACCTCAGTAATGTGCGCGTTGATGCGAAGCAACGAATTCCTCGGCTACTGGA ACCGGTTCCCGAAGCCTTGAGGCTTATACACGGTTTCAGGAACGGCTACGTTAGATTCCATACACTGCTCGGTGTTGAGGTATAAAGTACGAGCACTGA AGTTGACGAACATAAGTGATTACCAAGATAGACTTGGAGCCCATGTCAATCATACGGCATGTAAAGTTTGTCAACGTAACCTTGTGTAATTAG
52504	B2	ATGGCTACATGCGATTATACCTTCACACACCAGCATGGGATTGTTTCCTTCTGCTCGGATTCTCTACTGGTCGGGTTGACCTTCCAATCAATCCAGCCGGC CATATACCAATTATAAGCGTCTCAACGACCAACCTGTGATCAAGTCATGCGAAATCCCTCTAGGAATACACGGCGTATCCCTCAAGCTCAAGATCCG TCATTACGACATAGTCCCTATCGATGCACTCCACATCGGCTGTCTCTATTACACAGCTGATTTCAAAGCGACCATCCAGCCGGAGAAAGCCGGAG ATACCAAAAGACTTCTATGAGAGATGCTTTGCTGAAGCAGCTGCCATTAACCTGTTTTTCCCGGATGTACGATCGGAGAACATCCGCATATGCATGACT GCGTGGCTGGCTGCGAACTGCGCGCGAGATGACACACTTGAGGCTATGCCACCTGCTGCGGCCAGCGTTGCTTGGAAAGACCATATCGAAGCTCC AAGGAAAGAAGGCAGACGTCTCCCGACAAACAAGTGCATCGATTCTTGTCTTTCAAGAATATTGCGTCCAGAAACTTGACCTCAGCGAAAG CACCTCCCGGAGCTGAGCAATGATATATGCGATATGTGCGAGGGCTTGGCTTGACGAGCTCCGATTTTCGACAGGGCATTCTGCCTGGGAACCTTGAA CTTACCTACAATTCGCGGTAAACAACTGGGTATCCATCCATTCTTAACCTGATTCGCGCACTGCACAAACCAATCGAGGGCGACTATCTCTCCAGA CTCCGAGATCTCCAACAGCGGGTTAGTTTGGTGTCTGGGCTGCAAAACGACCTTGTGTCCTGGAGAAAGATCGTCGGAAATGGCGAGCCTATGAATG CAGTCTTGGTATCCTTGAAGGAACAAGCGGGGACGGATGTGGATCACATGAATGACATATTGCCGAGAACGATCGATGAGGTGTGCGGTATCCACAA TCGTTGTCTTTCAGCTGCGGTGGAGATATACGAAGGACTCCACGTTTCATTGAAATGGGGATATCCATGACCCGATTCTGGAGACTGCAGTCTCGGCT TTGCAGAAACCCATCTGAAATGGTGGCATCTGTCCAAAAGATATCAGGCAAAGGTAGAATAG
275231	C2	ATGTTGCGGATTGATGGATGGCGTGAATATGAGAAAAGGGAAACCAACCCCGAAATTCAAAAACAAATGTTACGCTAAAGAAGAAGAAGGATATA TCACTTGGATTTCGACATACCCTATCATGGCCTAGTCAGCATCCCCCTTGAGGAAGGTTTCAATTTGGTGAAATAGATATCACTCAGAGTCTTGAAGAA TGTCGTGAGTATCCACGGAAGGTTTCAGGATTGCAATGCAAGGAATGGTCCAAAGGGCAAACCGAAGAGCTATTGGAGCAAGCGGTTGGCGAAGTAAA AGCCATTGGCCTATTTTCCCGGATGTTGAGCTGAACATGTCCGGATATGATGGCTGCATGGCTCGCAGTACTCTGCAGCTTTGATGATTACTCGA GTTGATGAGTCCCGACGAAGCAGAGAATGCTATCCAGTGAACCATCGCTGCTTGGAGGTAACAAGGAATAATGACGGGATATGCGATTCTTTCGA ACCAAAGCGCAGCTTTGTTTCCTTATGTCATTGTTTCAGAAGCACTGTGCGCAGTATCTCTCGGATGCACTGTGAAAGATTTCTTTTCGAGAAATCTGC ACTGTTTTCGCGGCTTTCTGTCAGGAGAATCAAGTTTCGACAGGGCTACTTGAACCGGATTTGAAGAATACTACATGGAGATTGCAACACGAACAATCG GAGTTGCGCCCTTTTTAGTCTTATCAAAAGCGATCTATTCCCCCTTGCTATTGTTTCAGACAATATCTGGCCATGCAAGGTAATGAAATTTGCGG TTGGCTTCAAAACGATCTTGTGGGCTAGAAAAAGACATCGATAAAAAAGAGTCCATGAATGCCGTACTTGTGAGTATGGAGGAGCCCTGGGGAAA GGAGGAAGCGCAGGGGACAAAGCTTGCAGACTATTGCACTGTTTGTGCAATTCATAACTCCAGTGTGCGGAGGCGGTACGGATACACCAGAGA ATTCTACAGGAGCAGAGGAATCGGAAATAATTGTGCAAACTCTCAGCTTCTCTACTGAGACCACTTCAATGGTGTACAACAGCCAAAGAGGT ATAGAACAAGCTCGAATAA
78169	D2	ATGTCCTCTAGCAAGTCAGAAAAGTTGAACATGTCTACCCAGCCAAGTGAAATTCAGGGATGAACGTCGCAAAAGAAGATATTTCTCCGCTATAC GGTCTTTCTCTCGAGGTGCTCGATTCCATATCGAAAAGTGCCCATGGATCTACCGTTTCTGGAATTATGCTACAACGAGGCAGCCAAACGAGGGTAC ACTGTGGAAGGCTCTGACCCACTGAAACCTTTCACTCCAGCAGGGGTACCATAGGCCATTAAGTGTATGCCCACTGAGAGACACCTCTGTTCAGT CTGGATTGCGCTTTTACAGCTGCGCAATTTACTGCGAAGACATTCACGAAGGACACTACTGCTATGCTGTTGCTTTCCTTGAACGACTCATACGTG GTAAACCACAAGAGGATCGTGCCTGGATGCTTTGAGATCTCATACTCGAGACTTCGCGCCATTTTCTAGATAAACAGCCAATTTTATGGTGTGCG TCGACGCTGAATTTGTCATGCACTGCTGTTGGAAGACCAACCAACCTTGAAGTTTCTACTACCGCGAGTAATTATCTGCATTCAGTGAAC TATGTCGGGTTTATCTGAGCTCTACGCTCTTGTCTATATTTCGAAGTACGAGACCTCTAGAATCTTTATACAAGCTTCTCCACCTTAATGGTGT ATCGTCAATGTCAATGACATTTCTCTATTTTATAAAGAAGAACTATACGGAGAAACTGTCAACCAATATCCCTGTTGGCGCTCTGCCGTGACCGGAG CAAAAGCCACGCATTTCTCTTATCCACGAAACAGTTGAGGCTCATGAAAAAATTATACGCATTTTGGAAACAGACAGGGAAGCATTGGATGCTT ATAAAGAGTTTGTCTGTTGTTATGTAGAGTTCCATGTTAGTCTGGACAGCAGGTATCGCCTGAATGAGCTTGAATTTCAACCGGCTTGTGCCATTCTTT AG
1222295	E2	ATGGCCGACAACGACTGGGCGGAGCGGTACGAGGGGTGCTGAAGAAGAGTCTTCACTCCCTCGATATGGCCACGCGCGTCTTGAGGTAGATGCC ATTTTCATGGGCTATGCCACGATGTTGCGCTGGCCAAAGGCTACATGGCGGACGGCTGGCTGTCTCTGGCCCCATACCTGGATTTCGGTATCGCTATGT GCCAACCCGCTTTTCGCGCATCTGGACAGCATGGACGCGCGTGTGTGGTTGGCATTATTTACCGCCTTCGGCGGCTGCCTGGACGATACATACCTGCTG GTGGAGGACACGGTCACTCTCTGTTGCTCTTCGCTTGACGTCCGCGCAGCGACAGCTACATCCAGACCTCGACCGCTTCGCGCGCATCTCCGCGA

Gene code	ID	Sequence (5'-3')
		ACTGCCGCAGCATATGGACGAGACCCGCTCGGGACTCGGCCTCGTCTCGATATTCAACTGGATCAGCAGTCTCCACATTGATCGGTTTCATGTCCGATA AAGAGATTCCGCGCGCCTCTGGGCGGTTCCGAGCCTGGTACGGCCTCCTGAGCGCGGCCAACGAGCTGTACGCGATGTTTCGCATTTCCTCCCGAGATC CCATTCCACATACTCGCACCTCGGATCCCGGACATGATGGTCGTCATCAGGAACAATAACGATGTGTTCTCGTTTATAAGGAGGAGCGGGCTGGAGA GAAAGCCACCCAAGTGTGTATCCTCGCGGATACAAACAGGGCTGACCAAACTCGATGCCCTGGACACAGTTGCCAGGGTAGCAGTCGATGCCACGAG AGCATATGTGACGCTCTCCGAGAGCACCTGTCTGCACGGACGGCCTGGCTGGCCTTCGCGAGAGGCTACGTAGAGTTCCACTTCGCTGCGAAGCGGT ATCGCCTCGATGAATTGGGCGTGGATCGGGAAGTCATTTGGACTAG
369867	A3	ATGCTTTTATTCACTGCTAGGGACGCAGAGGACTGGGATTGGTCACCAAGAATCATTAAATACAGATAAACTACTTGCCACACAGCGCTCTGACATT AATACTTGTGTAACTTTGGTTGTCTTATTGAAGTGGCCTCGTAAGGACCCAACCTTTGTTATTGGAAATATATCGCCAGGGGATACCGTGTCAAAGA CCAGTTATGAGGATATCGTTAATGGATTCTTATGTGACATTTCCCTCAAGCCGCCAAATTGTCAGTTAGACATGGGTCTAAAGACCAGGGTTGAAAC TCCCTTGAGTCCCACGGAATATCCCGTGGGTTAATAACTCAGATCGGTACATGTATAAATACAGCCACCAAGATTACTTCTGTGCATACCCATTTCGTC AGCCACGAGGTTCAGGAAGCAATAGCTTTATACGCAAGCTACGTAATCGCGATAGACGACTTGACCTCTGACATATTGCCAGACTTGAAGTCTATGT AATTCACTAATTCTGGGCAGGTCTCACAAAGCAGAGCTGTTGTCGCGGATTTACAAGTTCCTTGGTACTCAGCAGCAGTTATTCGGCCAGTTTGGTG GAGATATTCGTTAAGGGATCCATGGAGTTCATTCTGACAGGCTATAGAACAGCAAGCAATAGATTAAACCTGCCACCCGAGGCTAGTGA CTATCTAGTCTTCTCCGTGCAAAAGACTGGGGTTCGCGAGCCCTTCGCAATCTTTTGCTTTCCCGAGGACATGAATCCAGAAGATAGAGACTTGGCCG GTTACATTGCTGCAATTCATCCATAATGCTTATTCTAGGTTACGTAACGATTTACTATCCTTTTATAAAGAGGAGCACAAGGCAGGCGACTGTCCTG GTTTCATACAAAGTCACGCTAAAGTGCACGGTTTAAACAAGCTGCACTTTGACAGCAGTTAAGGATCGAGACTAAGAAAGGTTAGAAAGGTCAG GTCGATCTTGTCAAACGATGCTGTCTTGGCTGAGCAGATTACCCGTTTATATATGGGTATATATTCTACCACTTATGTTCCGGTAGGTATAAGTTGGA CGAGTTAAACATCCCCGCTGCGGTCCAAGCTAGGCGTCAATTCACGAGATGCCAACAGTCAGAAGGTA
1249602	B3	ATGAAGGTCCTTGCTAAACACTACCCCGGACCCACTTGGTGGATTGCCCATTAACCCGGAATTGCAAACACAGGATGTTGCTGTCAAAGAAATAGTAC ATAAGTTCTTAACCGAGTGCAGGACACAGTATTTACCGTGCCGTTATTCCTATGATTGTCATGCATATGCATACAGGACGCTATCCGTAGAGGTTGCG GTTATGGACGGAGAATTTGGAATGCAGCACTGCATCCCAAGTGGTGTGTTATGCGCATGACTGCATTCGCCCATTTAGATTAACCAAGCAACTAGGAT CTTAATTGCGCTTTATACAGCTTTCTTAATCTTCTGGATGACTCTTTCCCGGAAGCAATTGAATTAGTTGCGTCAATTTAATTGAGATTCATAAGGAA AGAAGTTCAGGGCCACGCCTCTCTGGATGCTTTCCGACAGATATTACGTGAGTTACCGCACCACTTCGGATCAGTTAGTGCTAATATTATGGTTTCCTC TTCGTTAAACTTCGTTACAGCACTAAGTCTAGAGCATGAAGTTCATCTATGACCATCAGCCCTTCGCTGGATCATTTTCAACATTTCAGTAGGGTGAT GTCAGGTGCTAGCGAGTGCTATGCGTTATTCACTTTCCACCGGAAGTCCCACTAAATAGCTTTGTTTCAGGCTCTTCTGACATGATGACCTATATAAAA CAACGGAATGATGTGCTTTCTGTTCTACGGGGAAGAATTGCTGGCGAGACCTCAAACAGGGTCTCCATCTTGGCCACTTGGCAAGAGGCGGAGCAAA CATGACATATTGTTAAGATTGGTTGCTGATGCTAGCGCATGTCAACACAGAGTTCTAGGAATTTCTAGCGCTGCACCCCGCTTCCTTAGAGGCGTACGG TAATTTTGTAGAGGTTACATAGGTTCCACACATCCACTCCGAGATACAAATTGAACGATTATTTTGTCTAA
7937	C3	ATGGCCCTTTTCCCTCTCAATTAGTCAACACAGCTAACATATCGAGCTACTTCTCTTTTCCCGTCAGGAGTTGCGAACTAGACGCACTTCCAAC ATTACGAGGCCTTAAACGACACAATCTATGTTGCACTGCGCCAGGTAGCAAAGAACGTAAGAAGGCCGAGTACAGACACACCAACCCAGCGGGC AACTTATTCGGTTTGTCGTTCCGATTGTGTGAAGCAGACCGTCTGGGCTATGTGGGTAAATTAATCGAATTCCTATGCATAGTGACGACGTAATGGA AGATTTACCGTTCCGGCAAGCCTGCATTGAGCAGCAATCTTACGTACGGCATTGTACGAGTCTCTATGATGACGATCAATATGGCGGAAGGGCTGTGC GAAAGATGAAATCATCTTGAGAGAGTTGAGGATGGAGTTAGCATCTTTAAATACCCCTTGACACCCGTGTTACTTAAACCTTTGGACACGCTTTTG AGGGACAGGACTCTGACGACGACAGAAATCCAACTTTAGCAGAGTACATTCCTTATCGTAAGACCAACTTCGATTACGACTTCGTATGTCACTTGGT CAGATGGGCGATGAACATTTCCGCCGAAAGTTCAAGAGAAGTTCATTGTAAGAAGTTACGAACACACCGTGGTGTGATAGTGGGCTTGACCAACGAC TACTTCTCCTGGGAAATGGAGCGTCAACAGCCTACGGACCGTATCAGGAATGCACTACCAAGTTCGATGAAAGACGATAGTATATCCGAGGTTCAAG CTAAGTCTATGCTGAAAGAGATCATAATTGGAGAAGAGTCAAAAGCTCGTGAGCTGAAGTCAGAAGTCATGAACCAGACTGAGGAAATCTTCGGAGT TCAGTTAGTCCCGACACGACAGTATGTAAGAAGAACATTTAG
335196	D3	ATGATGTGCGCAGGGATTCCCGACCGATTACTATTTCAACTGCGTCGCTCGTTTACTAGACACTATCCAGTACAACGATACCAACTACACCCCGAGGA AAGACTAGCTAAGCTACATTACGTATATACTAAGACAGCAAAGCACTTCGCGCGTCACGATCGTCAGAAACATATCAAGGTGAGTCCCTCAAACCTT CAACCGCTACTTCAAGCCATGGTTACGATGGTTGTTTACTCCTGGGCTACTGCTGCTGATGATGGTAATGGTTGACCTGAGTATACACTACACATATAT GTTGATTTTGGCAGATTCTCAGCAGGAGGACCTACCGAGTCAATGAAATCCTTGTATAATACAATGTTAGCTAGTTTACCGCAGGGAACACCCCTTGGT GGCAGATGGTAAATTCGCATTTCCCCAGGTCTTAAGGCATTACGGCCCTACTGCGGTTTAAACATAGTACGTAGTACAACAGATTTCTTCGAAGGG TGTTGGGTTGAGCAACATAATTTCAATGGGACCGCTCGTTCTCTGAACAGTCCCATCCCGCTGCGTCAGATGAACGCTTTGGGTCAATGCATCGGTGG ATCTTCTTCCCAAGGAGCGTTTCGACGAGAATGAATTTTCGCGGAGATAACCGGTGCCATCAGACAGCTGCCATGACAGCAAGGATAGCAATTTGTAA GATCTATTGTCTTTTATAGGGAGTTCGGTAAGCCTAGAGACCAACCTCTCTTGTTAATAACTACGCAGCTTGCAACGAGATAACCCCTGAGGAAGC CCTTAAGAACTGACCAATGACGCTATCGTCGGTCTGAGCACTTTTGGGTGCTCTTATGGGACAAGGATCTTTCAATTTTGCATACGTTAAGGACCTT TTGTCAAGGCTATGTCACTTGGCATTGTGCGACCCTAGGTACAGGCTACAGGAAGTTACAGCTTAACTAAAGAGTCCGCTGAGGCTCTCTCAAAGT TCCATAGGTATCTTCAAAGTGCTAAGGAAGCGGGCTCCGTGAATCCAAAGTCATGGGCATATCCAGTGTAGCATCTCTGGCCGACAGCAGCTGGCT TATTGGAGTGATGGCTATAA
85296	E3	ATGGGTGCCCCGAGTTAGAAGTATTCAAGCTACCAGACACCCTGTGTAGATGGGCTTGGAGCAGGAAGCTTAACCCGAATTATAAGAAAGTGAAG AGGAAAGCAAAGCTTGGTTAGAGAGCTTCCACGCGTTTCCACAAAAGCTCAAAACGCTTTCAATAGATGCGACTTTTCCGAACTGCCACCGATAGCT GTTGTGACACACCTAAGGACAGGTTGTGACCTAATGAACCTGTTCTTCGTGTTCGACGAGTACACTGATGTGTGCGACGCGTTGAGGCACGTAAGTT AGCTGACGTAGTGATGGACGCTTTGAGAAACCCGGACCTGCCAAGACCGCAAAACGAGTCCATAGTTGGTGAAATCACTAGACAGTTTGGATGTTG GCTCGTAAAGACAAGCGTCAAGTCAGCTCAAGAGGTTTTCATCGAGACACTTCGATACCTATACCTACCACTGTAGTGAGCAGGCAGGATCGGTGTA CTCATCACATTAGAGACGTAGAGTCTTACTTAAAGTGGCGTGTGATACATTGTGAGGATAAACCTTCTTCGCTGTTACACTTGGGCTCTGACTTGC CAGAGGGTTTCTTTTAAACCCATTGGTTGAGTGCATTTCAACCTTAACGATTGACTTGTTAATTATTGGCAATGACATTTGTTTCATATAACGTGGAGC AGGCTAGGGGGTGATGACGCACATAATATGGTTACTATAGTGATGAATTCATGCTGTTGAACCTGGACGAAGCACTATTATGGATATCTAATTTACAC GACATTTGGTATGGTTTCGTAGAGACATAGACACAACTCCGTGCTGAGTTGCCATGCTAAAGATACAGATGCTGTGCGGTATATCAATGG ATTGGGCAACTGGGTGAGAGCAACGATTGCTGGTCTGAGTGCAGAGAGATATTTCGGTCAGAAAGGAGCGGGTATTCAATCTACCAGGACTGTA GAGCTACTACCTCGTAAGCACACGATGATTACCTCAGACTTCAGGCTAGCTACCGGTGAGTTATGCTCCATGATTTTGGTGAAACCAATCCCGCCGAG CTTGGCAGAGGTACACTAG
3508618	F3	ATGGAAGCCTACACCTTCTGTCTTAGGCTTATTTATTGGAAGTGTCTACCTCTTACGCCATTTCCGGTCTTGGCGACAAGAGGGCACAAACGTGTGCTA GACGGCTTAGAGTAGTTTCTGAAAAGTCTCGGTTACGGAAAAAGGGGACAGGCGAATGCAGGCCTAAAGTCTTTATCCCTCGTACTCTGGCAAGAT GGCCTTGGCCAAAGGCGTATCTCTCAACACTACTCAGTAGTTAAGGAAAGCGCATCGGCTGGCTGCTCATTTAGAGGCTTTTCGCCCCAACGCACGAG GCCTTTGACAGGGGTGCTTTCGGCCATTTAGCATGCATGGCCTACCTAATGCAAAGAAAGTATTCTGTGATTGATGAATACACCGATGTTAGTAATCC TTTTCGAGGTTCAGAAGCAGAAGGATATTATCATGGATGCCTTCGTTAAACCCACACAAGCCAAGGCCGAGTGGGTAGGTTGGTGAGATGACT

Gene code	ID	Sequence (5'-3')
		AGACAGTTTTGGGAGTTAAACAATAAGAAACGCGAGCGAAACAATCACAACTAGATTTATAGCAGGTTTCGATGAATATCTAGAGCGGTGGTTCAAC AGGCTATTGACAGGTCCGGCTCTCACATCAGAGACATCCAGTCATATATCGATATTCGTAGGGACACGATAGGCGCAAAACCCGGCTTCGCCCTAATC GAATTAGGGCTGGAGATTCGGGACGAAGTAATGTCTCACCTGCAATTCAAGAGATGGTAGTGGCTGCAATAGACATGATCTGCATTAGTAACGACA TTACCTCTTATAACATCGAGCAGTCTTGTGGGAATGATAGTCATAACGTTGTCACTATAGTCATGAAAGAATTGGGCACAAACGTCAACGGGGCAATG TTGTGGGCACAGGACTGTACACAGAGTTGGAGCAGAAATTCACGCGCAATGGCAGCTTTACCTGAGCTTGATGAGCCTCTTAACACAGGTAA AGAGTACTGCAACGGGTTGGCCAGTGGGTGAGGGCAATGACGACTGGGAGTTTGAGGGTGGCCGTTATTTCCGGGATAAAGGAAGAGAGATACG TGAGAAAAGGTGGATGTTCTGTCTGCCAAAAGAGAGAAAGAATGGTGCACAAGAGTCGGACCTATTCTGGTTGATGGGTGCTTATTGTGA
348993	G3	ATGCAGGTTGCTTCTAGTAGTTCGCCGCTTCTGTCTATACCAAGTGTCTATGAGACCGTTTCAGGGTTACAAGGTCCTGTATTCTTTTCATGTTACCCG GCGGCATTAGTATCGCTGCTCGTAAGGTGGAAGACGCACTAAGGAAAATTGACGAAGAGGCATCCGAAGAGGGGACCAGGGAGAGAAGACGTGC AGTCATCCGTACACAAAATCCGTACGACGTTACAGAGGAAATGGACCATATCTCAGCCTGCAGAGAACATGACGAGTTGGCCAAGGTTTTGCAAAATA GACATAGACCATCTATATTGAAAGCAGGGAACGTAAGACAGAAAGCGTTGGCAGTAGTTTTGAGGAAGGCGATCGATATCGATCTAAACAGGCC CTGACATGATCGAGATGTTGAGAAAGTACTTTGGTACTTTTGACAACATTGGCGGAGATTTACAGAGGATGGAAGAGTACATGCCCTATAGAATCTC AACTCCGGTACGTGATGAGTTCTTACTTTATTAGATGGGGAATGGACATGACCTTCAGAAAGAGGATTACCGCAAGATTCGAACAGTTTGATATTG CAATGGGTAATGTCTTGGGTCTTACAAACGACTACTTCTCTTGGAAATATAGAGAAGGACCAGGAGACGGATCGTGTAGAAACGGTGTGTGTGTGA ATGAAGGAGTATCATACCACCGCTGACGCAGCGAAAATGATATTGTTGGGTGTGATTGTTGAGCAGGAGTCATTGGCTGCCAACTGAAGAGGAAA GACTGAAGAAGCCGGCTTCCAAGGAAATCTTGCACTTTCGAAGCCATCGAGCTTTACGTTGGTGAAGTTGCTACTGGCATTCGACAGCTCCCAGA TACCAAGTCTTTGAATAA
11954	H3	ATGGGTTTAGGAGTCAATGCACATCACCGTTTGGCGGACCTGGCGGTTAGGTCCTGTAGACACAGGAAACAGGTTACTACCGAGACGAAGAAATGGT TGTTTCAGGGGACAAATCTAGTTGGAAGATACGTGACAGGTATCATGGTTTAAACTGCGGTCTACTGGCAGCATATCTCAGTTGCGTGTGTTGCAAC GACTTCTTGACCTACCTATTCCATCTTGACAACTTATCAGATGACATGGACAATCAGGTAACAATGACTACGGCAGATGTGGTTATGAACACGCTATA CCATCCCGAGAACCATCAGTCCCAAGGATAGCGAAAATGCAAGGACTATTGGAAGAGGCTAGATCCAGGCCAGCCCTGGCGTACAGCAGCGT TTCATGGAACACTAGACTTCTTCTTTCAGTCGGTTACAGAGCAGGCGAAGCATCGTCAAGAGGGAGCCATTCCAGACCTGGAAGCTATATTGCCCT GAGAAGGGACACATCGGGCTGCAAACTTCATTGCTGTGATAGAGTACAGTAACAATTTGGATATACCTGACGAGGTAATGGAACACCCGCTGATA AGATCGCTTGGTGAGGCTGCTAACGACTTAGTTACATGGTCCCAATGACATCTTCAGTTACAATGTTGAACAAAGCAAGGGCGATACACACAACATGA TTCCTGTTATCATGAATCATCACTCGTTAGATCTACAGCCGGCGGTAGACCTGGTCGGGCAGTTGTGTCAGCAGAGTATTGACAGGTTTATGAACGAC AGGGCCAGCTTCCTTCTTGGGGCCAGAAAATTGATAGGCAGGTTGAGATTTATGTGACTGGACTAGAGGACTGGATGGTAGGTTCTTCTACTGGTC CTTCGAGTCAGAAAGATATTTCGCAAGGGCGGTAGGGAAGTAAAGAAGTCAAGGGTTGTAACCTATTGCCAAAGAGAGCTTAA
514883	I3	ATGCCTAGAATAGTGAATACCAATTACCAAAACATAGACTTGTTATTAGACCGTCAATCGAGGCCAAACTATAAGGTCAATCCGTACCATCCGGAGCT AGAAACCGAATTTATTTCTTTCTAGACAAAATCGACCACGCCATTAGTCTTTTCAGAGGAAGGAAATTGAGGGAGCACAGTTACCGCTATGTGCAG CCGCTTACGCACCCGACGCCGACATGCCCCAGCTGCGTGTATACTTAGACTACTTTGTTCTGCTATTACTAGTAGAGGTCATGACCGATGTCTCGAAT GCGCAGGCTGGACAGTTGATCTAGACGACATGATAAATCTGTTAAATACATGCAGCGATAACGAGCCITTCGACACAAGTAAGCAGAAGGATCCCT TAGTTACCTAGCCGCGAAGTTTGCAAGAAAGTGCTTCCCGCGGTTATCCCGGTTGTATAGATATGATGCCACCCAGGTAGATTCTTACCTTGCACAT AGAAGAAAGGTAGTAGCCATTATCCAGTAATCGCTTTTCAACAGGTGGATGTCACAGGTAATAATGATCCGGAGATACCTTCAGACACCCGCTTCG AGGGGTGTTGGAGAGTACTATTGATTTGATCGCCCTACCAATGACCTTATATCTTATAAGAAAGAAAAGTCCCAGGTGAAGATATACATAACCTA GTATCGGTTGCAATGGTAGACCCAGCTATCCCTGTAAAGTCAGGCGACATACAGGCGGCTTTGGACTTTGCAGGAGTTAAGATATCTGAGGCATATCT AAGATTTCAACAGTGTAAATTGGACGCTATAGCCAGACCGCGCTGGAGACTCTGAGACAAGATCAGAGATTATGCGTTACGTAGACATTCTTCTG AAAACGATTGACGGTAACATTGTCTGGCAATCGACCCAGAACTGTTAGGTACCATGTATTTCGAGACGCTGAAGAGAGGGCACAGTGGCTAGTAT CTATCTGCATATGA
9278	J3	ATGCATAGGTTTCAAAGGGCATGTATCTGGAAGAGGCTGGCAAATGAACGTAGATTTGAGGTGAAATTGGCAGTGAATAGACCGATTTTCAGCTCTC ACATTACTCTATCAACAACTGGCGAGACAGTATTCGACTCTAGAAACGTTCCCTGCGAGATACCCGCTCGTCAGTTGGAATCCAAAACACATCCT TTAGGGTGGAAACAAATTCAGGCCAGGTTGAACCTACTTTGGAATAAAGTGAACCTTTGCTAGTGAAGAGGAGAAGAAGGGCTTCTTATGCATTG GTTTTAGCAGAGCTTTCTACATTTGTTCCCTTTAAACACTGGACGACAGGATTAACCTAACTTGCAAAATGTTGTATCTTTCTTTGCTTATCGACGATCA ATTGGAAGAAGATGTCTTTACACAGATGCTAAGCTACAGAGATCGTGTTATGACAATTGCAATTGGGGAGCTGCTTCCCCGACAGATCAATTTGCCTTG AATGGATGCTACACGACACTGTGTTAGCGATGAGAAGTATGGACGAGGTTCTAGCAATGACGTTGCACAAGGCTTTTGTGCTATTATGCAGGCACA GACATCTCAAGAAAGAGTACCATCAAGACACTAGGATCATATCTTAAAGTTTAGGGAGATCGATGTAGGTAGACCTTCTACACTGCCCTAATACGTT TCGGCGCAAGTTATACTTAGCCACGCCGAGTTGAAGAAAAGCGCGGCACTTGAGTCGACCGCTTCAGGCACGTCAGTGTAATGAACGACATATA CTCTTGGGAAAAGAGAATGGAAGGTGTACCAAGCAAAACCTACCGATGGCGCTAGGCCCTTTTCCGCGATATACATACTAGCTAATGAACAGGCTTG CCTTACGCTGCGTGCAAAAGGCTAATGATTCTACTGTAGGGAGCTGGACGCTGGCATTGAAACAGTCTACCCGACGAGATAAGACATAAAGTTCGATGG GCAGTTAACATACGAATTGGAATGTACATCAAGGGCTGGAGTACTTCATGTGCGGTATCGAGCTATGGAGTCAGTGACGCGGAGGTATAGGCA ATAA

Table S5. Primer sequences for colony PCR, sequencing, and Q5U PCR (USER). For some genes, two sets of primers were available (two forward and two reverse). Note that constructs for batch 1 were prepared before this thesis project, hence the primers are not listed.

Gene code	ID	Primers	Primers (USER)
87916	A2	Forward (1) ATGCCTGTAAATATCAGAGACATTG Reverse (1) GTAGCGACTTGGGGAAGGCTAGCA Forward (2) TGTTACTTCACTATTATGAGCGTC Reverse (2) TCAATTACAAAGGTTCCCTTTGACAT	Forward ATCAACGGGUAAAAATGCCTGTAAATATCAGAGACATTG Reverse CGTGCGAUTCAATTACAAAGGTTCCCTTTGACAT

Gene code	ID	Primers	Primers (USER)
52504	B2	<i>Forward (1)</i> ATGGCAACTTGTGACTACACATTTA <i>Reverse (1)</i> TTGAACACAGTACTCCTGGAAGAAG <i>Forward (2)</i> AAGCTGATGTTAGTCCAACATAA <i>Reverse (2)</i> TCACTCCACTTTAGCTTGGTACCTC	<i>Forward</i> ATCAACGGGUAAAAATGGCAACTTGTGACTACACATTTA <i>Reverse</i> CGTGCGAUTCACTCCACTTTAGCTTGGTACCTC
275231	C2	<i>Forward (1)</i> ATGTTAAGAATCGACGTTGGAGAG <i>Reverse (1)</i> CTTAACCGCAGCGTCAGATAAGTAT <i>Forward (2)</i> GTGATGTATGCTTCTTAATGTCCT <i>Reverse (2)</i> CTACTCCAATTTTGCCTGTATCT	<i>Forward</i> ATCAACGGGUAAAAATGTTAAGAATCGACGTTGGAGAG <i>Reverse</i> CGTGCGAUCTACTCCAATTTTGCCTGTATCTT
78169	D2	<i>Forward (1)</i> ATGAGCCCATCAAAGAGTGAGAAAC <i>Reverse (1)</i> TGTGGAAGATACCATGAAGTTTGC <i>Forward (2)</i> TTGGAAACATCTAGGCACTTCCCGC <i>Reverse (2)</i> TTACAGGATACTTAACACGGGCTG	<i>Forward</i> ATCAACGGGUAAAAATGAGCCCATCAAAGAGTGAGAAAC <i>Reverse</i> CGTGCGAUTTACAGGATACTTAACACGGGCTGG
1222295	E2	<i>Forward (1)</i> ATGGCAGATAATGATTGGGCGGAAAG <i>Reverse (1)</i> CTTGGTATTTCTTGTCCGACATA <i>Forward (2)</i> GGTGGTATCTATCTTTAATTGGA <i>Reverse (2)</i> TTAATCAAGGATAACCTCTCTGTC	<i>Forward</i> ATCAACGGGUAAAAATGGCAGATAATGATTGGGCGGAAAG <i>Reverse</i> CGTGCGAUTTAATCAAGGATAACCTCTCTGTCT
369867	A3	<i>Forward</i> TCGGCCAGTTTGGTGAGATATGA <i>Reverse</i> CACTAGCCTCGGGTGGCAGGTTTAAT	<i>Forward</i> ATCAACGGGUAAAAATGCTTTTATTCAGTCTAGGGAC <i>Reverse</i> CGTGCGAUTTACCTTCTGACTGTTGGCATCTC
1249602	B3	<i>Forward</i> ACGTGAGTTACCGCACCACTTCGGAT <i>Reverse</i> ATGGTCATAGATGGAACCTTCATGCTC	<i>Forward</i> ATCAACGGGUAAAAATGAAGGCTTGTCTAAACACTACC <i>Reverse</i> CGTGCGAUTTAGCAAAATAAATCGTTCAATTT
7937	C3	<i>Forward</i> TGTCGGAAAGATGAAATCATACTTGA <i>Reverse</i> CAAAGTTTAAAGTAACAGGGGTGT	<i>Forward</i> ATCAACGGGUAAAAATGGCCCTTTCCCTCTCAATTA <i>Reverse</i> CGTGCGAUCTAAATGTTCTTTTACATACTGT
335196	D3	<i>Forward</i> CCTCGTTCTCTGAACAGTCCCATC <i>Reverse</i> TCGAAACGCTCCTTTGGGAATATAG	<i>Forward</i> ATCAACGGGUAAAAATGATGTCGAGGGATTCCCGACC <i>Reverse</i> CGTGCGAUTTATAGGCCATCACTCCAATAAGC
85296	E3	<i>Forward</i> TATACTACCACTGTAGTGGAGCAGG <i>Reverse</i> AATAGTATCACGACGCACTTTTAAG	<i>Forward</i> ATCAACGGGUAAAAATGGGTGCCCCGAGTTAGAAGTA <i>Reverse</i> CGTGCGAUCTAGTGTACCTCTGCCAAGCTCGG
3508618	F3	<i>Forward</i> CGAGCGAACAATCACAACGTAGATT <i>Reverse</i> TGTGAGAGCCGGACCTGTCAATAGC	<i>Forward</i> ATCAACGGGUAAAAATGGAAGCCTACACCTTCTGTCT <i>Reverse</i> CGTGCGAUTCACAATAACGACCCATCAACCAG
348993	G3	<i>Forward</i> ACAGGCCCTGACATGATCGAGATG <i>Reverse</i> GGCATGTACTCTCCATCCTCGTGA	<i>Forward</i> ATCAACGGGUAAAAATGCAGGTTGCTTCTAGTAGTTCC <i>Reverse</i> CGTGCGAUTTATTCAAAGACTTGGTATCTGGG
11954	H3	<i>Forward</i>	<i>Forward</i>

Gene code	ID	Primers	Primers (USER)
		GAGCCATTCCAGACCTGGAAAGCT <i>Reverse</i> ATTGTTACTGTACTCTATCAACACG	ATCAACGGGUAAAAATGGGTTTAGGAGTCAATGCACAT <i>Reverse</i> CGTGCGAUTTAAGCTCTCTTTGGCAATAGGTT
514883	I3	<i>Forward</i> TGCCACCCAGGTAGATTCTTACCTT <i>Reverse</i> TTACCTGTGACATCCACCTGTTGA	<i>Forward</i> ATCAACGGGUAAAAATGCCTAGAATAGTGCAATACCAA <i>Reverse</i> CGTGCGAUTCATATGCAGATAGATACTAGCCA
9278	J3	<i>Forward</i> GACTGCTTCCCCGACAGATCAATTG <i>Reverse</i> TCATTGCTAGAACCTCGTCCATACT	<i>Forward</i> ATCAACGGGUAAAAATGCATAGGTTTCAAAGGGCATGT <i>Reverse</i> CGTGCGAUTTATTGCCTATACCTCGGCGTCCA

Table S6. The GC-MS program used to produce chromatograms for G1 and O1, before this project work. Results are borrowed for this report.

Step	Rate [°C/min]	Temperature [°C]	Hold time [min]
1	-	60	-
2	10	300	3.5

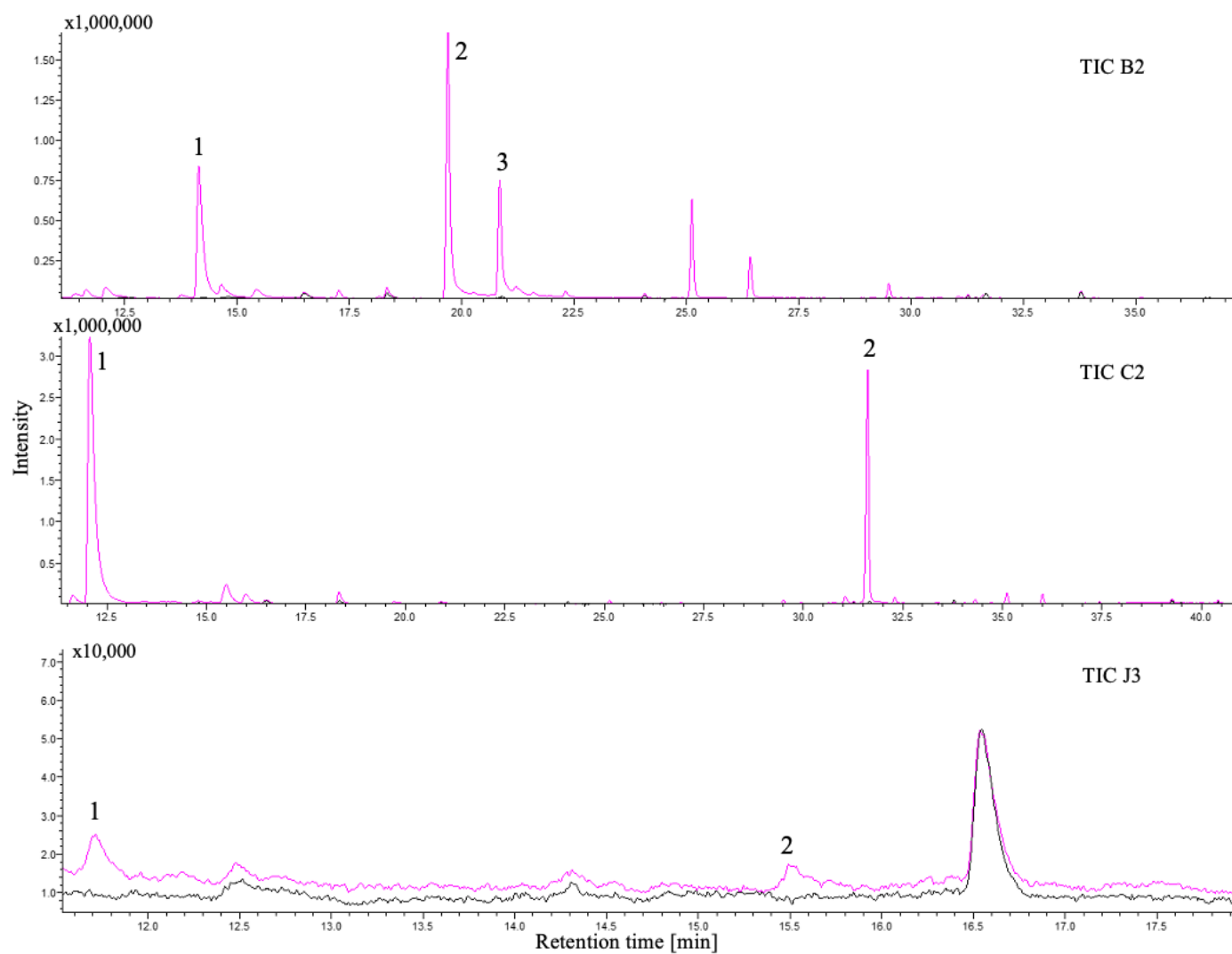


Figure S1. Chromatogram (TIC) for the TPSs B2, C2, and J3 that produced mainly monoterpenes in the in vivo HS-SPME analysis. Program A is used for all three. The peaks are labeled and listed in Section 5.1.1.1.

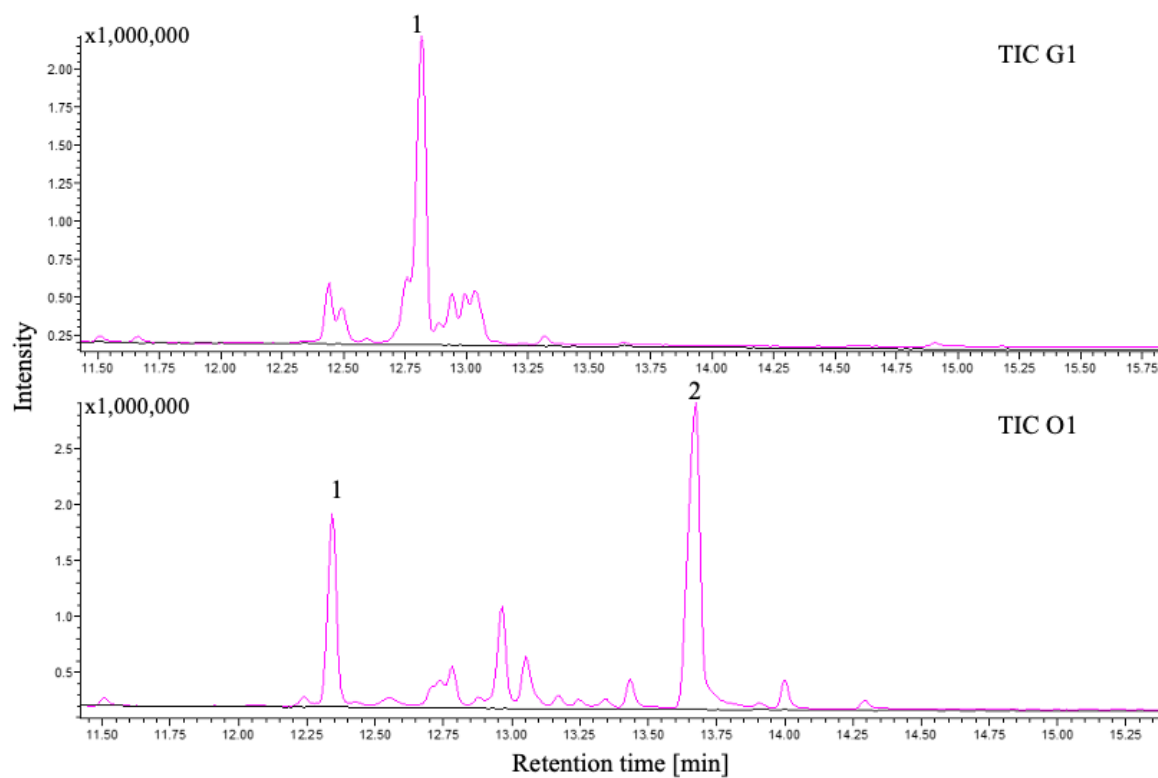


Figure S2. Chromatogram (TIC) for TPSs G1 and O1 that produced mainly sesquiterpenes in the *in vivo* HS-SPME analysis, produced by Chi Zhang. Program in Table S6 is used. The peaks are labeled and listed in Section 5.1.1.2.

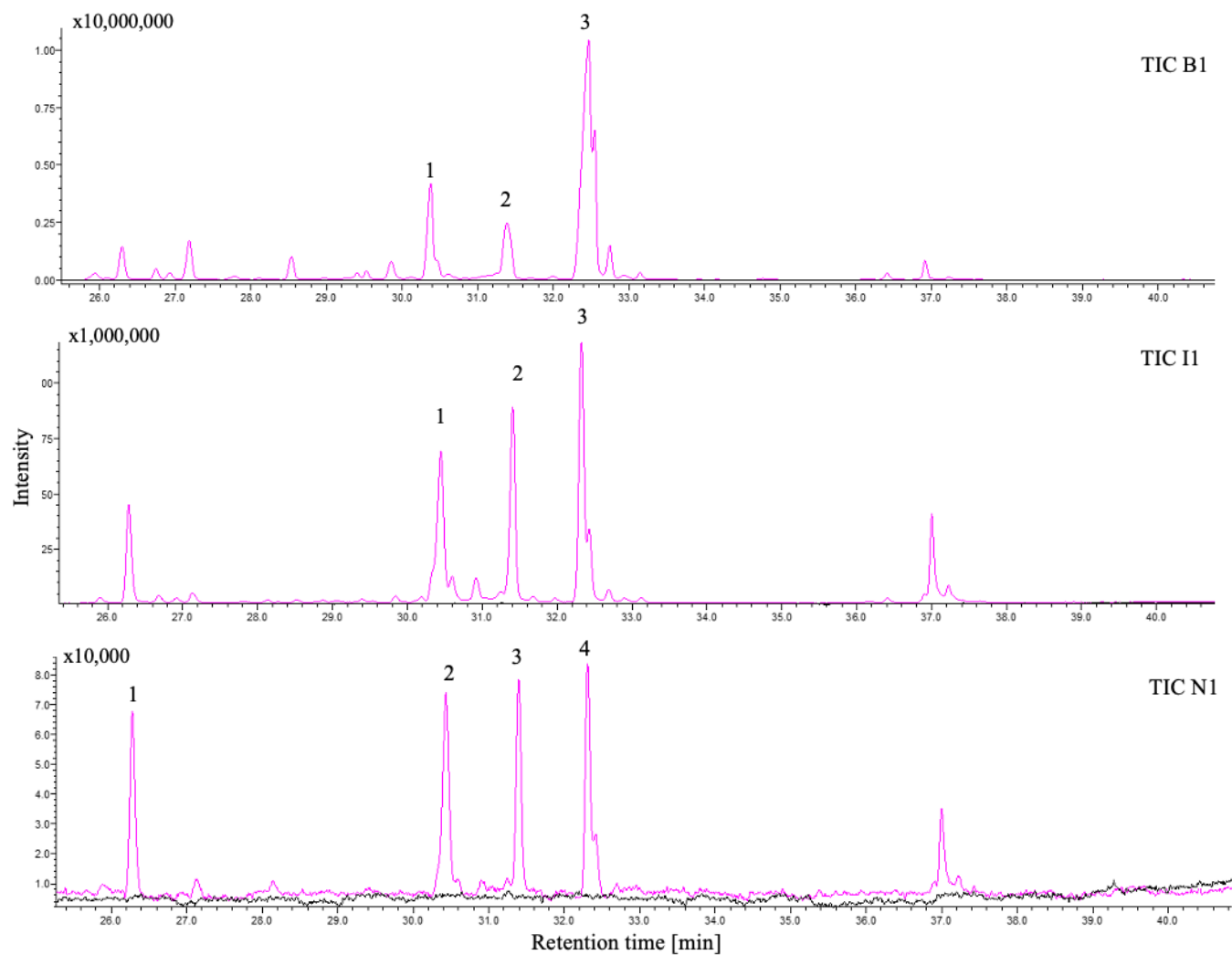


Figure S3. Chromatogram (TIC) for TPSs B1, I1, and N1 that produced mainly sesquiterpenes in the in vivo HS-SPME analysis (1). Program A is used for all three. The peaks are labeled and listed in Section 5.1.1.2.

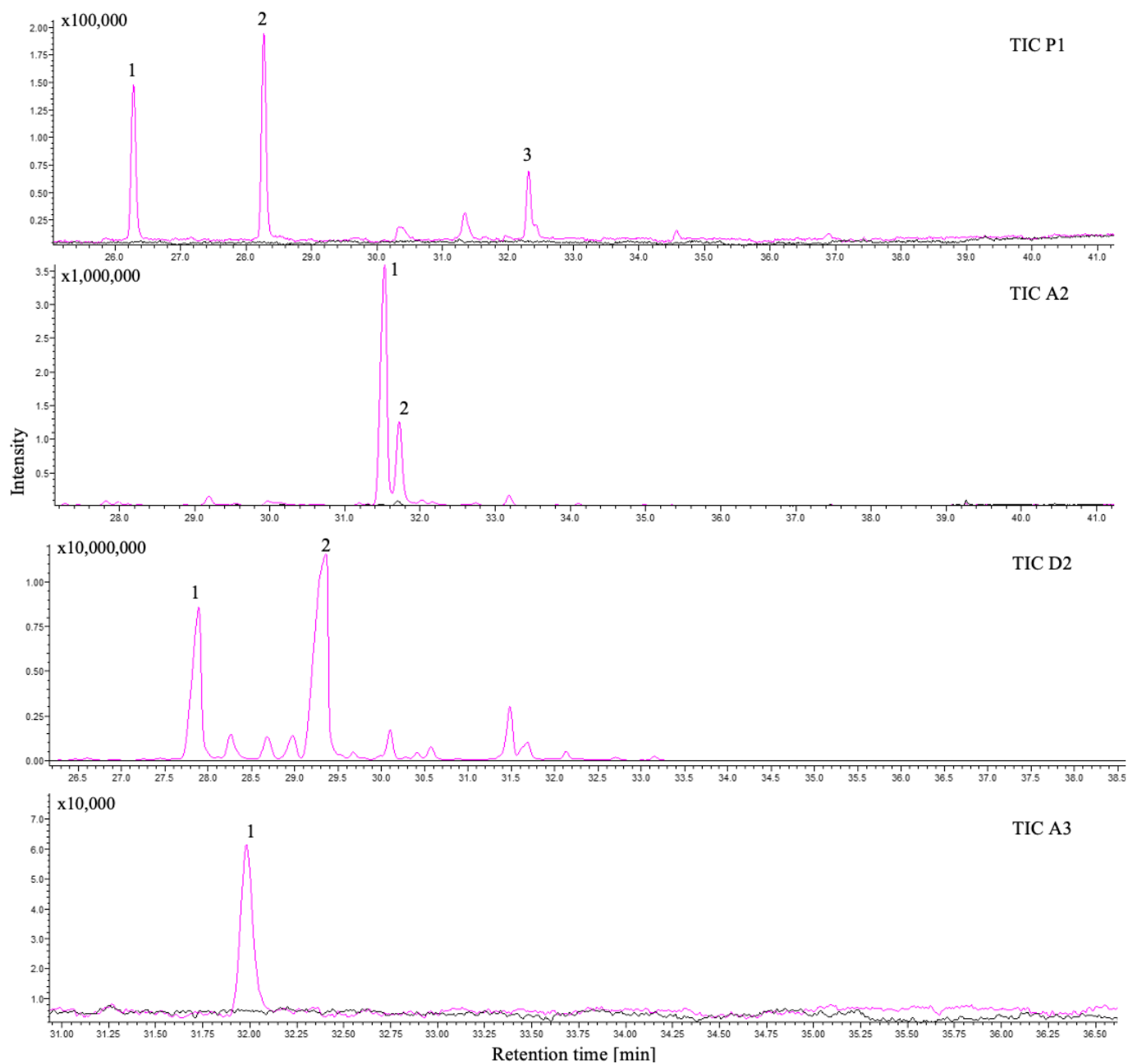


Figure S4. Chromatogram (TIC) for TPSs P1, A2, D2, and A3 that produced mainly sesquiterpenes in the *in vivo* HS-SPME analysis (2). Program A is used for all three. The peaks are labeled and listed in Section 5.1.1.2.

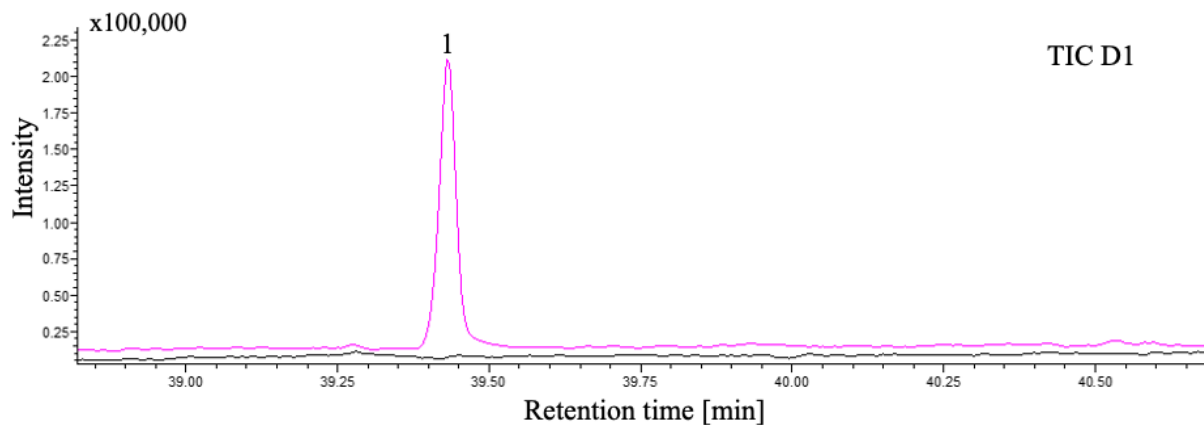


Figure S5. Chromatogram (TIC) for the TPS D1 that produced diterpenes in the *in vivo* HS-SPME analysis. Program A is used. The peaks are labeled and listed in Section 5.1.1.3.

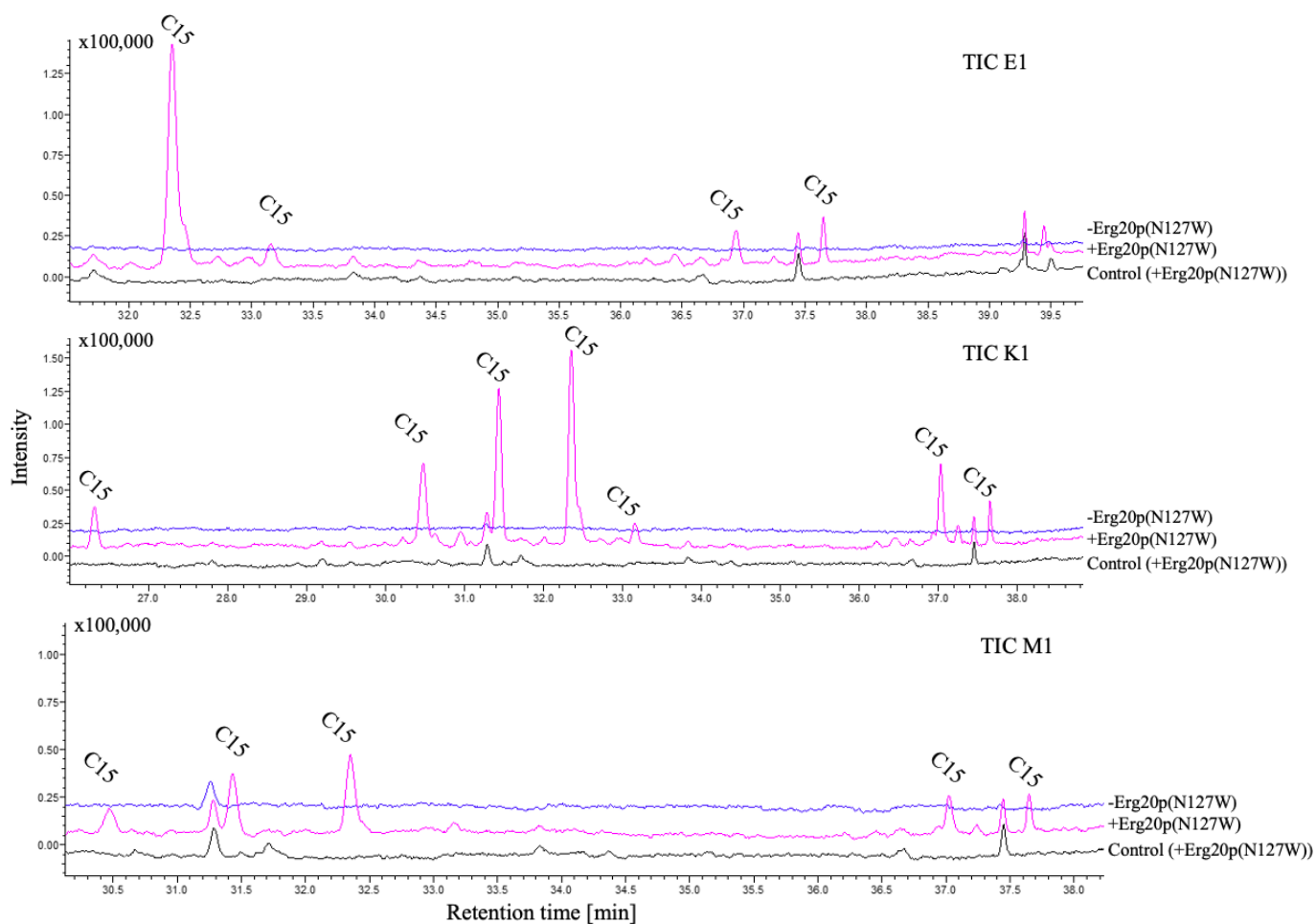


Figure S6. Spectra for E1, K1, and M1, presenting peaks from transformation without Erg20p(N127W) and with Erg20p(N127W). Sesquiterpene peaks (labeled C15) are visible for all three at the co-transformation (+erg20p(N127W)).