**DEPARTMENT FOR INNOVATION IN BIOLOGICAL,**
**AGRO-FOOD AND FOREST SYSTEMS - DIBAF**
***Università degli Studi della Tuscia***
***Via S. Camillo de Lellis ,01100 Viterbo, Italy***

# MSc THESIS

**MSc COURSE: "FORESTRY AND ENVIRONMENTAL SCIENCES" (CLASSE LM-73)**

*INTERNATIONAL CURRICULUM WITH MULTIPLE DIPLOMAS – SPECIALIZATION: «MEDITERRANEAN FORESTRY AND NATURAL RESOURCES MANAGEMENT - MEDFOR»*

# "ISOLATION AND CHARACTERIZATION OF MONOTERPENE SYNTHASE GENES FROM *PINUS NIGRA* SUBSP. *LARICIO* IN CALABRIA"

SUPERVISOR: Prof. Mario Ciaffi

CANDIDATE: Osadolor Samson

Matr. SFA/LM/214

CO-SUPERVISOR: Dr.ssa Enrica Alicandri

Dr.ssa Anna Rita Paolacci

**ACADEMIC YEAR 2017/18**

DOUBLE DIPLOMA ISSUED IN COLLABORATION WITH

MEDfOR
European Erasmus Mundus MSc

Erasmus Mundus

## ABSTRACT

*Pinus nigra* subsp. *laricio* (Poiret) is one of the six subspecies of *Pinus nigra* J.F. Arnold (black pine); it is found in Corsica and in southern Italy with a natural range extending from Calabria to Sicily. In Calabria, where it is considered an endemic species, it grows on the Sila and Aspromonte mountains, and it represents an essential element of the forest landscape which plays an important role not only in soil conservation and watershed protection, but also in the local forest economy.

Conifer trees, including *P. laricio*, produce complex mixtures of mono-, sesqui-, and diterpenoid specialized (i.e. secondary) metabolites that are components of the oleoresin secretions and volatile emissions and can act as a physical and chemical defense against insect and pathogen attack. The oleoresin terpenoids also serve as a large-volume, renewable resource for industrial bioproducts, including solvents, flavors, fragrances, coatings, and resins. These functions of terpenoid continue to increase interest in the molecular and biochemical mechanism regulating their synthesis. Terpenoids are biosynthesized by a large family of catalytically diverse terpene synthases (TPSs) that contribute to a wide array of different compounds that can be produced by a single tree. The many mono-, sesqui-, and diterpene synthases of conifer specialized metabolism form the gymnosperm-specific TPS-d subfamily. In addition, conifer diterpene synthases of general (i.e. primary) gibberellin metabolism belong to the TPS-c and TPS-e/f subfamilies, which also include orthologous genes of angiosperms.

In the framework of the ALForLab (Public-private Laboratory for the Environment-Wood-Forest Chain) project, we have been carrying out the first attempt, to our knowledge, to gain insight into the ecological and functional roles of volatile terpenes emitted from endemic coniferous species in Calabria. Here, we report for the first time about the isolation and characterization of monoterpene synthase (MTPS) cDNA sequences from the needles of *Pinus nigra* subsp. *laricio* populating the forest stands of the river Bonis basin, in the so-called “Greek” Sila.

The strategy adopted was based on the PCR amplification of cDNA sequences by using specific primers designed on conserved regions of MTPSs previously isolated from different *Pinus* species. The putative sequences of the pine genes coding for MTPS were identified by BLAST searches in the National Center for Biotechnology Information (NCBI) database, using as queries selected MTPS from different conifer species. These searches detected 74 putative full-length (FL) MTPS cDNAs from 26 different *Pinus* species. However, a detailed analysis of the retrieved FL cDNAs allowed us to classify as true MTPSs only 32 of the 74 identified sequences. The 42 remaining FL cDNAs isolated from 18 different *Pinus* species encoded the 2-Methyl-3-buten-2-ol (MBO) synthase, which use dimethylallyl diphosphate as a substrate to produce hemiterpenes.

The phylogenetic analysis allowed us to divide the 74 MTPS and MBOS sequences into seven distinct groups, some of which contain functionally related proteins from different pine species. The deduced amino acid and nucleotide sequences belonging to the different groups were aligned, in order to identify highly conserved regions that were used to design specific primers for the isolation by RT-PCR of partial transcripts coding for MTPSs in *P. nigra* subsp. *laricio*. By using such strategy, we were able to isolate and sequence partial MTPS transcripts of putative

*P. laricio* orthologous genes belonging to five out of seven phylogenetic groups. Moreover, four partial *P. laricio* transcripts of groups 1, 2, 5 and 7, were used as a template for isolating FL MTPS cDNAs by 5' and 3' RACE extensions. The four FL cDNAs contained ORFs of 1845, 1857, 1908 and 1890 bp encoding proteins of 614, 618, 635 and 629 aa, respectively. The group 4 partial transcript of 1132 bp in length encodes an incomplete protein of 376 aa. Since putative orthologous genes for the phylogenetic group 3 were not found in the transcriptome of needles of *P. laricio*, we isolated a genomic fragment of 2630 bp that extended from the 5' to the 3' ends of the coding region. The genomic sequence contained ten exons (with the first and the tenth incomplete) and nine introns, consistent with the previously characterized genomic sequences of conifer MTPSs, and hold a partial nucleotide sequence potentially translated to having 1517 bp coding for an incomplete protein of 505 aa.

Phylogenetic analysis of the six isolated sequences with the 74 pine MTPSs and MBOSs identified in the NCBI database placed the *P. laricio* predicted proteins in six of the seven phylogenetic groups, thus confirming the validity of the approach used for their isolation. All the six *P. laricio* predicted proteins contained highly conserved and characteristic regions of plant MTPSs. For instance, each of the four FL predicted proteins included sequences for a putative transit peptide ranging from 40 to 56 aa for import of mature proteins into plastids upstream a conserved $RRX_8W$ domain, which is shown to be essential for catalysis of monoterpene cyclization. Moreover, all the six *P. laricio* predicted proteins had a conserved Asp-rich domain, DDxxD, that coordinates substrate binding via the formation of divalent cation salt bridges.

The study of the MTPS gene family in *P. laricio* and the functional characterization of their members will further help to understand the chemical diversity of terpenoids in this species, as affected by the interactions with its native environment.

Keywords: Terpenoid, Terpene synthase, cDNA, MTPS, *Pinus laricio,* Transcript, Nucleotide, Amino acid.

**ACKNOWLEDGMENT**

Foremost, I would like to express my sincere gratitude to my thesis supervisor, Professor Mario Ciaffi, for accepting me to his research team, also for his enthusiasm, unreserved constructive insight and scientific guidance, consistent and invaluable advice which helped me a lot all through the research period.

My sincere gratitude also goes to Professor Elena Kuzminsky for her follow up, positive criticisms and valuable suggestions in the project. And also for being a mother to we the MEDFOR student. I remain eternally grateful.

My sincere thanks goes to the directors of the MEDFOR programme: The coordinator, Professor José Borges from University of Lisbon; Professor Felipe Bravo from University of Valladolid; Professor Paolo De Angelis who aided my arrival to the University of Tuscia and making the completion of my masters possible with continuous provision of valuable information; The co-coordinator, Dr Catarina Tavares for her excellent service in response to student welfare and administrative concerns.

I am highly grateful to the members of the lab where the research was carried out, Anna Rita Paolacci and most especially Enrica Alicandri who made the lab a fun place to be with her positive vibes, and also for her patience in getting me up to speed with the methods and techniques required to be effective in the project. I say thank you.

I extend my gratitude to the Erasmus incoming officer, Vanessa Torri, for always being available when I needed her despite my worrisome requests from time to time, I remain very grateful.

Last but not least, I would like to thank my family: my parents for giving birth to me at the first place and supporting me spiritually throughout my life, and the one above all of us, the omnipresent God, for answering my prayers and giving me the strength to hurdle all obstacles in the completion of the master's programme. Thank you so much Dear Lord.

# TABLE OF CONTENTS

# LIST OF TABLES

# LIST OF FIGURES

**INTRODUCTION**



**MATERIALS AND METHODS**



**RESULTS AND DISCUSSION**

# 1. INTRODUCTION

## 1.1 Terpenoids: natural functions and industrial uses

Terpenoids, also known as isoprenoids or terpenes, are the largest and most diverse class of natural products (Bharat Singh and Sharma, 2015; Tholl, 2015; Jia, 2016; Abbas *et al.*, 2017) produced by a variety of plant, insects, fungi and bacteria. The term "terpene" is derived from the word "turpentine", which is a fluid distilled from resin produced mainly by pine trees (Jia, 2016). Terpenoids which encompasses more than 40,000 individual compounds (John, 2004; Muhleman *et al*., 2014) are all biosynthesized from two fundamental C5 isoprene units: Isopentenyl diphosphate (IPP) and it's isomer, dimethylallyl diphosphate (DMAPP). Based on the number of isoprene units they contain, terpenoids can be classified as hemiterpenoids(C5), monoterpenoids (C10), sesquiterpenoids (C15), diterpenoids (C20), triterpenoid (C30), tetraterpenoid (C40) or polyterpenoids (C5n) (Tholl and Lee, 2011). These compounds are dominant in plants, as found in lemons (limonene), eucalyptus (eucalyptol), pine (pinene), mint (menthol), geranium (geraniol) or roses (ß-damascenone).

Terpenoids, which have been extensively studied in plants are known to have a wide range of biological functions, as they act as primary metabolites, essentially for growth and development. However, the vast majority of terpenoids found in plants are known to be secondary metabolites; these are compounds whose synthesis have evolved as a result of selection for increased fitness through better adaptation to the environment of each species (Chen *et al.*, 2011).

As primary metabolites, terpenes form key components of membrane structures (sterols, C30), photosynthetic pigments (carotenoids, C40, phytol, C20, and plastoquinones, C45), plant hormones (abscisic acid, C15, gibberellins, C20 and cytokinins, C5), and ubiquinones which are involved in mitochondrial electron transport (Zhou, 2011; Jia, 2016; Abbas *et al.*, 2017). For example, strigolactones, a carotenoid derived compound, act as internal signals in plants, they function as growth and developmental hormones by suppressing shoot branching (Li and Yeh, 2002). Other processes involving their signaling functions include root growth and development, stem elongation, secondary growth, leaf expansion and senescence, and responses to drought and salinity (Brewerm *et al.*, 2013; Waldie *et al.*, 2014).

Terpenoids secondary metabolites have been recognized for their range of specialized roles in plant/environment and plant/plant interactions. These specialized compounds with low-molecular-weight which are volatile, semivolatile or nonvolatile at ambient temperatures, are very important in the protection of plants against abiotic stresses and in various biotic interactions above- and below-ground. Plant produces an array of terpenoids that attract other organisms which, while feeding on the plants, provide benefits to the plants (e.g., pollinators, nitrogen-fixing bacteria, and mycorrhizal fungi) (Hartmann, 2007). Many terpenes have also been demonstrated to be toxins, growth inhibitors or deterrents to microorganisms and herbivores. The emissions of terpenoids such as isoprene and monoterpenes from various vascular and nonvascular plants have been associated with the protection against thermal stress. This protective function is due to the temporary storage of these compounds in voids of photosynthetic membranes (Velikova *et al.*, 2014). The combination of the aforementioned functions of terpenes has led to the widely held belief of their ecological roles in antagonistic or mutualistic interactions among organisms (Gershenzon and Dudareva, 2007).

Not every organism in contact with plants are enemies. Many are partners involved in mutually beneficial interactions such as with pollinators and seed dispersal agents (animals, insects, mammals, birds and bats). Terpenoids which are the major components of fruit and floral volatiles are involved in mutualistic interactions with plant pollinators and seed dispersal agents. Volatile terpenoids released from flowers and fruits serve as an announcement to attract pollinators and dispersal agents. Terpenes play a critical role in mediating these interactions by serving as a medium of communication. Many studies have proven the role of plant terpenoids in communication between plants and pollinators (Baldwin *et al.*, 2006). An interesting example is the floral scent, composed mainly of terpenoid compounds, which serves as an important source of communication between pollinators and flowering plants for their evolution, particularly during long-range communication (Dudareva and Pichersky, 2000; Farré-Armengol *et al.*, 2013). The information sent by the floral volatiles causes a specific behavioral response base on the prevailing context and composition of the emission sent to the respective pollinator. Long distance floral scent emission mostly contributes to guiding pollinators to flowers, especially for night-emitting plants, for which the production of scent intensity is high to prevail over the low conspicuousness of flowers under low illumination. Potential pollinators quickly identify and locate the scented flowers, thus, promoting

the association between pollinators and plants through individual ratios of general compounds or special compounds (Wright and Schiestl, 2009). Monoterpenes and sesquiterpenes, which are the major components of floral volatiles, are good carriers of information over distances because they are low-molecular-weight, lipophilic molecules with high vapor pressures at ordinary temperatures (Tholl, 2015; Abbas *et al.*, 2017).

Herbivores feeding on foliage also induces the emission of blends of volatiles in which terpenes are major components (Dicke *et al.*, 1990). In relation to floral and fruit volatile terpenoids, that from foliage usually is a call for help, attracting predators and parasitoids to attack herbivores (Kessler and Baldwin, 2001). Plant uses terpenes in this case as an indirect defense strategy for the protection of photosynthetic tissues against pathogens and herbivores. Previous studies have also reported some direct defense response such as the repellent function of monoterpene volatiles that are emitted by leaves of *Chrysanthemum morifolium* and also the emission of hemiterpenes with some herbivore-deterrent functions (Laothawornkitkul *et al.*, 2008; Wang *et al.*, 2008).

The diverse functions of floral volatile terpenoids, which includes their role in the defense of floral tissues against microbial pathogens, have been demonstrated in flowers of Arabidopsis mutants, which lack the emission of (E)-β-caryophyllene from their stigmatic tissue. The mutant flowers were more susceptible to infection by *P. syringae*, which resulted in lighter and often misshaped seeds suggesting reduced plant fitness.

In relation to the aforementioned aboveground functions of terpenoids, they also serve similar functions belowground in plants. This includes mediating interactions such as allelopathic activity on the germination, growth and development of competitive nearby plants (Abbas *et al.*, 2017). Terpenoid volatiles emitted from roots possess antimicrobial activity, hence protecting the plants against pathogen attack (Li and Yeh, 2002). Terpenoids in root zone such as rhizathalenes (semivolatile diterpene hydrocarbons) function as local antifeedants by reducing herbivore damage on root cell layers. There are also beneficial interactions mediated by terpenes below ground, as demonstrated for strigolactones, a carotenoid derived compound that play an important role in mediating recruitment of arbuscular mycorrhizal fungi in the rhizosphere, in form of exogenous signals (Akiyama *et al.*, 2005).

In addition to their importance in plant physiology and ecology, terpenoids are significant because of their numerous applications in the pharmaceutical, food and cosmetic industries (Abbas *et al.*, 2017). The diverse collection of terpenoid structures and functions continues to increase interest in their commercial use resulting in some with established medical applications being registered as drugs in the market (Ludwiczuk *et al.*, 2017), mainly because several terpenoids are biologically active and have proven to be effective in the fight against cancer, malaria, inflammation, and a variety of infectious diseases (Mbaveng *et al.*, 2014). The antimicrobial activity of terpenoids is particularly important because of the severe increase in bacterial resistance to antibiotics, which is a growing cause of concern globally (Islam *et al.*, 2003; Singh and Sharma, 2015). Furthermore, many terpenoids are commercially important due to their wide application in the industrial sector such as fibers, polymers and rubbers. The most extensively used terpene by human beings is rubber, a polyterpene composed of repeated isoprene subunits (Abbas *et al.*, 2017).

The food industries exploit terpenoids effectiveness and potential as a flavour enhancer. In a bid to boost crop production in the agricultural industry, plant breeders are currently using genetic engineering alongside biotechnology to enhance the defensive role of terpenes in crops that show less resistance toward pathogen and herbivore attacks. In livestock, the addition of terpenes can replace conventional antibiotics, as they can slow down the resistance of bacteria against antibiotics (Abbas *et al.*, 2017).

### 1.2 Synthesis of terpenoids in plants

The first step of terpenoid biosynthesis is the generation of the two inter-convertible C5 units isopentenyl diphosphate (IPP) and dimethylallyl diphosphate (DMAPP) (Abbas *et al.*, 2017). They are derived from two independent pathways that are localized in different cellular compartments: the mevalonic acid (MVA) pathway which operates mainly in the cytosol, but also in the endoplasmic reticulum and peroxisomes and the methylerythritol phosphate (MEP or Non-mevalonate) pathway, localized in the plastids (Fig. 1.1 in Tholl, 2015). The MVA pathway predominantly provides the precursors for the cytosolic biosynthesis of sesquiterpenoids, polyprenols, phytosterols, brassinosteroids, and triterpenoids,

and for terpenoid biosynthesis in mitochondria (e.g., ubiquinones, polyprenols), while the five-carbon units derived from the MEP pathway are preferably used for the biosynthesis of hemiterpenoids (e.g., isoprene), monoterpenoids, diterpenoids, carotenoids and their breakdown products, cytokinins, gibberellins, chlorophyll, tocopherols, and plastoquinones (Hemmerlin *et al.*, 2003; Leivar *et al.*, 2005; Dudareva *et al.*, 2006; Carrie *et al.*, 2007; Merret *et al.*, 2007; Simkin *et al.*, 2011; Markus Lange and Ahkami, 2013). This compartmentally separated biosynthetic pathways are known to have some metabolic "crosstalk" (Flügge and Gao, 2005; Orlova *et al.*, 2009) but very little is known about the exchange mechanism by which they are finely cooperated (Hemmerlin *et al.*, 2003).

It has become evident that both pathways are heavily regulated at multiple levels (Tholl, 2015). In addition to the transcriptional regulation of MVA and MEP pathway genes and their different paralogues, isoprenoid-pathway fluxes are controlled at post-transcriptional/translational levels and by feedback regulation (summarized in Fig. 1.1). Recent studies have given a more global view of the dynamics and networks of the core isoprenoid pathways and the regulation of metabolic flux during plant development and in response to external stimuli (reviewed in Vranová *et al.*, 2012, 2013).

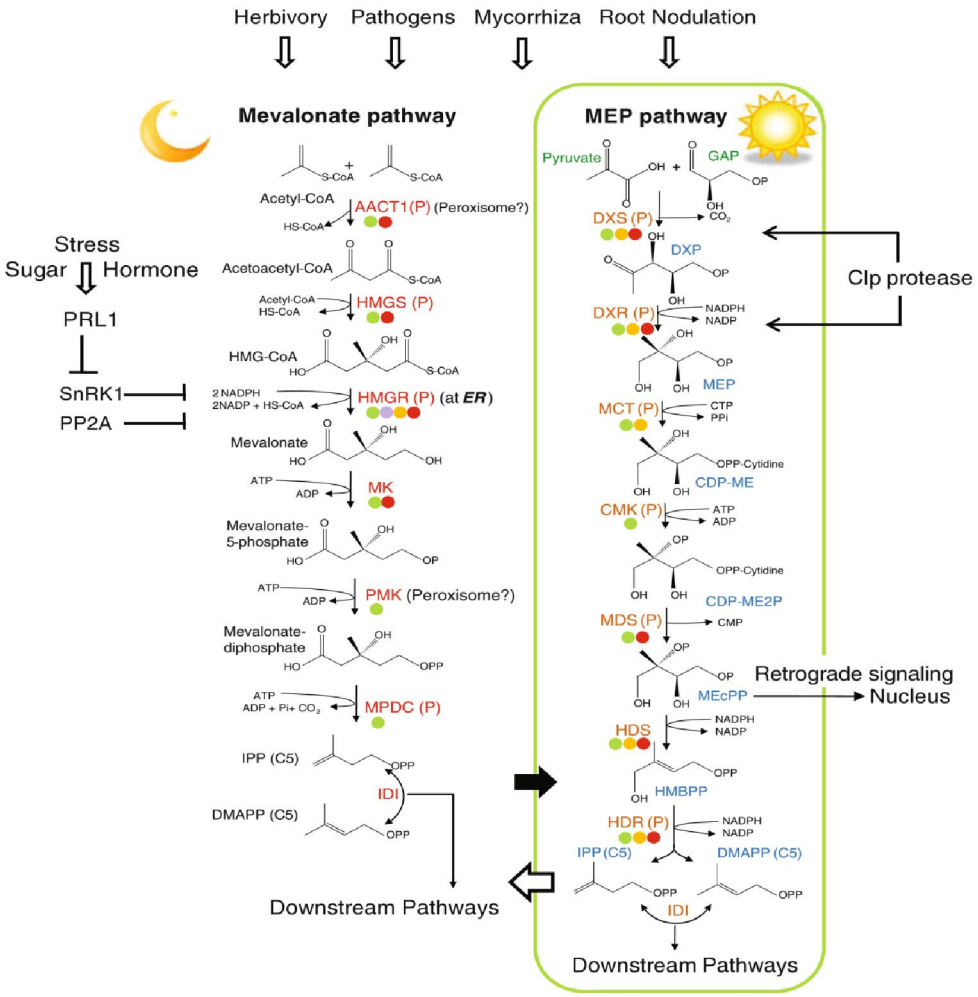



Figure 1.1. Enzymatic steps of the MVA and MEP pathways and their regulation in isoprenoid precursor biosynthesis (Tholl, 2015). Colored dots indicate the different levels of regulation for each enzyme according to the current status of knowledge (Hemmerlin, 2013): green—transcriptional, purple—posttranscriptional, yellow—translational, red—posttranslational including feedback modulation. One or more gene paralogues as described from different plant species (Hemmerlin *et al.*, 2012) are indicated by (P). Arrows indicate preferred trafficking of isoprenoid precursors between the cytosol and plastids in light (white) and dark (black) exposed tissues.

The MVA pathway in plants (Fig. 1.1) consists of six steps that lead to the production of IPP, starting with the condensation of two molecules acetyl-CoA to acetoacetyl-CoA (AcAc-CoA), catalyzed by acetoacetyl-CoA thiolase (AACT). In a subsequent condensation reaction catalyzed by HMG-CoA synthase (HMGS), AcAc-CoA is combined with a third molecule of acetyl-CoA, leading to the synthesis of 3-hydroxy-3-methylglutaryl CoA (HMG-CoA). In the following rate-limiting step, HMG-CoA

reductase (HMGR) catalyzes the conversion of S-HMG-CoA to R-mevalonate in two NADPH-dependent reduction steps. MVA produced by HMGR is finally converted into IPP via three enzymatic steps: two ATP-dependent phosphorylation steps, catalyzed by mevalonate kinase (MK) and phosphomevalonate kinase (PMK), and an ATP-driven decarboxylative elimination catalyzed by mevalonate diphosphate decarboxylase (MVD or MPDC) (Fig. 1.1). IPP can further be converted to its isomer, DMAPP, in a reaction catalyzed by the IPP/DMAPP isomerase (IDI) (Tholl, 2015; Jia, 2016; Abbas *et al.*, 2017).

The MEP pathway consists of seven enzymatic steps (Fig. 1.1). In the first reaction, 1-deoxy-D-xylulose 5-phosphate (DXP) is formed by DXP synthase (DXS) from (hydroxyethyl) thiamine diphosphate, which is derived from pyruvate, and glyceraldehyde-3-phosphate (GAP) in a transketolase-like condensation (Tholl, 2015). Numerous studies have confirmed that DXS functions as an important regulatory and rate-limiting enzyme in the biosynthesis of plastidial terpenes (Lois *et al.*, 2000; Walter *et al.*, 2000; Estévez *et al.*, 2001).

In the second step of the MEP pathway, DXP is converted into 2-C-methyl-D-erythritol 4-phosphate (MEP) by the enzyme DXP-reductoisomerase (DXR), which causes an intramolecular rearrangement of DXP into 2-C-methyl-D-erythrose 4-phosphate, followed by an NADPH-dependent reduction. This reaction catalyzed by DXR is in some cases considered a rate-limiting step depending on the species, tissue, and developmental stage (Tholl, 2015).

In the following step, MEP is converted to 4-Diphosphocytidyl-2-C-methylerythritol (CDP-ME) in a CTP-dependent reaction catalysed by the enzyme 4-diphosphocytidyl-2C-methyl-D-erythritol synthase (MCT or IspD). Phosphorylation of CDP-ME by the enzyme 4-diphosphocytidyl-2-C-methyl-D-erythritol kinase (CMK) then leads to the formation of 4-diphosphocytidyl-2-C-methyl-D-erythritol 2-phosphate (CDP-MEP), which is subsequently cyclized by 2-C-methyl-D-erythritol 2,4-cyclodiphosphate synthase (MDS) into 2-C-methyl-D-erythritol 2,4-cyclodiphosphate (MEcPP) upon loss of CMP. In a subsequent step, the enzyme, 4-hydroxy-3-methylbut-2-en-1-yl diphosphate synthase (HDS) catalyzes the reduction of MEcPP to 4-hydroxy-3-methylbut-2-enyl diphosphate (HMBPP) and finally, through a branching step, HMBPP is converted by 4-hydroxy-3-methylbut-2-enyl diphosphate reductase (HDR) to a mixture of IPP and DMAPP with an approximate ratio of 5 or 6:1 (Tholl, 2015).

In the final major stage of terpenoid biosynthesis, IPP and DMAPP units are fused by the catalytic activity of prenyltransferases (isoprenyl diphosphate synthases) to form prenyl diphosphate as the linear central precursors of all terpenoids. This reaction involves a head-to-tail (1'–4) condensation of IPP with the allylic co-substrate, DMAPP, based on an ionization–condensation–elimination mechanism to produce a C10-allylic diphosphate, the geranyl diphosphate (GPP). Additional rounds of head-to-tail condensation of the allylic product with more IPP units lead to the formation of short-chain (C15–C20) such as farnesyl diphosphate (FPP, C15) and geranylgeranyl diphosphate (GGPP, C20). Longer prenyl diphosphate are also synthesized and some are ultimately added (with the loss of the diphosphate moiety) as components to form such compounds as chlorophylls, phylloquinones, and ubiquinones.

GPP, FPP and GGPP are the major substrates in reactions catalyzed by various type of terpene synthases (TPSs) to produce monoterpenes, sesquiterpenes and diterpenes, respectively (Fig. 1.2). These terpenoid compounds can be further modified through secondary enzymatic reactions such as hydroxylation, methylation, glycosylation, acylation, peroxidation or cleavage to produce biologically active end products of even larger structural diversity (Jia, 2016).

Some terpenoids produced by the activity of prenyltransferases and TPSs are found in all plants and therefore belong to 'primary,' or general, metabolism. For example, head-to-head condensation of two FPP molecules leads to the production of squalene (with the loss of both diphosphate groups), the precursor of sterols. A similar condensation of two GGPP molecules gives phytoene, the precursor of carotenoids. GGPP is also a precursor of gibberellins; a pair of structurally related TPS enzymes in angiosperms and gymnosperms, copalyl diphosphate synthase (CPS) and kaurene synthase (KS), convert GGPP first to copalyl diphosphate (CPP), then to ent-kaurene, the precursor of all plant gibberellins (Fig. 1.2).

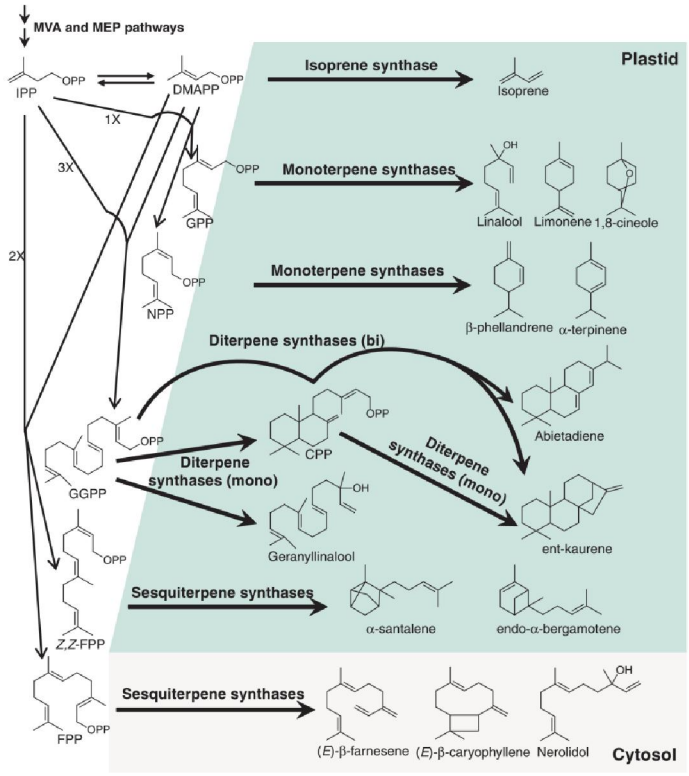



Figure 1.2. An outline for the formation of terpenoids catalyzed by various types of terpene synthases (TPSs) (Chen *et al.*, 2011). Isopentenyl diphosphate (IPP) is the common precursor of all terpenes. It is synthesized by both the cytosol-localized mevalonic acid (MVA) pathway and the MEP pathway in plastids. IPP is isomerized to give dimethylallyl pyrophosphate (DMAPP). DMAPP either serves as the substrate for hemiterpene biosynthesis or fuses with one IPP unit to form geranyl diphosphate (GPP). The condensation of one GPP molecule with one IPP molecule gives farnesyl diphosphate (FPP), and the condensation of one FPP molecule with one IPP molecule will give geranylgeranyl diphosphate (GGPP). GPP, FPP and GGPP are the precursor for monoterpenes, sesquiterpenes and diterpenes, respectively. Various types of TPSs such as isoprene synthase, monoterpene-, sesquiterpene-, and diterpene-synthases convert DMAPP, GPP, FPP, and GGPP to isoprene, monoterpenes, sesquiterpenes and diterpenes, respectively. In general, the biosynthesis of isoprene, monoterpenes, and diterpenes occurs in the plastid and the biosynthesis of sesquiterpenes occurs in the cytosol.

## 1.3 Plant terpene synthases

As described above, all terpenoids are produced from allylic diphosphates by the action of terpene synthases (TPSs), which constitute a superfamily and contribute to the tremendous diversity of terpenoid carbon skeletons. TPS enzymes facilitate adaptations of terpene metabolism to the changing environment because their promiscuous activity often results in the production of more than a single compound and TPS proteins easily acquire new catalytic properties by minor structural changes (Tholl, 2015). TPS enzymes have therefore become a focus point in planta and heterologous metabolic engineering of terpenoid end products with use as pharmaceuticals, flavors, biofuels, or plant chemical defenses. TPSs are present in multiple domains of life, but their protein sequences show a relatively low level of conservation even within the same domain (Bohlmann and Steele, 1997). However, the elucidated protein structures of several TPSs (Hyatt *et al.*, 2007; Köksal *et al.*, 2010, 2011; McAndrew *et al.*, 2011; Zhou *et al.*, 2012) revealed a high similarity in their tertiary structures consisting of an N-terminal domain and a catalytically active C-terminal domain, which implies a similarity in their catalytic mechanisms.

### 1.3.1 Structure of plant terpene synthases

Mechanistically, TPS proteins are divided into class I and class II enzymes. This classification is based on the chemical strategies employed for initiating cyclization reactions. The reaction catalyzed by class I TPSs starts with the ionization of the prenyl diphosphate substrate by a divalent cation-dependent subtraction, the produced intermediate then enters different reactions that can include cyclization, hydride shifts and rearrangements prior to a termination of the reaction by proton loss or the addition of a nucleophile such as water (Davis and Croteau, 2000). By contrast, class II TPSs, which include diterpene synthases, catalyzes the ionization of their substrate by adding a proton to an epoxide ring or via protonation at the 14,15-double bond of GGPP, respectively (Tholl, 2015). Class II diterpene synthases that fall into this category are ent-copalyl diphosphate (CPP) synthases (CPSs), which are involved in gibberellin and phytoalexin biosynthesis. In the gibberellin biosynthetic pathway, CPSs catalyze a protonation-induced bicyclization of the substrate GGPP to form ent-CPP, which is further ionized and converted to ent-kaur-16-ene by a class I ent-kaurene synthase (KS) activity.

Class I TPSs has a functional C-terminal domain (also referred to as α-domain or class I fold) that adopts the α-helical protein fold and contains two metal binding motifs, highly conserved DDXXD and less well conserved NSE/DTE motifs located on opposing helices near the entrance of the active site (Jia, 2016). DDxxD is the best known structural motif of the terpene synthase family and it has been found in virtually all isolated plant terpene synthases as well as in isoprenyl diphosphate synthases and microbial terpene synthases. Site-directed mutagenesis, as well as X-ray structural analysis, revealed that this region is involved in binding divalent metal ions which in turn interact with the diphosphate moiety of the substrate (Tarshis *et al.*, 1996; Lesburg *et al.*, 1997; Starks *et al.*, 1997). The location of the DDxxD motif at the entrance of the catalytic site appears to be critical in positioning the substrate for catalysis. In an event of mutations in this region, this could result to a decreased catalytic activity and the appearance of abnormal products which can be attributed to altered substrate binding (Little and Croteau, 2002; Rynkiewicz *et al.*, 2002; Seemann *et al.*, 2002; Prosser *et al.*, 2004). The NSE/DTE motif is an additional metal cofactor binding motif located on the opposite sides of the entrance of the catalytic site. The DDxxD and the NSE/DTE motifs both help to position the diphosphate substrate by binding of a trinuclear magnesium cluster to trigger the ionization of the isoprenoid substrate and initiate the cyclization reaction. In some sesquiterpene synthases, the NSE/DTE motif is replaced by a second DDxxD motif (Steele *et al.*, 1998) which was also shown to be involved in catalysis (Little and Croteau, 2002). Many TPSs requires potassium for enzyme activity to occur (Green *et al.*, 2007). Recently, the H-α1 loop, which is located in direct proximity to the NSE/DTE motif, was identified as the potassium binding region in these enzymes. It is assumed that potassium ions stabilize this loop region for optimal substrate binding (Green *et al.*, 2009).

In contrast, class II TPSs possesses a functional N-terminal domain (or β-domain), which together with a third "insertion" γ-domain forms the class II fold. Enzymes in this class contain a conserved DXDD motif, which is located in the β-domain and responsible for the protonation-initiated cyclization (Christianson, 2017). The γ fold exhibits a similar topology with the β fold (KÖksal *et al.*, 2011) and it carries a highly acidic EDXXD-like Mg2+/diphosphate binding motif that also contributes to the activity of class II TPSs (Cao *et al.*, 2010).

The available protein crystal structures suggest that most of plant diterpene and some sesquiterpene synthases possess all the three domains γ, β, and α (Trapp and Croteau, 2001; Cao *et al.*, 2010). However, usually only one domain is functional. In angiosperms, all diterpene synthases that have been characterized to date are monofunctional, with loss of activity in one domain or the other. Some of these monofunctional diterpene synthases (class II) catalyze the formation of ent-CPP, syn-CPP, or hydroxy-CPP from GGPP, while others (class I) use one of these intermediates to make a diterpene hydrocarbon (Cho *et al.*, 2004; Otomo *et al.*, 2004; Prisic, 2004; Falara *et al.*, 2010). Moreover, in both gymnosperms and angiosperms, some diterpene synthases have been found that have retained class I activity only (with or without the actual loss of the N-terminal class II domain) and use GGPP to directly produce a diterpene without a CPP intermediate (Herde *et al.*, 2008; Köksal *et al.*, 2011). However, some gymnosperm diterpene synthases, such as abietadiene synthase from grand fir (*Abies grandis*), retain both class I (KS-type) and class II (CPS-type) functional domains and bifunctional properties, catalyzing the formation of an enzyme-bound CPP from GGPP and then converting CPP to a diterpene (Peters *et al.*, 2000).

In contrast to the diterpene synthases, which exist both in the form of bifunctional and monofunctional enzymes, all known mono- and sesquiterpene synthases are believed to be monofunctional, having retained only one active site corresponding to the α-type domain or class I activity. In most cases, the β-domain is present but it is rendered inactive due to the loss of conserved DXDD motif (Whittington *et al.*, 2002)

Terpene synthases are approximately 550–860 amino acids long and 50–100 kDa in their molecular masses (Figure 1.3), differentiated by the combinations of domains and motifs. In general, sesquiterpene synthases are 550-600 amino acids (aa) long and 50–70 aa shorter than monoterpene synthases, which contain the N-terminal plastid-targeting peptides. Diterpene synthases are even longer than monoterpene synthases due to the additional insertion sequences (γ-domain) in their N-terminal. Many terpene synthases also carry a highly conserved RR(x)8W motif downstream of the N-terminal transit peptide, which is essential for catalysis of monoterpene cyclization (Whittington *et al*., 2002; Hyatt *et al*., 2007) and is also conserved with variations in most sesquiterpene- and diterpene synthases.

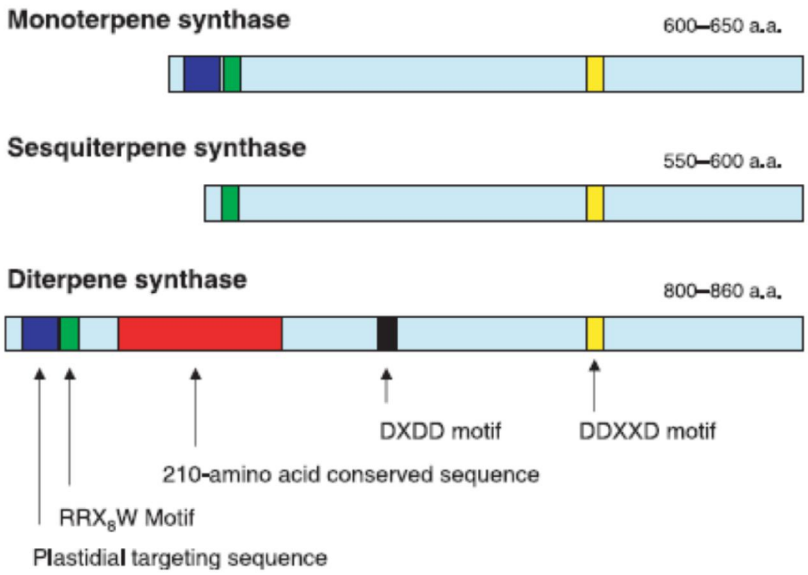



Figure 1.3. Schematic representation of terpene synthase enzymes. a.a., amino acids (Keeling, 2006).

### 1.3.2 Origin and evolution of plant terpene synthase genes

Previous studies proposed that plant terpene synthase gene ancestor, resembling a conifer diterpene synthase associated with gibberellin biosynthesis, emerged prior to the divergence of angiosperms and gymnosperms (Trapp and Croteau, 2001). Later, the bifunctional diterpene synthase (PpCPS/KS) identified in the bryophyte *Physcomitrella patens*, was hypothesized to be the common ancestor (Hayashi *et al.*, 2010; Keeling *et al.*, 2010). Interestingly, in bacterium *Bradyrhizobium japonicum*, the ent-copalyl diphosphate synthase (BjCPS) and ent-kaurene synthase (BjKS) share some similarity with the β-domain (CPS activity) and α-domain (KS activity) of the plant and fungal TPS genes, respectively. This, together with the fact that plant terpene synthase genes are longer and the two types of monoterpene synthases (CPS/KS) are of roughly equal length, suggested a common ancestral diterpene synthase gene shared by plants, fungi, and bacteria (Morrone *et al.*, 2009).

More insight to the origin and evolution of the plant TPS enzymes has been gained from the analysis of an increasing number of crystal structures including those from an isoprene synthase (Köksal *et al.*, 2010), monoterpene synthases (Whittington *et al.*, 2002; Whittington *et al.*, 2002; Hyatt *et al.*, 2007; Kampranis *et al.*, 2007), sesquiterpene synthases (Starks *et al.*, 1997; Gennadios *et al.*, 2009), a class I diterpene synthase, taxadiene synthase (Köksal *et al.*, 2011), a class II CPP synthase

(KÖksal *et al.*, 2011; Köksal *et al.*, 2014), and a class II/I diterpene synthase, abietadiene synthase (Zhou *et al.*, 2012). Comparisons of the assembly of a class I type α-domain and class II type β and γ domains led to the prediction of an evolutionary scenario according to which an ancestral bifunctional classII/classI diterpene synthase (consisting of all three domains with a functional α- and β-domain that contain a transit peptide) similar to the gymnosperm abietadiene synthase (which in turn resembles *P. patens* CPS/KS enzyme) gave rise to class II type diterpene synthases (consisting of all three domains with a functionally active β-domain and an inactive α-domain) and class I type TPSs (consisting of a nonfunctional β-domain and a functionally active α-domain) (Cao *et al.*, 2010; Gao *et al*., 2012). Variations of the three-domain structure in many of the well-characterized enzymes of the plant TPS family can be explained by the loss of a particular domain such as the KS- (β) or CPS (α) -type domain or by the loss of activity associated with a particular domain in different lineages of TPS evolution. These variations also account for the variable length of plant TPSs of approximately 600–900 aa (Chen *et al*., 2011).

The model of common ancestry of all plant TPSs, as reconstructed on the level of protein domain structure by Cao *et al*. (2010) and supported by recent x-ray structure analysis of different terpene synthases (Gao *et al*., 2012), is also supported on the level of genomic TPS sequences with patterns of conserved intron positions across many TPSs from bryophytes, gymnosperms and angiosperms representing all major TPS subfamilies (Trapp and Croteau, 2001; Aubourg *et al.*, 2002; Keeling *et al.*, 2010; Martin *et al.*, 2010), although exceptions from conserved intron-exon structures have been reported (Lee and Chappell, 2008).

Further evolutionary modifications leading to diversification of product profiles have not only been associated with changes in active domain structure, but isoprene and monoterpene synthases have lost the γ-domain, while sesquiterpene synthases, the target peptide and in most cases the γ-domain (Hillwig *et al.*, 2011; KÖksal *et al.*, 2011; Rajabi *et al*., 2013). Recently, the analysis of several proteins with mixed substrate specificity allows for developing novel hypotheses about the timing of major evolutionary modifications, the loss of γ-domain and transit peptide in TPSs with different substrate specificity (Pazouki and Niinemets, 2016) (Figure 1.4). For instance, analysis of the structure of bi-domain, α-β, kaurene like diterpene synthase from *Triticum aestivum* (TaKSL5) that can use both *ent*-copalyl diphosphate to produce *ent*-kaurene and (*E,E*)-FDP to produce (*E*)-nerolidol (Hillwig *et al.*, 2011),

suggests that evolution of sesquiterpene synthesis can occur first by loss of γ-domain followed by changes in subcellular localization by loss of transit peptide and further diversification and loss of capacity for use of C20 substrate. Such a possibility is underscored by the occurrence of multi-substrate (*E*)-nerolidol/(*E,E*)-geranyllinalool synthase in *V. vinifera* (VvPNLNGl1 and VvCSENerGl) that have both C15 and C20 substrate use capacity, but lack both the γ-domain and the transit peptide (Martin *et al.*, 2010). The terpene synthase from *T. aestivum* (TaKSL5) and the two from *V. vinifera* (VvPNLNGl1 and VvCSENerGl) represent three putative intermediates of the evolution of sesquiterpene synthases directly from diterpene synthases by γ-domain loss, that is predicted to precede the loss of the transit peptide. On the other hand, in gymnosperms there are three *Abies grandis* C10/C15 multi-substrate TPSs that lack the transit peptide, but the (*E*)-α-bisabolene synthase (AF006194) is a tri-domain, α-β-γ, protein, while δ-selinene (AGU92266) and γ-humulene synthases (AGU92267) are bi-domain, α-β, proteins (Bohlmann *et al.*, 1998). This suggests that in the evolution toward sesquiterpene synthesis in gymnosperm proteins, the transit peptide could have been lost, followed by the loss of the γ-domain (Figure 1.4).

In angiosperms, there is also evidence of the evolution of sesquiterpene synthases from monoterpene synthases (Fig. 1.4). It has been suggested that *Lavandula angustifolia* (*E*)-α-bergamotene synthase (LaBERS) (DQ263742) has evolved from a monoterpene synthase by the loss of the plastidial signal peptide and by broadening its substrate spectrum (Landmann *et al.*, 2007). LaBERS is similar to an α-zingiberene synthase of sweet basil (*Ocimum basilicum*) in that the latter has greater similarity to monoterpene synthases than to sesquiterpene synthases (Landmann *et al.*, 2007). Analogously, a vestigial activity of santalene synthases with GDP suggests that these enzymes may have evolved from a monoterpene synthase ancestor through loss of the plastid signal peptide and then adaptation of the active site to (*E*,*E*)-FDP (Jones *et al.*, 2011). For example, snapdragon (*Antirrhinum majus*) has two C10/C15 (*E*)-nerolidol/linalool synthases (AmNES/LIS-1 and AmNES/LIS-2) (Nagegowda *et al.*, 2008), *Arabidopsis thaliana* has two C10/C15 (*E,E*)-α-farnesene/(*E*)-β-ocimene synthases (AtTPS02 and AtTPS03) (Huang *et al.*, 2010) and *Fragaria ananassa* has two C10/C15 (*E*)-nerolidol synthases (FaNES1 and FaNES2; Aharoni *et al.*, 2004). In all these three cases, one terpene synthase protein lacks the target peptide in the N terminus (AmNES/LIS-1, AtTPS03, FaNES1), while the other

has it (AmNES/LIS-2, AtTPS02, FaNES2), which further suggest that sesquiterpene synthases might have evolved from monoterpene synthases (Fig. 1.4).

Recently, a new class of plant terpene synthase gene, the microbial-type class I terpene synthase genes containing only an α-domain, has been discovered in the spikemoss *Selaginella muellendorffi* (Li *et al.*, 2012). This class of TPS genes is more similar to microbial TPS genes than other plant TPS genes and is probably integrated into plant genomes from microbes via horizontal gene transfer, indicating a different evolutionary path rather than evolving from the three-domain diterpene synthase gene (Jia, 2016).

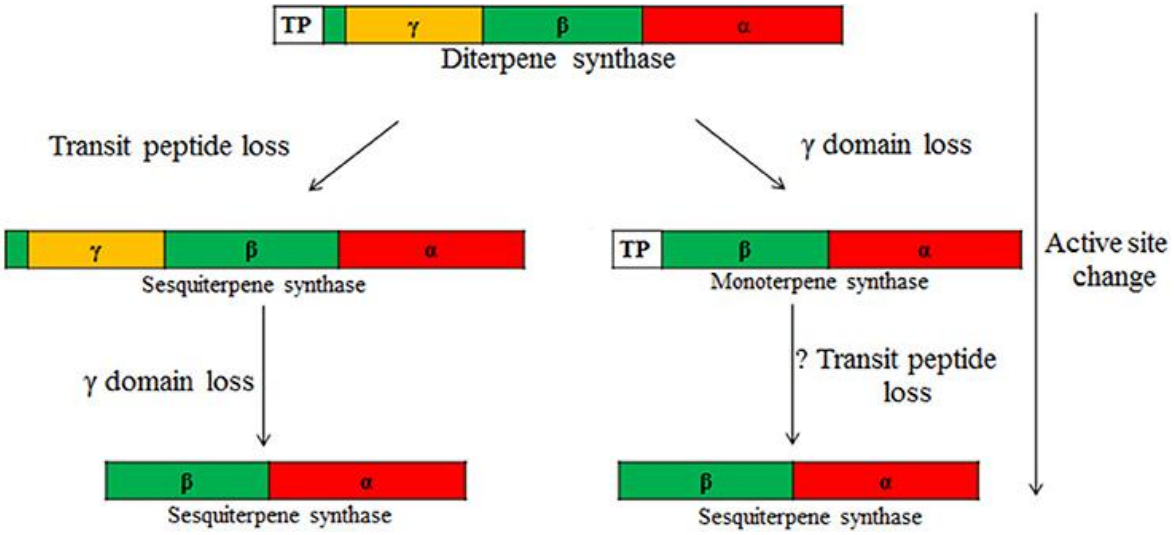



Figure 1.4. Hypothesis of the evolution of multi-substrate sesquiterpene and monoterpene synthases according to two potential routes (Pazouki and Niinemets, 2016). Ancient terpenoid synthases underlying the diversity of terpene synthases in plants are tri-domain, α-, β,- and γ-domain proteins with two active sites, one in the α-domain (class I activity) and the other in the β-domain (class II activity). The γ-domain without an active site is inserted between the first and second helices of the β-domain. These ancient proteins also carry a transit peptide (TP) at the N terminus targeting these proteins to chloroplasts. Through evolution, these complex enzymes have undergone considerable simplification, resulting in changes in catalysis, enzyme subcellular localization, and product and substrate specificities. Class II activity seems to have been lost first (not shown in the figure) and is missing in all confirmed multi-substrate enzymes. A tri-domain terpene synthase functionally active in the cytosol is formed through the loss of the transit peptide from a diterpene synthase. This can be eventually followed by γ-domain loss, resulting in formation of a bi-domain cytosol-active synthase (left). While the transit peptide is maintained, γ-domain loss can first lead to formation of a bi-domain diterpene synthase and ultimately to a monoterpene synthase (right). Loss of the transit peptide can further lead to a cytosol-active enzyme. Changes in substrate specificity are typically also associated with changes in active center size, and thus, the capacity for the use of multiple substrates will critically depend on whether the active center cavity can accommodate substrates of varying size.

Although the TPS represents a mid-size gene family, they are too few in numbers when compared to the great number of compounds in this group, which is mainly due to the functional plasticity of most TPSs. Many terpene synthases can produce more than one compound via only very few substitutions. In order to survive in the rapidly changing environment, TPS genes must evolve quickly to generate the required terpenoid profile. In this context, it is worth noting that recent studies, in the characterization of genes and enzymes responsible for substrate and end product biosynthesis as well as efforts in metabolic engineering, have demonstrated existence of a number of multi-substrate plant terpene synthases (Pazouki and Niinemets, 2016). Multi-substrate use could lead to important changes in terpene

product profiles upon substrate profile changes under perturbation of metabolism in stressed plants as well as under certain developmental stages. It is therefore argued that multi-substrate use can be significant under physiological conditions and can result in complicate modifications in terpene profiles. Perturbation of terpenoid metabolism under stress conditions can lead to enhanced substrate exchange between cytosol and plastids (Rasulov *et al.*, 2015) as well as modifications in the expression of enzymes responsible for product pool sizes (Steele *et al.*, 1998), and thus, also favor synthesis of terpenoids according to non-conventional pathways (Pazouki and Niinemets, 2016).

The existence of medium-size to large TPS families in the sequenced genomes of several plant species strongly supports the notion that TPS genes evolve by gene duplication and neofunctionalization (Chen *et al.*, 2011). The number of genes coding for TPSs ranges from 2 and 13 in *Physcomitrella patens* and *Selaginella moellendorffii* to 113 and 85 in *Eucalyptus grandis* and *Vitis vinifera*, with *Populus trichocarpa* (59), *Oryza sativa* (51), *Sorghum bicolor* (47) and *Arabidopsis thaliana* (34) possessing an intermediate but large number of genes (Myburg *et al.*, 2014). These data suggest that the TPS family has undergone significant expansion during the evolution of land plants. TPS genes are found in all five chromosomes in Arabidopsis, and in all but one chromosome in both rice and sorghum, whose chromosome number is 12 and 10, respectively (Chen *et al.*, 2011). In poplar, with 19 chromosomes, 12 chromosomes contain TPS genes (Irmisch *et al.*, 2014). In grapevine, the TPS genes are localized on seven of the 19 grapevine chromosomes, but 18 TPSs remained unmapped (Martin *et al.*, 2010).

A significant number of TPS genes in the genome of the angiosperm plants occur in tandem arrays of two or more genes (sometimes separated by one or a few unrelated genes). In Arabidopsis, rice, poplar, grapevine and sorghum, 42, 64, 59, 85 and 66% of TPS genes, respectively, occur in such tandem arrays (Chen *et al.*, 2011). These tandem arrays are likely the consequence of local gene duplication by unequal crossover. Consistent with this hypothesis, the genes in the tandem arrays are typically highly homologous to each other. For example, AtTPS23 and AtTPS27 are two Arabidopsis TPS genes located in such a cluster, and they are identical to each other both in the coding region and intron sequences, and thus represent a very recent TPS gene duplication (Chen, 2004). The tandem arrays of TPS genes are in some cases quite extensive, as in grapevine, where 45 VvTPS genes are organized as an extremely dense TPS gene cluster across a stretch of 690 kb on chromosome

18 (Martin *et al.*, 2010) and in rice, where 14 TPS genes occur in a 480 kb stretch on chromosome 4 (Chen *et al.*, 2011).

Based on phylogenetic analyses plant TPSs can be divided into seven clades or subfamilies: a, b, c, d, g, e/f and h (Bohlmann *et al.*, 1998; Dudareva, 2003; Martin *et al.*, 2004; Chen *et al.*, 2011) (Fig. 1.5). Function and taxonomic distribution of the seven plants TPS subfamilies are summarized in Table 1.1. As previously reported recently, a new class of terpene synthases (subfamily x) was found in *S. moellendorffii* which showed sequence similarity to microbial terpene synthases and were designated as microbial terpene synthase-like (MTPSL) genes (Li *et al.*, 2012).

The a, b and g subfamilies are angiosperm specific, while the d subfamily is gymnosperm-specific. The subfamily c encodes mono and bi-functional di-terpene synthases and contains the single terpene synthase gene in the moss *Physcomitrella patens*, genes from angiosperms and gymnosperms and also genes in *Selaginella moellendorffii*. The subfamily e/f contains ent-kaurene synthases (KSs) from angiosperms and gymnosperms. The subfamily h is specific to *Selaginella moellendorffii* and members of this subfamily contain both DxDD and DDxxD motifs, which are signatures of bifunctional diterpene synthases (Table 1.1).

Table 1.1. Function and taxonomic distribution of plant TPS subfamilies (Chen *et al.*, 2011). DiTPS, diterpene synthase; IspS, isoprene synthase; MonoTPS, monoterpene synthase; SesquiTPS, sesquiterpene synthase

| Subfamily | Groups | Functions | Distributions |
|---|---|---|---|
| TPS-a | TPS-a-1<br>TPS-a-2 | SesquiTPS<br>SesquiTPS | Dicots<br>Monocots |
| TPS-b | | MonoTPS, IspS | Angiosperms |
| TPS-C | | CPS/KS, CPS, other DiTPS | Land plants |
| TPS-d | TPS-d-1<br>TPS-d-2<br>TPS-d-3 | Primarily MonoTPS, SesquiTPS<br>SesquiTPS<br>Primarily DiTPS, SesquiTPS | Gymnosperms<br>Gymnosperms<br>Gymnosperms |
| TPS-e/f | | KS, other DiTPS,<br>monoTPS,SesquiTPS | Vascular plants |
| TPS-g | | MonoTPS, SesquiTPS, DiTPS | Angiosperms |
| TPS-h | | Putative bifunctional DiTPS | *Selaginella moellendorffii* |

The angiosperm-specific TPS-a, TPS-b and TPS-g clades have substantially diverged from other TPS clades (Fig. 1.5). Many of the genes in these three clades have been functionally characterized in model- and non-model systems and based on current knowledge, these three clades comprise entirely of genes of specialized mono-, sesqui- or di-terpene synthases with roles in ecological plant interactions rather than roles in primary plant metabolism (Chen *et al.*, 2011). TPS-a clades contain predominantly sesquiterpene synthases, TPS-b and TPS-g clades consist mostly of monoterpene synthases (Irmisch *et al.*, 2014).

Analysis of several flowering plants, whose genome is sequenced, suggests that the TPS genes in TPS-a subfamily account for more than half of the angiosperm TPS genes. TPS-a can be further divided into two groups, TPS-a-1 and TPS-a-2, with the

first being monocot-specific and the second dicot-specific (Fig. 1.5). Generally, the TPS-a clade appears to be highly divergent in all seed plants characterized to date.

The other angiosperm specific clade TPS-b, contain either monoterpene synthases (including all Arabidopsis monoterpene synthase except linalool synthase) or isoprene synthases (Chen *et al.*, 2011). The first member of this clade was discovered as (–)-limonene synthase in *Mentha spicata* (Colby *et al.*, 1993). While the majority of TPS-b genes are from dicots, two TPSs from sorghum also belong to this group. However, none of the rice TPS genes fall into the TPS-b clade. The enzymes of the TPS-b group produce cyclic monoterpenes and hemiterpenes (Külheim *et al.*, 2015). Many of the specific monoterpene synthase functions represented in the distantly related angiosperm-specific TPS-b and the gymnosperm-specific TPS-d-1 clades appear to have evolved convergently in the angiosperms and gymnosperms (Chen *et al.*, 2011).

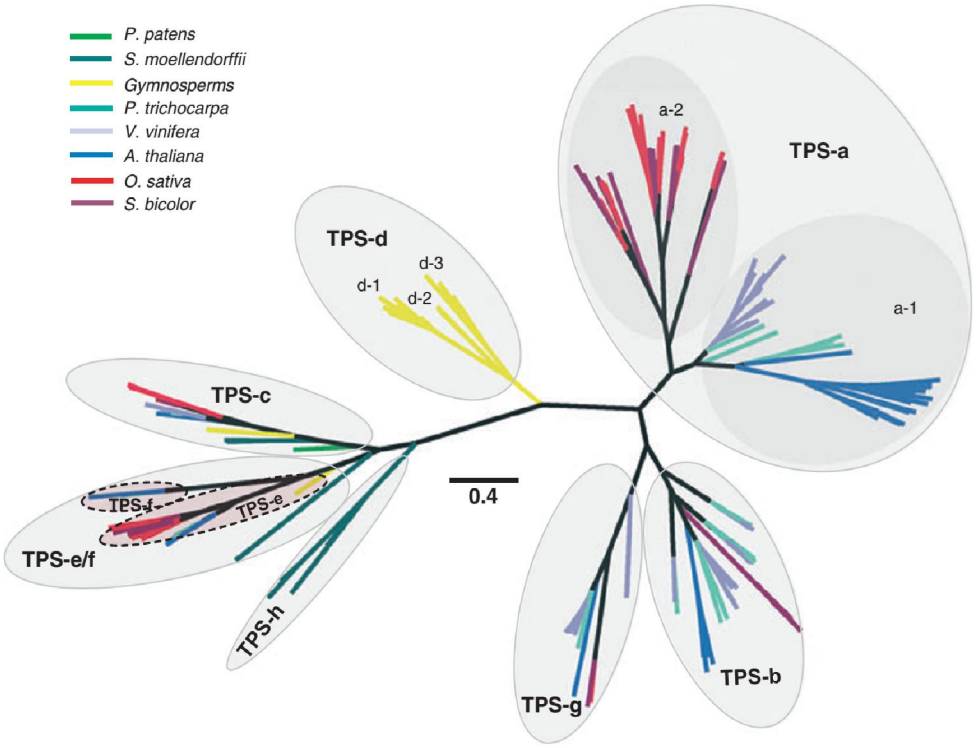



Figure 1.5. Phylogeny of putative full-length TPSs from seven sequenced plant genomes and representative characterized TPSs from gymnosperms (Chen *et al.*, 2011).

TPS-g is a clade closely related to TPS-b and was first defined by monoterpene synthases that produce the acyclic floral scent compounds myrcene and ocimene in snapdragon (Dudareva *et al.*, 2006). Two members of this clade identified in

Arabidopsis (At1g61680; Chen, 2003) and rice (Os02g02930; Yuan *et al.*, 2008), confirms that they produce an acyclic monoterpene, namely linalool synthases. A prominent feature of members from TPS-g group is the prevalence of acyclic products and the lack of the RRX8W motif which is highly conserved near the N-terminus of monoterpene synthases (mostly cyclases) of the angiosperm TPS-b clade and the gymnosperm TPS-d-1 clade (Chen *et al.*, 2011)

The TPS-c clades includes the bifunctional diterpene synthase CPS/KS (class I/II) of *Physcomitrella patens*, some CPS proteins from gymnosperms (spruce PgCPS and PsCPS), and angiosperms (Arabidopsis CPS, and rice OsCPS1, OsCPS2 and OsCPSsyn) as well as three TPSs from *Selaginella moellendorffii* that contain only the 'DXDD' motif but not the 'DDXXD' motif, suggesting that they are monofunctional CPS (Chen *et al.*, 2011). Previous studies suggest that all TPS genes evolved from the ancestral bifunctional CPS/KS gene, thus, It is plausible that the TPS-c clades represent the base of the phylogenetic tree of TPS genes. As discussed previously, evidence suggests that all plant TPS genes evolved from an ancestral bifunctional CS/KS gene, it is also plausible that the TPS-c clade represents the base of phylogenetic trees of TPS family (Fig. 1.5), and furthermore, that mono-functionalization occurred in the TPS family very early in land plant evolution.

Closely related to the TPS-c subfamily are the TPS-e and TPS-f subfamilies. TPS-e contains class I (KS) proteins from gymnosperms and angiosperms involved in the primary metabolism (Gibberellin synthesis). In the phylogenetic analysis of Chen *et al.* (2011) three *S. moellendorffii* TPSs form a subclade that is located near the bifurcation node of the TPS-c and the TPS-e clades (Fig. 1.5). The presence of the 'DDXXD' motif but not the 'DXDD' motif in the proteins encoded by these *S. moellendorffii* genes suggests they function as class I TPS, probably KS. Therefore, this branch of *S. moellendorffii* TPSs are placed in the TPS-e subfamily.

Previously, TPS-f were separated from the TPS-e subfamily, but using additional sequences in the phylogenetic analysis of Chen *et al.* (2011), it is clear that TPS-f is derived from TPS-e, hence, these two subfamilies have been combined into one clade designated as TPS-e/f. Genes assigned to the TPS-f includes AtTPS04, a diterpene synthase making geranyl linalool in Arabidopsis (Herde *et al.*, 2008), uncharacterized TPSs from poplar and grapevine (Martin *et al.*, 2010) as well as CbLIS, a monoterpene synthase producing S-linalool from *Clarkia breweri* flowers (Dudareva, 1996) and two unusual TPSs from Solanum. The TPS-f clade is probably

dicots specific because no TPS-f like genes have been found to date in monocot species.

*S. moellendorffii* TPSs that do not belong to the previously defined subfamilies of TPS-c and TPS-e/f, form a new clade designated as TPS-h (Fig. 1.5). In contrast to *S. moellendorffii* from other groups and one TPS gene from TPS-h that is presently missing the sequence at the region containing the 'DDXXD' motif, all *S. moellendorffii* TPSs in the TPS-h subfamily contain both 'DXDD' and 'DDXXD' motifs.

Interestingly, among all the TPS genes identified to date in angiosperm species, none contains both 'DXDD' and 'DDXXD' motifs. Conversely, several functionally characterized bifunctional diterpene synthases of specialized metabolism in the gymnosperms contain both the 'DXDD' and the 'DDXXD' motif, similar to the bifunctional PpCPS/KS (Keeling *et al.*, 2010; Hall *et al.*, 2013). These gymnosperm bifunctional diterpene synthases belong to the TPS-d subfamily (Figure 1.5). It is likely that gymnosperm bifunctional diterpene synthases evolved from a CPS/KS ancenstral PS (probably before the gymnosperm-angiosperm split, since neither lineage appears to contain a CPS/KS) (Keeling *et al.*, 2010). Similarly, it is likely that the putative bifunctional TPSs in the newly defined subfamily TPS-h evolved from PpCPS/KS and may be involved in specialized metabolism in *S. moellendorffii.*

In the gymnosperms, it is possible to see a clear phylogenetic separation of TPS genes of specialized metabolism and TPS genes of primary gibberellin metabolism. As previously described, known gymnosperm CPS and KS of gibberellin metabolism belong respectively to the TPS-c and TPS-e/f clades like their counterparts in the angiosperms (Fig. 1.5). All gymnosperm TPSs for specialized metabolism belong to the gymnosperm-specific subfamily TPS-d. The TPS-d subfamily can be further divided into TPS-d-1, TPS-d-2 and TPSd-3 (Martin *et al.*, 2004). TPS-d-1 contains all known gymnosperm monoterpene synthases for the production of a large array of conifer defense compounds, in addition to a few TPSs which produce the simple acyclic sesquiterpene (E,E)-a-farnesene. Most known gymnosperm sesquiterpene synthases, including those enzymes that produce large arrays of multiple terpenoids (Steele *et al.*, 1998), belong to the TPS-d-2 group. The TPS-d-3 contains primarily diterpene synthases and several known sesquiterpene synthases (Chen *et al.*, 2011). Well characterized diterpene synthases in the TPS-d-3 group are the single-product and multiproduct enzymes of conifer diterpene resin acid biosynthesis (Peters *et al.*, 2000; Keeling *et al.*, 2008) and taxadiene synthase from Taxus (Wildung and Croteau, 1996). Although sesquiterpene synthases are found in all three groups,

they can be distinguished by gene structures: the sesquiterpene synthases in TPS-d-1 and TPS-d-2 are approximately 600 amino acids in length and those in TPSd-3 are approximately 800 amino acids.

The presence of the *S. moellendorffii*-specific TPS-h subfamily, the gymnosperm-specific TPS-d subfamily and the angiosperm-specific subfamilies of TPS-a, TPS-d and TPS-g indicates lineage-specific expansion of the TPS family (Fig. 1.5). In angiosperms, the TPS-a subfamily is the major determinant of the size of the TPS family of individual species. Apparently, the expansion of the TPS-a family occurred after the split of the monocot and dicot lineages (Fig.1.5). Moreover, the positions of Arabidopsis TPS genes on the branches of clade TPS-a-1 indicate that many of them arose by gene duplications that occurred after the divergence of the Arabidopsis lineage from the *V. vinifera* and *P. trichocarpa* lineages (Chen *et al.*, 2011).

On the basis of phylogenetic and functional analyses, a general model for the description of the TPS family became apparent. For example, all monoterpene synthases were thought to be localized in the plastids and using GPP as substrate. Sesquiterpene synthases were thought to be generally localized in the cytosol where they use all-trans FPP as their substrate, with the use of GPP occurring only in vitro. Finally, the larger diterpene synthases had their place in plastids with all-trans GGPP as their substrate.

While much of this general model still holds true, new discoveries continue to reveal new facets of the TPS gene family. Even with this initial model, it was clear that the TPS family possesses a remarkable flexibility to evolve enzymes with new subcellular localization and substrate specificity. For example, the topology of the phylogenetic tree in the analysis of Chen *et al.* (2011) (Fig. 1.5) suggests that mono- and sesquiterpene synthases evolved from diterpene synthases independently in gymnosperms and angiosperms. As previously discussed, a change from diterpene synthase to sesquiterpene synthase involves not only a change in substrate specificity but also in subcellular localization (by the loss of a transit peptide). A relative recent example of such an event may exist with the closely related diterpene synthases and sesquiterpene synthases of the gymnosperm TPS-d-3 clade. Here, the sesquiterpene synthase genes encoding (E)-a-bisabolene synthases in grand fir and Norway spruce (Bohlmann *et al.*, 1998; Martin *et al.*, 2004) still share with the conifer diterpene synthases, the sequence encoding the ancestral 200

amino acid motif (the γ-domain), but have apparently lost the region encoding the N-terminal transit peptide.

As described above, the recently duplicated Arabidopsis AtTPS02 and AtTPS03 genes (Huang *et al.*, 2010) represent another good example of neofunctionalization of duplicated TPS genes involving a change in subcellular localization, where AtTPS02 produces (E)-b-ocimene in the plastid and the transit peptide-lacking AtTPS03 produces (E,E)-a-farnesene in the cytosol, although both enzymes can synthesize both compounds (from the respective substrates) *in vitro* (Huang *et al.*, 2010).

These and other new findings that were not reported (Chen *et al.*, 2011), highlights the fascinating potential of the TPS gene family to evolve surprising variations of biochemical functions and subcellular localization, and strongly establish that substrate specificity, product profiles and localization of a TPS gene cannot be predicted based on association with a specific TPS subfamily or general sequence similarity. As different subfamilies expand in different lineages by gene duplication and divergence, as has happened for example in the TPS-a, TPS-b, and TPS-g in angiosperms, TPS-d in gymnosperms, and TPS-h in *S. moellendorffii*, it can be expected that proteins with altered subcellular localization and new substrate specificities would have evolved.

### 1.3.3 Terpene synthase genes in gymnosperm

Conifer trees produce complex mixtures of terpenoids, most prominently in the form of oleoresin, that can act as a physical and chemical defense against insect and pathogen attack (Zulak and Bohlmann, 2010). The diversity of conifer terpenoids includes predominantly monoterpenes, sesquiterpenes and diterpenes, which originate from the activity of a family of terpene synthases (TPSs), and other enzymes, such as cytochromes P450, that may functionalize some of the terpenes (Ro *et al.*, 2005; Keeling and Bohlmann, 2006).

As previously described, the many mono-, sesqui-, and diterpene synthases of conifer specialized metabolism form the gymnosperm-specific TPS-d subfamily (Chen *et al.*, 2011). The functional diversity of conifer TPSs appears to have evolved through events of repeated gene duplication and further sub- and neofunctionalization, leading to an expansion of the TPS-d multigene family (Martin *et al.*, 2004; Chen *et al.*, 2011) that is the central player behind the chemical complexity of conifer specialized terpenes. In contrast, conifer diTPSs of general (i.e.

primary) GA metabolism belong to the TPS-c and TPS-e/f subfamilies, which also include orthologous genes of angiosperms (Keeling *et al.*, 2010; Chen *et al.*, 2011).

Just over 10 years ago, our knowledge of the size, functional diversity and phylogeny of gymnosperm TPSs was based on targeted cDNA cloning and characterization in two conifer species, grand fir (*Abies grandis*) and Norway spruce (*Picea abies*), along with a few TPSs in other gymnosperms (Keeling and Bohlmann, 2006). In grand fir, 11 different TPS genes have been functionally characterized (Bohlmann *et al.*, 1999). Martin *et al.* (2004) described a set of 9 different TPSs in Norway spruce (*P. abies*) and examined the phylogeny of 29 gymnosperm TPSs, all of which fell into the gymnosperm-specific TPS-d subfamily.

More insight to diversity and functional complexity of TPS family in gymnosperm has been gained from recent analysis of transcriptome in several spruce (*Picea*) species (Keeling *et al.*, 2011) and lodgepole pine (*Pinus contorta*) and jack pine (*P. banksiana*) (Hall *et al.*, 2013).

The availability of extensive transcriptome resources in the form of expressed sequence tags (ESTs) and full-length cDNAs in several spruce (*Picea*) species allowed Keeling *et al.* (2011) to identify 69 unique and transcriptionally active TPS genes in white spruce (*P. gluaca*), 55 in Sitka spruce (*P. sitchensis*) and 20 in hybrid white spruce (*P. glauca* × *P. engelmannii*). The estimate of at least 69 TPSs in white spruce is comparable to the number of putatively active TPS genes found in the sequenced genomes of angiosperms and is perhaps a good approximation of the total number of transcriptionally active TPS genes in a conifer species. Keeling *et al.*, 2011 functionally characterized a total of 21 spruce TPSs: 12 from Sitka spruce, 5 from white spruce, and 4 from hybrid white spruce, which included 15 monoterpene synthases, 4 sesquiterpene synthases, and 2 diterpene synthases (Table 1.2). The functional diversity of these characterized TPSs parallels the diversity of terpenoids found in the oleoresin and volatile emissions of Sitka spruce and provides a context for understanding this chemical diversity at the molecular and mechanistic levels.

Table 1.2. Gene name, origin, accession numbers, and functional annotation of spruce TPS identified by Keeling *et al.* (2011)

| Gene | Clone ID (genotype) | Functional Annotation* | NCBI Accession |
|---|---|---|---|
| MONOTERPENE SYNTHASES | | | |
| PgxeTPS-Car1 | WS0063_F08 (Fa1-1028) | (+)-3-Carene synthase | HQ426152 |
| PsTPS-Car1 | WS02910_I02 (FB3-425) | (+)-3-Carene synthase | HQ426167 |
| PgTPS-Cin | WS02628_N22 (PG29) | 1,8-Cineole synthase | HQ426160 |
| PgxeTPS-Cin | WS00921_D15 (Fa1-1028) | 1,8-Cineole synthase | HQ426156 |
| PsTPS-Cin | WS0291_H24 (FB3-425) | 1,8-Cineole synthase | HQ426165 |
| PgTPS-Lin | WS0054_P01 (PG29) | (-)-Linalool synthase | HQ426151 |
| PsTPS-Lin-1 | WS0285_L07 (FB3-425) | (-)-Linalool synthase | HQ426164 |
| PsTPS-Lin-2 | WS02915_K02 (FB3-425) | (-)-Linalool synthase | HQ426168 |
| PsTPS-Phel-1 | WS02729_A23 (FB3-425) | (-)-β-Phellandrene synthase | HQ426162 |
| PsTPS-Phel-2 | WS0296_I22 (FB3-425) | (-)-β-Phellandrene synthase | HQ426169 |
| PsTPS-Phel-3 | WS0276_M12 (FB3-425) | (-)-β-Phellandrene synthase | HQ426163 |
| PsTPS-Phel-4 | WS01042_E12 (Gb2-229) | (-)-β-Phellandrene synthase | HQ426159 |
| PgTPS-Pin-1 | WS00725_G07c1 (PG29) | (-)-α/β-Pinene synthase | HQ426153 |
| PgTPS-Pin-2 | WS00725_G07c2 (PG29) | (-)-α/β-Pinene synthase | HQ426154 |
| PsTPS-Pin | WS0291_K15 (FB3-425) | (-)-α/β-Pinene synthase | HQ426166 |
| SESQUITERPENE SYNTHASES | | | |
| PgxeTPS-Far/Oci | WS00926_E08 (Fa1-1028) | (*E*,*E*)-α-Farnesene/(*E*)-β-ocimene synthase | HQ426157 |
| PgTPS-Hum | WS0074_O16 (PG29) | α-Humulene synthase | HQ426155 |
| PgxeTPS-Lonf | WS00927_M20 (Fa1-1028) | Longifolene synthase | HQ426158 |
| PsTPS-Lonp | WS02712_A08 (FB3-425) | α-Longipinene synthase | HQ426161 |
| DITERPENE SYNTHASES | | | |
| PsTPS-Iso | pSW06061903 (Haney 898) | Isopimaradiene synthase | HQ426150 |
| PsTPS-LAS | WS0299_C21 (FB3-425) | Levopimaradiene/abieta-diene synthase | HQ426170 |

*Functional annotation is based on the main terpenoid product(s) of recombinant enzymes expressed in *E*. *Coli*

As reported above, the TPS-d subfamily has been subdivided into three clades TPS-d1 through TPSd3 based on a previous phylogeny of 29 gymnosperm TPSs (Martin *et al.*, 2004; Chen *et al*., 2011). Keeling *et al*. (2011) expanded the phylogeny of functionally characterized gymnosperm TPSs to a total of 72 members, of which 41 are from spruce species including the 21 new functionally characterized (Fig. 1.6).

The diversity of newly characterized spruce TPSs represents the three major clades (TPS-d1, TPS-d2 and TPS-d3) of the TPS-d subfamily, and help for the identification of groups of likely orthologous TPS genes across the spruce species. Examples for such groups of orthologous TPSs in the TPS-d1 clade are the (-)-a/b-pinene synthases, the (-)-linalool synthases, (E,E)-a-farnesene synthases; in the TPS-d2 clade are the longifolene synthases; and in the TPS-d3 clade are the levopimaradiene/abietadiene synthases and isopimaradiene synthases. These groups represent genes whose functions had apparently evolved prior to speciation of the spruce genus. In the TPS-d3 group of conifer diterpene synthases, the basal function of a multi-product levopimaradiene/abietadiene synthase had apparently

evolved prior to conifer speciation, as this function exists in a group of closely related genes from the genera Abies, Pinus and Picea.

In contrast to the many duplicated TPS-d genes of terpenoid specialized metabolism, the related spruce TPS genes of general gibberellin phytohormone biosynthesis, specifically ent-copalyl diphosphate synthase (TPS-c) and ent-kaurene synthase (TPS-e), appear to be expressed as single copy genes (Keeling *et al.*, 2010). As reported above, these primary metabolism TPS genes are basal to the specialized metabolism genes and are the descendants of an ancestral plant diterpene synthase similar to the one found in the non-vascular plant *Physcomitrella patens* (Keeling *et al.*, 2010). The mechanisms that suppress manifestation or retention of TPS gene duplication in diterpenoid primary metabolism and those that enhance TPS gene duplication and functional diversification in specialized metabolism in a conifer genome are not known but are worthy of future investigation.

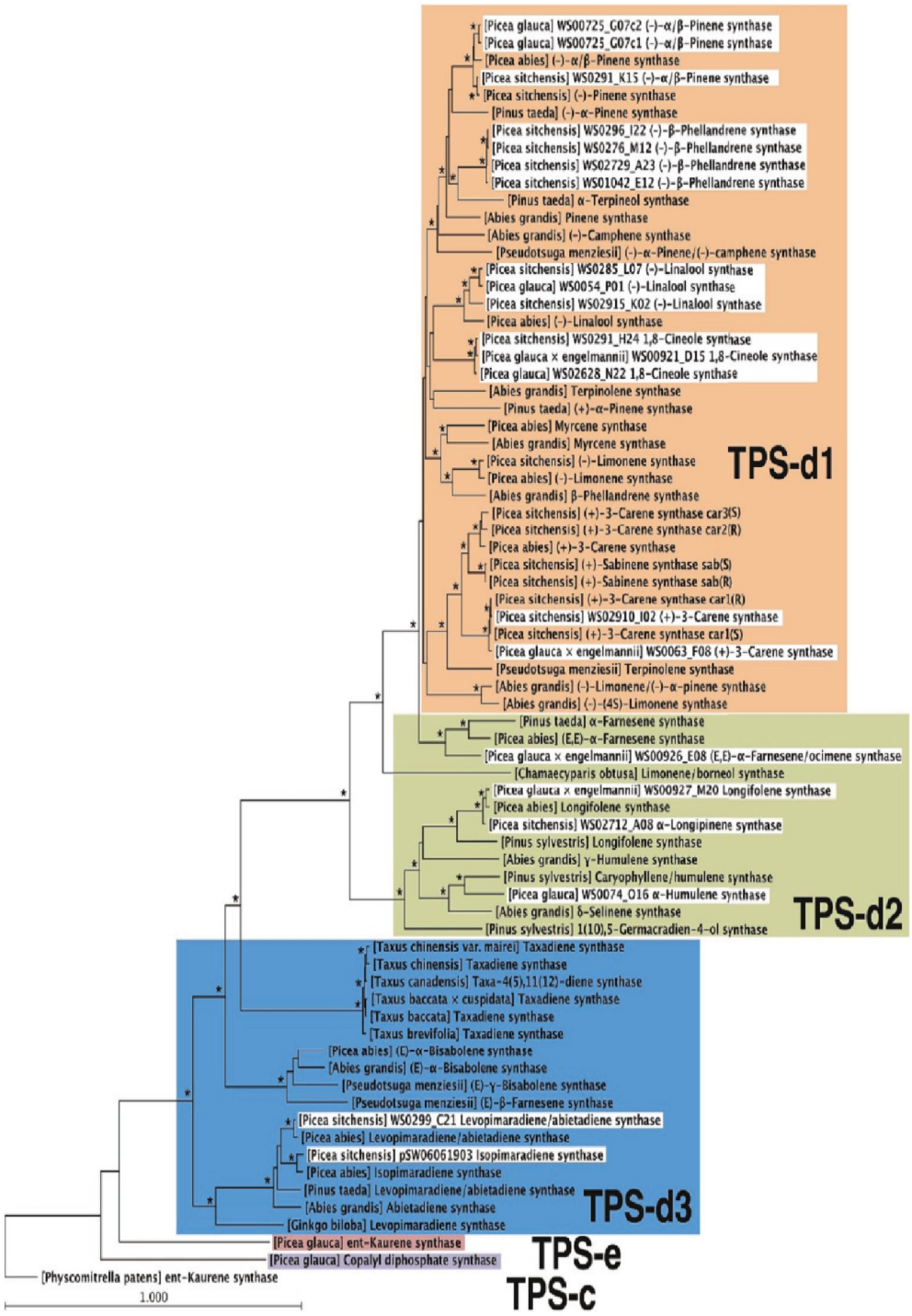



Figure 1.6. Phylogeny of 72 functionally characterized gymnosperm TPSs (Keeling *et al.*, 2011). The 21 identified TPS genes by Keeling *et al.* (2011) are shown with white background.

Despite the economic and ecological importance of pines and the importance of oleoresin terpenes in pine defense and as bioproducts, until 2013, only three mono-TPSs, which form either (+)-α-pinene, (−)-α-pinene, or (−)-α-terpineol as major

products, and a single bifunctional diTPS (PtLAS), which produces abietadiene, neoabietadiene, levopimaradiene, and palustradiene, have been characterized from loblolly pine (*Pinus taeda)* (Phillips *et al.*, 2003; Ro and Bohlmann, 2006).

More recently, using transcriptome sequence resources developed by a combination of Sanger, 454, and Illumina sequencing of complementary DNA (cDNA) libraries made from sapling stem tissues, Hall *et al.* (2013) cloned and functionally characterized nine different jack pine (*Pinus banksiana*) and eight different lodgepole pine (*P. contorta*) mono-TPSs. The newly identified lodgepole pine and jack pine include (+)-α-pinene synthases, (-)-α-pinene synthases, (-)-β-pinene synthases, (+)-3-carene synthases, and (-)-β phellandrene synthases from each of the two species (Hall *et al.*, 2013).

The phylogenetic analysis carried out considering several previously characterized conifer mono-TPSs, placed the lodgepole pine and jack pine mono-TPSs within the TPS-d1 clade. Many of the pine monoTPSs, including the genes responsible for (+)-3-carene and (−)-α-pinene biosynthesis, group phylogenetically with functionally similar mono-TPSs from loblolly pine, grand fir and spruce (Fig. 1.7). This functional conservation across species suggests that considerable gene duplication and functionalization occurred prior to the speciation of pine, fir and spruce.

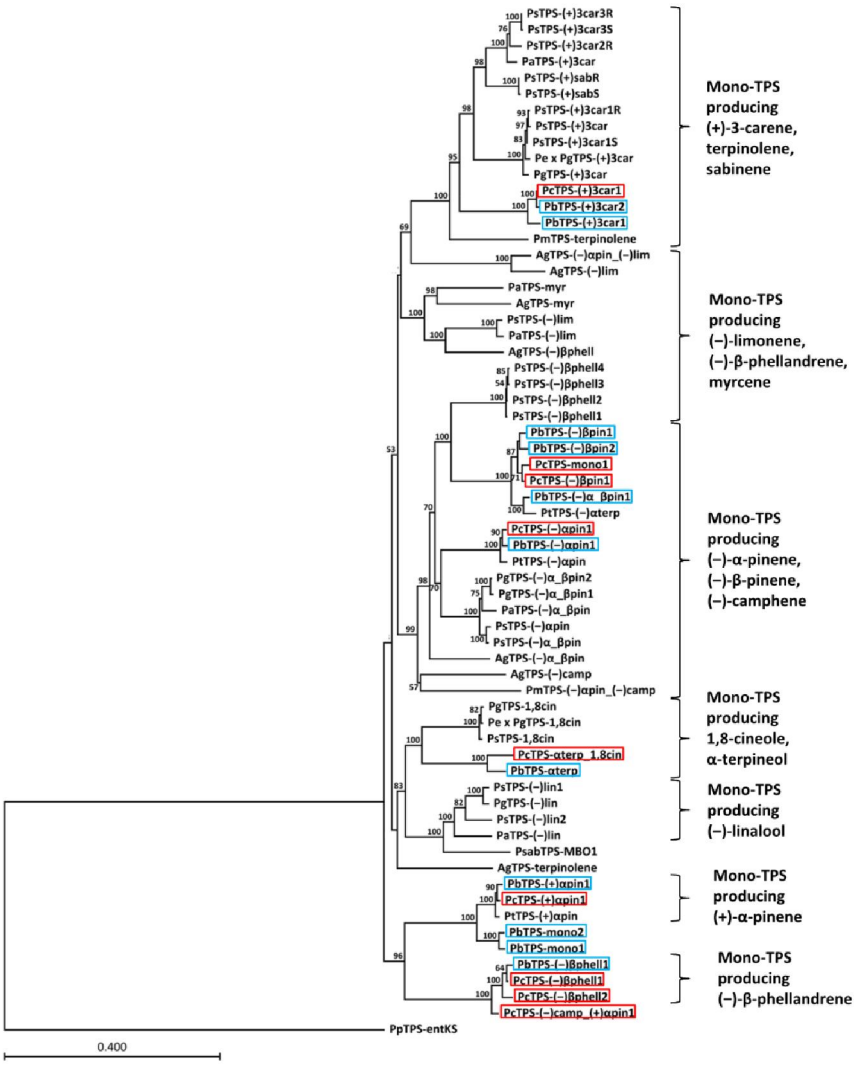



Figure 1.7. Phylogeny of full length mono-TPSs from lodgepole pine (Pc, red) and jack pine (Pb, blue) isolated by Hall *et al*. (2013a) with previously functionally characterized mono-TPSs from eight conifer species: 18 from *Picea sitchensis* (Ps), 7 from *Abies grandis* (Ag), 5 from *Picea abies* (Pa), 5 from *Picea glauca* (Pg), 3 from *Pinus taeda* (Pt), 2 from *Picea engelmannii* x *Picea glauca* (PexPg), 2 from *Pseudotsuga menziesii* (Pm) and 1 from *Pinus sabiana* (Psab). The ent-kaurene synthase (PpTPS-entKS) *from Physcomitrella patens* was used as an outgroup.

The jack pine and lodgepole pine (+)-α-pinene synthases and (−)-β-phellandrene synthases grouped together with the previously characterized loblolly pine (+)-α-pinene synthase (Phillips *et al.*, 2003) is a unique and apparently Pinus specific subclade within the TPS-d family (Fig. 1.7). The jack pine and lodgepole pine (−)-β-phellandrene synthases grouped separately from the Sitka spruce (Keeling *et al.*, 2011) and grand fir (−)-β-phellandrene synthases (Bohlmann *et al.*, 1999), highlighting the multiple origins of (−)-β-phellandrene biosynthesis in conifers. Genes that produce (+)-α-pinene as their major product have not been identified in

any conifer genus other than Pinus, suggesting this function may have evolved in the pine lineage after the separation from spruce and firs.

Three proteins from jack pine and lodgepole pine shared 91-93% sequence identity with the previously characterized loblolly pine (−)-α-terpineol synthase. Based on sequence identity, one may have predicted that the jack pine and lodgepole pine proteins would similarly produce α-terpineol. Surprisingly, these proteins produced 75-81% (−)-β-pinene and no α-terpineol. Previous reports demonstrate that a few amino acid substitutions are sufficient to alter the product profiles of mono-TPSs from grand fir (Katoh *et al.*, 2004; Hyatt and Croteau, 2005). The high level of sequence identity between these functionally distinct proteins from jack pine, lodgepole pine and loblolly pine serves as an example of the functional plasticity observed in conifer mono-TPS (Hall *et al.*, 2013). Hall and collaborators using the same high-throughput transcriptome sequencing approach reported the above discovered 11 diTPS from jack pine and lodgepole pine (Hall *et al.*, 2013). Three of the diTPS-like sequences (named PbLAS1, PcLAS1, and PcLAS2) showed 98% to 99% amino acid sequence identity to each other and to a previously characterized loblolly pine LAS (PtLAS; Ro and Bohlmann, 2006). These sequences contained the class I and class II active site functional motifs, suggesting that they were bifunctional class I/II diTPSs, resembling the known conifer diTPSs of diterpene resin acids (DRA) biosynthesis.

Unexpectedly, the eight remaining putative diTPS sequences contained only the class I signature motifs (NSE/DTE, DDxxD), but lacked either the conserved middle Asp residue (PcmISO1 and PbmISO1) or the first and last Asp residues (the remaining six PcmdiTPS1-3, PcPIM1, PbmdiTPS1 and PbPIM1) of the DxDD motif, which were previously shown to be critical for class II catalysis (Peters and Croteau, 2002). These eight sequences showed 66% to 73% amino acid identity to jack pine PbLAS1, lodgepole pine PcLAS1 and PcLAS2, and functionally characterized Norway spruce (*Picea abies*) PaISO and PaLAS of DRA specialized metabolism (Martin *et al.*, 2004). Although representing putative monofunctional diTPSs, the eight sequences only showed 33% to 34% protein sequence identity to the monofunctional white spruce (*Picea glauca*) class II *ent*-copalyl diphosphate synthase (PgECPS) and class I *ent*-kaurene synthase (PgEKS) of GA metabolism (Keeling *et al*., 2010), suggesting roles in specialized as opposed to general metabolism. With 99% amino acid sequence identity to each other, PcmISO1 and PbmISO1, and likewise PcmPIM1 and PbmPIM1, presumably represent two pairs of orthologous genes from jack pine and lodgepole

pine. The remaining class I diTPS candidates (PcmdiTPS1, PcmdiTPS2, PcmdiTPS3, and PbmdiTPS1), though highly similar among each other (97% to 98% protein sequence identity), showed a lower identity of 71% to 75% to the other pine diTPS candidates.

Seven of the 11 diTPS identified sequences were functionally characterized via expression of recombinant proteins in *E. coli* (Hall *et al.*, 2013). Functional characterization of the three jack pine and lodgepole pine LAS enzymes (PbLAS1, PcLAS1, and PcLAS2) identified the diterpene olefins abietadiene, levopimaradiene, and neoabietadiene as the three major products, consistent with the three principal products of the previously characterized PtLAS (Ro and Bohlmann, 2006). Functional characterization of four of the eight putative monofunctional class I diTPS identified the orthologous pair of PbmPIM1 and PcmPIM1 as single-product pimaradiene synthases, while the orthologous pair of PbmISO1 and PcmISO1 are isopimaradiene synthases, which also produced small amounts of sandaracopimaradiene. Monofunctional class I diTPSs of specialized DRA metabolism have not been previously reported. The only other known example of a monofunctional class I diTPS of specialized metabolism in a gymnosperm is taxadiene synthase (Wildung and Croteau, 1996), which converts GGPP directly into the macrocyclic taxadiene backbone without a bicyclic diphosphate intermediate. The monofunctional pimaradiene synthases described by Hall *et al*. (2013) are the first reported gymnosperm enzymes that produce predominantly pimaradiene and extend the scope of known conifer diTPS functions involved in DRA formation beyond the previously known ISO and LAS enzymes.

Based on their monofunctional activity and role in DRA biosynthesis, the newly discovered class I diTPSs of jack pine and lodgepole pine introduce a new functionality to the catalytic landscape of specialized conifer diTPSs. To gain a deeper understanding of their evolution, Hall *et al.* (2013) performed a phylogenetic analysis, including mono- and bifunctional diTPS of DRA biosynthesis, monofunctional class I and class II diTPSs of GA biosynthesis, and selected conifer monofunctional class I mono-, sesqui-, and diTPSs. It was previously shown that bifunctional class I/II diTPSs of conifer DRA biosynthesis belong to a gymnosperm-specific TPS-d3 subfamily, while the monofunctional gymnosperm diTPSs of GA biosynthesis cluster together with angiosperm diTPSs of GA biosynthesis in the TPS-c and TPS-e/f subfamilies (Keeling *et al.*, 2010; Chen *et al*., 2011). The phylogenetic analysis of Hall *et al*. (2013) placed the jack pine and lodgepole pine mono- and

bifunctional diTPSs in the gymnosperm-specific TPS-d3 clade, containing primarily diTPSs and few γβα-domain sesqui-TPSs, clearly distant from GA biosynthetic ECPS and EKS (Fig. 1.8). The bifunctional class I/II PbLAS1, PcLAS1, and PcLAS2 are closely related to previously characterized bifunctional LAS and ISO enzymes from grand fir (Vogel *et al.*, 1996), norway spruce (Martin *et al.*, 2004), loblolly pine (Ro and Bohlmann, 2006), sitka spruce (Keeling *et al.*, 2011), and balsam fir (Zerbe *et al.*, 2012). The eight monofunctional diTPSs (PcmPIM1, PbmPIM1, PcmISO1, PcmISO1, PbmdiTPS1, PcmdiTPS1, PcmdiTPS2, and PcmdiTPS3) form a distinct branch within the TPS-d3 family adjacent to the bifunctional class I/II diTPSs of DRA biosynthesis, but situated distantly from other monofunctional gymnosperm TPS, such as *Taxus* spp. taxadiene synthase (Wildung and Croteau, 1996), γβα-domain gymnosperm sesqui-TPS (Bohlmann *et al.*, 1998; Martin *et al.*, 2004; Huber *et al.*, 2005; McAndrew *et al.*, 2011), and other βα-domain conifer mono-TPSs. Furthermore, the orthologous pairs of PcmPIM1/PbmPIM1 and PcmISO1/PcmISO1, for which Hall *et al*. (2013) showed biochemical functions, are separated from the four remaining diTPS candidates (PbmdiTPS1, PcmdiTPS1, PcmdiTPS2, and PcmdiTPS3) for which no activity was observed.

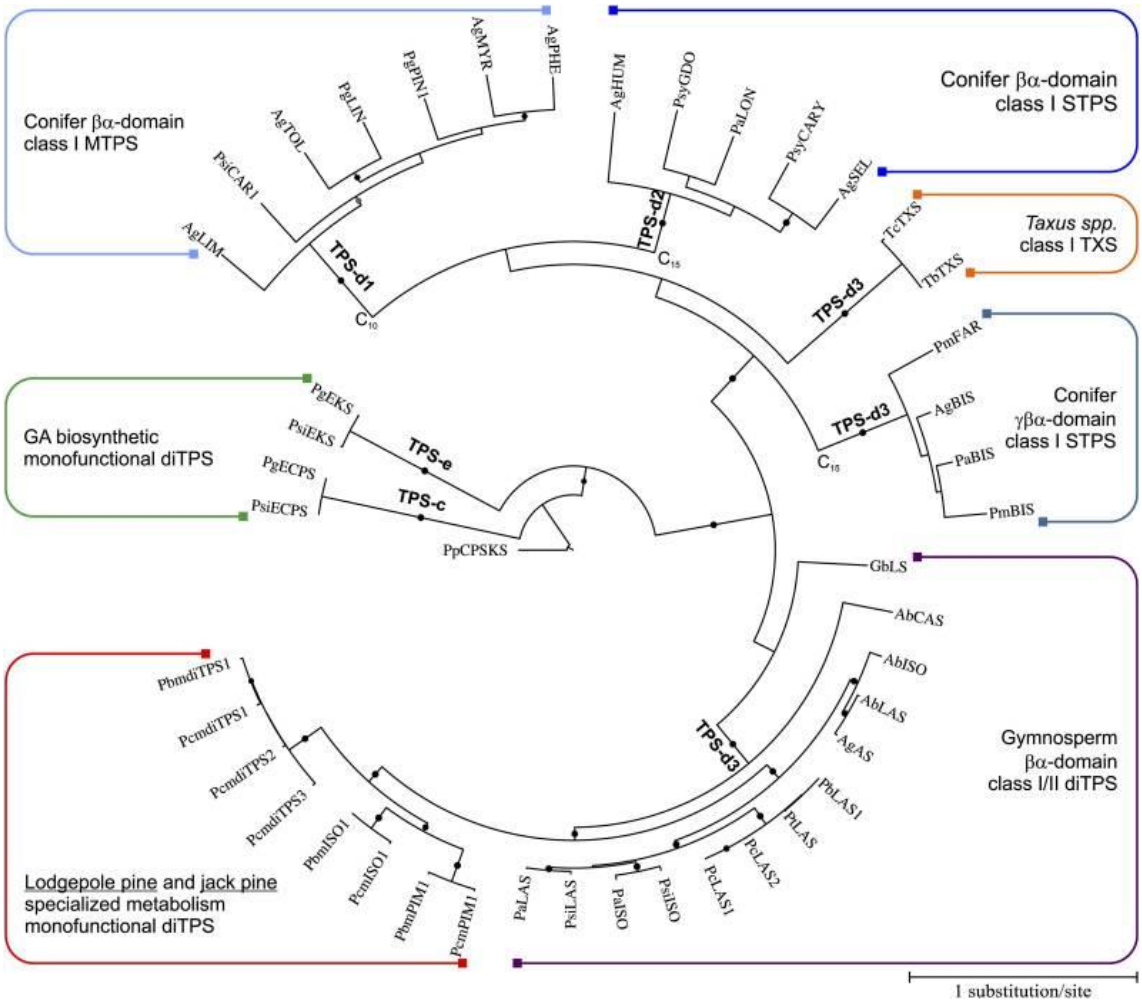



Figure 1.8. Phylogenetic tree of diTPS from jack pine and lodgepole pine with previously characterized members of the gymnosperm-specific TPS-d family (Hall *et al*., 2013b). *P. patens ent*-kaurene/kaurenol synthase (PpCPS/KS; accession no. BAF61135) was used to root the tree. Abbreviations and NCBI accession numbers are as follows: PcmISO1 (JQ240314), PbmISO1 (JQ240313), PcmPIM1 (JQ240315), PbmPIM1 (JQ240316), PcmdiTPS1 (JQ240318), PbmdiTPS1 (JQ240317), PcmdiTPS2 (JQ240319), PcmdiTPS3 (JQ240320), PtLAS (Q50EK2), PcLAS2 (JQ240311), PcLAS1 (JQ240310), PbLAS1 (JQ240312), PsiLAS (ADZ45517), PaLAS (Q675L4), PsiISO (ADZ45512), PaISO (Q675L5), AbLAS (JN254805), AgAS (Q38710), AbISO (JN254806), AbCAS (JN254808), GbLS (Q947C4), TcTXS (ABC25488), TbTXS (Q41594), AgBIS (O81086), PaBIS (AAS47689), PmBIS (Q4QSN4), PmFAR (ADX42737), PsiEKS (ADB55710), PgEKS (ADB55708), ΨECPS (ADB55709), PgECPS (ADB55707), AgHUM (O64405), AgSEL (AAC05727), AgLIM (AAB70907), AgTOL (AAF61454), AgPHE (Q9M7D1), AgMYR (AAB71084), PaLON (AAS47695), PgPIN1 (ADZ45507), PgLIN (ADZ45500), PsiCAR1 (ADZ45511), PsyCARY (ABV44452), and PsyGDO (ABV44453).

The topology of the phylogenetic tree obtained by Hall *et al.* (2013b) (Fig. 1.8) suggests that the monofunctional class I diTPSs of specialized metabolism in lodgepole pine and jack pine have evolved relatively recently through gene duplication of a bifunctional diTPS followed by loss of the class II activity and

additional functional diversification. While the bifunctional LAS enzymes of lodgepole pine and jack pine have orthologs in other conifers within and outside of the pine genus. For example, in loblolly pine, spruce, and firs, monofunctional class I diTPSs of specialized metabolism have not been found in other conifers. It is possible that they represent a lineage-specific clade of the TPS-d3 group that originated in a common ancestor of the closely related species of lodgepole pine and jack pine, perhaps after the separation from loblolly pine and after the separation of the pine, spruce, and fir genera.

The phylogenetic analysis carried out by Hall *et al.* (2013b) indicates that events of monofunctionalization (i.e. a form of subfunctionalization from a duplicated bifunctional ancestor) have occurred independently on at least three separate occasions in the evolution of gymnosperm diTPS, one being the evolution of monofunctional diTPSs of GA biosynthesis in gymnosperms and angiosperms, the second being the evolution of a monofunctional taxadiene synthase, and the third being the emergence of newly described monofunctional diTPSs of pine DRA biosynthesis (Fig. 1.8). The latter type of monofunctional class I diTPS appears to have evolved by loss of functionality of the class II active site that remained intact in the similar class I/II LAS and ISO enzymes. Beyond the diTPSs, monofunctionalization ultimately also led to the large family of bidomain conifer mono- and sesqui-TPSs of the TPS-d1 and TPS-d2 groups.

## 1.4 *Pinus nigra* J.F. Arnold

*Pinus nigra* J.F. Arnold known as European black pine or black pine, is one of the most economically important native conifers in southern Europe. It is a tertiary relictual species belonging to the group of Mediterranean pines (Bogunic *et al.*, 2007) and is one of the oldest European pine species, descending from a group that already existed in the Cretaceous (100 million years ago) (Gernandt *et al.*, 2008). *P. nigra* is a widespread species, with a discontinuous range that extends from North Africa through the northern Mediterranean and eastwards to the Black Sea (Isajev *et al.*, 2004; Enescu *et al.*, 2016). It has also become naturalized in some areas in North America. It is subdivided into several distinct subspecies and its taxonomic status is still a subject of debate among specialists (Enescu *et al.*, 2016).

### 1.4.1 Botanical description and biology

Black pine (*Pinus nigra* J.F. Arnold) is a large coniferous evergreen tree with an average height of 30 m at maturity, but at times, it is capable of attaining heights up to 40 m (Banfi and Consolino, 2011). Its bark is usually a dark greyish brown to black (giving rise to its Latin name "nigra") and becomes increasingly fissured with age (Enescu *et al.*, 2016). The crown is broadly conical on young trees and umbrella-shaped on older trees (Enescu *et al.*, 2016)**.** Their needles are usually 8-15 cm long, 1-2 mm thick, finely serrated (Banfi and Consolino, 2011) and could persist on the tree for 3-4 years (In rare cases, up to 8 years) (Enescu *et al.*, 2016). Black pine is a monoecious wind-pollinated conifer, and its seeds are wind dispersed. Flowering occurs annually although seed yield is abundant only once every 2–4 years. Trees reach sexual maturity at 15–20 years in their natural habitat. Flowers appear in May. Female inflorescences are reddish, and male catkins are yellow (Fig. 1.9). Cones are sessile and horizontally spreading, 4–8 cm long, 2–4 cm wide, and yellow-brown in colour (Fig. 1.9). They ripen from September to October of the second year, and open in the third year after pollination. Its cone contains 30–40 seeds with usually 50% viability rate. Seeds are grey, 5–7 mm long, with a wing 19–26 mm long. Germination can occur without stratification although this technique is often used in forest nurseries (30–60 day moist +5°C treatment).

Figure 1.9. Yellow male flowers clustered at the top of the shoot (left) and yellow-brown maturing cone (right) of *Pinus nigra*.

### 1.4.2 Distribution

The distribution of black pine populations particularly in the Mediterranean region, have faced serious consequences of climate warming and increased human activities. This resulted to the current fragmented distribution of black pine extending from North-Western Africa through southern Europe to Asia Minor (Isajev *et al.*, 2004; Enescu *et al.*, 2016). In particular, the discontinuous range of *P. nigra* spans across SW, S and SE Europe; N Algeria; N Morocco; Cyprus; Turkey; from the Krym (Crimea) in Ukraine along the Black Sea coast eastwards to Krasnodar in the Caucasus (Fig. 1.10).

Currently, black pine covers a large expanse of over 3.5 million hectares (Isajev *et al.*, 2004), making it one of the most widespread conifer species in the Balkans and Asia Minor. Its widest distribution worldwide is in Turkey, with more than 2.5 million hectares (Enescu *et al.*, 2016). Outside Europe, it has become naturalized in the midwestern states of the U.S, normally south of the normal native ranges of native pines, where it is known as Austrian pine and also in northern states in New England, around the Great Lakes and in the Northwest (Enescu *et al.*, 2016). As a result of the threats arising from climate warming, the future distribution of black pine is thought to change considerably but with varying response that depends on

the geographic region. In Mediterranean regions, climate warming increases water stress and thus has a negative influence on the growth of this species, whereas in central Europe climate amelioration is thought to lead to an expansion (Enescu *et al.*, 2016).

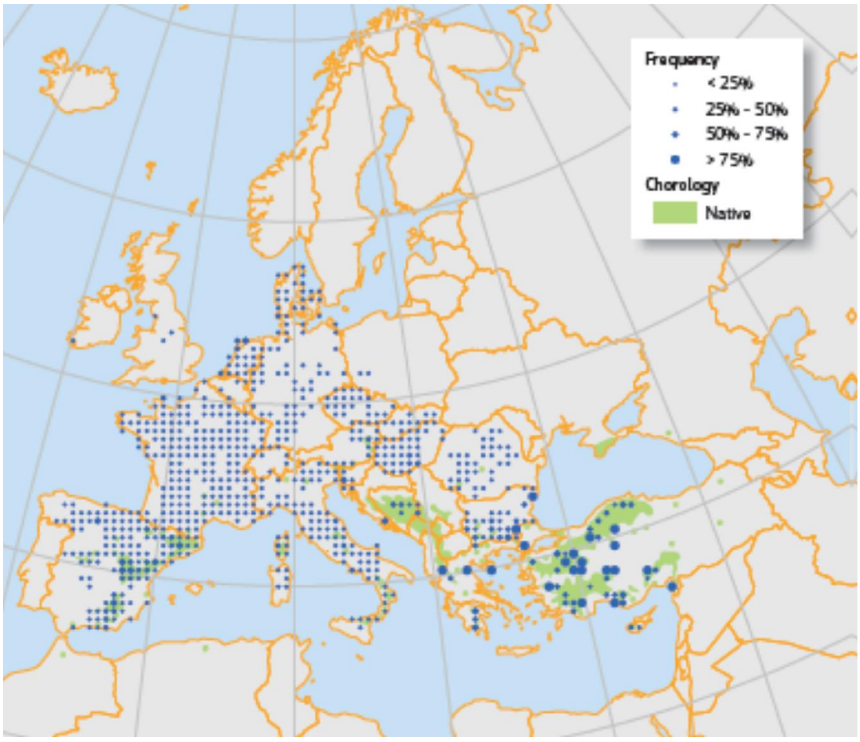



Figure 1.10. Plot distribution and simplified chorology map for *Pinus nigra*. Frequency of *Pinus nigra* occurrences within the field observations as reported by the National Forest Inventories. The chorology of the native spatial range for *P. nigra* is derived after EUFORGEN (Enescu *et al.*, 2016).

### 1.4.3 Taxonomy

The fragmented distribution of black pine has led to morphological variations, which are difficult to interpret and have resulted in several diverse classifications. As previously reported, relatively little is known about its long history, although in the Tertiary black pine was more widespread than today and it seems to have shifted over time from coastal areas to its current mountain locations, since the dry cold climate of these areas resembles that of bygone glacial periods (Naydenov *et al.*, 2006). Migratory movements, which occurred during interglacial warm periods, brought the emergence of various hybrid populations that later became genetically isolated, thus contributing to its difficult taxonomic characterization (Afzal-Rafii and

Dodd, 2007). Because of that, the systematic subdivision of this species is very controversial and different authors group the different geographically distinct ecotypes to the rank of subspecies or varieties.

Some authors (Christensen, 1993; Gymnosperm Database) divided *P. nigra* in two subspecies: *Pinus nigra* subsp. *salzmannii* and *Pinus nigra* subsp. *nigra*. Each subspecies is further subdivided into different varieties.

*Pinus nigra* subsp. *salzmannii* is distributed in the east of the range from Morocco and Spain to South France and Corsica. This subspecies includes three varieties:

*P. nigra* subsp. *salzmannii* var. *salzmannii* (Pyreneean pine), located in Pyrenees, from the Southern France to the Northern Spain;

*P. nigra* subsp. *salzmannii* var. *corsicana* (Corsican pine), occurring in Corsica, Sicily, and Southern Italy;

*P. nigra* subsp. *salzmannii* var. *mauretanica* (Atlas Mountains black pine), occurring in Morocco and Algeria.

*P. nigra* subsp. *nigra* occurred in the east of the range, from Austria, northeast and central Italy through Balkans up to Turkey and Crimea Peninsula. This subspecies also includes three varieties:

*P. nigra* subsp. *nigra* var. *nigra* (Austrian pine), located in Austria and Balkans (except southern Greece);

*P. nigra* subsp. *nigra* var. *pallasiana* (Crimean pine), occurring in Crimea;

*P. nigra* subsp. *nigra* var. *caramanica* (Turkish black pine) located in Turkey, Cyprus and southern Greece.

According to the recent classification of Euro+Med Plantbase (Raab-Straube, 2014), several authors recognize six main subspecies of black pine (Isajev *et al.*, 2004; Caudullo *et al.*, 2017): *Pinus nigra* subsp. m*auretanica, Pinus nigra* subsp. s*alzmannii, Pinus nigra* subsp. *laricio, Pinus nigra* subsp. *nigra, Pinus nigra* subsp. *dalmatica and Pinus nigra* subsp. *pallasiana* (Fig. 1.11).

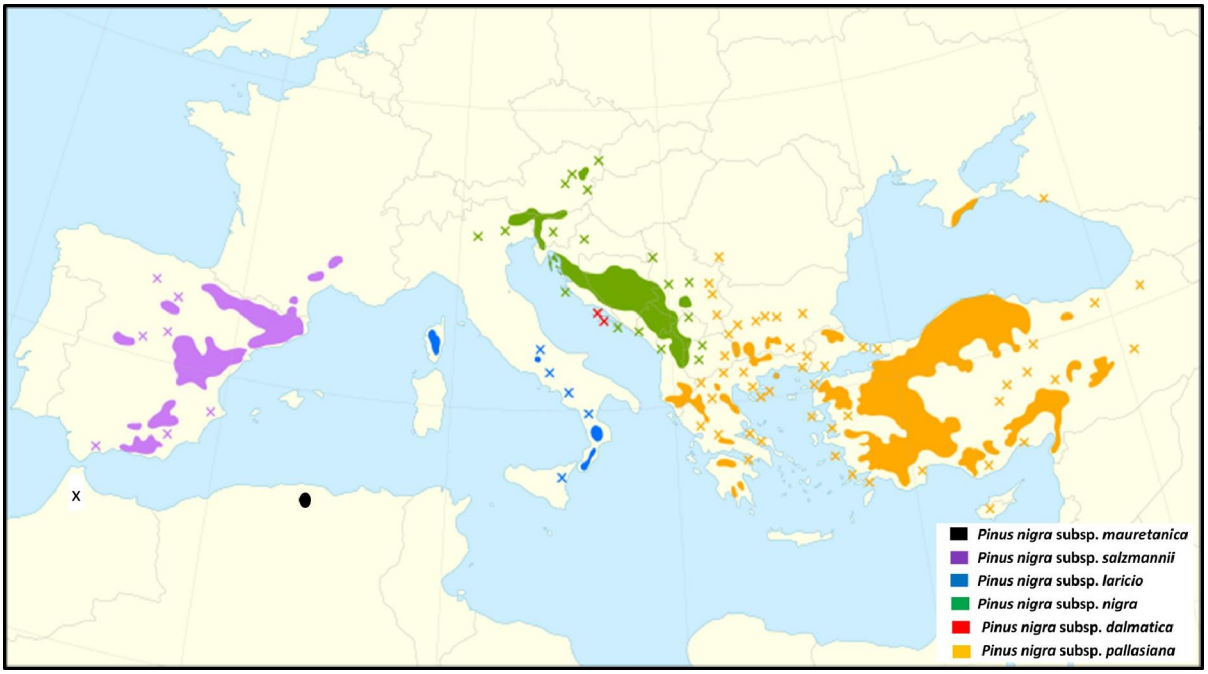



Figure 1.11. Native distribution of *Pinus nigra* subspecies according to the classification of Euro+Med Plantbase (modified from Caudullo *et al.*, 2017).

*P. nigra* subsp. *mauretanica* (Maire et Peyerimh.) Heywood covers only a few hectares in the Rif mountains of Morocco and the Djurdjura mountains of Algeria.

*Pinus nigra* subsp. *salzmannii* (Dunal) Franco covers extensive areas in Spain (over 350,000 ha from Andalucia to Catalunia and on the southern slopes of the Pyrenees) and is found in a few isolated populations in the Pyrenees and Cévennes in France. It is sometimes referred to as the Pyrenean pine.

*Pinus nigra* subsp. *laricio* (Poiret) Maire is found in Corsica (referred to as Corsican pine) covers over 22000 ha, in Calabria (where it is also recognized as the Calabrian pine) and in Sicily.

*Pinus nigra* subsp. *nigra* (syn: *P. nigra austriaca* Höss, *P. nigra nigricans* Host, the Austrian pine) is found from Italy, in the Apennines, to northern Greece through the Julian Alps and the Balkan mountains, covering more than 800,000 ha.

*Pinus nigra* subsp. *dalmatica* (Vis.) Franco, the Dalmatian pine, is found on a few islands off the coast of Croatia and on the southern slopes of the Dinaric Alps.

*Pinus nigra* subsp. *pallasiana* (Lamb.) Holmboe covers extensive areas, mostly in Greece and Turkey (2.5 million ha, 8% of total forest area) and possibly as far west

as Bulgaria. It can also be found in Cyprus and the Crimea. It is sometimes referred to as the Crimean pine.

### 1.4.4 Habitat and ecology

Most black pine subspecies (see Distribution and Taxonomy) grow in a Mediterranean-type climate, except *P. nigra* subsp. *nigra* which is more typically temperate. Bioclimatic conditions range from humid (800–1000 mm annual rainfall) as in subspecies *mauretanica* and *laricio*, to sub-humid (600–800 mm) as in subsp. *pallasiana* in Cyprus, to semi-arid (400–600 mm) as in subsp. *pallasiana* in Anatolia. European black pine stands is found at elevations ranging from 350 m in Italy to 2200 m in the Taurus mountains, the optimal altitudinal range being between 800 to 1500 m (Isajev *et al.*, 2004; Enescu *et al*., 2016). They grow on a wide variety of soil types, from podzolic sands to limestone, often dependent on geographic region and climate (Farjon, 2013). The subsp. *nigra* (Austrian pine) tolerates better, exposed chalk and limestone than subsp. *laricio* (Corsican pine) (Enescu *et al.*, 2016). However, Corsican pine is more often found in coastal areas due to its relatively higher resistance to salt wind than most other pine species (Farjon, 2013). Black pine is a light-demanding species, but it shows higher shade tolerance than Scots pine (*Pinus sylvestris*) (Isajev *et al.*, 2004; Enescu *et al*., 2016). Moreover, it is more resistant to drought and wind than other *Pinus* species (Isajev *et al.*, 2004). It grows in pure stands or in association with other broadleaved or conifer species, in particular *Pinus sylvestris* (Enescu *et al.*, 2016).

### 1.4.5 Importance and use

Black pine is an important timber-producing tree in southern Europe, especially that from Corsica. The wood of black pine is moderately hard and straight-grained, rich in resin and easy to process. It is used extensively throughout the Mediterranean region for general construction, fuelwood, and other purposes for which pine timber is needed. As a result of its climatic adaptation and growth performance on a wide variety of soils, it is one of the most widely used tree species for reforestation worldwide (Isajev *et al.*, 2004; Enescu *et al*., 2016) and it is considered a good prospect among indigenous coniferous species in Central Europe under future climate scenarios (Thiel *et al.*, 2012)

Black pine have been planted extensively in cold, semi-arid, exposed coastal regions for protection and sand dune fixation because of its capacity to withstand drought,

to grow on light, dry sandy soils of low productivity, and to tolerate fill (Dallimore and Bruce Jackson, 1966). Despite its relatively narrow native range, the broad European distribution range of black pine covers several areas with high erosion rates such as the European mountain systems. This frost-hardy, wind firm, and light-demanding species have also been widely used for nearly 100 years in windbreaks and roadside plantings throughout the eastern Great Plains of the United States, where its dense foliage and stiff branches withstand wind, ice, and heavy snow. In some part of the United States, it was reported as having good potential for revegetating sites denuded by heavy metal pollution from smelter emissions (Caborn, 1965).

Black pine, in particular the subsp. *nigra*, is also valued for landscaping, both in parks (isolated trees or in groups) and in urban and industrial areas because of its tolerance to pollution (Isajev *et al.*, 2004; Farjon, 2013; Enescu *et al.*, 2016). It is one of the most common introduced ornamentals in the USA and Canada.

### 1.4.6 Genetic diversity

As previously reported (see distribution and taxonomy) the ice cycles that shaped the Quaternary period in Europe are believed to have been responsible for the currently very discontinuous range of black pine. This geographic separation did not result in mating barriers, and all subspecies are interfertile under experimental conditions (Isajev *et al.*, 2004). Studies using morphological and genetic markers have confirmed the common phylogenetic origin of all black pines (Vidakovic, 1974; Nikolic and Tucik, 1983; Lauranson-Broyer and Lebreton, 1995). The most divergent and genetically original European groups are *P. nigra* subsp. *salzmanii* and *P. nigra* subsp. *laricio*, with the subspecies *nigra*, *dalmatica* and *pallasiana* that appear quite similar (Nikolic and Tucik, 1983; Lauranson-Broyer and Lebreton, 1995). The amount of genetic diversity is also high within populations (Nikolic and Tucik, 1983). Experiments measuring adaptive traits have revealed strong within- and among-population variability for traits such as vigour, form and drought, frost and disease resistance (Isajev *et al.*, 2004). It is this huge adaptive plasticity that has made black pine such a favourite for reforestation projects over a wide range of environments.

Intraspecific hybridization is easily performed among all black pine subspecies (a further proof of phylogenetic relatedness), but has not contributed any outstanding genotypes to breeding programmes so far. Interspecific crosses seem to be possible at a low survival rate with *P. sylvestris* (Isajev *et al.*, 2004).

Black pine is not recognized as a threatened species (Farjon, 2013), although some of its sub-Mediterranean endemic populations constitute priority habitats under the EU Natura 2000 directive (Habitat Directive n° 92/43/CEE, May 21st, 1992). The only subspecies under threat is P*. nigra* subsp. *dalmatica*, which occupies a tiny part of the entire range of the species, precisely in Croatia with a small and declining population (see taxonomy) (Farjon, 2013).

As previously reported (see importance and use), extensive plantations were often made across Europe in the past two centuries with material from unknown and/or very distant sources for which no historical traces currently exist. This has probably resulted in the extensive mixing of local and imported gene pool all over the distribution area of black pine. Therefore, in areas where *P. nigra* occurs in small isolated populations, major risks come from any factor that may provoke local extinction, either through illegal cutting and fires or through hybridization ('genetic pollution') from planted black pines belonging to other subspecies.

### 1.5 *Pinus nigra* subsp. *laricio* in Calabria

As one of the six recognized subspecies of *Pinus nigra* (see taxonomy), the laricio pine (*Pinus nigra* subsp. *Laricio* (Poiret) Maire) includes 2 varieties: Corsican pine (var. *Corsicana*) and Calabrian pine (var. *Calabrica*) (Debazac, 1965). Corsican pine which is the dominant species naturally exist between 1000 and 1800 m a.s.l. on the mountains of Corsica, where it covers approximately 45000 ha, of which 21000 ha are pure Corsican pine stands (Nicolaci *et al.*, 2014). Calabrian pine is endemic to southern Italy with a natural range extending from Calabria to Sicily. In Italy, it is known also in a small area of Tuscany in the province of Pisa that hosts a formation of *Pinus nigra* subsp. *laricio*, which probably originated from a medieval program of reforestation. In addition, it is worth noting their presence in the Abruzzo National Park of a transitional entity that comprises between *laricio* and *nigra* subspecies called "Villetta Barrea's black pine" (or *Pinus nigra* subsp. *nigra* var. *Italica*).

At present, there is no evidence of decline of the two varieties of the subsp. *laricio*, although undoubtedly, stands have been logged and replaced by other land use or woodland types in the past (Farjon, 2013). Several major stands are within protected areas representing the three major locations Corsica, Calabria and Sicily.

Calabrian pine grows in Sicily in fragmented areas on the slopes of Mount Etna between 1000 and 2000 m a.s.l., covering approximately 3000 ha (Poli Marchese,

1982; Barreca *et al.*, 2009)**.** In Calabria, it grows on the Sila and Aspromonte mountains, covering approximately 114000 ha, of which more than 50% are pure stands of both natural and artificial origin (Nicolaci *et al.*, 2015). The latter originated from extensive reforestation projects carried out between 1950 and 1970 following specific State laws (Iovino and Menguzzato, 2002). The largest area covered by Calabrian pine is in the Sila mountain range (about 80,000 ha), where this species characterizes the forest landscape.

Together with soil conservation and watershed protection, Calabrian pine has an important role in the local forest economy. In public properties (townships and State forests), management of Calabrian pine has usually been based on various types of clear felling (strip or patch), whereas on private properties, pine stands have generally been managed according to traditional and locally developed forms of selection cutting, which have contributed to the maintenance of pure pine stands with complex structures (Ciancio *et al.*, 2006). More recently, in State forests within the Sila National Park, management has been limited to felling only dead or dying trees. In these forests, and in other areas where for various reasons active management of Calabrian pine stands has stopped, broadleaved trees are spreading under the older pine trees (Iovino and Menguzzato, 1996). Thus, the primary environmental factors determining the occurrence and competitiveness of Calabrian pine are strongly influenced by anthropogenic activity.

## 2. OBJECTIVES

In the framework of the AlForLab (public-private laboratory for the Environment-Wood-Forest Chain) project, we have been carrying out the first attempt to our knowledge, to gain insight into the ecological and functional roles of terpenes from *P. nigra* subsp. *laricio*, an endemic coniferous species in Calabria. Since a preliminary study have identified several monoterpenes such as (−)-β-pinene, (−)-α-pinene, (+)-α-pinene, (+)-3-carene, and (−)-β-phellandrene as the most abundant among terpenoids in the needles of *P. laricio*, the main objective of the thesis was the isolation and characterization of cDNA sequences encoding MTPSs potentially involved in the synthesis of the monoterpenes in this species. The strategy adopted was based on PCR amplification of cDNA sequences by using specific primers designed on conserved regions of pine MTPS belonging to distinct phylogenetic groups.

In particular, the research activity concerned the following phases:

1) Identification of the putative sequences of the genes coding for TPSs in *Pinus* species by BLAST searches in the National Center for Biotechnology Information (NCBI) database;
2) Phylogenetic analysis of the identified sequences to better understand the evolution of the family of TPSs in the *Pinus* genus;
3) Identification of distinct phylogenetic groups in the TPS-d1 clade containing mainly MTPS;
4) Isolation of MTPS *P. laricio* cDNA and genomic sequences using specific primers designed on conserved regions of pine members of each identified phylogenetic group in the TPS-d1 clade;
5) Analysis and characterization of the predicted proteins from the isolated sequences.

# 3. MATERIALS AND METHODS

## 3.1 Plant material

The sampling of needles from *Pinus nigra* subsp. *laricio* was carried out on 10/10/2015 within the Bonis basin, near Acri (CS) in Calabria (Fig. 3.1). This is an area located in the so-called "Sila Greca Cosentina" with an extension of 139 hectares. The site was identified, designed and equipped in 1986 with the aim of establishing a permanent laboratory for the study of the hydrological balance in forest populations in relation also to forest management and possible disturbances. Approximately 80% of the area of the basin is occupied by *Pinus laricio* stands, most of artificial origin, made in the period between 1955 and 1970 (Fig. 3.2).

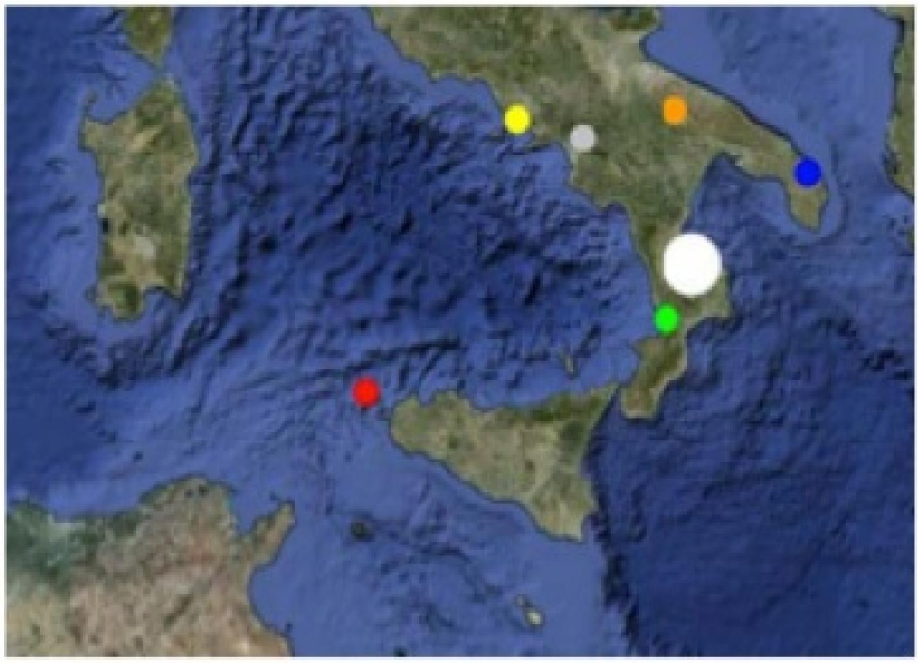


Figure 3.1. Geographic location of Bonis basin (large white circle) in Calabria.

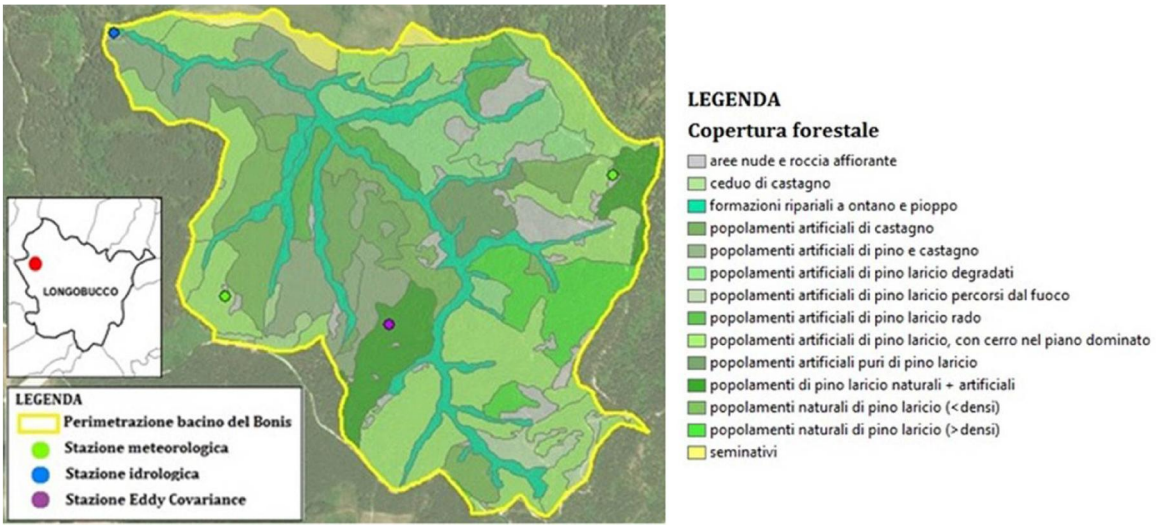



Figure 3.2. Bonis basin with forest cover and location of the Eddy Covariance station (Collalti *et al.*, 2017).

The needle samples were collected from five individuals located in the area near the Eddy Covariance Station (Fig. 3.2). The selected plants indicated by the letters A, B, C, D, and E had height of 10-15 m and were placed at a distance of 1-6 m from the dirt road (Fig. 3.3). For each plant, three needle samples were collected (3-6 g for each sample) on branches located between 2.5 and 4 m of height. The needle samples were immediately frozen in liquid nitrogen and stored at -80 ° C until use for DNA and RNA isolation.

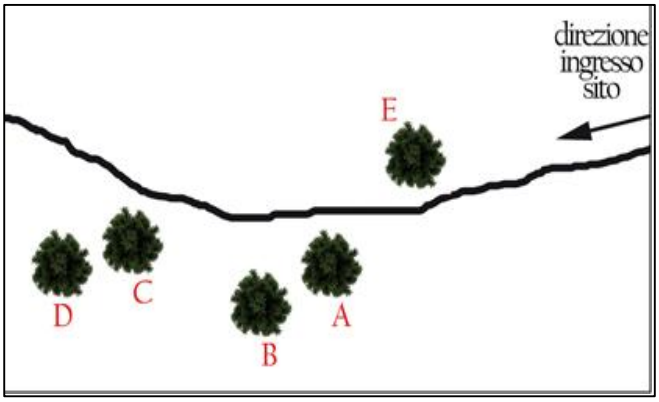



Figure 3.3. Schematic representation of the location of the selected plants for needles sampling.

### 3.2 DNA extraction

Total genomic DNA was extracted using NucleoSpin® Plant II kit (Macherey-Nagel, Düren, Germany) according to the manufacturer's instructions. The integrity and concentration of DNA were determined by 0.8 (w/v) agarose gels stained with ethidium bromide (0.001%) using known concentrations of unrestricted lambda DNA as control. All DNA samples were stored at -20 °C until use.

### 3.3 RNA extraction and cDNA preparation

Total RNA was extracted from needles, following the method described by Gambino *et al.*, (2008), with some modifications. Needle samples (250 mg) were ground in a mortar with liquid nitrogen, and immediately transferred to a microcentrifuge tube containing 900 µL of pre-warmed (65 °C) extraction buffer (2% cetyltrimethyl ammonium bromide [CTAB]; 2.5% PVP-40; 2 M NaCl; 100 mMTris-HCl, pH 8.0; 25 mM EDTA, pH 8.0; and 2% β-mercaptoethanol), vortexed for 2 min and incubated for 10 min at 65 °C. An equal volume of chloroform:isoamyl alcohol (24:1 v/v) was added, and the tube was inverted vigorously and centrifuged at 11,000 g for 10 min at 4 °C. The supernatant was recovered and a second extraction with chloroform:isoamyl alcohol was performed. The supernatant was then transferred to a new micro-centrifuge tube, and LiCl (3 M final concentration) was added to the mixture, which was left overnight at 4 °C. The RNA was precipitated by centrifugation at 21,000 g for 30 min at 4 °C. The pellet was resuspended in 500 µL of SSTE buffer (1 M NaCl; 1% SDS; 10 mMTris-HCl, pH 8.0; 1 mM EDTA, pH 8.0) and pre-heated at 65 °C. An equal volume of chloroform:isoamyl alcohol was added and the mixture was centrifuged at 11,000 g for 10 min at 4 °C. The supernatant was transferred to a new microcentrifuge tube and the RNA was precipitated with 0.7 volumes of cold isopropanol, and immediately centrifuged at 21,000 g for 15 min at 4 °C. The pellet was washed with ethanol (70%), dried and re-suspended in 100 µL of 0.1% diethyl pyrocarbonate (DEPC)-treated sterile water.

The RNA samples were treated with RNase-free DNase I (Promega, Madison, WI, USA), according to the manufacturer's protocol. Following digestion, nucleotides were removed from RNA using a G50 Sepharose buffer exchange column (Amersham, Pittsburgh, PA, USA). The RNA concentration and integrity were checked, using a Nano Drop ND-1000 spectrophotometer (Lab tech, East Sussex, UK). Only RNA samples with a 260/280 ratio (an index of protein contamination)

between 1.9 and 2.1, and a 260/230 ratio (an index of reagent contamination) greater than 2.0, were used for cDNA synthesis. The quality of RNA samples was also assessed by electrophoresis on 1% formaldehyde agarose gels.

First-strand cDNA was synthesized from 3 µg of total RNA using Expand Reverse Transcriptase (Roche Diagnostics, Milano, Italy), according to the manufacturer's protocol, and the resulting cDNA was used for PCR analyses (see chapter 3.6).

### 3.4 Identification of TPS gene sequences belonging to *Pinus* species

The putative sequences of the genes coding for mono-, sesqui- and di-terpene synthases (MTPSs, STPSs and DTPSs, respectively), and for ent-copalyl diphosphate- and ent-kaurene-synthases (CPS and KS) in *Pinus* species were identified by a BLAST search in the National Center for Biotechnology Information (NCBI) database, using the available and functionally characterized sequences of TPSs from different conifer species (Table 3.1). The search was restricted to the TPSs sequences in NCBI database that correspond to the taxid:3337 (*Pinus*). For each identified gene, the corresponding mRNA and protein sequences were retrieved (see Table 4.1 in the Results and discussion section).

### 3.5 Alignment and phylogenetic analysis of the identified sequences

A multiple sequence alignment of the identified TPS deduced proteins in Pinus species (see Table 4.1 in the Results and discussion section) was performed by ClustalX version 1.83 (Thompson *et al.*, 1997), using the Gonnet series as protein weight matrix and parameters set to 10 gap open penalty, 0.2 gap extension penalty, negative matrix on and divergent sequences delay at 30%. The ent-kaurene synthase from *Physcomitrella patens* (BAF61135) was also included in the analysis as outgroup. A phylogenetic tree was generated with the Neighbor-Joining method (Saitou and Nei, 1987) using MEGA7 software (Kumar *et al.*, 2016). The evolutionary distances were computed using the JTT matrix-based method and are in the units of the number of amino acid substitutions per site. The rate variation among sites was modeled with a gamma distribution (shape parameter = 1). Reliability of the tree obtained was tested using bootstrapping with 1000 replicates.

Two different phylogenetic trees were computed using amino acid sequences of 1) The whole set of the 93 TPSs (MTPS, STPS, DTPS, CPS and Ks) identified by the BLAST search in NCBI database in *Pinus* species and the ent-kaurene synthase from *P.*

*patens* as outgroup; 2) A set including only the 74 MTPS from *Pinus* species and the ent-kaurene synthase from *P. patens* as outgroup.

The protein sequences belonging to the different groups identified by the phylogenetic analysis of the 72 MTPS from Pinus species were aligned in order to identify highly conserved regions that were used to design specific primers for the isolation of partial transcripts coding for mono-TPSs in *Pinus nigra* subsp. *laricio* (see the Results and discussion section).

Table 3.1. Functionally characterized TPSs from different conifer species used in the BLAST search in the NCBI database

| Species | Function | Type of TPS | Accession | Reference |
|---|---|---|---|---|
| *Pinus contorta* | (+)-3-carene synthase | MTPS | JQ240307 | Hall et al (2013a) |
| | (–)-β-phellandrene synthase | MTPS | JQ240301 | Hall et al (2013a) |
| | (-)-β-pinene synthase | MTPS | JQ240293 | Hall et al (2013a) |
| | levopimaradiene/abietadiene synthase | DTPS | JQ240310 | Hall et al (2013b) |
| | monofunctional diterpene synthase | DTPS | JQ240318 | Hall et al (2013b) |
| | monofunctional isopimaradiene synthase | DTPS | JQ240314 | Hall et al (2013b) |
| *Pinus sylvestris* | longifolene synthase | STPS | EF679332 | Köpke et al (2008) |
| | β-farnese synthase | STPS | GU248335 | Köpke et al (2008) |
| *Pinus taeda* | (-)-α-pinene synthase | MTPS | AF543527 | Phillips et al (2003) |
| | α-terpineol synthase | MTPS | AF543529 | Phillips et al (2003) |
| *Picea abies* | E,E- α-farnese synthase | STPS | AY473627 | Martin et al (2004) |
| | E- α-bisabolene synthase | STPS | AY473619 | Martin et al (2004) |
| | (-)-limonene synthase | MTPS | AY473624 | Martin et al (2004) |
| | Isopimara-7,15-diene synthase | DTPS | AY473620 | Martin et al (2004) |
| *Picea glauca* | Copalyl diphosphate synthase | CPS | ACY25274 | Keeling et al (2010) |
| | ent-kaurene synthase | KS | ACY25275 | Keeling et al (2010) |
| | (–)-linalool synthase | MTPS | ADZ45500 | Keeling et al (2010) |
| | α-humulene synthase | STPS | HQ42615 | Keeling et al (2010) |

MTPS= mono-terpene synthase, STPS= sesqui-terpene synthase, DTPS= di-terpene synthase, CPS= ent-copalyl diphosphate, KS= ent-kaurene synthase

### 3.6 Amplification of partial cDNAs coding for monoterpene synthases in *Pinus nigra* subsp. *laricio*

RT-Polymerase chain reaction (PCR) was used to amplify partial cDNAs coding for MTPS in *P. laricio*, using forward and reverse primers designed in conserved regions among MTPS sequences of *Pinus* species of different groups identified by the phylogenetic analysis (see tables 4.3 in the Results and discussion section).

PCR reactions were performed in a total volume of 50 µl containing 2 µL of RT reaction (see chapter 3.3), 0.4 µM of each forward and reverse primer and 25 µl of UPTA$^{TM}$ TaqPCR Master Mix, 2× (Biotechrabbit, Hennigsdorf, Germany) which includes pure biotechrabbit UTpa*Taq* DNA Polymerase, dNTPs and optimized PCR buffer. All reactions were carried out in an Eppendorf Thermal Cycler (Master cycler Gradient) with the following parameters: initial denaturation at 95 °C for 5 min, 35 cycles of amplification, each at 95 °C for 1 min, 58-62 °C (depending on the annealing temperature of the primers) for 1 min, 72 °C for 3 min, and a final extension at 72 °C for 5 min.

### 3.7 Rapid Amplification of cDNA of either 5’ and 3’ Ends (5’/3’ RACE) of partial transcripts coding for monoterpene synthases in *Pinus nigra* subsp. *laricio*

The partial MTPS cDNAs from *P. laricio* were used as templates for 5' and 3' RACE (Rapid Amplification of cDNA Ends) extensions using the 5'/3' RACE kit from ROCHE following manufacturer's instructions. The 5’/3’ RACE kit contains Transcriptor Reverse Transcriptase and recombinant Terminal Transferase. Transcriptor Reverse Transcriptase transcribes full-length cDNA for the highly sensitive and rapid amplification of either 5’ or 3’ cDNA fragments up to 14 kb and, due to its thermostability (up to +65 °C) to work with GC-rich templates with high secondary structure. High sensitivity can be achieved using Transcriptor Reverse Transcriptase, resulting in highly efficient cDNA synthesis and the generation of long RACE products. Recombinant Terminal transferase is used to add a homopolymeric A-tail to the 3’ end of the cDNA. The poly(A)$^{+}$ tail decreases the likelihood of inappropriate truncation by the oligo(dT)-anchor primer and overcomes the weaker A/T compared to the G/C hybridization. Moreover, long stretches of A residues are required before the oligo(dT)-anchor primer can hybridize to an internal site and can truncate the amplification product. Tailed cDNA is amplified by PCR using a gene-specific primer and the oligo(dT)-anchor primer (5’-GACCACGCGTATCGATGTGTCGACTTTTTTTTTTTT TTTTV-3’, where V = A, C, or G). The obtained cDNA is further amplified by a second

PCR using a nested specific primer and the PCR-anchor primer 5'–GACCACGCGTATCGATGTGTCGA-3', allowing RACE products to be cloned into an appropriate vector for subsequent studies. In the following paragraphs, the two protocols for 5' and 3' RACE are described in detail, while their general overviews are reported in the Figures 3.4 and 3.5.

### 3.7.1 Experimental protocol for 5' RACE

This protocol covers consecutively the procedures for the synthesis of first-strand cDNA, poly(A) tailing of first-strand cDNAs and PCR amplification of dA-tailed cDNA. The synthesis of first-strand cDNAs starts with preparing a mixture with a total volume of 20 µl containing: cDNA synthesis buffer (4 µl), dNTP mixture (2 µl), 2 µg of total RNA (10 µl), Transcriptor reverse transcriptase (1 µl), 12.5 µM of specific primer (SP1) (1µl) and double-distilled water (2 µl), 10 µl; This mixture was incubated first for 60 min at +55$^{O}$C, then for 5 min at 85$^{O}$C and briefly spin down before purification by the High Pure PCR Product Purification Kit (Roche). To add the homopolymeric A-tail to the 3' end of the newly synthesized first-strand cDNA, 19 µl of the purified cDNA sample was used to prepare a reaction mixture containing the reaction buffer (2.5 µl) and 2 mM dATP (2.5 µl). This reaction mixture was incubated for 3 min at 94$^{o}$C, immediately chilled on ice, mixed with 1 µl of terminal transferase (80 U/ µl), and then incubated again at 70$^{o}$C for 20 min.

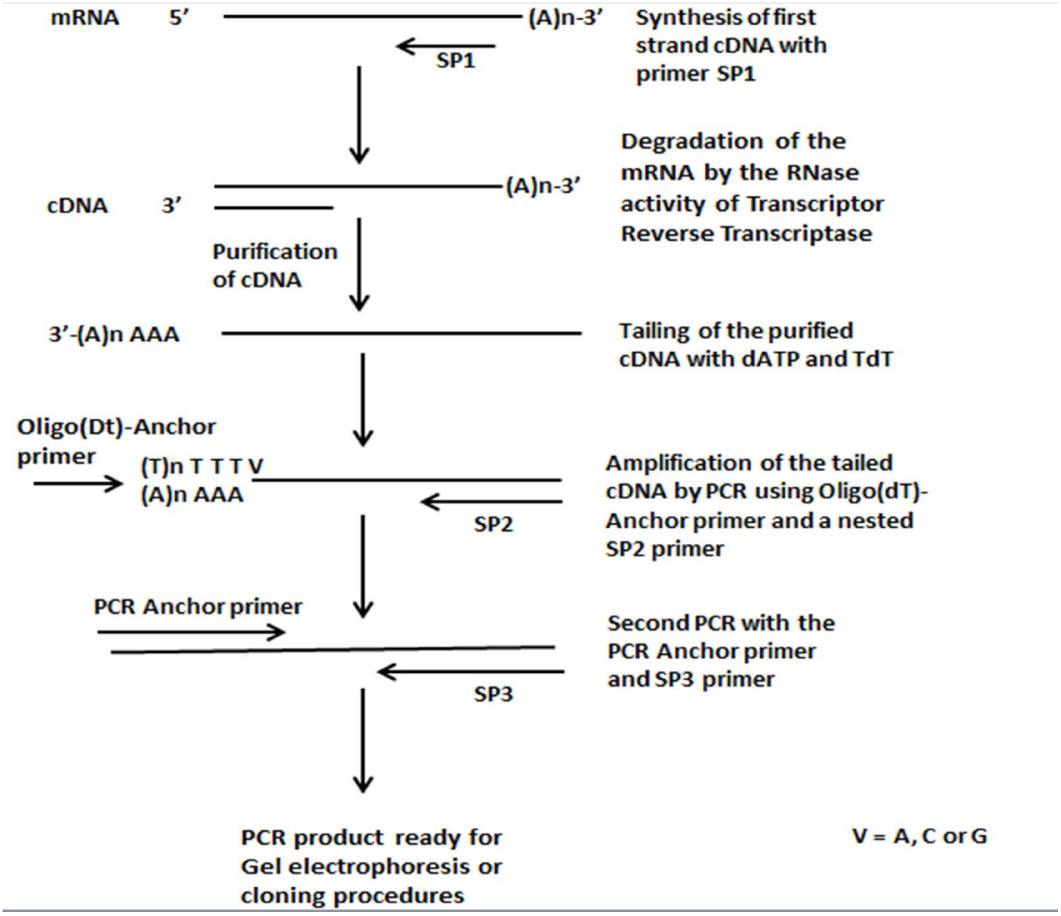



Figure 3.4. Overview of 5' RACE

Amplification of the dA-tailed cDNA was carried out in a reaction mixture with a total volume of 50 µl containing: dA-tailed cDNA (5 µl), Oligo dT-Anchor primer (1 µl), 12.5 µM of specific primer SP2 (1 µl), dNTP mixture (1 µl), 1 µl of Expand Long Template enzyme mix (ROCHE), 5 µl of Expand Long Template PCR System buffer, and 36 µl of double-distilled water. PCR reactions were carried out in an Eppendorf Thermal Cycler (Mastercycler Gradient) with the following parameters: initial denaturation at 94 °C for 2 min, 35 cycles of amplification, each at 94 °C for 15 sec, 58-62 °C for 30 sec (depending on the optimal annealing temperature of the different primer employed), 68 °C for 1 min, and a final elongation at 68 °C for 7 min. Ten µl of the first amplification reaction were diluted to 1:20 in double-distilled water, and 1 µl was amplified using a nested specific primer (SP3) and the PCR-anchor primer with the same protocol described above. The specific primers (SP1-SP3) used for the 5' extension of each *Pinus laricio* MTPS partial transcript are reported in table 4.4 in the Results and discussion section.

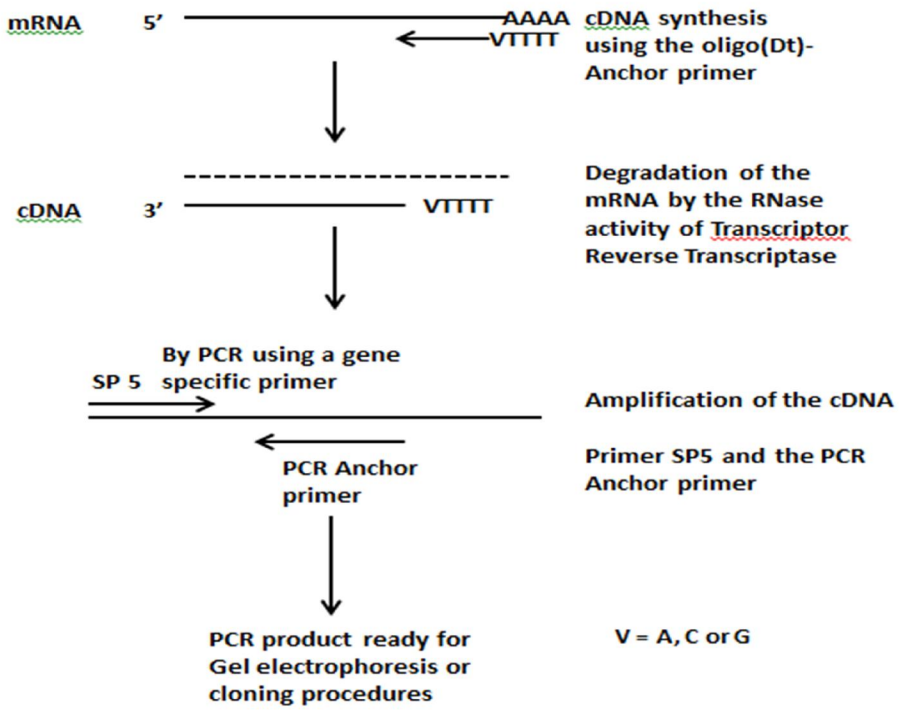



Figure 3.5. Overview of 3' RACE.

### 3.7.2 Experimental protocol for 3' RACE

This protocol covers the synthesis of the first-strand cDNA and its specific amplification by PCR. The synthesis of first-strand cDNAs is carried out in a mixture with a total volume of 20µl containing: 4 µl of cDNA synthesis buffer, 2 µl of dNTP mixture, 1 µl of oligo dT-anchor primer, 2 µg of total RNA (10 µl), 1 µl of Transcriptor reverse transcriptase, and 2 µl of double-distilled water. This mixture was incubated first for 60 min at 55$^{O}$C, then for 5 min at 85$^{O}$C. Amplification of the synthesized cDNA was carried out in a reaction mixture with a total volume of 50 µl containing: 1 µl of cDNA product, 1 µl of PCR Anchor primer, 12 µM of specific primer (SP5), 1 µl of dNTP mixture, 1 µl of Expand Long Template enzyme mix, 5 µl of Expand Long Template PCR System buffer, and 40 µl of double-distilled water. PCR conditions were the same as described in section 3.7.1. The specific primer (SP5) used for the 3' extension of each *Pinus laricio* MTPS partial transcript are reported in table 4.4 in the Results and discussion section.

### 3.8 Amplification of genomic DNA

When it was not possible to obtain the RT-PCR products for the partial transcripts of the *Pinus laricio* MTPS, the specific primers designed in conserved regions among MTPS sequences of *Pinus* species of different groups identified by the phylogenetic analysis were used to amplify the genomic DNA. The PCR reactions and conditions were the same as described in section 3.6.

### 3.9 Cloning and sequencing of cDNA, RACE and genomic amplification products

Samples (5-10 µl) of the amplification products of RACE, partial cDNAs and genomic DNA were separated on 1.2 % (w/v) agarose gels and visualized under UV radiation after staining with ethidium bromide (0.001%) and analyzed using the UVITEC Essential V6 Gel Imaging and Documentation System (Cleaver Scientific, Rugby, United Kingdom).

PCR products of expected size were excised from the gel, purified using the High Pure Purification kit (ROCHE) according to manufacturer's instructions, and cloned into the pGEM-T easy plasmid vector (Promega, Madison, WI, USA) following the manufacturer's instructions.

For the transformation, competent cells of *Escherichia coli* (DH5α strain) were used. The transformed cells were plated on LB (Luria Bertani) culture medium containing ampicillin (100 µg / ml), necessary to select cells containing the plasmid, since the latter carries an Ampr gene for resistance to the antibiotic. IPTG and X-gal, a chromogen analogous to galactose, were added to the LB culture medium, which allowed the selection of cells containing recombinant plasmids by screening for white/blue staining. From two to five positively transformed colonies (white) were picked for each cloned fragment. The extracted and purified recombinant plasmids were digested with the EcoRI and NotI enzymes, whose restriction sites are present at the two ends of the plasmid vector pGEM-T Easy Vector, and separated on an agarose gel to evaluate the size of the inserted fragments.

Plasmid DNA for sequencing reaction was prepared from 3 ml overnight cultures using a plasmid mini-prep kit (QIAGEN). Sequencing was performed by a private company (MWG, Biotech AG, Germany). Recombinant positive plasmids were sequenced on both strands by the ABI PRISM 377 capillary sequencer (PE Applied Biosystem) using an ABI Prism Dye Terminator sequencing kit (PE Applied Biosystem) and vector-specific primers (SP6 and T7).

### 3.10 Analysis of the nucleotide and deduced amino acid sequences

All sequences were analyzed by DNAMAN Sequence Analysis Software (Version 3, LynnonBiosoft) and their homologies were scored using the BLASTX program through the (NCBI) database. The predicted protein sequences were analyzed by searching for conserved motifs in CDD (Conserved Domain Database in the NCBI) and SMART (Simple Modular Architecture Research Tool, EMBL, UniversitatHeidelberg) databases; their subcellular locations were predicted by Target P1.1 [TargetP 1.1 Server (http://www.cbs.dtu.dk/services/TargetP/)], Chloro P1.1, [ChloroP 1.1 Server (http://www.cbs.dtu.dk/services/ChloroP/)], and Predotar (https://urgi.versailles.inra.fr/predotar/predotar.html).

## 4. RESULTS AND DISCUSSIONS

### 4.1 Identification and phylogeny of TPSs gene sequences in *Pinus* species

The identification of full length (FL) cDNAs coding for putative terpene synthases (TPSs) in *Pinus* species was based on the BLAST search of the NCBI database using selected and functionally characterized TPSs from different conifer species (Table 3.1). Several BLAST searches allowed the identification of 93 FL cDNA sequences coding for putative TPSs from 28 different *Pinus* species (Table 4.1).

BLAST searches using as queries the ent-copalyl diphosphate- and ent-kaurene-synthases (CPS and KS) from *P. glauca* (Table 3.1) detected orthologous FL cDNA sequences only in *P. tabuliformis* (Pta_CS1 and Pta_KS1 in Table 4.1). It is worth noting that CPS and KS gene sequences involved in the general (primary) metabolism of gibberellin hormones have been previously isolated and characterized in gymnosperms only in *P. glauca* and *P. sitchensis* (Keeling *et al*., 2010). Moreover, we identified in the NCBI database four sesquiterpene synthases (STPSs) in *Pinus* species, all from *P. sylvestris* (Ps_STPS_1-4 in Table 4.1.). The first three FL cDNAs have been expressed heterologously in *Escherichia coli* to produce (*E*)-β-caryophyllene and α-humulene (Ps_STPS_1), 1(10), 5-germacradiene-4-ol (Ps_STPS_2), and longifolene and α-longipinene (Ps_STPS_3) as their principal products (Köpke *et al.*, 2008). Ps_STPS_1, Ps_STPS_2 and Ps_STPS_3 contained ORFs of 1743, 1878 and 1728 bp, corresponding to polypeptides of 580, 625 and 575 aa, respectively. In comparison with each other, the three *P. sylvestris* deduced amino acid sequences had identities ranging from 60 to 65% (Köpke *et al*., 2008).

Ps_STPS_4 that produced (E)-β-farnesene as principal terpenoid, encoded a protein of 811 aa and according to its different structure, showed only low amino acid identity to the other STPS sequences isolated from *P. sylvestris* (from 35 to 39%) (Köpke *et al.*, 2010). In contrast, Ps_STPS_4 showed an amino acid identity of 75–78% to two other conifer sesquiterpene synthases, (E)-α-bisabolene synthase from *Abies grandis* and (E)-β-farnesene synthase from *Pseodotsuga menziesii* (Köpke *et al*., 2010).

Table 4.1. Characteristics and functions of the 93 FL cDNA sequences coding for putative TPSs in Pinus species identified in NCBI database (see next page)

| Species | Function | Abbreviation | Accession mRNA sequence | ORF (bp) | Accession protein sequence | Amino acid (aa) |
|---|---|---|---|---|---|---|
| *Pinus banksiana* | (-)-alpha pinene synthase | Pb_MTPS_1 | JQ240304 | 1890 | AFU73856 | 629 |
| | (-)-beta-pinene synthase | Pb_MTPS_2 | JQ240291 | 1887 | AFU73843 | 628 |
| | (-)-beta-pinene synthase (TPS-(-)Bpin2) | Pb_MTPS_3 | JQ240292 | 1884 | AFU73844 | 627 |
| | (-)-alpha/beta-pinene synthase | Pb_MTPS_4 | JQ240290 | 1872 | AFU73842 | 623 |
| | alpha terpineol synthase | Pb_MTPS_5 | JQ240308 | 1881 | AFU73860 | 626 |
| | (+)-3-carene synthase | Pb_MTPS_6 | JQ240306 | 1881 | AFU73858 | 626 |
| | (+)-3-carene synthase | Pb_MTPS_7 | JQ240305 | 1881 | AFU73857 | 626 |
| | (+)-alpha pinene synthase | Pb_MTPS_8 | JQ240298 | 1887 | AFU73850 | 628 |
| | monoterpene synthase-like | Pb_MTPS_9 | JQ240296 | 1887 | AFU73848 | 628 |
| | monoterpene synthase like | Pb_MTPS_10 | JQ240297 | 1887 | AFU73849 | 628 |
| | (-)-beta-phellandrene synthase | Pb_MTPS_11 | JQ240302 | 1866 | AFU73854 | 621 |
| | levopimaradiene/abietadiene synthase | Pb_DTPS_LAS1 | JQ240312 | 2574 | AFU73864 | 857 |
| | putative monofunctional diterpene synthase | Pb_MDTPS_1 | JQ240317 | 2559 | AFU73869 | 852 |
| | monofunctional isopimaradiene synthase | Pb_DTPS_mISO1 | JQ240313 | 2631 | AFU73865 | 876 |
| | monofunctional pimaradiene synthase | Pb_DTPS_mPIM1 | JQ240315 | 2607 | AFU73867 | 868 |
| *Pinus contorta* | (-)-alpha pinene synthase | Pc_MTPS_1 | JQ240303 | 1890 | AFU73855 | 629 |
| | (-)-beta-pinene synthase | Pc_MTPS_2 | JQ240293 | 1884 | AFU73845 | 627 |
| | monoterpene synthase like | Pc_MTPS_3 | JQ240294 | 1884 | AFU73846 | 627 |
| | (+)-3-carene synthase | Pc_MTPS_4 | JQ240307 | 1881 | AFU73859 | 626 |
| | (+)-alpha pinene synthase | Pc_MTPS_5 | JQ240295 | 1887 | AFU73847 | 628 |
| | alpha terpineol /1,8-cineole synthase | Pc_MTPS_6 | JQ240309 | 1851 | AFU73861 | 616 |
| | (-)-camphene / (+)-alpha-pinene synthase | Pc_MTPS_7 | JQ240299 | 1860 | AFU73851 | 619 |
| | (-)-beta-phellandrene synthase | Pc_MTPS_8 | JQ240301 | 1866 | AFU73853 | 621 |
| | (-)-beta-phellandrene synthase | Pc_MTPS_9 | JQ240300 | 1875 | AFU73852 | 624 |
| | levopimaradiene/abietadiene synthase | Pc_DTPS_LAS1 | JQ240310 | 2574 | AFU73862 | 857 |
| | levopimaradiene/abietadiene synthase | Pc_DTPS_LAS2 | JQ240311 | 2553 | AFU73863 | 850 |
| | putative monofunctional diterpene synthase | Pc_MDTPS_1 | JQ240318 | 2559 | AFU73870 | 852 |
| | putative monofunctional diterpene synthase | Pc_MDTPS_2 | JQ240319 | 2559 | AFU73871 | 852 |
| | putative monofunctional diterpene synthase | Pc_MDTPS_3 | JQ240320 | 2559 | AFU73872 | 852 |
| | monofunctional isopimaradiene synthase | Pc_DTPS_mISO1 | JQ240314 | 2631 | AFU73866 | 876 |
| | monofunctional pimaradiene synthase | Pc_DTPS_mPIM1 | JQ240316 | 2607 | AFU73868 | 868 |
| *Pinus massoniana* | (-)-alpha pinene synthase | Pm_MTPS_1 | KF547035 | 1890 | AGW25369 | 629 |
| | alpha-terpineol synthase | Pm_MTPS_2 | KJ803197 | 1863 | AIL88641 | 620 |
| *Pinus tabuliformis* | alpha-pinene synthase | Pta_MTPS_1 | EF608499 | 1890 | ABY65904 | 629 |
| | copalyl diphosphate synthase | Pta_CPS1 | KJ158966 | 2391 | AHW42450 | 796 |
| | ent-kaurene synthase | Pta_KS1 | KJ158985 | 2232 | AHW42469 | 743 |
| *Pinus pinaster* | alpha-pinene synthase | Pp_MTPS_1 | KP780394 | 1890 | ALB78130 | 629 |
| | alpha-pinene synthase | Pp_MTPS_2 | KP780395 | 1890 | ALB78131 | 629 |
| *Pinus pinea* | alpha-pinene synthase | Ppinea_MTPS_1 | KR011842 | 1890 | ALD18902 | 629 |
| | alpha-pinene synthase | Ppinea_MTPS_2 | KR011841 | 1890 | ALD18901 | 629 |
| *Pinus kesiya var. langbianensis* | alpha-pinene synthase | Pk_MTPS_1 | KX394684 | 1956 | AQZ36562 | 651 |
| | alpha-pinene synthase | Pk_MTPS_2 | KM382173 | 1875 | AIY22674 | 624 |
| *Pinus contorta var. murrayana* | 2-methyl-3-buten-2-ol synthase | Pmur_MBOS_1 | JN039217 | 1845 | AFJ73537 | 614 |
| | 2-methyl-3-buten-2-ol synthase | Pmur_MBOS_2 | JN039216 | 1845 | AFJ73536 | 614 |
| | 2-methyl-3-buten-2-ol synthase | Pmur_MBOS_3 | JN039221 | 1845 | AFJ73541 | 614 |
| | 2-methyl-3-buten-2-ol synthase | Pmur_MBOS_4 | JN039218 | 1845 | AFJ73538 | 614 |
| | 2-methyl-3-buten-2-ol synthase | Pmur_MBOS_5 | JN039219 | 1845 | AFJ73539 | 614 |
| | 2-methyl-3-buten-2-ol synthase | Pmur_MBOS_6 | JN039220 | 1845 | AFJ73540 | 614 |
| *Pinus teocote* | 2-methyl-3-buten-2-ol synthase | Pteo_MBOS_1 | JN039258 | 1845 | AFJ73576 | 614 |
| | 2-methyl-3-buten-2-ol synthase | Pteo_MBOS_2 | JN039260 | 1845 | AFJ73578 | 614 |
| | 2-methyl-3-buten-2-ol synthase | Pteo_MBOS_3 | JN039259 | 1845 | AFJ73577 | 614 |
| *Pinus greggii* | 2-methyl-3-buten-2-ol synthase | Pg_MBOS_1 | JN039230 | 1845 | AFJ73549 | 614 |
| *Pinus pseudostrobus* | 2-methyl-3-buten-2-ol synthase | Pps_MBOS_1 | JN039254 | 1845 | AFJ73572 | 614 |
| *Pinus attenuate* | 2-methyl-3-buten-2-ol synthase | Pa_MBOS_1 | JN039215 | 1845 | AFJ73535 | 614 |
| *Pinus pseudostrobus var. apulcensis* | 2-methyl-3-buten-2-ol synthase | Papu_MBOS_1 | JN039240 | 1845 | AFJ73559 | 614 |
| | 2-methyl-3-buten-2-ol synthase | Papu_MBOS_2 | JN039242 | 1845 | AFJ73561 | 614 |
| | 2-methyl-3-buten-2-ol synthase | Papu_MBOS_3 | JN039241 | 1845 | AFJ73560 | 614 |
| | 2-methyl-3-buten-2-ol synthase | Papu_MBOS_4 | JN039239 | 1845 | AFJ73558 | 614 |
| *Pinus torreyana* | 2-methyl-3-buten-2-ol synthase | Ptor_MBOS_1 | JN039263 | 1845 | AFJ73581 | 614 |
| | 2-methyl-3-buten-2-ol synthase | Ptor_MBOS_2 | JN039262 | 1845 | AFJ73580 | 614 |
| | 2-methyl-3-buten-2-ol synthase | Ptor_MBOS_3 | JN039261 | 1845 | AFJ73579 | 614 |
| *Pinus coulteri* | 2-methyl-3-buten-2-ol synthase | Pcou_MBOS_1 | JN039227 | 1845 | AFJ73546 | 614 |
| | 2-methyl-3-buten-2-ol synthase | Pcou_MBOS_2 | JN039229 | 1845 | AFJ73548 | 614 |
| | 2-methyl-3-buten-2-ol synthase | Pcou_MBOS_3 | JN039228 | 1845 | AFJ73547 | 614 |
| *Pinus montezumae* | 2-methyl-3-buten-2-ol synthase | Pmon_MBOS_1 | JN039234 | 1845 | AFJ73553 | 614 |
| *Pinus hartwegii* | 2-methyl-3-buten-2-ol synthase | Ph_MBOS_1 | JN039232 | 1845 | AFJ73551 | 614 |
| | 2-methyl-3-buten-2-ol synthase | Ph_MBOS_2 | JN039231 | 1845 | AFJ73550 | 614 |
| *Pinus arizonica var. cooperi* | 2-methyl-3-buten-2-ol synthase | Par_MBOS_1 | JN039226 | 1845 | AFJ73545 | 614 |
| | 2-methyl-3-buten-2-ol synthase | Par_MBOS_2 | JN039225 | 1845 | AFJ73544 | 614 |
| | 2-methyl-3-buten-2-ol synthase | Par_MBOS_3 | JN039224 | 1845 | AFJ73543 | 614 |
| *Pinus ponderosa var. scopulorum* | 2-methyl-3-buten-2-ol synthase | Ppon_MBOS_1 | JN039246 | 1845 | AFJ73564 | 614 |
| | 2-methyl-3-buten-2-ol synthase | Ppon_MBOS_2 | JN039248 | 1845 | AFJ73566 | 614 |
| *Pinus jeffreyi* | 2-methyl-3-buten-2-ol synthase | Pj_MBOS_1 | JN039233 | 1845 | AFJ73552 | 614 |
| *Pinus sabiniana* | 2-methyl-3-buten-2-ol synthase | Psab_MBOS_1 | JF719039 | 1845 | AEB53064 | 614 |
| *Pinus pseudostrobus var. estevezii* | 2-methyl-3-buten-2-ol synthase | Pest_MBOS_1 | JN039251 | 1845 | AFJ73569 | 614 |
| | 2-methyl-3-buten-2-ol synthase | Pest_MBOS_2 | JN039252 | 1845 | AFJ73570 | 614 |
| | 2-methyl-3-buten-2-ol synthase | Pest_MBOS_3 | JN039250 | 1845 | AFJ73568 | 614 |
| | 2-methyl-3-buten-2-ol synthase | Pest_MBOS_4 | JN039249 | 1845 | AFJ73567 | 614 |
| *Pinus muricata* | 2-methyl-3-buten-2-ol synthase | Pmuri_MBOS_1 | JN039235 | 1845 | AFJ73554 | 614 |
| *Pinus muricata* | 2-methyl-3-buten-2-ol synthase | Pmuri_MBOS_2 | JN039236 | 1845 | AFJ73555 | 614 |
| *Pinus radiata* | 2-methyl-3-buten-2-ol synthase | Prad_MBOS_1 | JN039257 | 1845 | AFJ73575 | 614 |
| | 2-methyl-3-buten-2-ol synthase | Prad_MBOS_2 | JN039256 | 1845 | AFJ73574 | 614 |
| *Pinus patula* | 2-methyl-3-buten-2-ol synthase | Ppat_MBOS_1 | JN039245 | 1845 | AFJ73563 | 614 |
| | 2-methyl-3-buten-2-ol synthase | Ppat_MBOS_2 | JN039243 | 1845 | AFJ73562 | 614 |
| *Pinus taeda* | (-)-alpha-pinene synthase | Pt_MTPS_1 | AF543527 | 1890 | AAO61225 | 629 |
| | alpha-terpineol synthase | Pt_MTPS_2 | AF543529 | 1884 | AAO61227 | 627 |
| | (+)-alpha-pinene synthase | Pt_MTPS_3 | AF543530 | 1887 | AAO61228 | 628 |
| | diterpene synthase | Pt_DTPS_LAS1 | AY779541 | 2553 | AAX07435 | 850 |
| *Pinus sylvestris* | longifolene synthase | Ps_STPS_1 | EF679332 | 1743 | ABV44454 | 580 |
| | 1(10),5-germacradien-4-ol synthase | Ps_STPS_2 | EF679331 | 1878 | ABV44453 | 625 |
| | caryophyllene/humulene synthase | Ps_STPS_3 | EF679330 | 1728 | ABV44452 | 575 |
| | e-beta farnesene synthase | Ps_STPS_4 | GU248335 | 2436 | ADH29869 | 811 |
| *Pinus densiflora* | putative abietadiene synthase | Pd_DTPS_ABS1 | EU439295 | 2577 | ACC54559 | 858 |

BLAST searches using as queries the four selected di-terpene synthases (DTPSs) from *P. contorta* and *P. abies* (Table 3.1) allowed the identification of 13 DTPSs in *Pinus* species, seven in *P. contorta* (Hall *et al*., 2013b), four in *P. banksiana* (Hall *et al*., 2013b), and one each in *P. taeda* (Ro and Bohlmann, 2006) and *P. densiflora* (Table 4.1). Five of the 13 identified DTPS-like sequences (Pc_DTPS_LAS1, Pc_DTPS_LAS2, Pb_DTPS_LAS1, Pt_DTPS_LAS1 and Pd_DTPS_ABS1) showed a high level of amino acid identity to each other (from 95 to 99%) and were orthologous to known conifer bifunctional classI/II DTPSs (levopiramene/abietadene synthases, LAS). The eight remaining FL cDNAs coding for putative DTPS were monofunctional class I enzymes that lacked functional class II active sites (Hall *et al*., 2013b). Functional characterization of four of the eight putative monofunctional class I DTPS showed that the orthologous pairs Pb DTPS mPIM1/Pc DTPS mPIM1 and Pb DTPS mISO1/Pc DTPS mISO1 (Table 4.1) converted (+)-copalyl diphosphate, but not GGPP, into isopimaradiene and pimaradiene as major products, respectively (Hall *et al*., 2013b).

BLAST searches using as queries the seven selected mono-terpene synthases (MTPSs) from *P. contorta, P. abies* and *P. glauca* (Table 3.1) detected 74 putative FL MTPS orthologs from 26 different *Pinus* species (Table 4.1). However, a detailed analysis of the retrieved FL cDNAs from this BLAST search allowed us to classify as true MTPS only 32 of the 74 identified sequences. The 42 remaining FL cDNAs isolated from 18 different *Pinus* species encoded the 2-Methyl-3-buten-2-ol (MBO) synthase. MBO is a five-carbon alcohol produced and emitted in large quantities by many species of pine (Lerdau and Gray, 2003). MBO is structurally and biosynthetically related to isoprene and can have an important impact on regional atmospheric chemistry. The gene for MBO synthase was first isolated from *P. sabiniana* (Psab_MBOS_1 in Table 4.1) and the functional characterization of the encoded protein indicated that MBO synthase is a bifunctional enzyme that produces both MBO and isoprene in a ratio of 90:1 (Gray *et al*., 2011). Phylogenetic analysis showed that MBO synthase falls into the TPS-d1 group, together with the gymnosperm MTPSs, and is most closely related to linalool synthase from *P. abies* (Gray *et al.*, 2011). MBO synthase and isoprene synthase comparison clearly demonstrate that hemiterpene synthase evolved independently in gymnosperms and angiosperms. The MBO synthase clusters with gymnosperm MTPSs, isoprene clusters with angiosperm MTPSs (Gray *et al*., 2011) and these gene families diverged between 250–290 million years ago (Martin *et al*., 2004).

The 42 FL MBOS sequences identified in NCBI database had a high level of homology to each other (from 99% to 93% amino acid sequence identity) as shown in the phylogenetic tree reported in Fig. 4.1. Therefore, for graphic clarity in the following phylogenetic trees that include all the TPSs or MTPS from *Pinus* species, only 15 of the 42 MBOS identified sequences will be considered.

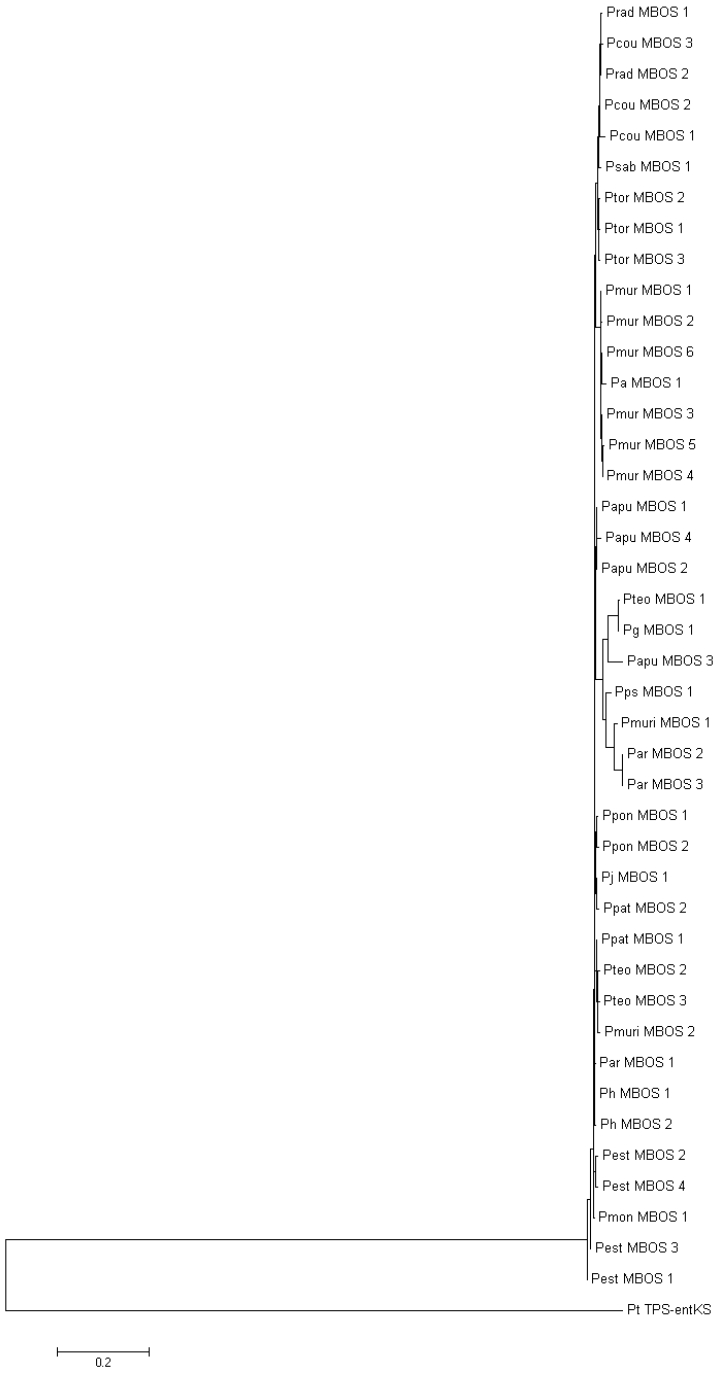



Figure 4.1. Phylogenetic tree for the deduced amino acid sequences of the 42 FL MBOSs from *Pinus* species identified in NCBI database (Table 4.1). The ent-kaurene synthase from *Physcomitrella patens* (Pt TPS-entKS, BAF61135) was included in the analysis as outgroup. The multiple alignments of protein sequences was performed by ClustalX 1.83 software and the phylogenetic tree was constructed by the neighbor-joining (NJ) method.

To gain a deeper understanding of the evolution of terpene synthases in the *Pinus* genus, we performed a phylogenetic analysis including all the identified MTPSs (32), STPSs (4) and DTPSs (13) from different *Pinus* species, the two *P. tabuliformis* monofunctional class I (KS) and class II (CPS) of gibberellin biosynthesis, and 15 selected MBOSs from 13 *Pinus* species, using the ent-kaurene synthase from *P. patens* (Pt TPS-entKS) as an outgroup (Fig. 4.2).

Consistent with previous phylogenetic analyses, all pine TPSs for specialized metabolism (MTPSs, STPSs and DTPSs) were clearly separated from the two *P. tabuliformis* TPSs of primary gibberellin metabolism (Pta KS1 and Pta CPS1). All known conifer TPSs of secondary metabolism are members of the gymnosperm-specific TPS-d subfamily, which is a distinct clade of the larger plant TPS gene family (Chen *et al.*, 2011). In contrast, known gymnosperm CPS and KS of gibberellin metabolism, belong respectively to the TPS-c and TPS-e/f subfamilies which also include orthologous genes of angiosperms (Chen *et al*., 2011).

The pine TPSs for specialized metabolism can be divided into three groups corresponding to the three TPS-d1, TPS-d2 and TPS-d3 clades in which the gymnosperm-specific TPS-d subfamily has been subdivided (Keeling *et al.*, 2011, Chen *et al*., 2011). TPS-d1 contains all the MTPSs identified in the *Pinus* species that, in agreement with previous phylogenetic analysis (Gray *et al*., 2011), cluster with the pine MBOSs (Fig. 4.2). The four identified STPSs in Pinus species belong to the TPS-d2 group. However, it is worth noting that Ps STPS4, that is 811 amino acids in length and likely adopting the three α-helical domains, α, β and γ, is clearly separated from the three shorter (approximately 600 amino acids) βα-domain STPSs (Ps STPS1-3) (Fig. 4.2). TPS-d3 contains all the mono- and bifunctional DTPS identified in *Pinus* species. Consistent with previous results (Hall *et al*., 2013b), the five bifunctional class I/II DTPS (Pc DTPS LAS1, Pc DTPS LAS2, Pb DTPS LAS1, Pt DTPS LAS1 and Pd DTPS ABS1) form a distinct branch within the TPS-d3 group adjacent to the eight monofunctional class I DTPS (Pb MDTPS1, Pc MDTPS1, Pc MDTPS2, Pc MDTPS3, Pc DTPS mISO1, Pb DTPS mISO1, Pc DTPS mPIM1 and Pb DTPS mPIM1). Furthermore, the putative orthologous pairs Pb DTPS mPIM1/Pc DTPS mPIM1 and Pb DTPS mISO1/Pc DTPS mISO1, for which Hall *et al*. (2013b) showed biochemical functions, are separated from the four remaining monofunctional DTPSs (Pb MDTPS1, Pc MDTPS1, Pc MDTPS2 and Pc MDTPS3) for which no activity was observed. While the pine bifunctional levopiramene/abietadene synthases have orthologs in other conifers, monofunctional class I DTPS of specialized metabolism

have not been found in other conifers. It is possible that they represent a lineage-specific clade of the TPS-d3 group that originated in a common ancestor of the Pinus species after the separation of the pine, spruce, and fir genera.

Figure 4.2. Phylogenetic tree of the deduced amino acid sequences of pine TPSs for specialized metabolism (32 MTPSs, 4 STPSs and 13 DTPSs), 15 selected pine MBOSs, and the two *P. tabuliformis* monofunctional class I (KS) and class II (CPS) of gibberellin biosynthesis. The ent-kaurene synthase from *Physcomitrella patens* (Pt TPS-entKS, BAF61135) was included in the analysis as outgroup. The multiple alignments of protein sequences was performed by ClustalX 1.83 software and the phylogenetic tree was constructed by the neighbor-joining (NJ) method and evaluated by bootstrap analysis (MEGA 7). The numbers on the main branches indicate bootstrap percentages higher than 50% for 1,000 replicates. The three distinct groups of TPSs involved in secondary metabolism (TPS-d subfamily) are highlighted with rectangles with different colors: TPS-d1 (red) containing MTPSs and MBOs, TPS-d2 (blue) including STPSs, and TPS-d3 (green) containing DTPSs (see next page).

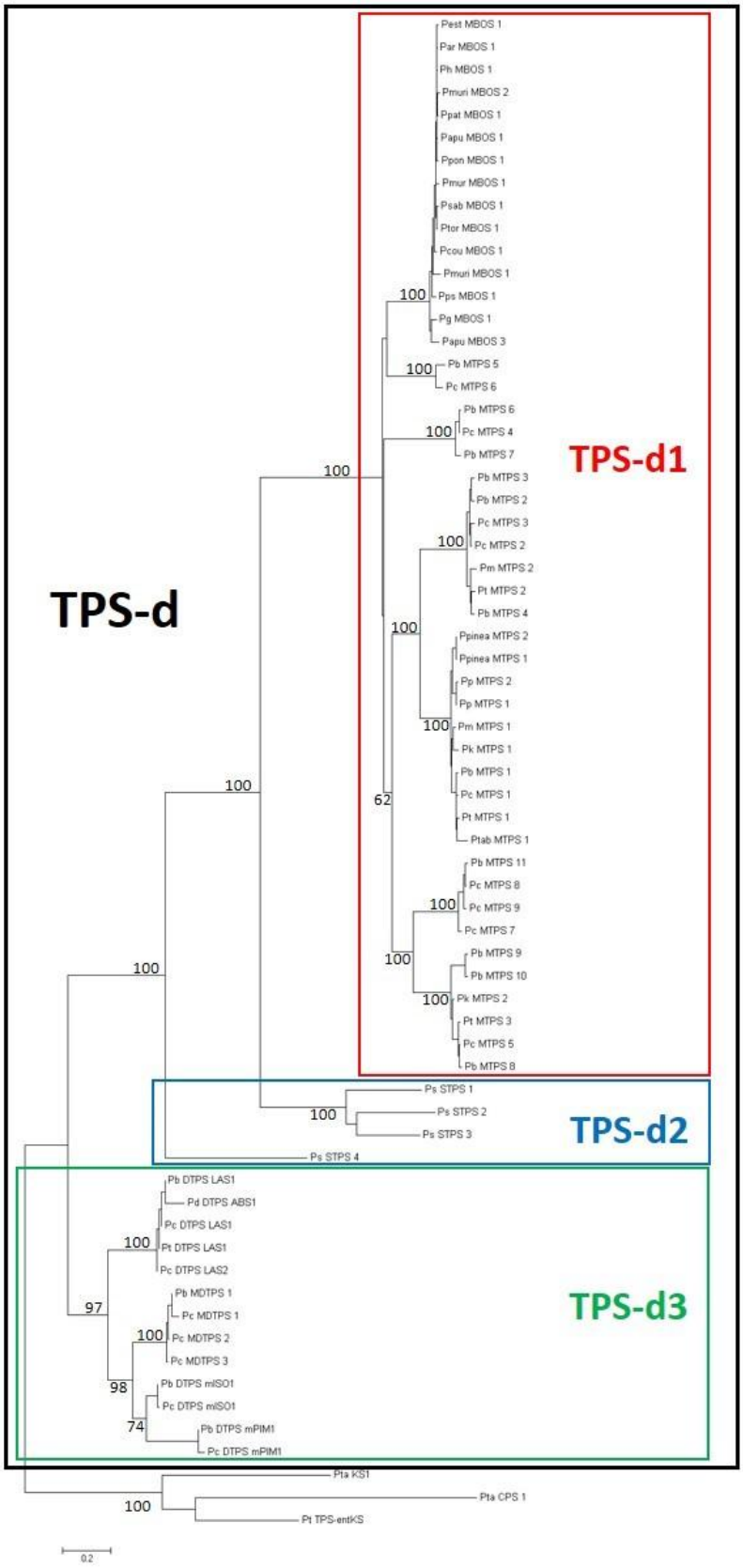



Figure 4.2. For the legend see the previous page.

## 4.2 Phylogenetic analysis of the pine members of TPS-d1 clade

Previous preliminary study have identified several monoterpenes such as (−)-β-pinene, (−)-α-pinene, (+)-α-pinene, (+)-3-carene, and (−)-β-phellandrene, as the most abundant terpenoids in the needles of *P. laricio*. Thus, we focused our attention in isolating cDNA sequences encoding MTPSs potentially involved in the synthesis of the aforementioned monoterpenes in this species. The strategy adopted was based on the PCR amplification of cDNA sequences by using specific primers designed on conserved regions of pine MTPS belonging to distinct phylogenetic groups. Therefore, to analyze in detail the relationships between the pine members of the TPS-d1 clade we performed a phylogenetic analysis including all the 32 MTPSs and the 15 selected MBOSs identified in different *Pinus* species (Fig. 4.2). The phylogenetic analysis aided the division of the 47 MTPS and MBOS sequences into seven distinct groups, some of which contain functionally related proteins from different pine species (Fig. 4.3).

Group 1 contains all the 15 selected MBOSs that use dimethylallyl diphosphate as a substrate to produce hemiterpenes, as was recently shown in *P. sabiniana* (Gray *et al*., 2011). On the basis of phylogenetic relationships and similarity in protein structure with the MTPSs, we can consider the members of this group as MTPS-like.

Group 2 includes only two proteins from *P. contorta* (Pc MTPS6) and *P. banksiana* (Pb MTPS5) (Fig. 4.3), that formed α-terpineol as the major product (Hall *et al*., 2013a). These proteins had only 62% sequence identity to Pt MTPS2, a *P. taeda* protein that also produced α-terpineol (Phillips *et al*., 2003) but assigned to the group 4 (Fig. 4.3). Previous phylogenetic analyses based on MTPSs from different conifer species indicated that Pc MTPS6 and Pb MTPS5 were more closely related (77% identity) to 1,8-cineole synthases from white spruce and a white spruce hybrid (Hall *et al*., 2013a).

Group 3 contains two *P. banksiana* proteins (Pb MTPS6-7) and one from *P. contorta* (Pc MTPS4), that were shown to produce (+)-3-carene as their major product (Hall *et al*., 2013a). It has been previously reported that these three proteins group phylogenetically with functionally similar MTPSs from *P. abies, P. glauca, and P. sitchensis* (Hall *et al*., 2013a), indicating that the genes involved in the synthesis of (+)-3-carene originated prior to the speciation of pine and spruce.

Table 4.2. Characteristics and functions of the 74 FL cDNA sequences coding for MTPSs and MBOSs in Pinus species identified in NCBI database. The 15 selected MBOSs used in phylogenetic analyses are underlines and reported in Italics (see next page).

| Species | Function | Abbreviation | NCBI N0 Accession mRNA sequence | ORF (bp) | NCBI N0 Accession protein sequence | Amino acid (aa) |
|---|---|---|---|---|---|---|
| *Pinus banksiana* | (-)-alpha pinene synthase | Pb_MTPS_1 | JQ240304 | 1890 | AFU73856 | 629 |
| | (-)-beta-pinene synthase | Pb_MTPS_2 | JQ240291 | 1887 | AFU73843 | 628 |
| | (-)-beta-pinene synthase (TPS-(-)Bpin2) | Pb_MTPS_3 | JQ240292 | 1884 | AFU73844 | 627 |
| | (-)-alpha/beta-pinene synthase | Pb_MTPS_4 | JQ240290 | 1872 | AFU73842 | 623 |
| | alpha terpineol synthase | Pb_MTPS_5 | JQ240308 | 1881 | AFU73860 | 626 |
| | (+)-3-carene synthase | Pb_MTPS_6 | JQ240306 | 1881 | AFU73858 | 626 |
| | (+)-3-carene synthase | Pb_MTPS_7 | JQ240305 | 1881 | AFU73857 | 626 |
| | (+)-alpha pinene synthase | Pb_MTPS_8 | JQ240298 | 1887 | AFU73850 | 628 |
| | monoterpene synthase-like | Pb_MTPS_9 | JQ240296 | 1887 | AFU73848 | 628 |
| | monoterpene synthase like | Pb_MTPS_10 | JQ240297 | 1887 | AFU73849 | 628 |
| | (-)-beta-phellandrene synthase | Pb_MTPS_11 | JQ240302 | 1866 | AFU73854 | 621 |
| *Pinus contorta* | (-)-alpha pinene synthase | Pc_MTPS_1 | JQ240303 | 1890 | AFU73855 | 629 |
| | (-)-beta-pinene synthase | Pc_MTPS_2 | JQ240293 | 1884 | AFU73845 | 627 |
| | monoterpene synthase like | Pc_MTPS_3 | JQ240294 | 1884 | AFU73846 | 627 |
| | (+)-3-carene synthase | Pc_MTPS_4 | JQ240307 | 1881 | AFU73859 | 626 |
| | (+)-alpha pinene synthase | Pc_MTPS_5 | JQ240295 | 1887 | AFU73847 | 628 |
| | alpha terpineol /1,8-cineole synthase | Pc_MTPS_6 | JQ240309 | 1851 | AFU73861 | 616 |
| | (-)-camphene / (+)-alpha-pinene synthase | Pc_MTPS_7 | JQ240299 | 1860 | AFU73851 | 619 |
| | (-)-beta-phellandrene synthase | Pc_MTPS_8 | JQ240301 | 1866 | AFU73853 | 621 |
| | (-)-beta-phellandrene synthase | Pc_MTPS_9 | JQ240300 | 1875 | AFU73852 | 624 |
| *Pinus massoniana* | (-)-alpha pinene synthase | Pm_MTPS_1 | KF547035 | 1890 | AGW25369 | 629 |
| | alpha-terpineol synthase | Pm_MTPS_2 | KJ803197 | 1863 | AIL88641 | 620 |
| *Pinus tabuliformis* | alpha-pinene synthase | Ptab_MTPS_1 | EF608499 | 1890 | ABY65904 | 629 |
| *Pinus pinaster* | alpha-pinene synthase | Pp_MTPS_1 | KP780394 | 1890 | ALB78130 | 629 |
| | alpha-pinene synthase | Pp_MTPS_2 | KP780395 | 1890 | ALB78131 | 629 |
| *Pinus pinea* | alpha-pinene synthase | Ppinea_MTPS_1 | KR011842 | 1890 | ALD18902 | 629 |
| | alpha-pinene synthase | Ppinea_MTPS_2 | KR011841 | 1890 | ALD18901 | 629 |
| *Pinus kesiya var. langbianensis* | alpha-pinene synthase | Pk_MTPS_1 | KX394684 | 1956 | AQZ36562 | 651 |
| | alpha-pinene synthase | Pk_MTPS_2 | KM382173 | 1875 | AIY22674 | 624 |
| *Pinus taeda* | (-)-alpha-pinene synthase | Pt_MTPS_1 | AF543527 | 1890 | AAO61225 | 629 |
| | alpha-terpineol synthase | Pt_MTPS_2 | AF543529 | 1884 | AAO61227 | 627 |
| | (+)-alpha-pinene synthase | Pt_MTPS_3 | AF543530 | 1887 | AAO61228 | 628 |
| *Pinus contorta var. murrayana* | 2-methyl-3-buten-2-ol synthase | Pmur_MBOS_1 | JN039217 | 1845 | AFJ73537 | 614 |
| | 2-methyl-3-buten-2-ol synthase | Pmur_MBOS_2 | JN039216 | 1845 | AFJ73536 | 614 |
| | 2-methyl-3-buten-2-ol synthase | Pmur_MBOS_3 | JN039221 | 1845 | AFJ73541 | 614 |
| | 2-methyl-3-buten-2-ol synthase | Pmur_MBOS_4 | JN039218 | 1845 | AFJ73538 | 614 |
| | 2-methyl-3-buten-2-ol synthase | Pmur_MBOS_5 | JN039219 | 1845 | AFJ73539 | 614 |
| | 2-methyl-3-buten-2-ol synthase | Pmur_MBOS_6 | JN039220 | 1845 | AFJ73540 | 614 |
| *Pinus teocote* | 2-methyl-3-buten-2-ol synthase | Pteo_MBOS_1 | JN039258 | 1845 | AFJ73576 | 614 |
| | 2-methyl-3-buten-2-ol synthase | Pteo_MBOS_2 | JN039260 | 1845 | AFJ73578 | 614 |
| | 2-methyl-3-buten-2-ol synthase | Pteo_MBOS_3 | JN039259 | 1845 | AFJ73577 | 614 |
| *Pinus greggii* | 2-methyl-3-buten-2-ol synthase | Pg_MBOS_1 | JN039230 | 1845 | AFJ73549 | 614 |
| *Pinus pseudostrobus* | 2-methyl-3-buten-2-ol synthase | Pps_MBOS_1 | JN039254 | 1845 | AFJ73572 | 614 |
| *Pinus attenuata* | 2-methyl-3-buten-2-ol synthase | Pa_MBOS_1 | JN039215 | 1845 | AFJ73535 | 614 |
| *Pinus pseudostrobus var. apulcensis* | 2-methyl-3-buten-2-ol synthase | Papu_MBOS_1 | JN039240 | 1845 | AFJ73559 | 614 |
| | 2-methyl-3-buten-2-ol synthase | Papu_MBOS_2 | JN039242 | 1845 | AFJ73561 | 614 |
| | 2-methyl-3-buten-2-ol synthase | Papu_MBOS_3 | JN039241 | 1845 | AFJ73560 | 614 |
| | 2-methyl-3-buten-2-ol synthase | Papu_MBOS_4 | JN039239 | 1845 | AFJ73558 | 614 |
| *Pinus torreyana* | 2-methyl-3-buten-2-ol synthase | Ptor_MBOS_1 | JN039263 | 1845 | AFJ73581 | 614 |
| | 2-methyl-3-buten-2-ol synthase | Ptor_MBOS_2 | JN039262 | 1845 | AFJ73580 | 614 |
| | 2-methyl-3-buten-2-ol synthase | Ptor_MBOS_3 | JN039261 | 1845 | AFJ73579 | 614 |
| *Pinus coulteri* | 2-methyl-3-buten-2-ol synthase | Pcou_MBOS_1 | JN039227 | 1845 | AFJ73546 | 614 |
| | 2-methyl-3-buten-2-ol synthase | Pcou_MBOS_2 | JN039229 | 1845 | AFJ73548 | 614 |
| | 2-methyl-3-buten-2-ol synthase | Pcou_MBOS_3 | JN039228 | 1845 | AFJ73547 | 614 |
| *Pinus montezumae* | 2-methyl-3-buten-2-ol synthase | Pmon_MBOS_1 | JN039234 | 1845 | AFJ73553 | 614 |
| *Pinus hartwegii* | 2-methyl-3-buten-2-ol synthase | Ph_MBOS_1 | JN039232 | 1845 | AFJ73551 | 614 |
| | 2-methyl-3-buten-2-ol synthase | Ph_MBOS_2 | JN039231 | 1845 | AFJ73550 | 614 |
| *Pinus arizonica var. cooperi* | 2-methyl-3-buten-2-ol synthase | Par_MBOS_1 | JN039226 | 1845 | AFJ73545 | 614 |
| | 2-methyl-3-buten-2-ol synthase | Par_MBOS_2 | JN039225 | 1845 | AFJ73544 | 614 |
| | 2-methyl-3-buten-2-ol synthase | Par_MBOS_3 | JN039224 | 1845 | AFJ73543 | 614 |
| *Pinus ponderosa var. scopulorum* | 2-methyl-3-buten-2-ol synthase | Ppon_MBOS_1 | JN039246 | 1845 | AFJ73564 | 614 |
| | 2-methyl-3-buten-2-ol synthase | Ppon_MBOS_2 | JN039248 | 1845 | AFJ73566 | 614 |
| *Pinus jeffreyi* | 2-methyl-3-buten-2-ol synthase | Pj_MBOS_1 | JN039233 | 1845 | AFJ73552 | 614 |
| *Pinus sabiniana* | 2-methyl-3-buten-2-ol synthase | Psab_MBOS_1 | JF719039 | 1845 | AEB53064 | 614 |
| *Pinus pseudostrobus var. estevezii* | 2-methyl-3-buten-2-ol synthase | Pest_MBOS_1 | JN039251 | 1845 | AFJ73569 | 614 |
| | 2-methyl-3-buten-2-ol synthase | Pest_MBOS_2 | JN039252 | 1845 | AFJ73570 | 614 |
| | 2-methyl-3-buten-2-ol synthase | Pest_MBOS_3 | JN039250 | 1845 | AFJ73568 | 614 |
| | 2-methyl-3-buten-2-ol synthase | Pest_MBOS_4 | JN039249 | 1845 | AFJ73567 | 614 |
| *Pinus muricata* | 2-methyl-3-buten-2-ol synthase | Pmuri_MBOS_1 | JN039235 | 1845 | AFJ73554 | 614 |
| | 2-methyl-3-buten-2-ol synthase | Pmuri_MBOS_2 | JN039236 | 1845 | AFJ73555 | 614 |
| *Pinus radiata* | 2-methyl-3-buten-2-ol synthase | Prad_MBOS_1 | JN039257 | 1845 | AFJ73575 | 614 |
| | 2-methyl-3-buten-2-ol synthase | Prad_MBOS_2 | JN039256 | 1845 | AFJ73574 | 614 |
| *Pinus patula* | 2-methyl-3-buten-2-ol synthase | Ppat_MBOS_1 | JN039245 | 1845 | AFJ73563 | 614 |
| | 2-methyl-3-buten-2-ol synthase | Ppat_MBOS_2 | JN039243 | 1845 | AFJ73562 | 614 |

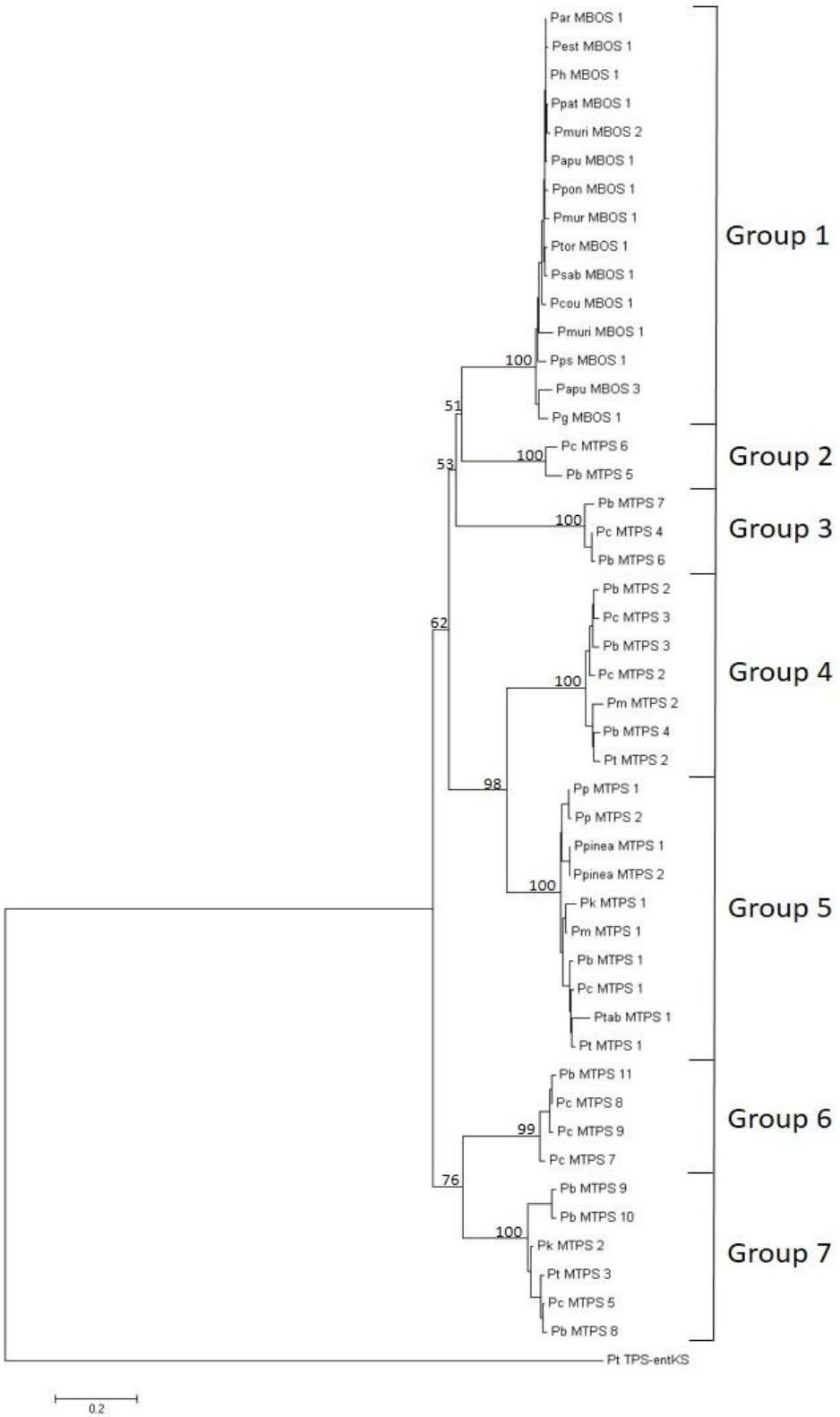



Figure 4.3. Phylogenetic tree of the deduced amino acid sequences of 32 MTPSs and 15 selected MBOSs identified in different *Pinus* species. The ent-kaurene synthase from *Physcomitrella patens* (Pt TPS-entKS, BAF61135) was included in the analysis as outgroup. The numbers on the main branches indicate bootstrap percentages higher than 50% for 1,000 replicates. The seven phylogenetic groups identified in the pine members of TPS-d1 clade are highlighted with square brackets.

Group 4 contains four MTPS from *P. contorta* and *P. banksiana* (Pc MTPS 2 and Pb MTPS 2-4) (Fig. 4.3) that were shown to produce (−)-β-pinene as their major product and also (−)-α-pinene, but in low amount (Hall *et al*., 2013a). These five MTPS are closely related to *P. taeda* protein (Pt MTPS2) that instead produced (−)-α-terpineol and no (−)-β-pinene or (−)-α-pinene (Phillips *et al.,* 2003). This demonstrates that in some cases, it is not possible to predict the function of a putative MTPS based on only their sequence identity. Previous reports showed that a few amino acid substitutions are sufficient to alter the product profiles of mono-TPSs from grand fir (Katoh *et al*., 2004; Hyatt *et al*., 2005). The high level of sequence identity between these functionally distinct proteins from *P. contorta*, *P. banksiana* and *P. taeda* serves as an example of the functional plasticity observed in conifer MTPSs. The second member of the group 4 from *P. contorta* (Pc MTPS3) did not show any activity with GPP, FPP or GGPP, either as FL or truncated protein lacking the putative plastid targeting sequence (Hall *et al.*, 2013a). Finally, the group 4 MTPS from *P. massoniana* (Pm MTPS2), although reported as (−)-α-terpineol synthase in the NCBI database most likely on the basis of the high sequence identity with Pt MTPS2, was not functionally characterized.

Group 5 includes 10 putative α-pinene synthases for which only three from *P. taeda* (Pt MTPS1), *P. contorta* (Pc MTPS1) and *P. banksiana* (Pb MTPS1) have been functionally characterized as producing (−)-α-pinene as their dominant product (Phillips *et al*., 2003; Hall *et al*., 2013a). Previous phylogenetic analyses showed that these three pine MTPS grouped most closely with spruce and fir enzymes that produce (−)-α-pinene (Hall *et al*., 2013a). This indicates that the genes involved in the synthesis of (−)-α-pinene originated prior to the speciation of pine, fir and spruce, as also occurred for the genes encoding for (+)-3-carene synthases.

Group 6 contains three proteins from *P. contorta* and *P. banksiana* (Pc MTPS8-9 and Pb MTPS11) that formed (−)-β-phellandrene as their major product (Hall *et al*., 2013a). A fourth group 6 member from *P. contorta* (Pc MTPS7) had 95% identity to both *P. banksiana* and *P. contorta* (−)-β-phellandrene synthases but produced 29% (−)-camphene and 26% (+)-α-pinene along with other minor products (Hall *et al*., 2013a).

Finally, group 7 contains three MTPSs from *P. taeda* (Pt MTPS3), *P. contorta* (Pc MTPS5) and *P. banksiana* (Pb MTPS8) that were shown to produce (+)-α-pinene as dominant product (Phillips *et al*., 2003; Hall *et al*., 2013a). Two additional *P. banksiana* proteins belonging to the group 7 (Pb MTPS9-10) showed no activity with

GPP, GGPP or FPP (Hall *et al*., 2013a). Finally, the other *P. kesiya* group 7 member, although reported as α-pinene in the NCBI database, was not functionally characterized.

Previous phylogenetic analyses showed that *P. contorta*, *P. banksiana* and *P. taeda* (+)-α-pinene synthases (group 7) and (−)-β-phellandrene synthases (Group 6) form a unique and apparently *Pinus* specific subclade within the TPS-d1 clade (Hall *et al*., 2013a). The pine (−)-β-phellandrene synthases grouped separately from the *P. sitchensis* and *A. grandis* (−)-β-phellandrene synthases (Hall *et al.*, 2013a), highlighting the multiple origins of (−)-β-phellandrene biosynthesis in conifers. Genes that produce (+)-α-pinene as their major product have not been identified in any conifer genus other than Pinus, suggesting this function may have evolved in the pine lineage after the separation from spruce and firs.

### 4.3 Isolation of partial and full length (FL) cDNAs coding for MTPS in *P. nigra* subsp. *laricio*

Deduced amino acid and nucleotide sequences of pine MTPSs belonging to each of the seven identified phylogenetic groups in the TPS-d1 clade (Fig. 4.3) were aligned in order to identify highly conserved regions among members of each group. The nucleotide sequences in the identified conserved regions for each group were then used to design specific primers for the isolation by RT-PCR of partial transcripts of orthologous genes in *P. laricio.* Figure 4.4 schematically outlines the FL cDNAs for six of the representative members of the seven phylogenetic groups, and the positions of their specific forward and reverse primers, the complete list of the same primers are reported in Table 4.3.

By using such strategy, we were able to isolate and sequence partial MTPS transcripts of putative *P. laricio* orthologous genes belonging to five out of seven phylogenetic groups. Moreover, four partial *P. laricio* transcripts of groups 1, 2, 5 and 7, were used as templates for isolating FL MTPS cDNA sequences by 5' and 3' RACE extensions (sequences of RACE primers are reported in Table 4.4).

Table 4.3. Forwards and reverse primers used for the isolation of partial cDNAs coding for MBOS and MTPS in *P. nigra* subs. *laricio*

| Phylogenetic groups | Forward primers | Reverse primers |
|---|---|---|
| **1** | **F1**: 5'-CATCATTCCAACCTCTGGGA-3'<br>**F2**: 5'-GGGCAGTTTGCATGTTCG-3'<br>**F3:** 5'-ATGACCTTCCCTCGGCAT-3' | **R1**: 5'-GACCGGAAGCTTTAGTATGGC-3'<br>**R2:** 5'-TTCTGGAAGCCACTCCGTC-3'<br>**R3**: 5'-AGGCACAGGCTCAATGACG-3' |
| **2** | **F2**: 5'-CCTTTCCATGGTCGATAGCA-3'<br>**F3:** 5'-GGTGGAAGGATTCAGGCTTC-3' | **R2:** 5'-ATAGGCTTCCCAAGCGTGT-3'<br>**R3:** 5'-ATTGGTGGCGACGCTGTA-3' |
| **3** | **F2:** 5'-TTCTAACCTGTTGGGACGACAA-3'<br>**F3:** 5'-TCCGCCAATGATACAGAGTTG-3' | **R1:** 5'-AATGTAGCATGTTCAGGCACC-3'<br>**R2:** 5'-TCTGTGCTTCTTTCGCCATT-3'<br>**R3:** 5'-GCATTGTTGTCCGGTCTAAGA-3' |
| **4** | **F2:** 5'-TTCTGTCAACGCCTTATGGG-3' | **R1:** 5'-CCACCATCTCAAGAGGGACTC-3'<br>**R2:** 5'-CGATCCACTTTGCTTCTTGC-3' |
| **5** | **F1:** 5'-AACTTGCAAAGTTGGAGTTCAAC-3'<br>**F2:** 5'ATTCCAACCTGTGGGACGATGAT-3' | **R2:** 5'-TTGATATGATTGAGAGCATCT-3' |
| **7** | **F1:** 5'-GGTTTCTGCTGTCCCGTTGG-3'<br>**F2:** 5'-TTCGAACTCTCCGACTACACG-3'<br>**F3:** 5'-CTTCCAGATTATATGAAAGG-3' | **R1:** 5'-CTTGTGAAAGACCGGAGGG-3'<br>**R2:** 5'-CCGTTGCTGTCGGGTCTAAGTAAC-3' |

Table 4.4. Specific primers used for 5' and 3' RACE (Rapid Amplification of cDNA Ends) extensions of partial transcripts of *P. laricio* of group 1, 2, 5 and 7.

| Phylogenetic groups | RACE 5' | RACE 3' |
|---|---|---|
| **1** | **R2:** 5'-ATCTGAAGACACCGGGTATTCC-3'<br>**R1:** 5'-TCGATTCCCAAACGTTCAA-3' | **F1:** 5'-TACCAGGCTGAGAGGAACCG-3' |
| **2** | **R2:** 5'-GACGTCCATGTAATGCCTTGC-3'<br>**R1:** 5'-AACGCTTGAAGACACCGGG-3' | **F1:** 5'-TTCAGTAGCTTGGCGGCTG-3' |
| **5** | **R2:** 5'-GAGGGAAGCCCGATATAAATT-3'<br>**R1:** 5'-ATCATCGTCCCACAGGTTGGAAT-3' | **F1:** 5'-ACTATGAGAACGGGAAAGTTAG-3' |
| **7** | **R2:** 5'-CCACTTTCTCTCCCACGTCC-3'<br>**R1:** 5'-CGGTGATGGAGGTCAGTGA-3' | **F1:** 5'-TCTCGCATAACCACGCTCG-3' |

The results of cloning and sequencing of cDNA and RACE amplification products obtained for putative *P. laricio* orthologous MTPS genes belonging to the five phylogenetic groups are described in detail in the following paragraphs.

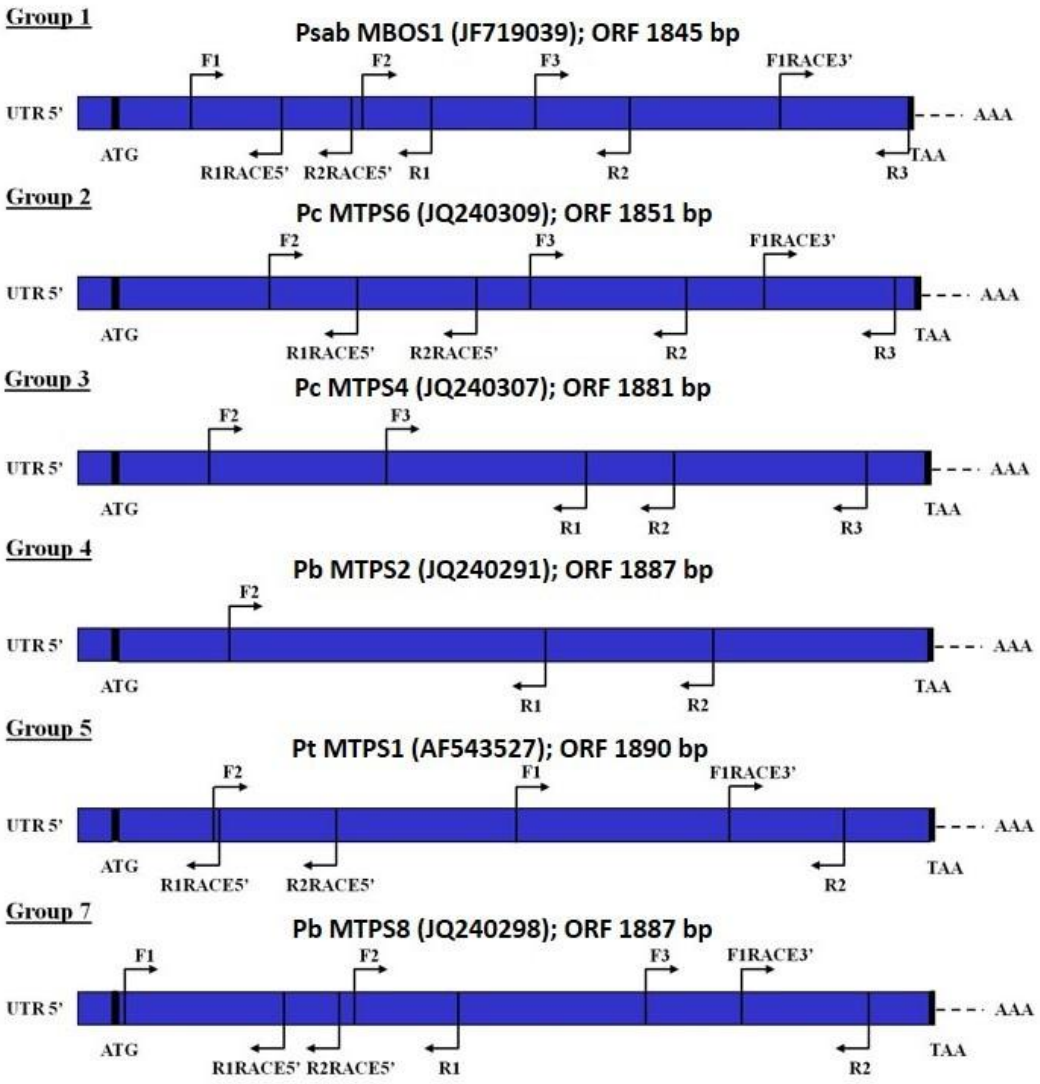



Figure 4.4. Schematic representation of the FL cDNAs for six of the seven representative members of the phylogenetic groups and the positions of their forward and reverse primers used for the isolation of the partial transcripts coding for orthologous genes in *P. nigra* subsp. *laricio*. The position of specific primers used for 5' and 3' RACE (Rapid Amplification of cDNA Ends) extensions of partial transcripts of *P. laricio* of group 1, 2, 5 and 7 are also indicated

Using the three forward and reverse primers designed in conserved regions of pine members of the phylogenetic group 1, we were able to amplify four distinct cDNA fragments from needles of *P. laricio* (Fig. 4.5). The largest fragment (named Pnl MBOS 1.1 F1-R3, about 1700 bp in length) extended from the 5' to the 3' ends of the coding region of the corresponding FL transcript, while Pnl MBOS 1.1-F1-R1, -F2-R2

and -F3-R3, are smaller overlapping regions respectively at the 5’ end, in the central part, and at the 3’ end of the coding region (Fig. 4.5). For each of the four cDNA fragments, 3 clones were sequenced.

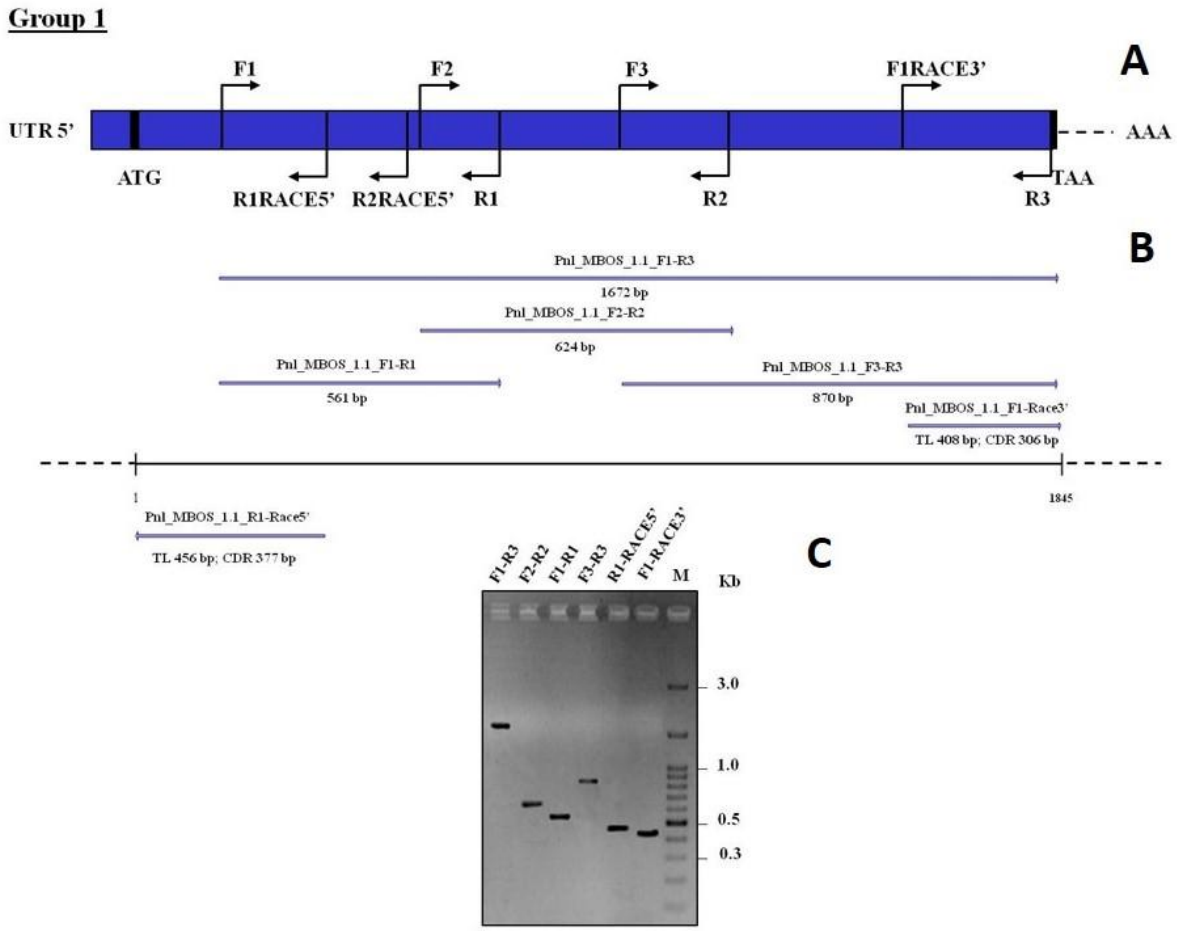



Figure 4.5. Partial cDNA fragments and corresponding 5’ and 3’ RACE extension amplification products of a putative *P. laricio* gene belonging to the phylogenetic group 1.

A) Schematic representation for the position of forward (F1, F2 and F3), reverse (R1, R2 and R3), and 5' and 3' RACE primers used for the isolation of partial and FL transcripts. B) Sequence assembly of identical cDNA fragments in the area of overlapping region and 5’ and 3’ extension products to obtain the FL transcript for which is reported only the length of the coding region. C) Agarose gel electrophoresis of cDNA and RACE amplification products.

In some cases, two slightly different sequences were recognized among the three clones analyzed for each of four cDNA fragments. These different clones for the same cDNA product showed high sequence identity (over 96%), most probably because they derived from transcripts of alleles of the same gene and/or from

transcripts of duplicated copies of the same gene. Mutations detected in these similar sequences consisted mainly of nucleotide substitutions, most of them synonymous (data not shown). However, among the 12 sequenced clones (3 each for the 4 different cDNAs) we identified identical sequences in the overlapping regions of the four cDNA fragments (Fig. 4.5 B), indicating that they are part of the same FL transcript. The assembled and unique sequence of 1672 bp in length was used as a template to isolate by RACE the corresponding 5' and 3' extensions and afterward, was subsequently validated by sequence analysis. The RACE 3’ product (F1-Race3’) of 408 bp in length, contained a sequence of 306 bp in the coding region (CDR), that is identical to the overlapping region with the cDNA products obtained with the primer pairs F1-R3/F3-R3, and a fragment of 102 bp in the untranslated region (UTR) at the 3’ end, while the RACE 5’ product (R1-Race5’) included a 377 bp sequence in the CDR, identical to the overlapping region with the cDNA products obtained with primer pairs F1-R3/F1-R1, and a short fragment of 79 bp in the 5’ UTR (Fig. 4.5). Therefore, the assembled FL cDNA, named Pnl MBOS 1.1, contained an ORF of 1845 bp coding for a protein of 614 aa.

Using the two forward (F2 and F3) and reverse (R2 and R3) primers designed in conserved regions of pine members of the phylogenetic group 2, we were able to amplify three distinct cDNA fragments from needles of *P. laricio* (reported as Pnl MTPS 1.2 F2-R2, F3-R3 and F2-R3 in Fig. 4.6), and the three sequenced clones for each of the three cDNA fragments exhibited identical sequences. Moreover, the sequences of the three partial cDNAs are identical in their overlapping regions, indicating that we isolated a unique transcript of the putative *P. laricio* orthologous genes belonging to the phylogenetic group 2. These results could support the hypothesis of the existence of a single MTPS gene belonging to the phylogenetic group 2 in the genome of *P. laricio*. The assembled and unique sequence of 1459 bp in length was then used as a template to isolate by RACE, the corresponding 5' and 3' extensions, which were then validated by sequence analysis. The RACE 3’ product (F1-Race3’) of 479 bp in length, contained a sequence of 375 bp in the CDR, that is identical to the overlapping region with the Pnl MTPS 1.2-F2-R3 and -F3-R3 cDNA products, and a fragment of 104 bp in the 3’ UTR, while the RACE 5’ product (R1-Race5’) included a 552 bp sequence in the CDR, identical to the overlapping region with the cDNA products obtained with the primer pairs F1-R3/F1-R1, and a fragment of 79 bp in the 5’ UTR (Fig. 4.6). Therefore, the assembled FL cDNA, named Pnl MTPS 1.2, contained an ORF of 1857 bp coding for a protein of 618 aa.

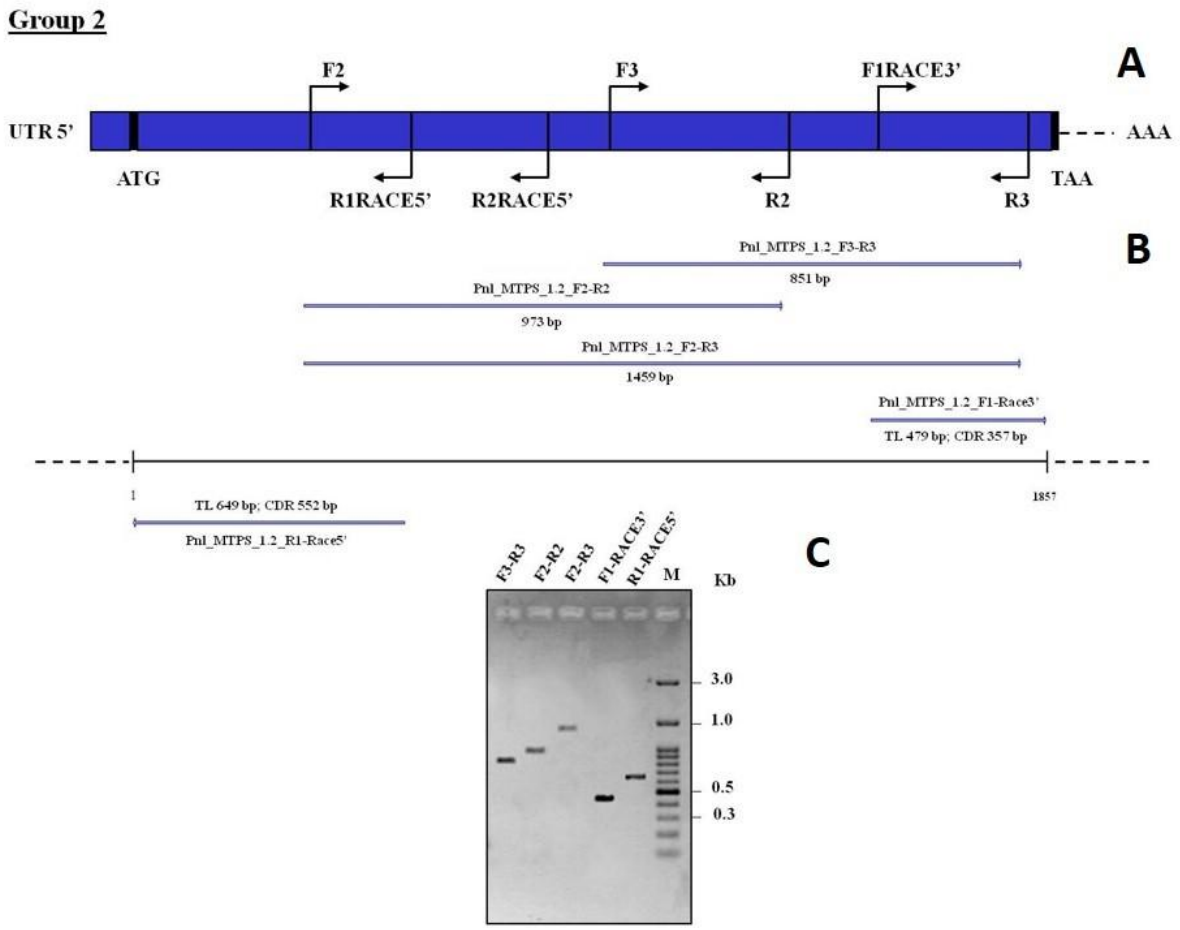



Figure 4.6. Partial cDNA fragments and corresponding 5’ and 3’ RACE extension amplification products of a putative *P. laricio* gene belonging to the phylogenetic group 2.

A) Schematic representation for the position of forward (F2 and F3), reverse (R2 and R3), and 5' and 3' RACE primers used for the isolation of partial and FL transcripts. B) Sequence assembly of identical cDNA fragments in the area of overlapping region and 5’ and 3’ extension products to obtain the FL transcript for which is reported only the length of the coding region. C) Agarose gel electrophoresis of cDNA and RACE amplification products.

For the group 5, we obtained two cDNA fragments from needles of *P. laricio* (Fig. 4.7). The largest fragment (named Pnl MTPS 1.5 F2-R2, about 1500 bp in length) extended from the 5' to the 3' ends of the CDR of the corresponding FL transcript, while Pnl MTPS 1.5 F1-R2 (about 800 bp in length) included a smaller region at the 3’ end of the CDR (Fig. 4.7). The sequences of the two partial cDNAs are identical in their overlapping regions, indicating probably the existence of a single MTPS gene belonging to the phylogenetic group 5 in the genome of *P. laricio*. The assembled and unique sequence of 1489 bp in length was then used as a template to isolate by

RACE, the corresponding 5' and 3' extensions which were validated by sequence analysis. The RACE 3’ product (F1-Race3’) of 510 bp in length, contained a sequence of 473 bp in the CDR that is identical to the overlapping region with the cDNA products obtained with the primer pairs F2-R3/F3-R3, and a fragment of 37 bp in the 3’ UTR, while the RACE 5’ product (R1-Race5’) included a 234 bp sequence in the CDR, identical to the overlapping region with the cDNA product Pnl MTPS 1.5 F2-R2, and a fragment of 86 bp in the 5’ UTR (Fig. 4.7). Therefore, the assembled FL cDNA, named Pnl MTPS 1.5, possessed an ORF of 1908 bp with a coding ability of 635 aa.

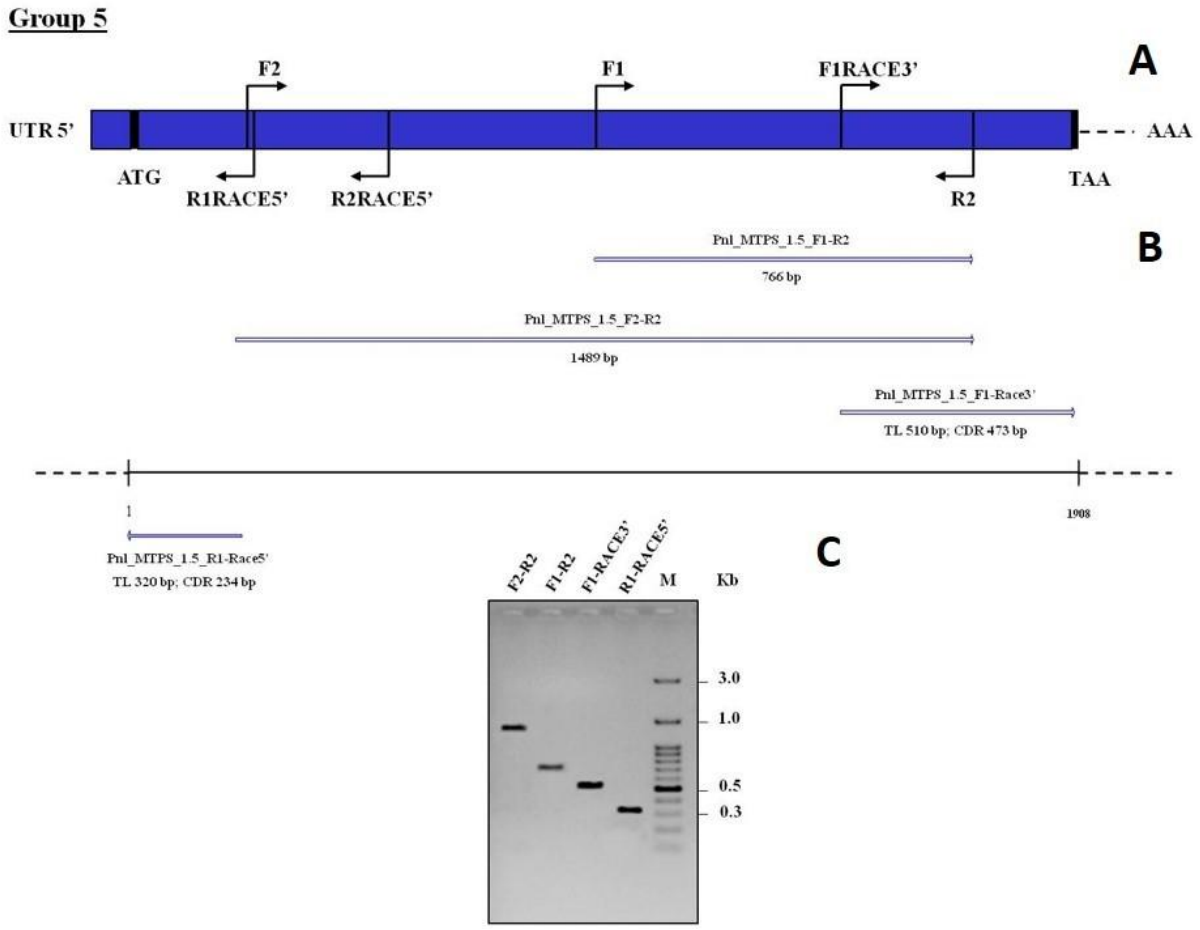



Figure 4.7. Partial cDNA fragments and corresponding 5’ and 3’ RACE extension amplification products of a putative *P. laricio* gene belonging to the phylogenetic group 5.

A) Schematic representation for the position of forward (F1 and F2), reverse (R2), and 5' and 3' RACE primers used for the isolation of partial and FL transcripts. B) Sequence assembly of identical cDNA fragments in the area of overlapping region and 5’ and 3’ extension products to obtain the FL transcript for which is reported only the length of the coding region. C) Agarose gel electrophoresis of cDNA and RACE amplification products.

Using the three forward (F1, F2 and F3) and two reverse (R1 and R2) primers designed in conserved regions of pine members of the phylogenetic group 7, we were able to amplify three distinct cDNA fragments from needles of *P. laricio* (indicated as Pnl MTPS 1.7-F1-R1, -F2-R2 and F3-R2 in Fig. 4.8). Two slightly different sequences were recognized among the three clones that were analyzed for each of the three cDNA fragments and as discussed before, indicating most likely, the presence of different alleles of the same gene or duplicated copies of the same gene. However, among the 9 clones sequenced (3 for 3 different cDNAs) we identified identical sequences in the overlapping regions of the cDNA fragments (Fig. 4.8 B), indicating that they are part of the same FL transcript. The assembled and unique sequence of 1737 bp in length was used as a template to isolate by RACE the corresponding 5' and 3' extensions and afterward, was subsequently validated by sequence analysis. The RACE 3' product (F1-Race3') of 530 bp in length, contained a sequence of 447 bp in the CDR that is identical to the overlapping region with the cDNA products obtained with the primer pairs F2-R2/F3-R2, and a fragment of 83 bp in the 3' UTR, while the RACE 5' product (R1-Race5') included a 380 bp sequence in the CDR, identical to the overlapping region with the Pnl MTPS 1.7-F1-R1 cDNA product, and a fragment of 110 bp in the 5' UTR (Fig. 4.8). Therefore, the assembled FL cDNA, named Pnl MTPS 1.7, contained an ORF of 1890 bp coding for a protein of 629 aa.

At the moment for the group 4, we obtained two *P. laricio* partial cDNA fragments that cover only part of the entire CDR of the FL transcript (Fig. 4.9). Also in this case, two slightly different sequences were recognized among the three clones that were analyzed for each of the two cDNA fragments, most likely because they are derived from transcripts of alleles of the same gene and/or from transcripts of duplicated copies of the same gene. The assembled and unique sequence of 1132 bp in length (Fig. 4.9) will be used as a template to isolate by RACE the corresponding 5' and 3' extensions.

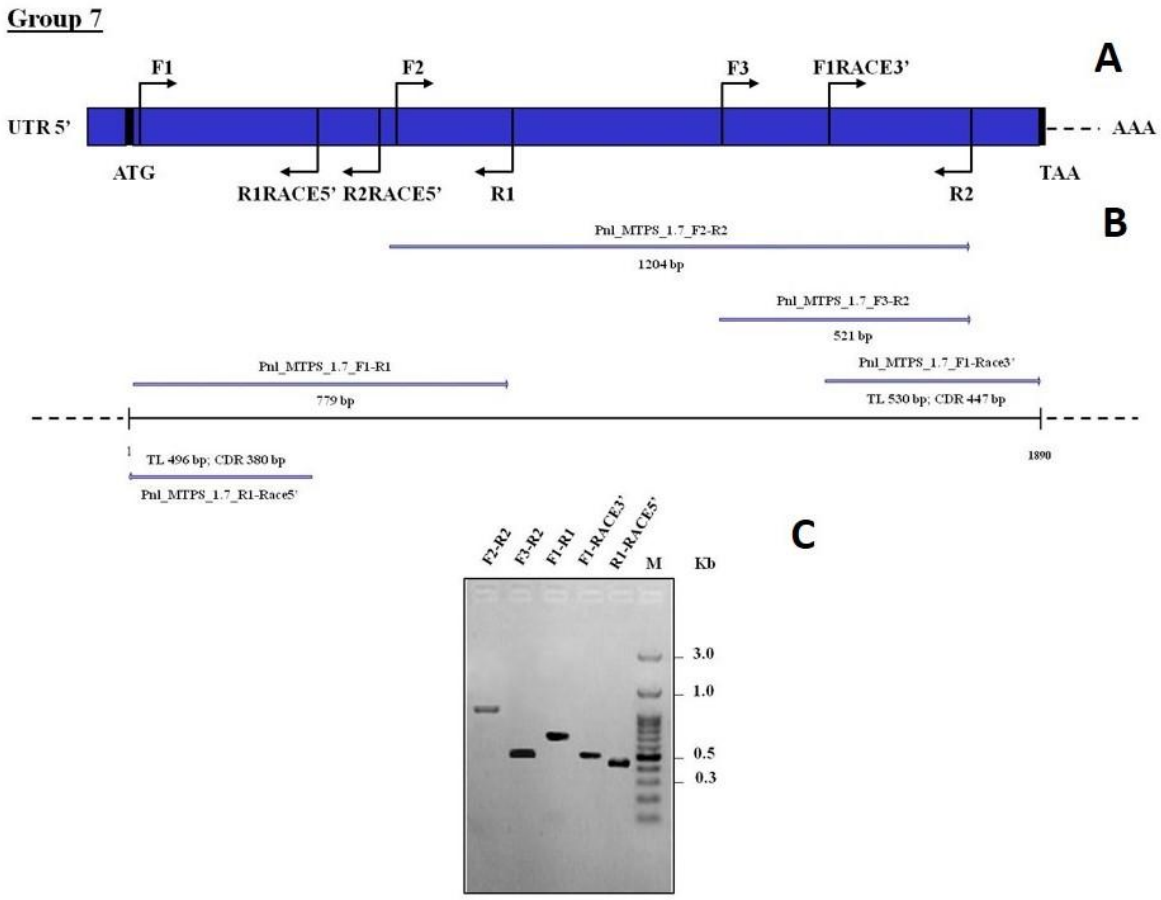



Figure 4.8. Partial cDNA fragments and corresponding 5' and 3' RACE extension amplification products of a putative *P. laricio* gene belonging to the phylogenetic group 7.

A) Schematic representation for the position of forward (F1, F2 and F3), reverse (R1 and R2), and 5' and 3' RACE primers used for the isolation of partial and FL transcripts. B) Sequence assembly of identical cDNA fragments in the area of overlapping region and 5' and 3' extension products to obtain the FL transcript for which is reported only the length of the coding region. C) Agarose gel electrophoresis of cDNA and RACE amplification products.

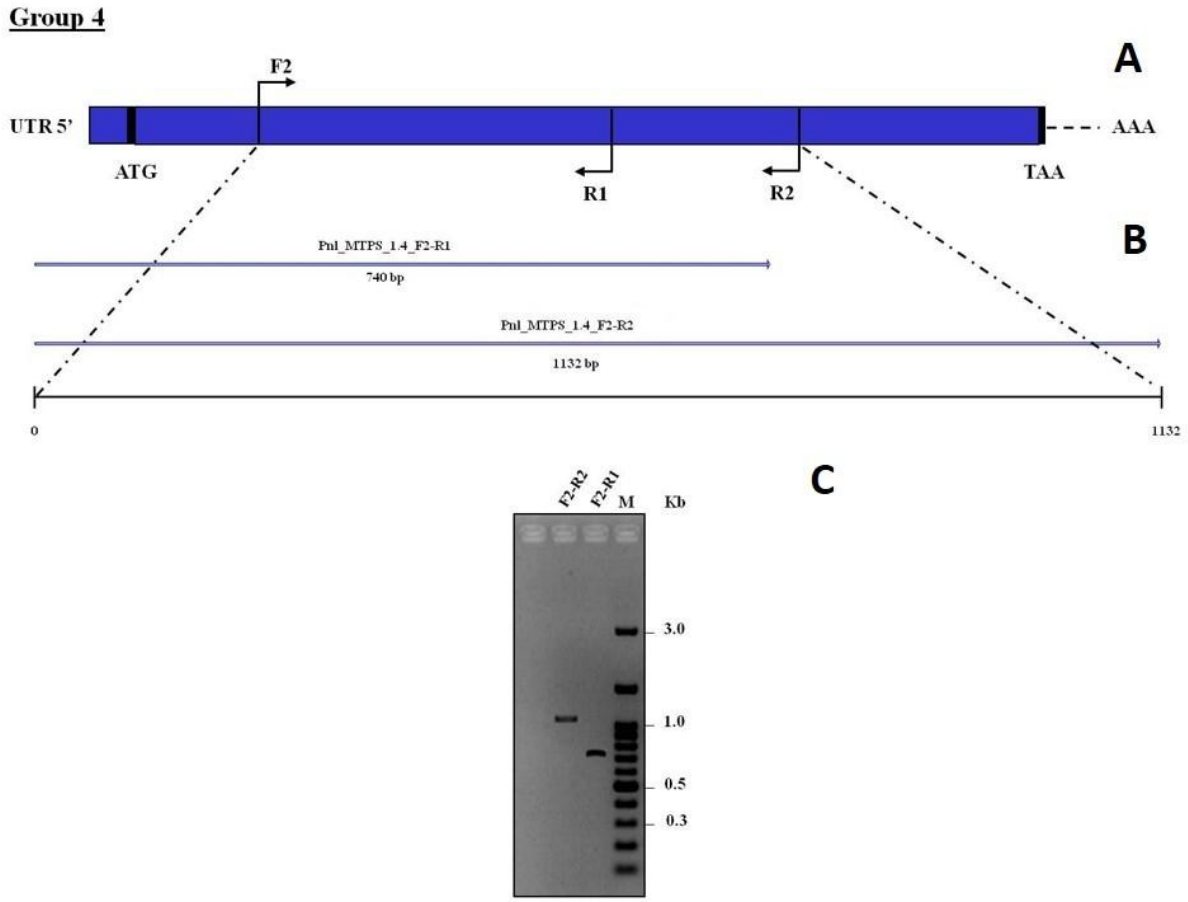



Figure 4.9. Partial cDNA fragments of a putative *P. laricio* gene belonging to the phylogenetic group 4. A) Schematic representation for the position of forward (F1) and reverse (R1 and R2) used for the isolation of partial transcripts. B) Sequence assembly of two identical cDNA fragments in the area of overlapping region. C) Agarose gel electrophoresis of cDNA amplification products.

## 4.4 Isolation of genomic sequences coding for *P. laricio* MTPS of phylogenetic group 3

Putative orthologous genes for the phylogenetic group 3 were not found in the transcriptome (i.e. cDNA) of needles of *P. laricio*, despite extensive efforts to amplify by PCR cDNA fragments of these genes, suggesting that they were not expressed in the tissue tested. To assess the presence of group 3 genes within the *P. laricio* genome, we used the primers designed in conserved regions of pine members of the phylogenetic group 3 (Fig. 4.10A) and the genomic DNA extracted from *P. laricio* needles as a template. PCR amplification using the three primer pairs F2-R1, F3-R2 and F3-R3 yielded three amplicons ranging from 1.2 to 1.9 Kb (Fig. 4.10C), confirming the presence of group 3 genes in the genome of *P. laricio*. All genomic amplification fragments were much larger than expected on the basis of pine cDNA

sequences of phylogenetic group 3, suggesting the presence of intron sequences. Sequencing of the three genomic fragments, named Pnl MTPS 1.5-F2-R1, -F3-R2 and –F3-R3, showed that they are identical in their overlapping regions, most likely because they are derived from the same gene. The assembly of the three DNA fragments of 1645 bp (Pnl MTPS 1.5-F2-R1), 1189 bp (Pnl MTPS 1.5-F3-R2) and 1854 bp (Pnl MTPS 1.5-F3-R3) resulted in a consensus sequence of 2630 bp (Fig. 4.10B), named Pnl MTPSG 1.5-F2-R3. Alignment of the nucleotide sequence of Pnl MTPSG 1.5-F2-R3 with the FL cDNAs of pine members of the phylogenetic group 3 indicated that the isolated genomic sequence contained ten exons (with the first and the tenth incomplete) and nine introns (Fig. 4.10D), consistent with the previously characterized genomic sequences of conifer MTPSs (Trapp and Croteau, 2001; Hamberger *et al*., 2009; Hall *et al*., 2011). In this context it is noteworthy, the high conservation of the exon size detected between genes coding for MTPS in *A. grandis* (Trapp and Croteau, 2001), *P. glauca* (Hamberger *et al*., 2009), *P. sitchensis* (Hall *et al*., 2011) and the *P. laricio* genomic sequence isolated in this study.

The structure of the putative *P. laricio* MTPS gene belonging to the phylogenetic group 3 was also confirmed by the identification of potential exon/intron junction sequences, which were consistent with the presence at the 5’ and 3’ ends of each intron of the conserved dinucleotide sequences GT and AG (the so-called universal rule GT..AG).

On the basis of the determined intron/exon structure (Fig. 4.10D), the assembled genomic sequence Pnl MTPSG 1.5-F2-R3 contained a partial nucleotide CDR sequence of 1517 bp coding for an incomplete protein of 376 aa.

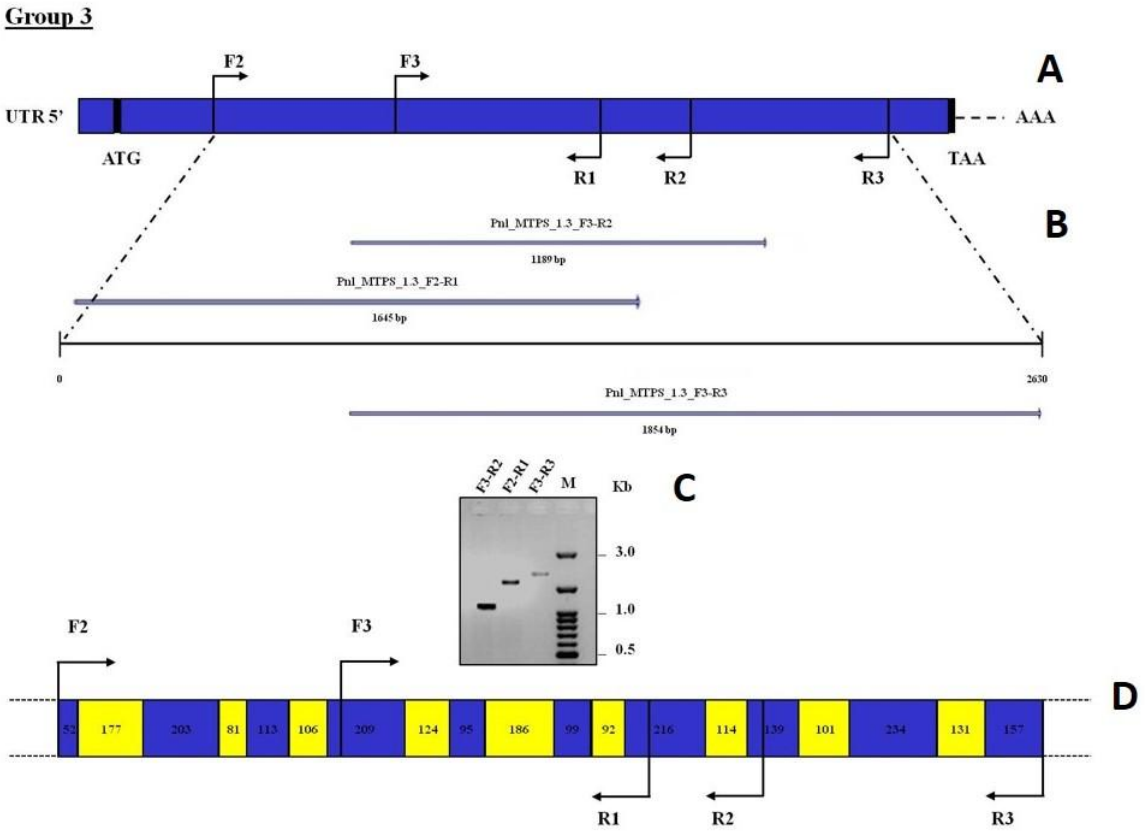



Figure 4.10. Genomic amplification products of a putative *P. laricio* gene belonging to the phylogenetic group 3.

A) Schematic representation for the position of forward (F2 and F3) and reverse (R1, R2 and R3) used in the amplification of genomic DNA. B) Sequence assembly of three genomic fragments identical in their overlapping regions. C) Agarose gel electrophoresis of cDNA and RACE amplification products. D) Intron/exon structure of the assembled genomic sequence. The exons are in blue, the introns in yellow. The position of the primers used to amplify the three genomic fragments are also indicated.

## 4.5 Characterization of the predicted MTPS proteins

As previously described, we isolated four FL cDNAs, namely Pnl MBOS-1.1, Pnl MTPS-1.2, -1.5 and -1.7, encoding proteins of 614, 618, 635 and 629 aa, respectively. Partial cDNA and genomic sequences for *P. laricio* members of phylogenetic groups 4 and 3 (indicated as Pnl MTPS-1.4 and -1.3), encoding incomplete proteins of 376 and 505 aa, respectively, were also obtained.

Pairwise sequences identities among predicted amino acids of the six *P. laricio* cDNA and genomic sequences are shown in Table 4.5. The identities of the six predicted proteins ranged from 55.6% (Pnl MTPS-1.3/-1.4) to 75.7% (Pnl MTPS-1.4/-1.5). It is worth noting that Pnl MBOS 1.1, the putative *P. laricio* enzyme that produce hemiterpenes, is over 65% identical to the five *P. laricio* MTPSs (Table 4.5).

Table 4.5. Amino acid sequence identity (%) matrix of P. laricio MTPS

| | **1.1** | **1.2** | **1.5** | **1.7** | **1.3** | **1.4** |
|---|---|---|---|---|---|---|
| **Pnl MBOS 1.1** | 100 | | | | | |
| **Pnl MTPS 1.2** | 66.8 | 100 | | | | |
| **Pnl MTPS 1.5** | 65.6 | 66.1 | 100 | | | |
| **Pnl MTPS 1.7** | 66.3 | 64.9 | 66.6 | 100 | | |
| **Pnl MTPS 1.3** | 65.5 | 61.4 | 64.3 | 59.4 | 100 | |
| **Pnl MTPS 1.4** | 66.3 | 63.4 | 75.7 | 63.5 | 55.6 | 100 |

The molecular weight of the deduced amino acid sequences of the four FL cDNAs ranged from 72.458 Da to 70.620 Da and as predicted, they have pI that range from 5.48 for Pnl MTPS 1.7 to 6.28 for Pnl MBOS 1.1. Pnl MTPS 1.5 and Pnl MTPS 1.2 possessed very similar pIs at 5.82 and 5.86, respectively.

All the six *P. laricio* predicted proteins contained highly conserved and characteristic regions of plant MTPSs (Fig. 4.11). For instance, each of the four FL predicted proteins included sequences for a putative transit peptide ranging from 40 to 56 aa for import of mature proteins into plastids and they are located upstream from a conserved $RRX_8W$ domain (Fig. 4.11), which is known to be essential for catalysis of monoterpene cyclization (Whittington *et al*., 2002; Hyatt *et al*., 2007). Moreover, all the six *P. laricio* predicted proteins had a conserved Asp-rich domain, DDxxD, that coordinates substrate binding via the formation of divalent cation salt bridges (Tarshis *et al.*, 1996; Lesburg *et al.*, 1997; Starks *et al.*, 1997).

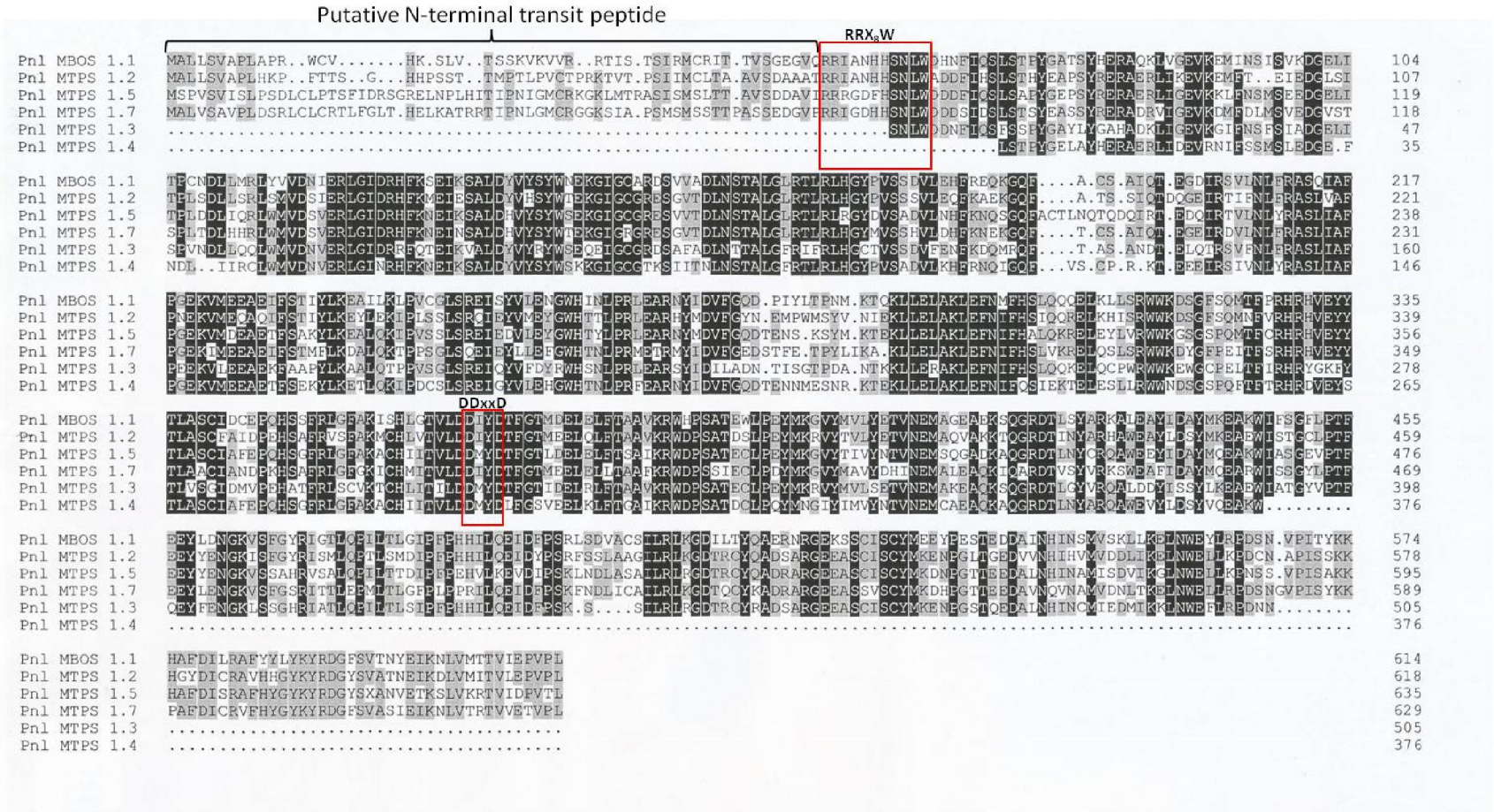



Figure 4.11. Alignment of deduced amino acid sequences of FL cDNAs (Pnl MBOS-1.1, Pnl MTPS-1.2, -1.5 and -1.7) and partial genomic and cDNA sequences (Pnl MTPS-1.3 and-1.4). Residues shaded black indicates highly conserved residues (identical in at least five of the six sequences), and gray shaded residues are identical in at least three of the six sequences shown. The horizontal line indicates the putative N-terminal transit peptide region. The RRX8W motif and the DDxxD motif, which are highly conserved in plant MTPSs and have known functions in the TPS reaction mechanism are indicated with red open rectangles.

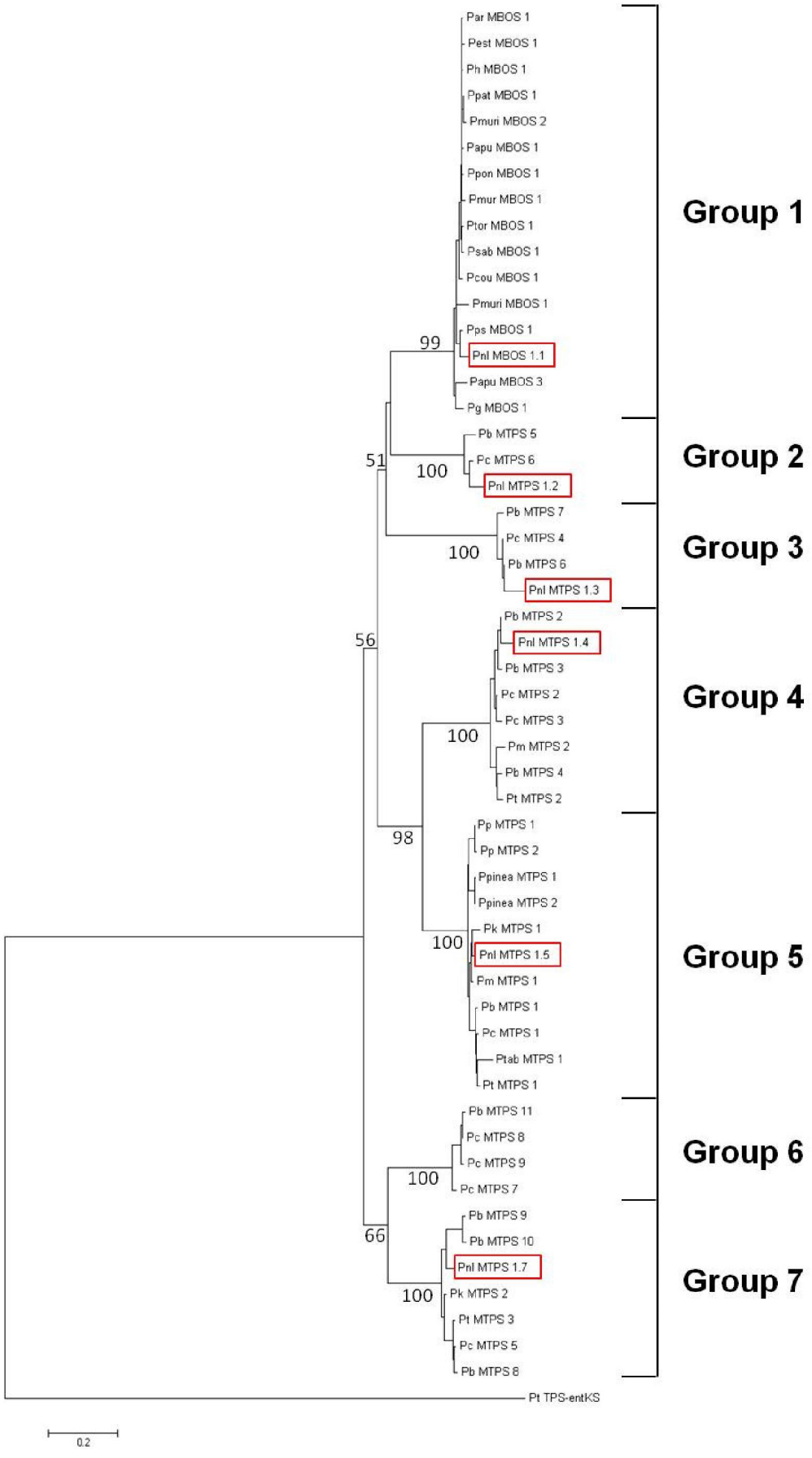
Phylogenetic analysis of Pnl MBOS-1.1, Pnl MTPS-1.2, -1-3, -1-4, -1.5 and -1.7 with the 47 pine MTPSs and MBOSs identified in the NCBI database (see Table 4.2 and Fig. 4.3) placed the six *P. laricio* predicted proteins in six of the seven phylogenetic groups by which the Pinus members of the TPS-d1 clade can be divided (Fig. 4.12), thereby confirming the validity of the approach used for their isolation.

Pnl MBOS 1.1 clustered with the 15 selected pine MBOSs of the phylogenetic group 1 (Fig. 4.12), which showed 93 to 97% amino acid sequence identity. This implies that we isolated a FL transcript from a putative MBOS orthologous gene in *P. laricio*.

Pnl MTPS 1.2 was closely related with two proteins belonging to the phylogenetic group 2 from *P. contorta* (Pc MTPS6) and *P. banksiana* (Pb MTPS5) (Fig. 4.12), that formed α-terpineol as their major product (Hall *et al*., 2013a). These three proteins, that showed 93 to 95% amino acid sequence to each other, had a lower identity of 60 to 62% to Pt MTPS2, a *P. taeda* protein that also produced α-terpineol (Phillips *et al*., 2003), but that was assigned to the phylogenetic group 4 (Fig. 4.12). As previously reported, phylogenetic analyses based on MTPSs from different conifer species indicated that Pc MTPS6 and Pb MTPS5 were more closely related (77% identity) to 1,8-cineole synthases from white spruce and a white spruce hybrid (Hall *et al*., 2013a). All this indicates that only the functional characterization of Pnl MTPS 1.2, via expression of recombinant protein in *E. coli*, may elucidate its true function regardless of the sequence identity detected with other conifer MTPSs.

The incomplete protein Pnl MTPS 1.3 was assigned to the phylogenetic group 3. It showed 92 to 94% amino acid sequence identity with the group 3 proteins from *P. banksiana* (Pb MTPS6-7) and *P. contorta* (Pc MTPS4), that were shown to produce (+)-3-carene as their major product (Hall *et al.,* 2013a). The most remarkable differences among the four putative orthologous pine genes coding for (+)-3-carene synthase was a deletion of 15 bp in the nucleotide sequence of Pnl MTPS 1.3, which determined the loss of 5 amino acids in its C-terminal region (Fig. 4.13).

Figure. 4.12. Phylogenetic tree of the deduced amino acid sequences of 47 MTPSs and MBOSs identified in different *Pinus* species, and the six from *P. laricio* isolated in this study. The ent-kaurene synthase from *Physcomitrella patens* (Pt TPS-entKS, BAF61135) was included in the analysis as outgroup. The numbers on the main branches indicate bootstrap percentages higher than 50% for 1,000 replicates. The seven phylogenetic groups identified in the pine members of TPS-d1 clade are highlighted with square brackets. The *P. laricio* sequences are outlined in red (see next page).


Par MBOS 1
Pest MBOS 1
Ph MBOS 1
Ppat MBOS 1
Pmuri MBOS 2
Papu MBOS 1
Ppon MBOS 1
Pmur MBOS 1
Ptor MBOS 1
Psab MBOS 1
Pcou MBOS 1
Pmuri MBOS 1
Pps MBOS 1
99
Pnl MBOS 1.1
Papu MBOS 3
Pg MBOS 1
Pb MTPS 5
51
Pc MTPS 6
100
Pnl MTPS 1.2
Pb MTPS 7
Pc MTPS 4
100
Pb MTPS 6
Pnl MTPS 1.3
Pb MTPS 2
56
Pnl MTPS 1.4
Pb MTPS 3
Pc MTPS 2
Pc MTPS 3
100
Pm MTPS 2
Pb MTPS 4
Pt MTPS 2
Pp MTPS 1
98
Pp MTPS 2
Ppinea MTPS 1
Ppinea MTPS 2
Pk MTPS 1
100
Pnl MTPS 1.5
Pm MTPS 1
Pb MTPS 1
Pc MTPS 1
Ptab MTPS 1
Pt MTPS 1
Pb MTPS 11
Pc MTPS 8
Pc MTPS 9
100
Pc MTPS 7
Pb MTPS 9
66
Pb MTPS 10
Pnl MTPS 1.7
100
Pk MTPS 2
Pt MTPS 3
Pc MTPS 5
Pb MTPS 8
Pt TPS-entKS
0.2
Group 1
Group 2
Group 3
Group 4
Group 5
Group 6
Group 7

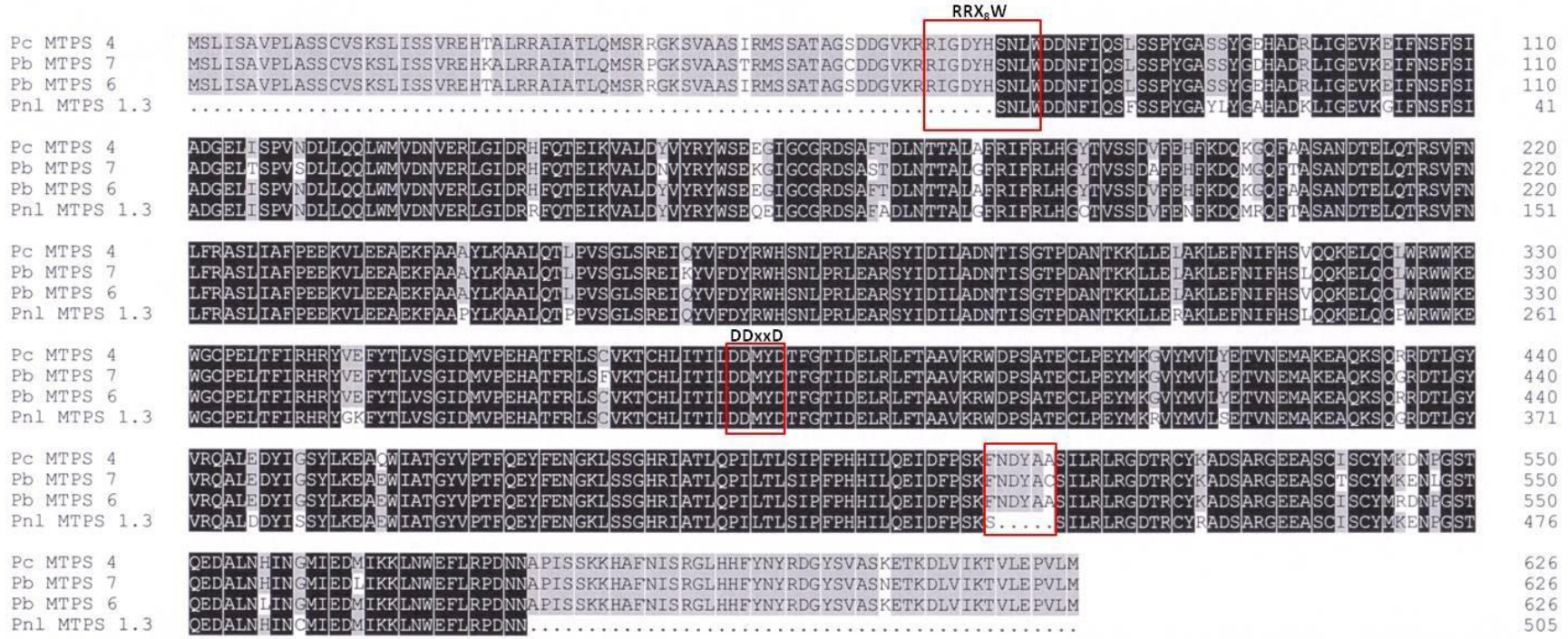



Figure 4.13. Alignment of deduced amino acid sequences of members of the phylogenetic group 3 (Pc MTPS 4, Pb MTPS 6-7 and Pnl MTPS 1.3). Residues shaded black indicates highly conserved residues (identical in all the four sequences), and residues shaded gray are identical in at least three of the four sequences shown). The RRX8W motif and the DDxxD motif, which are highly conserved in plant MTPSs and have known functions in the TPS reaction mechanism, are shown in red open rectangles. The deletion of five aa in the C-terminal region of Pnl MTPS 1.3 is also indicated.

In the phylogenetic group 4, we discovered that the incomplete protein Pnl MTPS 1.4 was closely related (90-94% sequence identity) to five functionally characterized MTPSs from *P. contorta* (Pc MTPS 2), *P. banksiana* (Pb MTPS 2-4) and *P. taeda* (Pt MTPS2) (Fig. 4.12). Pc MTPS 2 and Pb MTPS 2-4 formed (−)-β-pinene as their major product (Hall *et al*., 2013a), whereas Pt MTPS2 produced (−)-α-terpineol and no (−)-β-pinene or (−)-α-pinene (Phillips *et al*., 2003). The high level of sequence identity between these functionally distinct proteins from *P. contorta*, *P. banksiana* and *P. taeda* indicate that only the isolation of FL cDNA Pnl MTPS 1.4 and the functional characterization of the expressed protein may elucidate its true function.

Pnl MTPS 1.5 clustered in the phylogenetic group 5 with ten putative α-pinene synthases for which only three from *P. taeda* (Pt MTPS1), *P. contorta* (Pc MTPS1) and *P. banksiana* (Pb MTPS1) have been functionally characterized as producing (−)-α-pinene as their dominant product (Phillips *et al*., 2003; Hall *et al*., 2013a). Pnl MTPS 1.5, although highly similar to the ten α-pinene synthases (92 to 97% protein sequence identity), showed an insertion of six aa in its N-terminal region not present in any members of the phylogenetic group 5 (Fig. 4.14). The potential involvement of the six aa insertion in modifying the product profile of the corresponding protein will be the subject of future studies.

Finally, Pnl MTPS 1.7 was assigned to the phylogenetic group 7 that contains putative MTPSs coding for (+)-α-pinene synthase (Fig. 4.12). As previously reported, genes that produce (+)-α-pinene as their major product have not been identified in any conifer genus other than *Pinus*, suggesting that this function may have evolved in the pine lineage after the separation from spruce and firs. Moreover, previous phylogenetic analyses showed that *P. contorta*, *P. banksiana* and *P. taeda* (+)-α-pinene synthases (group 7) and (−)-β-phellandrene synthases (Group 6) form a unique and apparently *Pinus* specific subclade within the TPS-d1 clade (Hall *et al*., 2013a).

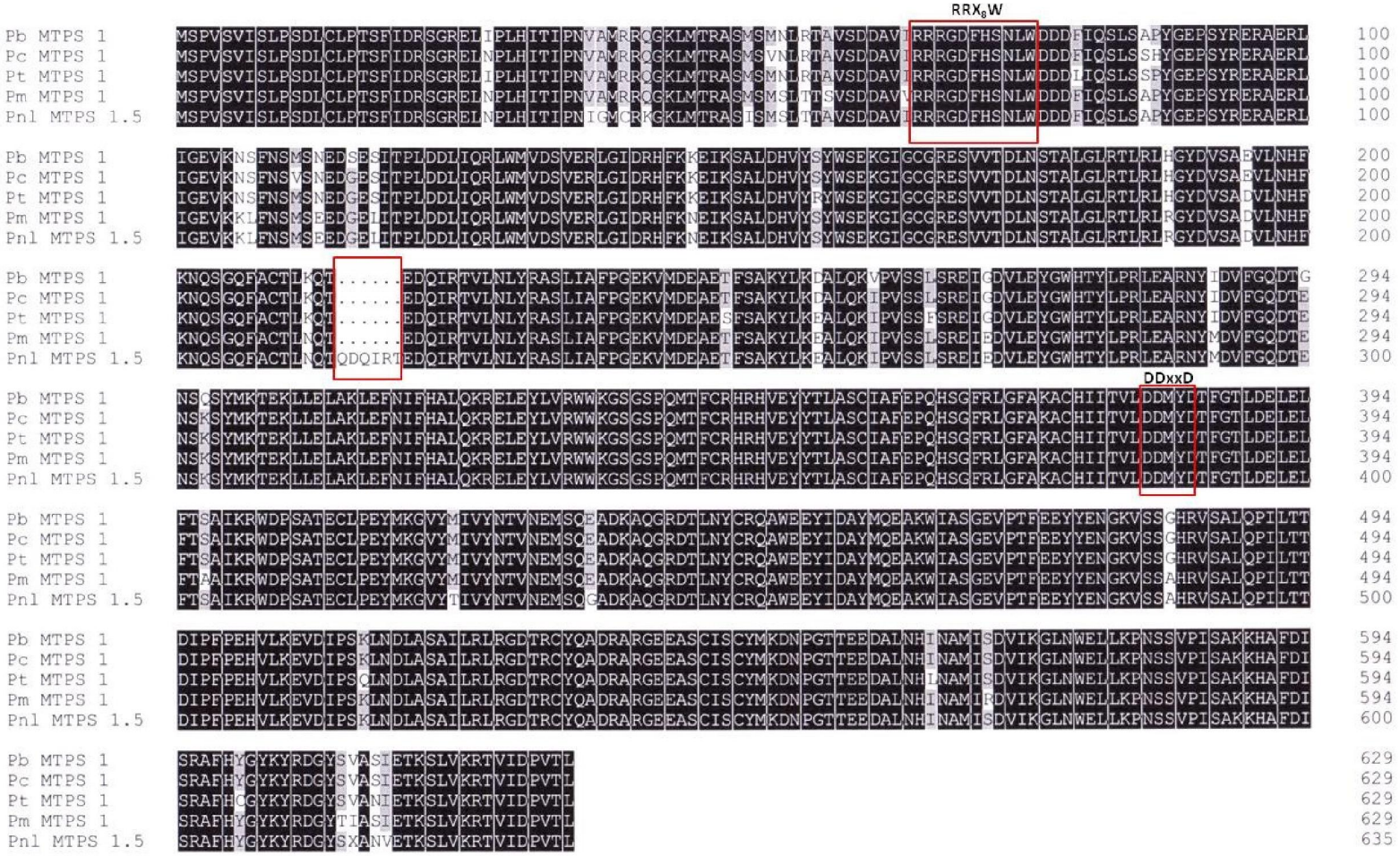



Figure 4.14. Alignment of deduced amino acid sequences of members of the phylogenetic group 5 (Pb MTPS 1, Pc MTPS 1, Pt MTPS1 and Pnl MTPS 1.5). Residues shaded black indicates highly conserved residues (identical in all the five sequences), and residues shaded gray are identical in at least four of the five sequences shown). The RRX8W motif and the DDxxD motif, which are highly conserved in plant MTPSs and have known functions in the TPS reaction mechanism, are indicated with red open rectangles. The insertion of six aa in the N-terminal region of Pnl MTPS 1.5 is also indicated.

## 5. CONCLUSIONS

The importance of terpenes in the physiological and ecological processes in plants like in conifers, as well as their use in the human society at large, cannot be overstated. In a bid to understand the molecular mechanism regulating terpene synthesis, we have faithfully isolated and characterized predictively, some MTPS genes with *Pinus laricio* as the subject, and in doing so, it was necessary that we understand first, the phylogeny of terpene synthase genes that have been identified in all *Pinus* species. This study represents the first attempt to trace the evolutionary history of the *Pinus* members of the large family of terpenes synthase genes of specialized metabolism belonging to the specific gymnosperm clade TPS-d. The phylogenetic analyses have enriched our understanding particularly with the variations that occur in their functions, among TPS members with lower or higher sequence identities.

Since the main aim of the thesis was the isolation and characterization of genes involved in the synthesis of monoterpenes, we further performed a phylogenetic analysis of *Pinus* members of TPS-d1 subclade, containing mainly MTPSs. This phylogenetic analysis allowed the recognition of seven distinct groups and by examining the members of each group for their conserved regions, we were able to design specific primers that were used to isolate partial transcripts belonging to five out of the seven identified phylogenetic groups. Furthermore, partial transcripts from four groups were used as templates to isolate full length (FL) cDNAs. Our success in isolating the FL transcripts coding for MTPSs using this strategy, assures us of the feasibility of employing the same approach to isolate TPS-d members producing diterpenes and sesquiterpenes, that are also very important in plant interactions with the environment. Also, our findings provides a foundation to further examine plant responses, in relation to the involvement or the expression of the MTPS genes, in the event of biotic and abiotic stress factors.

Although we were able to predict the potential functions for the MTPS genes isolated from needles of *P.larico*, it is important to mention that these functions cannot always be relied on, due to the apparent lack of structure-function correlation as was observed in proteins from group 2 members, *P. contorta* (Pc

MTPS6) and *P. banksiana* (Pb MTPS5), and a group 4 member, *P.taeda* (Pt MTPS2) that formed α-terpineol as their major product regardless of having only 62% sequence identity for their proteins.

As we observed from the phylogenetic tree of the deduced amino acid sequences of 47 MTPSs and MBOSs identified in different *Pinus* species, and the six from *P. laricio* isolated in this study, Pnl MTPS 1.5 which clustered in the group 5, showed an insertion of six amino acids in its N-terminal region, which were not present in other members of the same group. Furthermore, we also observed a deletion of five amino acids in the C-terminal region of Pnl MTPS 1.3, assigned to the phylogenetic group 3. As in both cases of Pnl MTPS 1.5 and Pnl MTPS 1.3, despite having very high sequence identities with pine members of the corresponding phylogenetic groups, the substitution, insertion or deletion of few amino acids can be sufficient to alter their product profiles as demonstrated in previous studies. For all these reasons, further characterization of the isolated sequences requiring their heterologous expression in bacterial systems and the subsequent incubation of the corresponding soluble recombinant enzymes with the three main precursors of terpenoids (GPP, FPP or GGPP), is essential to elucidate their true functions.

In some cases, clones of the same cDNA fragment, showed slightly different sequences due to few nucleotide substitutions, as was observed for some partial transcripts obtained for Pnl MBOS 1.1, Pnl MTPS 1.7, Pnl MTPS 1.4, and as earlier discussed, this indicates most likely, the presence of different alleles of the same gene or duplicated copies of the same gene. The implication of this is that we could have as many more MTPS closely related genes belonging to each phylogenetic groups in *P. laricio*, as observed in other *Pinus* species. This possibility will be tested in future studies.

The MTPS genes identified in this study provide a foundation to further investigate the complexity of TPS gene family in *P. laricio*. For instance, the genes coding for enzymes involved in the synthesis of sesqui- and di-terpenes remain to be discovered. The study of the TPS gene family in *P. laricio* and the functional characterization of their members will further help to understand the chemical

diversity of terpenoids in this species, as affected by the interactions with its native environment.

## References


Abbas, F. *et al.* (2017) 'Volatile terpenoids : multiple functions , biosynthesis , modulation and manipulation by genetic engineering', *Planta*. Springer Berlin Heidelberg, 246(5), pp. 803–816. doi: 10.1007/s00425-017-2749-x.

Afzal-Rafii, Z. and Dodd, R. S. (2007) 'Chloroplast DNA supports a hypothesis of glacial refugia over postglacial recolonization in disjunct populations of black pine (Pinus nigra) in western Europe', *Molecular Ecology*, 16(4), pp. 723–736. doi: 10.1111/j.1365-294X.2006.03183.x.

Aharoni, A. *et al.* (2004) 'Gain and Loss of Fruit Flavor Compounds Produced by Wild and Cultivated Strawberry Species', *THE PLANT CELL ONLINE*, 16(11), pp. 3110–3131. doi: 10.1105/tpc.104.023895.

Akiyama, K., Matsuzaki, K. I. and Hayashi, H. (2005) 'Plant sesquiterpenes induce hyphal branching in arbuscular mycorrhizal fungi', *Nature*, 435(7043), pp. 824–827. doi: 10.1038/nature03608.

Aubourg, S., Lecharny, A. and Bohlmann, J. (2002) 'Genomic analysis of the terpenoid synthase (AtTPS) gene family of Arabidopsis thaliana', *Molecular Genetics and Genomics*, 267(6), pp. 730–745. doi: 10.1007/s00438-002-0709-y.

Baldwin, I. T. *et al.* (2006) 'Volatile signaling in plant-plant interactions: "Talking trees" in the genomics era', *Science*, pp. 812–815. doi: 10.1126/science.1118446.

Banfi, E. and Consolino, F. (2011) *Guide Compact - Alberi. Conoscere e riconoscere tutte le specie piu diffuse di Alberi spontane e ornamentali*.

Barreca, L. *et al.* (2009) 'Struttura delle pinete di laricio dell'Etna: il caso della pineta Ragabo [Stand structure of Etna Calabrian pine forests: the Ragabo pine forest case].', in Ciancio, O. (ed.) *Proceedings of the "Terzo Congresso Nazionale di Selvicoltura"*. Taormina (ME, Italy): Accademia Italiana di Scienze Forestali, Firenze, pp. 95–100.

Bogunic, F. *et al.* (2007) 'Genome size stability among five subspecies of Pinus nigra Arnold s.l.', *Environmental and Experimental Botany*, 59(3), pp. 354–360. doi: 10.1016/j.envexpbot.2006.04.006.

Bohlmann, J. *et al.* (1999) 'cDNA cloning, characterization, and functional expression of four new monoterpene synthase members of the Tpsd gene family from grand fir (Abies grandis)', *Archives of Biochemistry and Biophysics*, 368(2), pp. 232–243. doi: 10.1006/abbi.1999.1332.

Bohlmann, J., Meyer-Gauen, G. and Croteau, R. (1998) 'Plant terpenoid synthases: Molecular biology and phylogenetic analysis', *Proceedings of the National Academy of Sciences*, 95(8), pp. 4126–4133. doi: 10.1073/pnas.95.8.4126.

Bohlmann, J. and Steele, C. (1997) 'Monoterpene synthases from grand fir (Abies grandis)', *Journal of Biological Chemistry*, 272(35), pp. 21784–21792. doi: 10.1074/jbc.272.35.21784.

Brewerm, P. B., Koltai, H. and Beveridge, C. A. (2013) 'Diverse roles of strigolactones in plant development.', *Molecular Plant*, 6, pp. 18–28.

Caborn, J. M. (1965) *Shelterbelts and windbreaks.* Faber and Faber, Ltd., London.

Cao, R. *et al.* (2010) 'Diterpene cyclases and the nature of the isoprene fold', *Proteins: Structure, Function and Bioinformatics*, 78(11), pp. 2417–2432. doi: 10.1002/prot.22751.

Cao, R., Zhang, Y. and Mann, F. M. e. al. (2010) 'Diterpene cyclases and the nature of the isoprene fold.', *Proteins: Structure, Function and Bioinformatics*, 78, pp. 2417–2432.

Carrie, C. *et al.* (2007) 'Nine 3-ketoacyl-CoA thiolases (KATs) and acetoacetyl-CoA thiolases (ACATs) encoded by five genes in Arabidopsis thaliana are targeted either to peroxisomes or cytosol but not to mitochondria', *Plant Molecular Biology*, 63(1), pp. 97–108. doi: 10.1007/s11103-006-9075-1.

Caudullo, G., Welk, E. and San-Miguel-Ayanz, J. (2017) 'Chorological maps for the main European woody species', *Data in Brief*, 12, pp. 662–666. doi: 10.1016/j.dib.2017.05.007.

Chen, F. (2003) 'Biosynthesis and Emission of Terpenoid Volatiles from Arabidopsis Flowers', *THE PLANT CELL ONLINE*, 15(2), pp. 481–494. doi: 10.1105/tpc.007989.

Chen, F. (2004) 'Characterization of a Root-Specific Arabidopsis Terpene Synthase Responsible for the Formation of the Volatile Monoterpene 1,8-Cineole', *PLANT PHYSIOLOGY*, 135(4), pp. 1956–1966. doi: 10.1104/pp.104.044388.

Chen, F. *et al.* (2011) 'The family of terpene synthases in plants: A mid-size family of genes for specialized metabolism that is highly diversified throughout the kingdom', *Plant Journal*, 66(1), pp. 212–229. doi: 10.1111/j.1365-313X.2011.04520.x.

Chen, F. and Tholl, D. (2011) 'The family of terpene synthases in plants : a mid-size family of genes for specialized metabolism that is highly diversified throughout the kingdom', pp. 212–229. doi: 10.1111/j.1365-313X.2011.04520.x.

Cho, E. M. *et al.* (2004) 'Molecular cloning and characterization of a cDNA encoding ent-cassa-12,15-diene synthase, a putative diterpenoid phytoalexin biosynthetic enzyme, from suspension-cultured rice cells treated with a chitin elicitor', *Plant Journal*, 37(1), pp. 1–8. doi: 10.1046/j.1365-313X.2003.01926.x.

Christensen, K. I. (1993) 'Comments on the earliest validly published varietal name for the Corsican pine', *Taxon*, 42(3), pp. 649–653.

Christianson, D. W. (2017) 'Structural and Chemical Biology of Terpenoid Cyclases', *Chemical Reviews*, pp. 11570–11648. doi: 10.1021/acs.chemrev.7b00287.

Ciancio, O. *et al.* (2006) 'Structure and growth of a small group selection forest of calabrian pine in Southern Italy: A hypothesis for continuous cover forestry based on traditional silviculture', in *Forest Ecology and Management*, pp. 229–234. doi: 10.1016/j.foreco.2005.12.057.

Colby, S. M. *et al.* (1993) '4S-limonene synthase from the oil glands of spearmint (Mentha spicata). cDNA isolation, characterization, and bacterial expression of the catalytically active monoterpene cyclase', *Journal of Biological Chemistry*, 268(31), pp. 23016–23024.

Collalti, A. *et al.* (2017) 'Protocollo di simulazione, calibrazione e validazione del modello 3D-CMCCCNR-FEM: il caso studio del bacino altamente strumentato del Bonis in Calabria', *@Forest*, 14, p. 247–256.

Dallimore, W. and Bruce Jackson, A. (1966) *A handbook of Coniferae and Ginkgoaceae.* 4th edn. Edited by S. G. Harrison. Edward Arnold, Ltd., London.

Davis, E. M. and Croteau, R. (2000) 'Cyclization enzymes in the biosynthesis of monoterpenes, sesquiterpenes, and diterpenes.', in *Topics in current chemistry: biosynthesis-aromatic polyketides, isoprenoids, alkaloids*. Springer, Heidelberg, pp. 53–95.

Debazac, E. F. (1965) 'Les pineraies de Calabre et de Sicile [Pine forests of Calabria and Sicily]. Revue Forestière Française', *Revue Forestière Française*, 17(10), pp. 662–673. doi: 10.4267/2042/24690.

Dicke, M. *et al.* (1990) 'Isolation and identification of volatile kairomone that affects acarine predator-prey interactions', *Journal of Chemical Ecology*, 16(2), pp. 381–396. doi: 10.1007/BF01021772.

Dudareva, N. (1996) 'Evolution of Floral Scent in Clarkia: Novel Patterns of S-Linalool Synthase Gene Expression in the C. breweri Flower', *THE PLANT CELL ONLINE*, 8(7), pp. 1137–1148. doi: 10.1105/tpc.8.7.1137.

Dudareva, N. (2003) '(E)-beta-Ocimene and Myrcene Synthase Genes of Floral Scent Biosynthesis in Snapdragon: Function and Expression of Three Terpene Synthase Genes of a New Terpene Synthase Subfamily', *THE PLANT CELL ONLINE*, 15(5), pp. 1227–1241. doi: 10.1105/tpc.011015.

Dudareva, N. *et al.* (2006) 'Plant volatiles: Recent advances and future perspectives', *Critical Reviews in Plant Sciences*, pp. 417–440. doi: 10.1080/07352680600899973.

Dudareva, N. and Pichersky, E. (2000) 'Biochemical and Molecular Genetic Aspects of Floral Scents', *Plant Physiology*, 122(3), pp. 627–633. doi: 10.1104/pp.122.3.627.

Enescu, C. M. *et al.* (2016) 'Pinus nigra in Europe: distribution, habitat, usage and threats.', in *European Atlas of Forest Tree Species*. Off. EU, Luxembourg, p. e015138+.

Estévez, J. M. *et al.* (2001) '1-Deoxy-D-xylulose-5-phosphate Synthase, a Limiting Enzyme for Plastidic Isoprenoid Biosynthesis in Plants', *Journal of Biological Chemistry*, 276(25), pp. 22901–22909. doi: 10.1074/jbc.M100854200.

Falara, V., Pichersky, E. and Kanellis, A. K. (2010) 'A Copal-8-ol Diphosphate Synthase from the Angiosperm *Cistus creticus* subsp. *creticus* Is a Putative Key Enzyme for the Formation of Pharmacologically Active, Oxygen-Containing Labdane-Type Diterpenes', *Plant Physiology*, 154(1), pp. 301–310. doi: 10.1104/pp.110.159566.

Farjon, A. (2013) 'Pinus nigra - Version 2014.3', in *The IUCN Red List of Threatened Species*, p. 42386/0. doi: citeulike-article-id:13621289.

Farré-Armengol, G. *et al.* (2013) 'Floral volatile organic compounds: Between attraction and deterrence of visitors under global change', *Perspectives in Plant Ecology, Evolution and Systematics*, pp. 56–67. doi: 10.1016/j.ppees.2012.12.002.

Flügge, U. I. and Gao, W. (2005) 'Transport of isoprenoid intermediates across chloroplast envelope membranes', *Plant Biology*, 7(1), pp. 91–97. doi: 10.1055/s-2004-830446.

Gambino, G., Perrone, I. and Gribaudo, I. (2008) 'A rapid and effective method for RNA extraction from different tissues of grapevine and other woody plants', *Phytochemical Analysis*, 19(6), pp. 520–525. doi: 10.1002/pca.1078.

Gao, Y., Honzatko, R. B. and Peters, R. J. (2012) 'Terpenoid synthase structures: A so far incomplete view of complex catalysis', *Natural Product Reports*, pp. 1153–1175. doi: 10.1039/c2np20059g.

Gennadios, H. A. *et al.* (2009) 'Crystal structure of (+)-δ-cadinene synthase from Gossypium arboreum and evolutionary divergence of metal binding motifs for catalysis', *Biochemistry*, 48(26), pp. 6175–6183. doi: 10.1021/bi900483b.

Gernandt, D. S. *et al.* (2008) 'Use of Simultaneous Analyses to Guide Fossil-Based Calibrations of Pinaceae Phylogeny', *International Journal of Plant Sciences*, 169(8), pp. 1086–1099. doi: 10.1086/590472.

Gershenzon, J. and Dudareva, N. (2007) 'The function of terpene natural products in the natural world', *Nature Chemical Biology*, pp. 408–414. doi: 10.1038/nchembio.2007.5.

Gray, D. W. *et al.* (2011) 'Biochemical characterization and homology modeling of methylbutenol synthase and implications for understanding hemiterpene synthase evolution in plants', *Journal of Biological Chemistry*, 286(23), pp. 20582–20590. doi: 10.1074/jbc.M111.237438.

Green, S. *et al.* (2007) 'Unusual features of a recombinant apple alpha-farnesene synthase.', *Phytochemistry*, 68(2), pp. 176–88. doi: 10.1016/j.phytochem.2006.10.017.

Green, S. *et al.* (2009) 'Defining the potassium binding region in an apple terpene synthase', *Journal of Biological Chemistry*, 284(13), pp. 8661–8669. doi: 10.1074/jbc.M807140200.

Hall, D. E. *et al.* (2013) 'Evolution of Conifer Diterpene Synthases: Diterpene Resin Acid Biosynthesis in Lodgepole Pine and Jack Pine Involves Monofunctional and Bifunctional Diterpene Synthases', *PLANT PHYSIOLOGY*, 161(2), pp. 600–616. doi: 10.1104/pp.112.208546.

Hall, D. E. *et al.* (2013) 'Transcriptome resources and functional characterization of monoterpene synthases for two host species of the mountain pine beetle , lodgepole pine ( Pinus contorta ) and jack pine ( Pinus banksiana )', *BMC Plant Biology*. BMC Plant Biology, 13(1), p. 1. doi: 10.1186/1471-2229-13-80.

Hartmann, T. (2007) 'From waste products to ecochemicals: fifty years research of plant secondary metabolism.', *Phytochemistry*, 68, p. 2831–2846.

Hayashi, K. -i. *et al.* (2010) 'Endogenous Diterpenes Derived from ent-Kaurene, a Common Gibberellin Precursor, Regulate Protonema Differentiation of the Moss Physcomitrella patens', *PLANT PHYSIOLOGY*, 153(3), pp. 1085–1097. doi: 10.1104/pp.110.157909.

Hemmerlin, A. *et al.* (2003) 'Cross-talk between the cytosolic mevalonate and the plastidial methylerythritol phosphate pathways in tobacco bright yellow-2 cells', *Journal of Biological Chemistry*, 278(29), pp. 26666–26676. doi: 10.1074/jbc.M302526200.

Hemmerlin, A. (2013) 'Post-translational events and modifications regulating plant enzymes

involved in isoprenoid precursor biosynthesis', *Plant Science*, pp. 41–54. doi: 10.1016/j.plantsci.2012.12.008.

Hemmerlin, A., Harwood, J. L. and Bach, T. J. (2012) 'A raison d'être for two distinct pathways in the early steps of plant isoprenoid biosynthesis?', *Progress in Lipid Research*, pp. 95–148. doi: 10.1016/j.plipres.2011.12.001.

Herde, M. *et al.* (2008) 'Identification and Regulation of TPS04/GES, an Arabidopsis Geranyllinalool Synthase Catalyzing the First Step in the Formation of the Insect-Induced Volatile C16-Homoterpene TMTT', *THE PLANT CELL ONLINE*, 20(4), pp. 1152–1168. doi: 10.1105/tpc.106.049478.

Hillwig, M. L. *et al.* (2011) 'Domain loss has independently occurred multiple times in plant terpene synthase evolution', *Plant Journal*, 68(6), pp. 1051–1060. doi: 10.1111/j.1365-313X.2011.04756.x.

Huang, M. *et al.* (2010) 'Variation of Herbivore-Induced Volatile Terpenes among Arabidopsis Ecotypes Depends on Allelic Differences and Subcellular Targeting of Two Terpene Synthases, TPS02 and TPS03', *PLANT PHYSIOLOGY*, 153(3), pp. 1293–1310. doi: 10.1104/pp.110.154864.

Huber, D. P. W. *et al.* (2005) 'Characterization of four terpene synthase cDNAs from methyl jasmonate-induced Douglas-fir, Pseudotsuga menziesii', *Phytochemistry*, 66(12), pp. 1427–1439. doi: 10.1016/j.phytochem.2005.04.030.

Hyatt, D. C. *et al.* (2007) 'Structure of limonene synthase, a simple model for terpenoid cyclase catalysis', *Proceedings of the National Academy of Sciences*, 104(13), pp. 5360–5365. doi: 10.1073/pnas.0700915104.

Hyatt, D. C. and Croteau, R. (2005) 'Mutational analysis of a monoterpene synthase reaction: Altered catalysis through directed mutagenesis of (-)-pinene synthase from Abies grandis', *Archives of Biochemistry and Biophysics*, 439(2), pp. 222–233. doi: 10.1016/j.abb.2005.05.017.

Iovino, F. and Menguzzato, G. (1996) 'A return to complex formations through forest management.', in Ciancio O ed (ed.) *The forest and man*. Florence, Italy.: Accademia Italiana di Scienze Forestali, Tipografia Coppini., pp. 211–221.

Iovino, F. and Menguzzato, G. (2002) 'Disboscamento e ripristino del manto boschivo nell'Appennino calabrese [Deforestation and retrieval of forest cover in the Calabrian Apennines].', in Lazzarini, A. and Angeli Storia, F. (eds) *Disboscamento montano e politiche territoriali. Alpi e Appennini dal Settecento al Duemila*, pp. 494–509.

Irmisch, S., Jiang, Y., Chen, F., Gershenzon, J. and Köllner, T. G. (2014) 'Terpene synthases and their contribution to herbivore-induced volatile emission in western balsam poplar ( Populus trichocarpa )', pp. 1–16.

Irmisch, S., Jiang, Y., Chen, F., Gershenzon, J. and Köllner, T. G. (2014) 'Terpene synthases and their contribution to herbivore-induced volatile emission in western balsam poplar (Populus trichocarpa)', *BMC Plant Biology*, 14(1). doi: 10.1186/s12870-014-0270-y.

Isajev, V. *et al.* (2004) *EUFORGEN Technical Guidelines for genetic conserva- tion and use for European black pine (Pinus nigra).*, *International Plant Genetic Resources Institute, Rome, Italy*.

Islam, A. *et al.* (2003) 'An antimicrobial terpenoid from Caesalpinia pulcherrima Swartz.: its characterization, antimicrobial and cytotoxic activities.', *Asian Journal of Plant Science*, 2, pp. 17–24.

J, B. (2004) *Dictionary of natural products web version*. Chapman and Hall, London. Available at: http://www.chemnetbase.com.

Jia, Q. (2016) 'Computational Identification of Terpene Synthase Genes and Their Evolutionary Analysis'.

Jones, C. G. *et al.* (2011) 'Sandalwood fragrance biosynthesis involves sesquiterpene synthases of both the terpene synthase (TPS)-a and TPS-b subfamilies, including santalene synthases', *Journal of Biological Chemistry*, 286(20), pp. 17445–17454. doi: 10.1074/jbc.M111.231787.

Kampranis, S. C. *et al.* (2007) 'Rational Conversion of Substrate and Product Specificity in a *Salvia* Monoterpene Synthase: Structural Insights into the Evolution of Terpene Synthase Function', *The Plant Cell*, 19(6), pp. 1994–2005. doi: 10.1105/tpc.106.047779.

Katoh, S., Hyatt, D. and Croteau, R. (2004) 'Altering product outcome in Abies grandis (-)-limonene synthase and (-)-limonene/(-)-α-pinene synthase by domain swapping and directed mutagenesis', *Archives of Biochemistry and Biophysics*, 425(1), pp. 65–76. doi: 10.1016/j.abb.2004.02.015.

Keeling, C. I. (2006) 'Genes , enzymes and chemicals of terpenoid diversity in the constitutive and induced defence of conifers against insects and pathogens *', pp. 657–675.

Keeling, C. I. *et al.* (2008) 'Functional plasticity of paralogous diterpene synthases involved in conifer defense', *Proceedings of the National Academy of Sciences*, 105(3), pp. 1085–1090. doi: 10.1073/pnas.0709466105.

Keeling, C. I. *et al.* (2010) 'Identification and Functional Characterization of Monofunctional ent-Copalyl Diphosphate and ent-Kaurene Synthases in White Spruce Reveal Different Patterns for Diterpene Synthase Evolution for Primary and Secondary Metabolism in Gymnosperms', *PLANT PHYSIOLOGY*, 152(3), pp. 1197–1208. doi: 10.1104/pp.109.151456.

Keeling, C. I. *et al.* (2011) 'Transcriptome mining , functional characterization , and phylogeny of a large terpene synthase gene family in spruce ( Picea spp .)'.

Keeling, C. I. and Bohlmann, J. (2006) 'Genes, enzymes and chemicals of terpenoid diversity in the constitutive and induced defence of conifers against insects and pathogens', *New Phytologist*, 170(4), pp. 657–675. doi: 10.1111/j.1469-8137.2006.01716.x.

Kessler, A. and Baldwin, I. T. (2001) 'Defensive function of herbivore-induced plant volatile emissions in nature', *Science*, 291(5511), pp. 2141–2144. doi: 10.1126/science.291.5511.2141.

Köksal, M. *et al.* (2010) 'Structure of isoprene synthase illuminates the chemical mechanism of teragram atmospheric carbon emission', *Journal of Molecular Biology*, 402(2), pp. 363–373. doi: 10.1016/j.jmb.2010.07.009.

Köksal, M. *et al.* (2011) 'Taxadiene synthase structure and evolution of modular architecture in terpene biosynthesis', *Nature*, 469(7328), pp. 116–122. doi: 10.1038/nature09628.

Köksal, M. *et al.* (2014) '1.55 Å-resolution structure of ent-copalyl diphosphate synthase and exploration of general acid function by site-directed mutagenesis', *Biochimica et Biophysica Acta - General Subjects*, 1840(1), pp. 184–190. doi: 10.1016/j.bbagen.2013.09.004.

KÖksal, M. *et al.* (2011) 'Structure and mechanism of the diterpene cyclase ent-copalyl diphosphate synthase', *Nature Chemical Biology*, 7(7), pp. 431–433. doi: 10.1038/nchembio.578.

Köpke, D. *et al.* (2008) 'Does egg deposition by herbivorous pine sawflies affect transcription of sesquiterpene synthases in pine?', *Planta*, 228(3), pp. 427–438. doi: 10.1007/s00425-008-0747-8.

Köpke, D. *et al.* (2010) 'Species-specific responses of pine sesquiterpene synthases to sawfly oviposition', *Phytochemistry*. doi: 10.1016/j.phytochem.2010.03.017.

Külheim, C. *et al.* (2015) 'The Eucalyptus terpene synthase gene family', *BMC Genomics*, 16(1). doi: 10.1186/s12864-015-1598-x.

Kumar, S., Stecher, G. and Tamura, K. (2016) 'MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets', *Molecular biology and evolution*, 33(7), pp. 1870–1874. doi: 10.1093/molbev/msw054.

Landmann, C. *et al.* (2007) 'Cloning and functional characterization of three terpene synthases from lavender (Lavandula angustifolia)', *Archives of Biochemistry and Biophysics*, 465(2), pp. 417–429. doi: 10.1016/j.abb.2007.06.011.

Laothawornkitkul, J. *et al.* (2008) 'Isoprene emissions influence herbivore feeding decisions.', *Plant Cell Environ*, 31, pp. 1410–1415.

Lauranson-Broyer, J. Lebreton, P. (1995) 'Flavonic chemosystematics of the specific complex Pinus nigra Arnold', in Baradat, P., Adams, W. T., and Müller-Starck, G. (eds) *Population genetics and genetic conservation of forest trees*. Amsterdam: SPB Academic Publishing, pp. 181–188.

Lee, S. and Chappell, J. (2008) 'Biochemical and Genomic Characterization of Terpene Synthases in Magnolia grandiflora', *PLANT PHYSIOLOGY*, 147(3), pp. 1017–1033. doi: 10.1104/pp.108.115824.

Leivar, P. *et al.* (2005) 'Subcellular localization of Arabidopsis 3-hydroxy-3-methylglutaryl-coenzyme A reductase.', *Plant physiology*, 137(1), pp. 57–69. doi: 10.1104/pp.104.050245.

Lerdau, M. and Gray, D. (2003) 'Ecology and evolution of light- dependent and light-independent phytogenic volatile organic carbon', *New Phytologist*, 157, pp. 199–211. doi: 10.1046/j.1469-8137.2003.00673.x.

Lesburg, C. A. *et al.* (1997) 'Crystal structure of pentalenene synthase: Mechanistic insights on terpenoid cyclization reactions in biology', *Science*, 277(5333), pp. 1820–1824. doi: 10.1126/science.277.5333.1820.

Li, C. and Yeh, F. (2002) 'QTLs for Western Gall Rust (Endocronartium Harknessii) Resistance in Lodgepole Pine (P. contorta Spp. Latifolia)', 9(2), pp. 137–144.

Li, G. *et al.* (2012) 'Nonseed plant Selaginella moellendorffi has both seed plant and microbial

types of terpene synthases.', *Proceedings of the National Academy of Sciences of the United States of America*, 109(36), pp. 14711–5. doi: 10.1073/pnas.1204300109.

Little, D. B. and Croteau, R. B. (2002) 'Alteration of product formation by directed mutagenesis and truncation of the multiple-product sesquiterpene synthases δ-selinene synthase and γ-humulene synthase', *Archives of Biochemistry and Biophysics*, 402(1), pp. 120–135. doi: 10.1016/S0003-9861(02)00068-1.

Lois, L. M. *et al.* (2000) 'Carotenoid biosynthesis during tomato fruit development: Regulatory role of 1-deoxy-D-xylulose 5-phosphate synthase', *Plant Journal*, 22(6), pp. 503–513. doi: 10.1046/j.1365-313X.2000.00764.x.

Ludwiczuk, A., Skalicka-Woźniak, K. and Georgiev, M. I. (2017) 'Terpenoids', in *Pharmacognosy*, pp. 233–266. doi: 10.1016/B978-0-12-802104-0.00011-1.

Markus Lange, B. and Ahkami, A. (2013) 'Metabolic engineering of plant monoterpenes, sesquiterpenes and diterpenes-current status and future opportunities', *Plant Biotechnology Journal*, pp. 169–196. doi: 10.1111/pbi.12022.

Martin, D., Fäldt, J. and Bohlmann, J. (2004) 'Functional Characterization of Nine Norway SpruceTPS Genes and Evolution of GymnospermTerpene Synthases of the TPS-d Subfamily', *Plant Physiology*, 135(August), pp. 1908–1927. doi: 10.1104/pp.104.042028.1908.

Martin, D. M. *et al.* (2010) 'Functional Annotation, Genome Organization and Phylogeny of the Grapevine (Vitis vinifera) Terpene Synthase Gene Family Based on Genome Assembly, FLcDNA Cloning, and Enzyme Assays', *BMC Plant Biology*, 10. doi: 10.1186/1471-2229-10-226.

Mbaveng, A. T., Hamm, R. and Kuete, V. (2014) 'Harmful and Protective Effects of Terpenoids from African Medicinal Plants', in *Toxicological Survey of African Medicinal Plants*, pp. 557–576. doi: 10.1016/B978-0-12-800018-2.00019-4.

McAndrew, R. P. *et al.* (2011) 'Structure of a three-domain sesquiterpene synthase: A prospective target for advanced biofuels production', *Structure*, 19(12), pp. 1876–1884. doi: 10.1016/j.str.2011.09.013.

Merret, R. *et al.* (2007) 'A serine involved in actin-dependent subcellular localization of a stress-induced tobacco BY-2 hydroxymethylglutaryl-CoA reductase isoform', *FEBS Letters*, 581(27), pp. 5295–5299. doi: 10.1016/j.febslet.2007.10.023.

Morrone, D. *et al.* (2009) 'Gibberellin biosynthesis in bacteria: Separate ent-copalyl diphosphate and ent-kaurene synthases in Bradyrhizobium japonicum', *FEBS Letters*, 583(2), pp. 475–480. doi: 10.1016/j.febslet.2008.12.052.

Muhlemann, J. K., Klempien, A. and Dudareva, N. (2014) 'Floral volatiles: From biosynthesis to function', *Plant, Cell and Environment*, 37(8), pp. 1936–1949. doi: 10.1111/pce.12314.

Myburg, A. *et al.* (2014) 'The genome of *Eucalyptus grandis*', *Nature*, 509, pp. 356–362. doi: 10.1038/nature13308.

Nagegowda, D. A. *et al.* (2008) 'Two nearly identical terpene synthases catalyze the formation of nerolidol and linalool in snapdragon flowers', *Plant Journal*, 55(2), pp. 224–239. doi:

10.1111/j.1365-313X.2008.03496.x.

Naydenov, K. D. *et al.* (2006) 'Structure of Pinus nigra Arn. populations in Bulgaria revealed by chloroplast microsatellites and terpenes analysis: Provenance tests', *Biochemical Systematics and Ecology*, 34(7), pp. 562–574. doi: 10.1016/j.bse.2006.01.011.

Nicolaci, A. *et al.* (2014) 'Ecological and anthropogenic drivers of Calabrian pine (Pinus nigra J.F. Arn. ssp. laricio (Poiret) Maire) distribution in the Sila mountain range', *IForest*, 8, pp. 497–508.

Nicolaci, A. *et al.* (2015) 'Ecological and anthropogenic drivers of Calabrian pine (Pinus nigra J.F. Arn. ssp. laricio (Poiret) Maire) distribution in the Sila mountain range', *IForest*, 8(AUG2015), pp. 497–508. doi: 10.3832/ifor1041-007.

Nikolic, D. and Tucik, N. (1983) 'Isoenzyme variation within and among populations of European black pines (Pinus nigra Arnold).', *Silvae Genetica*, 32(3–4), pp. 80–89.

Orlova, I. *et al.* (2009) 'The Small Subunit of Snapdragon Geranyl Diphosphate Synthase Modifies the Chain Length Specificity of Tobacco Geranylgeranyl Diphosphate Synthase in Planta', *The Plant Cell*, 21(12), pp. 4002–4017. doi: 10.1105/tpc.109.071282.

Otomo, K. *et al.* (2004) 'Biological functions of ent- and syn-copalyl diphosphate synthases in rice: Key enzymes for the branch point of gibberellin and phytoalexin biosynthesis', *Plant Journal*, 39(6), pp. 886–893. doi: 10.1111/j.1365-313X.2004.02175.x.

Pazouki, L. and Niinemets, Ü. (2016) 'Multi-Substrate Terpene Synthases: Their Occurrence and Physiological Significance', *Frontiers in Plant Science*, 7(July), pp. 1–16. doi: 10.3389/fpls.2016.01019.

Peters, R. J. *et al.* (2000) 'Abietadiene synthase from grand fir (Abies grandis): Characterization and mechanism of action of the "pseudomature" recombinant enzyme', *Biochemistry*, 39(50), pp. 15592–15602. doi: 10.1021/bi001997l.

Peters, R. J. and Croteau, R. B. (2002) 'Abietadiene synthase catalysis: Conserved residues involved in protonation-initiated cyclization of geranylgeranyl diphosphate to (+)-copalyl diphosphate', *Biochemistry*, 41(6), pp. 1836–1842. doi: 10.1021/bi011879d.

Phillips, M. A. *et al.* (2003) 'cDNA isolation, functional expression, and characterization of (+)-α-pinene synthase and (-)-α-pinene synthase from loblolly pine (Pinus taeda): Stereocontrol in pinene biosynthesis', *Archives of Biochemistry and Biophysics*, 411(2), pp. 267–276. doi: 10.1016/S0003-9861(02)00746-4.

Poli Marchese, E. (1982) 'Poli Marchese E (1982). Zonation altitudinale de la végétation de l'Etna comparée avec celle d'autres hauts volcans', in *Workshop NATO*. Saint Maximin (Marseille, France), pp. 16–20.

Prisic, S. (2004) 'Rice Contains Two Disparate ent-Copalyl Diphosphate Synthases with Distinct Metabolic Functions', *PLANT PHYSIOLOGY*, 136(4), pp. 4228–4236. doi: 10.1104/pp.104.050567.

Prosser, I. *et al.* (2004) 'Enantiospecific (+)- and (-)-germacrene D synthases, cloned from goldenrod, reveal a functionally active variant of the universal isoprenoid-biosynthesis

aspartate-rich motif', *Archives of Biochemistry and Biophysics*, 432(2), pp. 136–144. doi: 10.1016/j.abb.2004.06.030.

Raab-Straube, E. (2014) '{Abies alba Mill.}', in *Euro+Med Plantbase - the information resource for Euro-Mediterranean plant diversity*. doi: citeulike-article-id:14266522.

Rajabi, M. H., Pazouki, L. and Niinemets, Ü. (2013) 'The biochemistry and molecular biology of volatile messengers in trees.', in Niinemetsand, Ü. and Monson, R. K. (eds) *Biology, Controls and Models of Tree Volatile Organic Compound Emissions*. Dordrecht: Springer, pp. 47–93.

Rasulov, B. *et al.* (2015) 'Bisphosphonate Inhibitors Reveal a Large Elasticity of Plastidic Isoprenoid Synthesis Pathway in Isoprene-Emitting Hybrid Aspen', *Plant Physiology*, 168(2), pp. 532–548. doi: 10.1104/pp.15.00470.

Ro, D.-K. *et al.* (2005) 'Loblolly pine abietadienol/abietadienal oxidase PtAO (CYP720B1) is a multifunctional, multisubstrate cytochrome P450 monooxygenase', *Proceedings of the National Academy of Sciences*, 102(22), pp. 8060–8065. doi: 10.1073/pnas.0500825102.

Ro, D. K. and Bohlmann, J. (2006) 'Diterpene resin acid biosynthesis in loblolly pine (Pinus taeda): Functional characterization of abietadiene/levopimaradiene synthase (PtTPS-LAS) cDNA and subcellular targeting of PtTPS-LAS and abietadienol/abietadienal oxidase (PtAO, CYP720B1)', *Phytochemistry*, 67(15), pp. 1572–1578. doi: 10.1016/j.phytochem.2006.01.011.

Rynkiewicz, M. J., Cane, D. E. and Christianson, D. W. (2002) 'X-ray crystal structures of D100E trichodiene synthase and its pyrophosphate complex reveal the basis for terpene product diversity', *Biochemistry*, 41(6), pp. 1732–1741. doi: 10.1021/bi011960g.

Saitou, N. and Nei, M. (1987) 'The neighbor-joining method: a new method for reconstructing phylogenetic trees', *Mol Biol Evo*, 4(4), pp. 406–425. doi: doi: 10.1093/.

Seemann, M. *et al.* (2002) 'Pentalenene Synthase. Analysis of Active Site Residues by Site-Directed Mutagenesis', *Journal of the American Chemical Society*, 124(26), pp. 7681–7689. doi: 10.1021/ja026058q.

Simkin, A. J. *et al.* (2011) 'Peroxisomal localisation of the final steps of the mevalonic acid pathway in planta', *Planta*, 234(5), pp. 903–914. doi: 10.1007/s00425-011-1444-6.

Singh, B. and Sharma, R. A. (2015) 'Plant terpenes : defense responses , phylogenetic analysis , regulation and clinical applications', pp. 129–151. doi: 10.1007/s13205-014-0220-2.

Singh, B. and Sharma, R. A. (2015) 'Plant terpenes: defense responses, phylogenetic analysis, regulation and clinical applications.', *3 Biotech*, 5(2), pp. 129–151. doi: 10.1007/s13205-014-0220-2.

Starks, C. M. *et al.* (1997) 'Structural basis for cyclic terpene biosynthesis by tobacco 5-epi-aristolochene synthase', *Science*, 277(5333), pp. 1815–1820. doi: 10.1126/science.277.5333.1815.

Steele, C. L. *et al.* (1998) 'Sesquiterpene synthases from grand fir (Abies grandis): Comparison of constitutive and wound-induced activities, and cDNA isolation, characterization, and bacterial expression of δ-selinene synthase and γ- humulene synthase', *Journal of Biological Chemistry*,

273(4), pp. 2078–2089. doi: 10.1074/jbc.273.4.2078.

Steele, C. L. *et al.* (1998) 'Sesquiterpene synthases from grand fir (Abies grandis). Comparison of constitutive and wound-induced activities, and cDNA isolation, characterization, and bacterial expression of delta-selinene synthase and gamma-humulene synthase.', *Journal of Biological Chemistry*, 273, pp. 2078–2089.

Tarshis, L. C. *et al.* (1996) 'Regulation of product chain length by isoprenyl diphosphate synthases.', *Proceedings of the National Academy of Sciences of the United States of America*, 93(26), pp. 15018–15023. doi: 10.1073/pnas.93.26.15018.

Thiel, D. *et al.* (2012) 'Uniform drought and warming responses in Pinus nigra provenances despite specific overall performances', *Forest Ecology and Management*, 270, pp. 200–208. doi: 10.1016/j.foreco.2012.01.034.

Tholl, D. (2015) 'Biosynthesis and biological functions of terpenoids in plants', *Advances in Biochemical Engineering/Biotechnology*, 148, pp. 63–106. doi: 10.1007/10_2014_295.

Tholl, D. and Lee, S. (2011) 'Terpene Specialized Metabolism in *Arabidopsis thaliana*', *The Arabidopsis Book*, 9, p. e0143. doi: 10.1199/tab.0143.

Thompson, J. D. *et al.* (1997) 'The CLUSTAL X windows interface: Flexible strategies for multiple sequence alignment aided by quality analysis tools', *Nucleic Acids Research*, 25(24), pp. 4876–4882. doi: 10.1093/nar/25.24.4876.

Trapp, S. C. and Croteau, R. B. (2001) 'Genomic organization of plant terpene synthases and molecular evolutionary implications.', *Genetics*, 158(2), pp. 811–832.

Velikova, V. *et al.* (2014) 'Genetic manipulation of isoprene emissions in poplar plants remodels the chloroplast proteome.', *J Proteome Res*, 13, pp. 2005–2018.

Vidakovic, M. (1974) 'Genetics of European black pine (Pinus nigra Arn.).', *Ann. Forest*, 6/3 JAZU Z, pp. 57–86.

Vogel, B. S. *et al.* (1996) 'Abietadiene Synthase from Grand Fir ( Abies grandis )', *Journal of Biological Chemistry*, 271(38), pp. 23262–23268. doi: 10.1074/jbc.271.38.23262.

Vranová, E., Coman, D. and Gruissem, W. (2012) 'Structure and dynamics of the isoprenoid pathway network', in *Molecular Plant*, pp. 318–333. doi: 10.1093/mp/sss015.

Vranová, E., Coman, D. and Gruissem, W. (2013) 'Network Analysis of the MVA and MEP Pathways for Isoprenoid Synthesis', *Annual Review of Plant Biology*, 64(1), pp. 665–700. doi: 10.1146/annurev-arplant-050312-120116.

Waldie, T., McCulloch, H. and Leyser, O. (2014) 'Strigolactones and the control of plant development: lessons from shoot branching.', *Plant Journal*, 79, pp. 607–622.

Walter, M. H., Fester, T. and Strack, D. (2000) 'Arbuscular mycorrhizal fungi induce the non-mevalonate methylerythritol phosphate pathway of isoprenoid biosynthesis correlated with accumulation of the "yellow pigment" and other apocarotenoids', *Plant Journal*, 21(6), pp. 571–578. doi: 10.1046/j.1365-313X.2000.00708.x.

Wang, H. *et al.* (2008) 'Experience-induced habituation and preference towards non-host plant

odors in ovipositing females of a moth.', *Journal of Chemical Ecology*, 34, pp. 330–338.

Whittington, D. A. *et al.* (2002) 'Insights into monoterpene cyclization reactions in biology: crystal structure of (+)-bornyl diphosphate synthase.', *Biochemistry*, 41, p. Biochemistry 41:8973.

Whittington, D. A. *et al.* (2002) 'Nonlinear partial differential equations and applications: Bornyl diphosphate synthase: Structure and strategy for carbocation manipulation by a terpenoid cyclase', *Proceedings of the National Academy of Sciences*, 99(24), pp. 15375–15380. doi: 10.1073/pnas.232591099.

Wildung, M. R. and Croteau, R. (1996) 'A cDNA clone for taxadiene synthase, the diterpene cyclase that catalyzes the committed step of taxol biosynthesis', *Journal of Biological Chemistry*, 271(16), pp. 9201–9204. doi: 10.1074/jbc.271.16.9201.

Wright, G. A. and Schiestl, F. P. (2009) 'The evolution of floral scent: The influence of olfactory learning by insect pollinators on the honest signalling of floral rewards', *Functional Ecology*, 23(5), pp. 841–851. doi: 10.1111/j.1365-2435.2009.01627.x.

Yuan, J. S. *et al.* (2008) 'Molecular and genomic basis of volatile-mediated indirect defense against insects in rice', *Plant Journal*, 55(3), pp. 491–503. doi: 10.1111/j.1365-313X.2008.03524.x.

Zerbe, P. *et al.* (2012) 'Bifunctional cis-abienol synthase from Abies balsamea discovered by transcriptome sequencing and its implications for diterpenoid fragrance production', *Journal of Biological Chemistry*, 287(15), pp. 12121–12131. doi: 10.1074/jbc.M111.317669.

Zhou, K. (2011) 'Structure and function of class I diterpene synthase', PhD.

Zhou, K. *et al.* (2012) 'Insights into diterpene cyclization from structure of bifunctional abietadiene synthase from Abies grandis', *Journal of Biological Chemistry*, 287(9), pp. 6840–6850. doi: 10.1074/jbc.M111.337592.