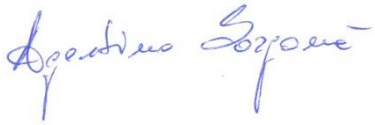
UNIVERSITY "MEDITERRANEA" OF REGGIO CALABRIA
AGRARIA DEPARTMENT
Ph.D. Course in Agricultural, Food and Forestry Sciences
Cycle XXXIV, DSS: AGR/13

Metabolomic and transcriptomic analysis of terpenes in *Pinus nigra* subsp. *laricio* in Calabria in response to the infestation of the pine processionary moth (*Thaumetopoea pityocampa*)

Ph.D. THESIS

Ph.D. Candidate
Enrica Alicandri

Tutor
Prof. Agostino Sorgonà



Ph.D. Coordinator
Prof. Marco Poiana

Academic Years 2018-2021

To my father

Fausto Alicandri

Table of contents

ABSTRACT

In conifers, the constitutive or induced production of terpenoids, either as oleoresin or emitted as volatile compounds, plays an important role in the physical and chemical defence responses against pathogens and herbivores. In addition to a wealth of traditional uses spanning over centuries, a considerable upsurge of interest has been recently observed for plant terpenoids as bio-products and green chemicals for many possible innovative applications, including biopesticides and biofuels.

Because structural and functional diversity of plant terpenoids mostly relies on the action of terpene synthases, the key enzymes in their biosynthetic pathway, novel and in-depth knowledge of the evolutionary diversification of members of terpene synthases family in the conifers, their modular structure, and their putative functions appear to be important not only for a deeper understanding of their physiological and ecological roles in the forests, but also to foster metabolic engineering and synthetic biology tools for the production of high-value terpenoid compounds.

The research activities carried out in the present Thesis aimed to gain insight into the ecological and functional roles of mono- and di- terpenes from a non-model conifer species, endemic of the southernmost Italian forests, namely *Pinus nigra* subsp. *laricio* (Poiret) Maire, Calabrian black pine, in terms of relationships with its biotic environment and overall plant performance.

In such context, an extensive characterization and phylogeny of all the known terpene synthases from different *Pinus* species were conducted, which, for such genus, can be seen as the first effort to explore the evolutionary history of a large family of genes and their products involved in specialised metabolism.

In addition, the quali-quantitative analysis of the di- and mono- terpenoid compounds in different Calabrian pine tissues was carried out. Concomitantly, we report about the first attempt to isolate full-length complementary DNAs and the corresponding genomic sequences encoding for diterpene- and monoterpene synthases involved in the specialized metabolism of Calabrian pine, by using a strategy based on the phylogeny of the available sequences of the corresponding genes in different *Pinus* species. The matching among gene expression analyses and the profiles of di- and mono- terpenoids from different tissues appeared to be consistent with the potential roles of the isolated diterpene- and monoterpene synthases in Calabrian pine.

Finally, to better understand the potential roles of terpenoids in plant defence, two experimental tests were conducted under natural or semi-natural conditions on Calabrian pine adult plants or saplings facing the infestation brought about by the lepidopteran *Thaumetopoea pityocampa*, whose *larvae* are popularly known as the pine processionary moths. In such contest, the

determination of the quali-quantitative profiles of terpene synthase genes and terpenoid metabolites was of help in obtaining a more realistic and reliable scenario of the plant-insect interaction dynamics, to better understand the defence mechanisms that plants deploy in their native environment against their natural enemies. In prospect, these knowledges could contribute for improving and diversifying the forest managements, as well as to develop the bio-control strategies in a range of possible contexts, including forestry, agroforestry and urban greenery, so complying with the commitments of the United Nations Agenda 2030 and its sustainable development goals.

RIASSUNTO

Nelle conifere, la biosintesi costitutiva o inducibile di composti terpenoidi, sia come oleoresina che come composti volatili emessi, gioca un ruolo importante nelle risposte di difesa di tipo fisico e chimico contro patogeni ed erbivori. In aggiunta agli usi tradizionali noti da secoli, è stata recentemente osservata una consistente ripresa di interesse verso i terpenoidi vegetali, mano a mano che la ricerca scientifica sia di base che applicativa ne suggeriva nuovi possibili impieghi aventi il pregio della ecocompatibilità e della sostenibilità, come nella produzione di biopesticidi e biocombustibili. Poiché la diversificazione strutturale e funzionale dei terpenoidi si basa sull'azione delle terpene sintasi, enzimi chiave nella loro biosintesi, una conoscenza nuova e approfondita della diversificazione evolutiva dei membri della famiglia delle terpene sintasi nelle conifere, della loro struttura modulare e delle loro funzioni presunte sembra essere importante non solo per una più profonda comprensione dei loro ruoli fisiologici ed ecologici, ma anche per favorire l'ingegneria metabolica e gli strumenti di biologia sintetica per la produzione di composti terpenoidi di elevato valore.

Le attività di ricerca del presente progetto di dottorato mirano ad approfondire i ruoli ecologici e funzionali dei mono- e di- terpeni in una conifera non-modello, endemica degli ecosistemi forestali meridionali, quale il pino nero calabrese o pino laricio (*Pinus nigra* subsp. *laricio* (Poiret) Maire), dal punto di vista dell'interazione con l'ambiente biotico e delle prestazioni complessive della pianta.

In tale contesto è stata condotta un'ampia caratterizzazione e filogenesi di tutte le terpene sintasi conosciute di *Pinus* spp., da considerare come il primo tentativo di esplorare la storia evolutiva della grande famiglia di geni che in questo genere sono coinvolti nel metabolismo specializzato dei terpenoidi.

È stata inoltre condotta un'analisi quali-quantitativa dei composti di- e mono- terpenoidi contenuti in diversi tessuti della specie studiata, portando a termine contestualmente il primo tentativo noto di isolare dal DNA complementare ottenuto dagli stessi tessuti i trascritti completi e le corrispondenti sequenze genomiche che codificano per le diterpene- e monoterpene sintasi coinvolte nel metabolismo specializzato dei terpenoidi. A tale scopo, è stata adottata una strategia basata sulla filogenesi delle di- e mono- terpeni sintasi disponibili in banca dati per diverse specie del genere *Pinus*. L'espressione tessuto-specifica dei geni isolati appariva correlata coi corrispondenti profili dei metaboliti di- e monoterpenoidi, rivelandosi coerente coi i ruoli potenziali delle diterpene- e monoterpene sintasi isolate in pino laricio.

Infine, per meglio comprendere il ruolo potenziale dei terpenoidi nella difesa delle piante, sono state condotte due prove sperimentali in condizioni naturali o semi-naturali, aventi come oggetto di studio le interazioni tra il pino laricio ed il lepidottero *Thaumetopoea pityocampa*, la cui larva è la ben nota processionaria del pino. In tale contesto, la determinazione dei profili quali-quantitativi dei geni delle terpene sintasi e dei corrispondenti terpenoidi ha consentito di definire scenari realistici ed affidabili sulle dinamiche delle interazione pianta-insetto, utili per una migliore comprensione dei meccanismi di difesa che le piante mettono in atto nel loro ambiente naturale. In prospettiva, l'aumento delle conoscenze in questi ambiti potrebbe contribuire a migliorare e diversificare le opzioni di gestione forestale, come anche promuovere strategie di bio-controllo da impiegare nei settori della selvicoltura, dell'agro forestazione e del verde urbano, in linea con gli obbiettivi di sviluppo sostenibile dell'Agenda 2030 delle Nazioni Unite.

KEYWORDS

Conifers; *Pinus nigra* subsp. *laricio* (Poiret) Maire; Calabrian pine; terpenoids; terpene synthase genes; pine processionary moth; *Thaumetopoea pityocampa* (Denis & Schiffermüller, 1775); plant defense; biotic stress in plants; oleoresins; volatile organic compounds.

KEY TO PLANT SPECIES

Abies balsamea (L.) Mill., 1768, balsam fir; *Abies grandis* (Douglas ex D. Don) Lindl., grand fir; *Juniperus* spp., junipers; *Physcomitrella patens* (Hedw.) Bruch & Schimp., spreading earthmoss; *Picea abies* (L.) H. Karst, Norway spruce; *Picea glauca* (Moench) Voss, white spruce; *Picea sitchensis* (Bongard) Carrière, 1855, Sitka spruce; *Pinus banksiana* Lamb., jack pine; *Pinus contorta* Douglas, lodgepole pine; *Pinus densiflora* Japanese red pine; *Pinus elliottii* Engelm., slash pine; *Pinus halepensis* Mill., Aleppo pine; *Pinus heldreichii* H. Christ, Heldreich pine*; Pinus mugo* Turra., dwarf mountain pine; *Pinus nigra* J.F.Arnold, Austrian pine or black pine; *Pinus nigra* subsp. *laricio* (Poiret) Maire, Calabrian pine; *Pinus pinaster* Aiton, maritime pine; *Pinus pinea* L., Italian stone pine; *Pinus radiata* D.Don, Monterey pine; *Pinus sabiniana* Douglas ex D.Don. towani pine, digger pine; *Pinus sylvestris* L., Scots pine*; Pinus strobus* L., eastern white pine; *Pinus taeda* L., loblolly pine; *Pseudolarix amabilis* (N. Nelson) Rehder, golden larch; *Taxus* spp., yew; *Thuja occidentalis* L., northern white cedar*; Tsuga canadensis* L. Carrière, eastern hemlock

GENERAL INTRODUCTION

TERPENOIDS IN PLANTS

➢ *Distribution and roles*

Terpenoids are one of the largest and more heterogeneous set of chemical substances produced by almost all plants. All the existing terpenoids (more than 40,000) are formed from the two basic five-carbon (C_5) isoprenoid units, namely isopentenyl diphosphate (IPP) and its isomer dimethylallyl diphosphate (DMAPP).

Based on the number of isoprenoid units, terpenoids can be divided into hemiterpenoids (C_5), monoterpenoids (C_{10}), sesquiterpenoids (C_{15}), diterpenoids (C_{20}), triterpenoid (C_{30}), tetraterpenoid (C_{40}) or polyterpenoids (C_{5n}) (Tholl and Lee, 2011).

In terms of storage of these compounds, plants can be divided into storing and non-storing ones (Kopaczyk *et al.* 2020). In the former case, terpenes are deposited into specific structures, such as resin ducts, resin blisters or resin cells. In the non-storing plants, instead, mostly volatile terpenoids are produced, which are emitted immediately after biosynthesis (Blanch *et al.* 2009; Kopaczyk *et al.* 2020).

Generally, terpenoids account for 1–2 % of tissue dry weight in plants, although in some species values as high as 15–20% have been measured (Blanch *et al.* 2009). Moreover, the profile of terpenes typically varies within species, or even among individual plants (Komenda and Koppmann, 2002).

Although terpenoids are not directly involved in growth and development (Bartwal *et al.* 2013), they play important roles in general (i.e., primary) metabolism as precursors of phytohormones, growth regulators (gibberellins, cytokinins, abscisic acid, brassinosteroids, and strigolactones), photosynthetic pigments (carotenoids), fundamental components of membranes (phytosterols) and electron carriers. On the other hand, in the specialised (i.e., secondary) metabolism, they are involved in plant/environment and plant/plant interactions (Zhou 2012; Tholl 2015; Abbas *et al.* 2017), by acting as essential components of the plants defence mechanisms to overtake stress.

Besides their role in the interaction with antagonists of plants, the constitutive and induced volatile terpenoids can act as allospecific or conspecific plant-to-plant signals, to promote defence responses in nearby plants, or even behave as systemic cues, to alert healthy organs or tissues of the same plant (Heil 2014; Tholl 2015).

Due to their functional diversity, terpenoid products provide plants with adaptative strategies for different environmental conditions, and bring benefits to long-term thanks to same co-operators, such as pollinators (Boncan *et al.* 2020). In this context, several studies demonstrated that volatile terpenoids emitted from flowers and fruits can be involved in mutualistic interactions with plant pollinators and seed dispersal agents (Abbas *et al.* 2017). Floral volatiles cause specific behavioural responses in the respective pollinators, based on the prevailing context and composition of the emissions (Wright and Schiestl 2009). Long-distance floral scent emissions mainly contribute to the guidance of pollinators to flowers, particularly in night-emitting plants, for which scent intensity will have to compensate for the limited visibility of flowers under low illumination (Dudareva *et al.* 2013).

There is increasing evidence that some tree species can cope with rapidly changing leaf temperatures through the production of isoprene and monoterpenes. The change in isoprene emission capacity through the canopy is similar to that of the xanthophyll cycle intermediates, suggesting that isoprene and monoterpene emission may protect plants against excess of heat, just as the xanthophylls do against excess of light (Loreto *et al.* 1998, Loreto & Velikova 2001). Using isoprene may be an effective way of changing membrane properties rapidly enough to track leaf temperature.

Isoprenoids can also mediate below-ground plant-host interactions. Indeed, as precursors of strigolactones, they have a crucial role as extracellular signals in the recruitment of arbuscular mycorrhizal fungi within the rhizosphere (Akiyama *et al.* 2005). Defensive interactions have also been reported, as in the case of the diterpene hydrocarbons known as rhizathalenes, produced by *Arabidopsis* roots as antifeedant against root-damaging herbivores (Vaughan *et al.* 2013).

➢ *Terpenoids and abiotic stress in conifers*

The terpene-related defenses undoubtedly contributed to the evolutionary success of conifer trees (Celedon and Bohlmann, 2019). It has been shown that some types of stress conditions can enhance or inhibit the production of terpenes (Kopaczyk *et al.* 2020). Biotic or abiotic *stimuli* from the environment (Fig. 1) may induce defense mechanism based both on the emission of volatile terpenoids (mono- and sesqui-), and/or on the production of oleoresins, i.e., fragrant mixtures of volatile and non-volatile terpenoids (mono-, sesqui- and di-terpenoids).

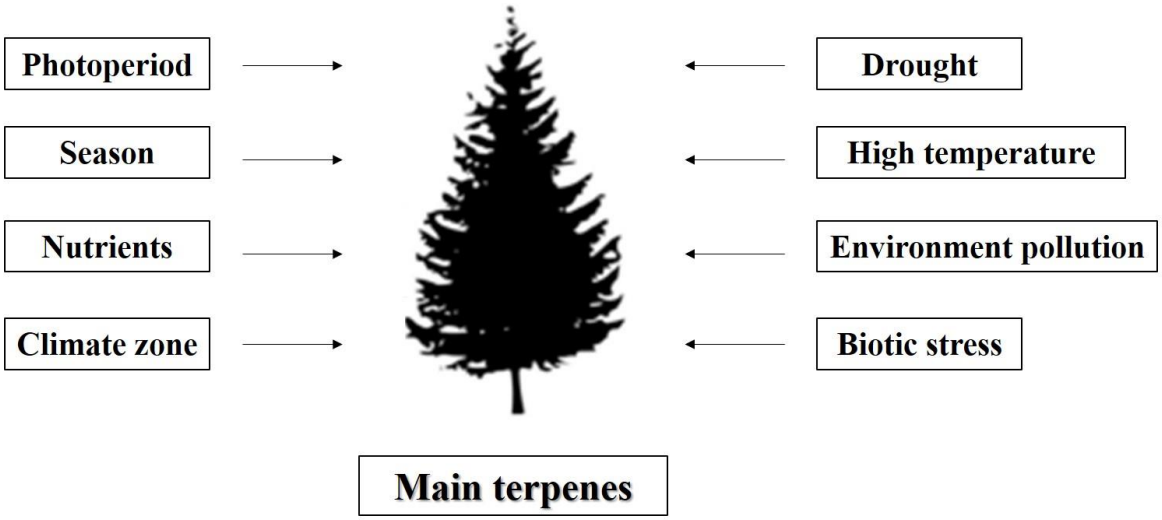


Fig. 1. *The principal environmental stimuli affecting the qualitative and quantitative terpene profiles of conifer species. (Adapted from Kopaczyk et al. 2020).*

Both primary and secondary metabolism could be affected by abiotic stress factors such as drought, air and soil pollution, high or low temperatures, high light intensity, salinity and mechanical damage (Bartwal *et al.* 2013; Kopaczyk *et al.* 2020).

Drought condition represents one of the most important stress factors that can cause a very strong impact on the physiology and survival of the plant. It appears that under typical drought stress conditions of hot Mediterranean summer days with decreased photosynthetic rates and stomatal conductance at midday, an important part of photosynthetic C fixation is still used for terpene production and emission (Blanch *et al.* 2009). Some conifer trees, such as *Pinus sylvestris* (Scots pine), are more tolerant to poor water supply conditions, as they usually grow in areas suffering from drought, while other species are more sensitive to such conditions, being accustomed to moist soil (e.g., *Picea abies*, Norway spruce) (Turtola, 2005). As a matter of fact, monoterpene concentration in *P. sylvestris* significantly increased under severe drought and their levels were almost twice as high as in control trees (Turtola *et al.* 2003). In a study conducted on *Pinus halepensis* (Aleppo pine), drought stress provoked a 54% increase in total terpene concentration (Blanch *et al.* 2009), and a negative correlation between terpene concentrations and relative water contents was observed. Taken together, the above findings suggest that an increase in terpene concentrations can be expected in conifers facing the warmer and drier conditions expected for the next decades in the Mediterranean region as consequences of global changes (Kopaczyk *et al.* 2020).

The biosynthesis and emission rate of terpenoids in conifers have also been shown to be modulated by the combination of environmental temperature and irradiance. In a study conducted by Gleizes and co-workers (1980) on *Pinus pinaster* (maritime pine), the synthesis of α- and β-pinene was found to be responsive to changes in irradiance, but not that of sesquiterpene hydrocarbons. In Japanese red pine (*Pinus densiflora*) monoterpenes emission from leaves was found to be dependent from both irradiance and temperature, because high levels of the former causes and increase of the latter (Yokouchi and Ambe, 1984). On the other hand, *Pinus elliottii* (slash pine) showed no significant differences in monoterpene emission rates under different irradiance conditions, and even in the dark (Tingey *et al.* 1979).

Focusing on temperature effects, it has been noted by Sallas *et al.* (2003) that terpenoid concentration is increased by elevated temperature in *P. sylvestris* seedling, showing major response than total phenolic compounds. Regarding low temperature, it is well known that chilling or freezing can influence primary metabolism by inhibiting photosynthesis in *P. sylvestris, Pinus banksiana* (jack pine) or *P. abies*, which may also affect the biosynthesis of "secondary" metabolites (Kopaczyk *et al.* 2020).

The availability of nitrogen and carbon nutrients in the environment, in combination with other abiotic and biotic stresses, can influence the synthesis of terpenoids in plants. For example, in needles of *P. sylvestris* the concentrations of resin acids (diterpenoids) were higher in nitrogen-fertilised plants than in the control ones of the same age (Björkman *et al.* 1998). In Aleppo pine (*P. halepensis*), plants growing on a calcareous soil showed a higher concentration of total monoterpenes (in particular sabinene, 3-carene and terpinene) than plants growing on siliceous soil (Ormeño *et al.* 2008). On the other hand, in a previous study on *P. halepensis* and *P. sylvestris*, fertilisation treatment or nutrient availability showed no significant effect on terpenoids levels (Heyworth *et al.* 1998; Blanch *et al.* 2009). In addition, by analysing the response of silver fir (*Abies alba*) to nitrogen fertilization, no effect was found on terpene concentrations in the needles (Ormeño and Fernandez, 2012). Taken together, the above results suggest that not all the secondary metabolites are influenced by soil nutrients availability. According to the ecology of each species, its tolerance to nutrients' deficiency may vary. Moreover, the variation of the final emission of volatiles may be due to possible interactions with other environmental factors, such as water deficit or light intensity, or biotic factors, e.g., leaf age or herbivore/pathogens attack (Ormeño and Fernandez, 2012).

Air pollution is definitely a negative environmental factor whose effects on plants' ecosystems structure and function have been amply demonstrated worldwide during the last decades. Different studies confirmed a strong relationship between air pollution and the changes of terpenoids

amounts in plant tissues. In *P. abies,* for example, the concentration of monoterpenes in needles exposed to air pollution was found to be significantly higher than in control needles exposed to clean air (Jüttner, 1988). The terpenoids concentrations in *P. sylvestris* trees from two industrial areas (oil refinery and cement factory) were studied in Lithuania (Kupcinskiene *et al.* 2008); the results showed the same trend for the two different polluted areas: in polluted trees terpenes were much more concentrated than in not polluted ones. Air pollution has an effect also on the type of terpenes produced by plants: in *Pinus nigra* (black pine) plants growimg in an urban environment only 12 terpenes were detected, whereas the number increased up to 15 in plants growing in a nearby rural site (Supuka *et al.* 1997). Undoubtedly, exposure of conifer trees to air pollutants alters the intensity of terpenoids emission, although the pollution-dependent molecular pattern of terpenoids biosynthesis is still far from being unravelled (Kopaczyk *et al.* 2020).

- *Terpenoids and biotic stress in conifers*

Terpenoids are also known to mediate plant-insect (e.g., pollinators, predators, parasites, and herbivores) interactions (Boncan *et al.* 2020). These secondary metabolites seem to be directly or indirectly involved in these ecological dynamics. As mentioned above, conifer species can react to biotic stress with the production of oleoresin and/or with the emission of volatile terpenoids, which work to the advantage of the plants and at the expenses of the pests. The bark beetles (*Scolytinae)* are among the most important parasites able to endanger plant stability and altering its terpenes concentration. For instance, infestation brought about by the white pine weevil (*Pissodes strobi*) on spruce trees (*Picea* spp.) caused a rapid increase of monoterpene olefin levels, traumatic resin duct formation (Byun-McKay *et al.* 2006; Nicole *et al.* 2006) and traumatic resin accumulation (Miller *et al.* 2005). Under weevil attack, changes in relative concentrations of most bark monoterpenes and sesquiterpenes were observed, showing a different response in bark tissues from needle ones (Nicole *et al.* 2006). The molecular and biochemical mechanisms underlying these phenomena have become an area of active research in recent years. There is evidence that the weevil-induced changes in the expression of terpene synthases are spatially differentiated and involve primarily cortical and traumatic resin ducts (Kopaczyk *et al.* 2020).

Others biotic stressing factor challenging plant survival are represented by pathogenic fungi. In *P. abies* inoculated with the blue stain fungus *Ceratocystis polonica* the average total terpenoids concentration in the phloem was significantly higher than in unwounded or only mechanically wounded plants parts (Viiri *et al.* 2001; Zhao *et al.* 2011; Mageroy *et al.* 2019). At variance with these results, in *Pinus spp.* seedlings inoculated with *Heterobasidion annosum,* the causal agent of

the annosus root rot, and subsequently with *Sphaeropsis sapinea*, the causal agent of the diplodia tip blight, the total terpene accumulation was suppressed by about 50% comparing to seedlings infected only with *H. annosum* (Bonello *et al.* 2008).

A study conducted by Steele *et al.* (1998) on grand fir (*Abies grandis*) showed that in response to wounding, oleoresin components are produced at different times: monoterpenes first, having a toxic effect on attacking insects, then sesquiterpenes and diterpenes (resin acids) to heal the wound.

- *Utilizations of terpenoids*

The interest for terpenoids stems from the important role they play not only in plants' metabolism and ecology, but also in technological terms. Indeed, gymnosperm diterpenoids are employed to produce solvents, flavours, fragrances, pharmaceuticals and a large selection of bioproducts (Zulak *et al.* 2010; Zerbe *et al.* 2013). Examples are the antitumor compound pseudolaric acid B obtained from the roots of *Pseudolarix amabilis* (golden larch; Mafu *et al.* 2017), the anticancer drug taxol present in yew (*Taxus* spp.; Croteau *et al.* 2006), and *cis*-abienol, a molecule of interest in the fragrance industry extracted from balsam fir (*Abies balsamea*; Zerbe *et al.* 2012).

Specific terpenes have been also identified as specialty biofuels that meet current industrial and chemical requirements, including viscosities, flash points and freezing points, high energy densities, and high volumetric net heats of combustion. Plants belonging to *Eucalyptus* genus were frequently used for such purpose (Mewalal *et al.* 2017).

Moreover, terpenoids could also have a potential in agroforestry biotechnology. In fact, terpenoids that attract herbivores could also function as bioherbicides, allowing management of weed growth in paddies or any agricultural fields (Boncan *et al.* 2020). In addition, with the support of nanotechnologies, these compounds could be also encapsulated and used to trap insect pests (Boncan *et al.* 2020). Another promising area of current research is about the use of terpenoids as biological pesticides for the control of pests and pathogens in the context of urban forestry and urban greenery (Foti *et al.* 2020).

CALABRIAN PINE [*Pinus nigra* subsp. *laricio* (POIRET) MAIRE]

➢ *Taxonomy*

Pinus nigra subsp. *laricio* (Poiret) is one of the six subspecies of *P. nigra* J.F. Arnold (black pine). On the basis of Euro+Med Plantbase (Raab-Straube, 2014) classification the six *Pinus nigra* subspecies are: *Pinus nigra* subsp. *mauretanica*, *Pinus nigra* subsp. *salzmannii*, *Pinus nigra* subsp. *laricio*, *Pinus nigra* subsp. *nigra*, *Pinus nigra* subsp. *dalmatica* and *Pinus nigra* subsp. *pallasiana*; their distributions are shown in Fig. 2.

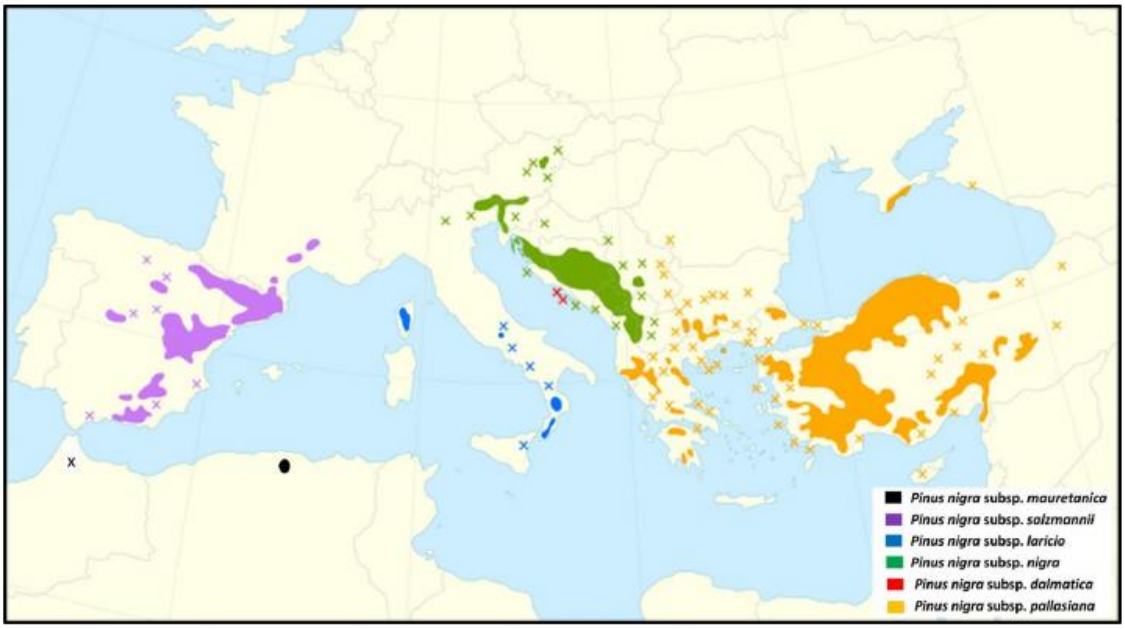


Fig. 2. *Distribution of Pinus nigra subspecies considering the classification of Euro+Med Plantbase.*

Pinus nigra subsp. *pallasiana* (Lamb.) Holmboe is present essentially in Turkey and Greece, covering a surface of around 2.5 milion ha.

Pinus nigra subsp. *nigra* is present in Italy (Appenines), northern Greece and Balkan Mountains, covering around 800,000 ha.

Pinus nigra subsp. *dalmatica* (Vis.) Franco is found in Croatia and Dinaric Alps.

Pinus nigra subsp. *mauretanica* (Maire et Peyerimh.) Heywood is localized in mountainous area of Morocco and Algeria.

Pinus nigra subsp. *salzmannii* (Dunal) Franco is distributed for more than 350.000 ha in Spain, but it is also present in the Pyrenees and the Cévennes in France with fragmented populations.

Finally, *Pinus nigra* subsp. *laricio* (Poiret) Maire is found in Corse, and in southern Italy with a natural range extending from Calabria to Sicily (Nicolaci *et al.* 2014).

- *Distribution, botanic traits and ecology*

Black pine has a very fragmented distribution range which is the results of different factors that have occurred over the time, such as the competition with other tree species, human activity and the climate warming at the onset of the Holocene. Glacial refugia were localized in Spain, Corse, North Africa, Balkans and Italy. After glaciation, black pine had a quickly expansion giving rise to the "age of pine". This expansion stopped around 5000 B.C. because of the diffusion of other species like spruce (*Picea* spp.) and beech (*Fagus* spp.); after this period, a regression and fragmentation of the black pine distribution range have occurred.

During the mid-19th century, the large use of such species for reforestation allowed the expansion of its distribution range. Currently, black pine covers a large expanse of over 3.5 million hectares (Isajev *et al.* 2004). Its widest distribution worldwide is in Turkey, with more than 2.5 million hectares (Enescu *et al.* 2016).

In Italy it is present in more than 200,000 ha (INFC, 2010) considering both reforestation and natural stands. Outside of Europe, it was also introduced in the United States (where it is known as Austrian pine) in 1759, and has now become naturalised in parts of New England and the Great Lake States.

Its altitudinal range goes from 350 m in Italy to 2200 m a.s.l. in the Taurus Mountains, even if the optimal altitudinal range is between 800 to 1500 m a.s.l.

Pinus nigra is an evergreen conifer species that at maturity reaches 15-25 (35) m of height. Its bark is usually dark greyish brown to black (from here its Latin name "nigra") and becomes increasingly fissured with age. Needles are in pairs 8-15 (19) cm long, 1-2mm in diameter, straight or curved, and finely serrated. They normally persist on the tree for 3-4 years (exceptionally up to 8); they are dark green coloured, with pointed apex but not always pungent (Enescu *et al.* 2016).

The canopy in young individual is pyramidal, while in mature trees becomes irregularly shaped. Macrosporophylls develop at the apex of the branch of the year, in group of 2-4 (5), with a green colour that at maturity becomes reddish. Cones are sessile and horizontally spreading, 4–8 cm long, 2–4 cm wide, and yellow-brown in colour; they ripen from September to October of the second year, and open in the third year after pollination.

Black pine can grow in either extremely dry or humid habitats, showing a considerable tolerance to temperature fluctuations. It is a light-demanding species, but it shows higher shade tolerance than Scots pine (*P. sylvestris*). It is resistant to drought and wind. Thanks to its ecological

flexibility, it is one of the most widely used tree species for reforestation worldwide. In addition, it is very efficient in the colonization of degraded soils, and its adventitious roots are suitable to be exploited for deep reinforcement and soil strength enhancement. It grows in pure stands or in association with other broadleaved or conifer species, in particular *P. sylvestris*. It is also commonly found in association with other pines such as dwarf mountain pine (*Pinus mugo*), Aleppo pine (*Pinus halepensis*), Italian stone pine (*Pinus pinea*) and Heldreich pine (*Pinus heldreichii*) (Enescu *et al.* 2016).

➢ *Pinus nigra subsp. laricio in Italy*

Pinus nigra subsp. *laricio* includes two varieties: var. *Corsicana* (the Corsican pine) and var. *Calabrica* (Calabrian pine). The first is found in Corse covering a surface of around 45.000 ha from 1000 to 1800 m a.s.l.. Calabrian pine is endemic to southern Italy with a natural range extending from Calabria to Sicily, even if a population is also present in Tuscany, in the province of Pisa, arising from a reforestation program. In Abruzzo National Park, in the locality of Villetta Barrea, there is a stand of Villetta Barrea's black pine that is a transition between *P. laricio* and *P. nigra* subspecies (*Pinus nigra* subsp. *nigra* var. *Italica*). In Sicily, Calabrian pine is present only on the Etna volcano, with an altitudinal range from 1200 up to 2000 m a.s.l..

In Calabria, Calabrian pine is mostly present in two different areas: the Sila mountains, covering around 37.000 ha, and the Aspromonte mountains, covering 3.000 ha within an altitudinal range 900-1600 m a.s.l.. In both Sicily and Calabria, annual rainfall, in areas where Calabrian pine is present, are from 1400 to 1800 mm, of which 80-120 mm falling during the summer.
The annual use of Calabrian pines in Calabria is around 60.000 m^3 and the plantations are managed with different strategies: selection cuttings that allow the formation of complex uneven-aged structure, but also strip (30x100m) or patch clear-cutting ("schiumarola cutting"). Artificial plantation or seeding are applied when natural regeneration is absent (Bernetti, 1995).

➢ *Importance and usage*

Black pine is one of the most economically important native conifers in southern Europe. The stems of black pine have been widely used in the past for naval construction because its wood is durable, rich in resin and easy to process. It is highly suitable for indoor flooring; in the Mediterranean area, it is used not only for general construction (doors, panelling, staircases, etc.) and furniture, but also as fuelwood, while the pulp is exploited for paper.

While *Pinus nigra* subsp. *laricio* is appreciated for building and roofing for its straightness and thin branches (if properly thinned the low amount of duramen make its wood suitable for carpentry and cabinetry), *Pinus nigra* subsp. *nigra* has a lower quality and thus, its use is restricted to lower-grade buildings wood and the making of crates.

Calabrian pine is also valued for landscaping, both in parks (isolated trees or in group) and in urban and industrial contexts thanks to its tolerance to environmental pollution; it is considered a good prospect among indigenous coniferous species in Central Europe under future climate scenarios (Thiel *et al.* 2012).

THE PINE PROCESSIONARY MOTH [*Thaumetopoea pityocampa* (DENIS & SCHIFFERMÜLLER, 1775)]

The pine processionary moth (PPM) *Thaumetopoea pityocampa* (Denis & Schiffermüller, 1775) (Lepidoptera: Notodontidae: Thaumetopoeinae) is one of the most important phytophagous that reproduces generally on species of the genus *Pinus* L. (Avtzis, 1986; Devkota & Schmidt, 1990; Zhang & Paiva, 1998) and, to a lesser degree, on other conifers, as cedars, and can hinder the growth of infested trees (Jacquet *et al.* 2012).

The Pine processionary moth is a univoltine species and is widespread in the Mediterranean, including southwest Europe, the Balkans and North Africa (Roques, 2015). The spread of this pest into new areas requires the presence of appropriate host plants and is facilitated by climate warming or a high population density (Hodar & Zamora, 2004; Battisti *et al.* 2005; Castagneyrol *et al.* 2016).

Its infestations, generally with different attack intensities (Devkota & Schmitd 1990; Masutti & Battisti 1990; Carus, 2004), usually represents not only a problem for the *Pinus* forests and for the forest economy, but also for the people in this environment or for anyone in direct contact with an infected plant. Indeed, the bristles produced and released in relation to daily activity (Werno & Lamy, 1994) of some larval stadiums of the insect are responsible for allergic pathologies in men and various animals (Ducombs *et al.* 1981, Lamy, 1990; Artola-Bordás, 2008).

Dispersal in *T. pityocampa*, as in other winged insects, besides the mechanisms linked to human activity, is mainly dependent on distance flights (Duan *et al.* 1998; Muirhead *et al.* 2006 Kerdelhué, 2006). Understanding flight characteristics might provide useful information for controlling the further spread of this moth and, eventually, for understanding the structure of the population. Generally, the insect flight is influenced by the habitat (Einhorn, 1983), by

atmospheric factors such as wind, air temperature and rainfall (Isard *et al.* 1999; Bonsignore & Bellamy, 2007) as well as by biological, genetic, physiological and enzymatic factors (Ferro *et al.* 1999; Perez-Mendoza *et al.* 1999; Zera *et al.* 1999; Auerswald & Gäde, 2002).

The female of the species is short-lived and has limited flight capabilities, dispersing over 3–4 km at most (Démolin, 1969). The male flight of Pine processionary moth can be checked by using traps baited with sex pheromone. They have been widely used since the sex pheromone was identified several years ago (Camps *et al.* 1981; Guerrero *et al.* 1981). This sex pheromone shows an activity similar to that displayed by virgin females in a wind tunnel (Quero *et al.* 1997) and provides a suitable tool for monitoring of Pine processionary moth populations (Tiberi & Niccoli, 1984; Jactel *et al.* 2006).

Pine processionary moth lays its eggs on pine needles during summer. The eggs hatch 30 to 45 days later (Aydın *et al.* 2018). Thus, the larvae feed throughout autumn and during winter and spring depending on temperature. The larvae are gregarious from early on in their development and spin conspicuous tents in tree tops. After completing their feeding, generally during late winter, larvae typically form a procession in which they walk head to tail in a column, to reach their pupation sites on the ground. Then, they burrow into the soil to a depth of a few centimetres, spin a silken cocoon in which they pupate. The duration of the period from larval emergence to pupation varies from year to year and site to site, and the total period may be up to 2 months (Demolin, 1971; Buxton, 1990; Bonsignore *et al.* 2015; Robinet *et al.* 2015). The larvae select sites well exposed to sun, e.g., the edge of a forest, where the thermal conditions and/or the microclimatic conditions in the soil are favourable (Dulaurent-Mercadal *et al.* 2011; Samalens & Rossi, 2011). The habitat influences pupation success and survival in Pine processionary moth (Markalas, 1989; Dulaurent Mercadal *et al.* 2011; Torres-Muros *et al.* 2017), and the pupae suffer additional mortality due to natural enemies, such as parasitoids (Bonsignore *et al.* 2015), predacious birds (Barbaro & Battisti, 2011) and entomopathogenic fungi (Er *et al.* 2007).

The community of antagonists is present with a modest number of species in all stages of development and can sometimes provide a significant contribution to the prevention of outbreaks, even though it is unable to exert a satisfactory control of the Pine processionary moth populations. In the pine trees various different natural enemies are present such as ants (Way *et al.* 2001), Tettigonidae (Ledesma 1971; Gonzalez-Cano 1981) some parasitoids (Buxton 1990; Tarasco, 1995) or insectivorous birds like the hoopoe (*Upupa epops* L.) (Battisti *et al.* 2000 that seem to play also a role in controlling the pest.

The adults emerge during summer, but some pupae can spend 2 or more years in diapause and the adults generally disperse over short distances (Bonsignore and Manti, 2013; Roques, 2015).

Pupation has a physiological basis as it is associated with reaching a particular size (D'Amico *et al.* 2001; Davidowitz *et al.* 2003). Traits such as pupal weight or size are considered important because they can affect the longevity of females and fitness of adults (Calvo & Molina, 2005; Fantinou *et al.* 2008). For example, pupal size may affect male phenotype, which could influence male mating success (Fox *et al.* 1995). The relationships between pupal traits and pupal density in gregarious Lepidoptera may vary according to the population dynamics of a species and are not easy to understand.

Pine processionary moth infestations are not evenly distributed within pine forests and mainly occur at the edges of pine stands (Samalens and Rossi, 2011). Several studies have shown that the adults are more likely to oviposit near the edges of pine stands (Samalens and Rossi, 2011) than more than 25 m from the edges (Parlak *et al.* 2019). Thus, clearings within forests are suitable sites for PPMs as they create edges and open areas where their processioning larvae can pupate in sun exposed soils. Pine forests with suitable pupation sites for supporting large aggregations of Pine processionary moth occur in southern Italy (Bonsignore *et al.* 2011, 2015, 2019).

The studies on the importance of flight in the spread of this insect are important for a better understanding of Pine processionary moth dispersal, necessary for developing techniques to prevent the spread of this insect from one *Pinus* stands to another. The facility with which these phytophagous attacks the *Pinus* woods mainly depends on the many situations of environmental stress to which numerous forest areas are submitted, and, in general, on the absence of correct techniques of forestry management. Another aspect about the diffusion of the caterpillar in the last years is linked to the increase of mean temperature with an important diffusion for the Pine processionary moth at higher elevations (Battisti *et al.* 2006).

The success of the spread of species is linked to the ability to find favourable conditions for development, such as temperature and host but also to its ability to spread into the surrounding flow more or less from areas where the species has found the optimum favourable temperature and host. The difference in homogeneity of an environment can affect the ability to spread and this is one of the major factors determining how individuals disperse within a landscape (Schultz & Crone, 2001, Ovaskainen & Cornell, 2003; Ovaskainen, 2004).

AIM AND ORGANIZATION

Plant secondary metabolites involved in environmental adaptation and stress tolerance can be broadly classified into phenolics, alkaloids and terpenoids (Boncan *et al.* 2020). While terpenoids are known to play essential primary functions as precursors of phytohormones, growth regulators and photosynthetic pigments, electron carriers and key components of membrane structures, the "secondary" terpenoid metabolites have been identified as having a range of specialized roles in plant/environment, plant/plant and plant/pathogen/pest interactions (Beran *et al.* 2019).

Most of the existing knowledge on the metabolism of terpenoids and their roles in tree-environment interactions came from well-known model species (whose genomes have been completely sequenced and annotated), the majority of which thrive under continental/temperate climates. Conversely, far less information exists on Mediterranean tree species which, owing to the global climate changes, are expected to increase their spread, along with the suite of biotic and abiotic constraints influencing their biology (Foti *et al.* 2020). A vast gap in knowledge concerns the ecological roles of terpenoid metabolism in the tree-environment interactions of conifer species representative of Mediterranean area such as *Pinus nigra* subsp. *laricio* that represents an essential element of Calabrian forest landscape. These Mediterranean pine forests are often attacked by caterpillars of *Thaumatopoea pityocampa,* one of the most important defoliators in the Mediterranean region causing large economic losses and ecological effects. Research on this topic may contribute to provide guidelines for the management, both in the field and in the context of urban greenery, taking into account biodiversity preservation, soil conservation, watershed protection, and urban welfare.

The aim of the present work was to assess the role played by terpenes in the interaction among Calabrian *Pinus nigra* subsp. *laricio* and its biotic environment in terms of ecological relationship and overall plant performance. To achieve these aims, the present thesis included five case-studies, organised in as many different chapters, in which, after an extensive characterization of all the known terpene synthase (TPS) genes and the corresponding metabolites from different *Pinus* species, a study about Calabrian pine-processionary moth biotic interaction was also conducted under field conditions.

The first chapter of the present thesis is a review on the biosynthesis and the biological functions of terpenoids that includes recent advances on the evolution and functional diversification of plant TPSs, with specific reference to gymnosperms. In such context, a detailed structural and functional analysis and phylogeny of all the TPSs found in *Pinus* genus is reported, which can be seen as the first effort to explore the evolutionary history of the large family of TPS

genes involved in specialized metabolism. Finally, a successful approach that can be extended to isolate other TPSs is also described, in which the phylogeny of TPSs in *Pinus* spp. has been exploited to isolate, for the first time, the mono-TPS sequences from *Pinus nigra* subsp. *laricio*.

The second chapter focuses on the quali-quantitative analysis of the diterpenoid compounds in five different Calabrian pine tissues, via a conventional GC-MS approach. In parallel, and by using a strategy based on the phylogeny of available diterpene synthases (DTPSs) from different *Pinus* species, full length cDNAs and the corresponding genomic sequences encoding for DTPSs have been isolated, for the first time in the studied species. Such complementary approach made it possible a tissue-specific matching among the profiles of diterpene metabolites and the expression of DTPS genes, which revealed to be consistent with the potential roles of the isolated DTPSs in the biosynthesis of the corresponding diterpene resin acids.

In the third chapter, genomics and metabolomics have been used in the above complementary manner for the isolation and characterization of monoterpenoids and monoterpene synthase genes in the same Calabrian pine tissues.

The fourth chapter concerns the qualitative and semi-quantitative characterization of the terpene volatile compounds emitted from Calabrian pine by means of gas chromatography/mass spectrometry (GC-MS) in two sites, namely Bova and Canolo, comprised within the premises of the Aspromonte National Park (ANP), Calabria Region, Southern Italy. At both sites, the emission of volatile terpenes from needles of infested plants has been monitored over time during the different phases of the trophic interaction with the *larvae* of the pine processionary moth. Nearby non-infested plants have been used for comparison.

Finally, in the last chapter, the temporal expression of the previously characterized mono- and di- terpene synthase genes and their corresponding metabolites in Calabrian pine trees infested by the pine processionary moth was also performed under natural conditions.

1. CHAPTER 1: ON THE EVOLUTION AND FUNCTIONAL DIVERSITY OF TERPENE SYNTHASES IN THE *Pinus* SPECIES: A REVIEW

Adapted from: Alicandri et al. (2020) **On the evolution and functional diversity of terpene synthases in the *Pinus* species: a review.** Journal of Molecular Evolution 88(3), 253-283 https://doi.org/10.1007/s00239-020-09930-8

Abstract

In the biosynthesis of terpenoids, the ample catalytic versatility of terpene synthases (TPS) allows the formation of thousands of different molecules. A steadily increasing number of sequenced plant genomes invariably show that the TPS gene family is mid to large in size, comprising from 30 to 100 functional members. In conifers, TPSs belonging to the gymnosperm-specific TPS-d subfamily produce a complex mixture of mono-, sesqui-, and di-terpenoid specialized metabolites that are parts of the volatile emissions and oleoresin secretions. Such substances can be harmful for plant pathogens and herbivores and is therefore conceivable they are used to combat biotic stress, although a role in contrasting abiotic stress is also envisaged. Humans have been known long since that terpenoids extracted from oleoresins can be profitably used in number of different fields, from traditional and modern medicine to chemical industry (fine chemicals, fragrances and flavors) and, in the last years, in bio-refinery too. In the present work, after summarizing the current views on the biosynthesis and biological functions of terpenoids, recent advances on the evolution and functional diversification of plant TPSs are reviewed, with a focus on gymnosperms. In such context, an extensive characterization and phylogeny of all the known TPSs from different *Pinus* species is reported, which, for such genus, represents the first effort to explore the evolutionary history of the large family of TPS genes involved in specialized metabolism. Finally, an approach is described in which the phylogeny of TPSs in *Pinus* spp. has been exploited to isolate and characterize for the first time mono-TPS sequences from *Pinus nigra* subsp. *laricio*, an ecologically important endemic pine in the Mediterranean area.

Keywords: Terpene synthase, Terpenoid biosynthesis, Plant specialized metabolism, Plant defense, Gymnosperms, Gene family evolution.

1.1. INTRODUCTION

Terpenoids, also referred to as terpenes or isoprenoids, make up the biggest and most diversified class of chemical substances discovered in plants, encompassing over 40,000 individual compounds (Tholl 2015; Singh and Sharma 2015; Abbas *et al.* 2017). The evolutionary success of the terpenoid metabolites largely depends on the flexibility of building molecules of various sizes. Indeed, terpenoids, arising from the two basic five-carbon (C5) isoprenoid units, isopentenyl diphosphate (IPP) and its isomer, dimethylallyl diphosphate (DMAPP), can be categorized as hemiterpenoids (C5), monoterpenoids (C10), sesquiterpenoids (C15), diterpenoids (C20), triterpenoid (C30), tetraterpenoid (C40) or polyterpenoids (C5n), based on the number of C5 units they contain (Tholl and Lee 2011). While a number of plants terpenoids are important for several fundamental functions in growth and development, most of them have specialized roles in plant-environment interactions (Tholl 2015; Singh and Sharma 2015).

The tremendous variety of terpenoid carbon structures may be ascribed mainly to the activity of terpene synthases (TPSs), the primary enzymes in terpenoid biosynthesis. The TPS genes constitute a mid-size to large family with approximately 30-100 functional members in the genomes of nearly all the plant species so far sequenced (Chen *et al.* 2011; Warren *et al.* 2015; Kumar *et al.* 2018). Based on their phylogenetic relationships, plant TPSs can be classified into seven clades or subfamilies: a, b, c, d, g, e/f and h (Chen *et al.* 2011). In conifers, the TPSs involved in specialized metabolism make up the gymnosperm-specific TPS-d subfamily, which, based on structural and catalytic properties, can be further split into three groups: TPS-d1, which includes mainly monoterpene synthases (MTPSs); TPS-d2, which comprises mainly sesquiterpene synthases (STPSs); TPS-d3, containing mainly diterpene synthases (DTPSs) (Martin *et al.* 2004; Keeling *et al.* 2011). Conversely, conifer DTPSs of primary metabolism (i.e., gibberellin biosynthesis) are members of the TPS-c and TPS-e/f subfamilies, which also comprise angiosperm orthologous genes (Keeling *et al.* 2010; Chen *et al.* 2011). The TPSs belonging to the gymnosperm-specific TPS-d subfamily produce a complex mixture of mono-, sesqui-, as well as di-terpenoid specialized metabolites that constitute the volatile emissions and oleoresin secretions. These specialized metabolites are involved in the defense against pathogens and herbivores and can help to protect against abiotic stress (Zulak and Bohlman 2010; Hall *et al.* 2011; Tholl *et al.* 2015). Oleoresin terpenoids are also important for humans in the production of flavors and fragrances, therapeutics, solvents, coatings and resins, and more recently have been taken into consideration as potential precursors of biofuel products (Bohlman and Keeling 2008; Zulak and Bohlman 2010; Hall *et al.* 2013a). Because of such wide functional diversification and versatility,

attempts are being made to decipher how terpenoids biosynthesis and metabolic routing are regulated in conifers.

In the present review, after summarizing recent progress in our comprehension of the biosynthesis and biological functions of terpenoids, the latest advances in research on the evolution and functional diversification of plant TPSs will be considered, focusing in particular on gymnosperms. In such context, an extensive characterization and phylogeny of all the known TPSs from different *Pinus* species will be reported, which, to the best of our knowledge, constitutes for such genus the first effort to explore the evolutionary history of the large family of TPS genes involved in specialized metabolism. Finally, we will report about our attempt to isolate and characterize for the first time MTPSs gene sequences from *Pinus nigra* subsp. *laricio*, an ecologically important endemic pine in the Mediterranean area, by using a strategy based on the phylogeny of all available MTPSs from different *Pinus* species.

1.2. BIOLOGICAL AND ECOLOGICAL FUNCTIONS OF PLANT TERPENOIDS: A SYNOPSIS

While terpenoids are known to play essential primary functions as precursors of phytohormones and growth regulators (gibberellins, cytokinins, abscisic acid, brassinosteroids, and strigolactones), photosynthetic pigments (carotenoids), electron carriers (ubiquinone and plastoquinone), and key components of membrane structures (phytosterols), "secondary" terpenoid metabolites (considered in particular here) have been identified as having a range of specialized roles in plant/environment and plant/plant interactions (Zhou 2012; Tholl 2015; Abbas *et al.* 2017). Low-molecular-weight terpenoids such as isoprene, monoterpenoids, sesquiterpenoids, and diterpenoids, which are volatile, semivolatile or nonvolatile at ambient temperature, respectively, are involved in plant defense from abiotic stress and in many above- and below-ground biotic interactions (Loreto *et al.* 2014; Tholl 2015; Abbas *et al.* 2017).

The emissions of terpenoids such as isoprene and monoterpenes from several plant species have been found to be strongly correlated with the prevention of temperature stress (Sharkey and Yeh 2001; Monson *et al.* 2013). This protective function is presumably due to the temporary storage of these volatile compounds into the photosynthetic membranes (Velikova *et al.* 2014). Other physiological functions of isoprene against abiotic stress in plants include tolerance to ozone and protection from oxidative stress (Loreto *et al.* 2001; Behnke *et al.* 2009; Schnitzler *et al.* 2010).

Plant volatile terpenoids play a role in the presence of biotic stress as well, being part of the constitutive and/or inducible defense line against pathogens and herbivores. For instance, insect-deterring effects have been observed for the monoterpene volatiles emitted by *Chrysanthemum morifolium* leaves (Laothawornkitkul *et al.* 2008) and for the sesquiterpenes accumulating in the glandular trichomes of wild tomato (Bleeker *et al.* 2011). Huang *et al.* (2012) showed that (*E*)-β-caryophyllene (a sesquiterpene) contributes to the reproductive success of *Arabidopsis* challenged with *Pseudomonas syringae*: in wild-type plants, volatile emission from the stigma limited bacterial growth, whereas non-emitting mutants showed a dense bacterial population on their flowers, resulting in lighter and often misshaped seeds compared to the wild-type.

Sesquiterpenes and diterpenes can function as phytoalexins in most plant species (Mumm *et al.* 2008). In cotton, for instance, gossypol and its related sesquiterpene aldehydes play a role in the inducible and constitutive defense responses against several pathogens (Townsend *et al.* 2005). Moreover, Prisic *et al.* (2004) isolated fourteen different diterpenes exhibiting antimicrobial properties from rice leaves challenged with the pathogenic blast fungus *Magnaporthe grisea.*

In conifers, the production of terpenoids, either as oleoresin or emitted as volatile compounds, play an important role in the physical and chemical defense responses against pathogens and herbivores (Zulak and Bohlman 2010). Oleoresin, whose main components are mono- and diterpenes (including diterpene resin acids, DRA), with lower quantities of sesquiterpenes, accumulates in specialized anatomical structures, such as resin ducts, which function as pressurized storage reservoirs. In case of wounding, the resin under pressure spreads out from the ducts and reaches the wounded area, acting as a physical and chemical weapon against invading organisms (Zulak and Bohlman 2010). The importance of terpenoids in the defense system of conifers against insect pests was confirmed by the study of Hall *et al.* (2011): the resistance to the white pine weevil (*Pissodes strobi*) in *Picea sitchensis* was found to be associated to the levels of the monoterpene (+)-3-carene, which in turn depended on the copy number and the extent of transcriptional activation of the gene coding for its biosynthetic enzyme, as well as on the amount and catalytic efficiency of the encoded protein.

The involvement of induced volatile terpenoid compounds in attracting natural enemies of pathogens and herbivores is also well documented (reviewed by Gols 2014 and Pierik *et al.* 2014). Such indirect defense strategy is used by plants to protect their photosynthetic tissues from pathogens and herbivores, as well as to limit insect oviposition. For instance, the deposition of eggs by the elm leaf beetle (*Xanthogaleruca luteola*) on the leaves of *Ulmus minor* results in the production of volatile compounds, including the irregular homoterpene, *(E)*-4,8-dimethyl-1,3,7-

nonatriene, which plays a key role in attracting the specialized egg parasitoid *Oomyzus gallerucae* (Büchel *et al.* 2011).

Besides their role in the interaction with herbivores and their antagonists, constitutive and induced volatile terpenoids can act as interspecific, intraspecific, and intraplant signals to promote defense responses in nearby plants or in healthy tissues of the same plant (Heil 2014; Tholl 2015). However, there is still a poor understanding of the molecular mechanisms involved in volatile-mediated plant-plant communications.

Not every organism in contact with plants are enemies, inasmuch as some of them, in fact, can act as partners involved in mutually beneficial interactions. In this context, several studies demonstrated that volatile terpenoids emitted by flowers and fruits can be involved in mutualistic interactions with plant pollinators and seed dispersal agents (Abbas *et al.* 2017). For instance, many studies have proven the role of volatile terpenoids as constituent of floral scent in communication between plants and pollinators (Baldwin *et al.* 2006; Abbas *et al.* 2017). Floral volatiles cause specific behavioral responses in the respective pollinators, based on the prevailing context and composition of the emissions (Wright and Schiestl 2009). Long-distance floral scent emissions mainly contribute to the guidance of pollinators to flowers, particularly in night-emitting plants, for which scent intensity will have to compensate for the limited visibility of flowers under low illumination (Dudareva *et al.* 2013). Monoterpenes and sesquiterpenes, as the major components of floral volatiles, are particularly suited as long-distance chemical messengers, because of their low-molecular-weight, lipophilic nature and high vapor pressure at ordinary temperatures (Tholl 2015; Abbas *et al.* 2017).

Although terpenes have been mostly studied in the aboveground tissues, similar functions in direct and indirect defense responses have been also identified in the belowground environment. For instance, *Arabidopsis* roots produce semi-volatile diterpene hydrocarbons, known as rhizathalenes, able to reduce root damage by acting as local antifeedant towards herbivores (Vaugham *et al.* 2013). Similarly, the triterpene saponins known as avenacins are powerful phytoalexins exuded by oat roots (Thimmappa *et al.* 2014). Indirect defense brought about by volatile terpenes is also seen belowground; for instance, maize roots attacked by the western corn rootworm (*Diabrotica virgifera*) emit (*E*)-β-caryophyllene, which acts as a volatile signal to attract predatory nematodes (Rasmann *et al.* 2005). Moreover, labdane-related volatile diterpenoids known as nomilactones, released by the roots of rice plants, exhibit allelopathic activity towards adjacent competing species (Xu *et al.* 2012). Finally, carotenoid-derived plant hormones such as strigolactones, in addition to their *in planta* primary functions, are also able to promote the beneficial root infection brought about by mycorrhiza (Akiyama, 2005).

1.3. BIOSYNTHESIS OF TERPENOIDS IN PLANTS

Isopentenyl diphosphate (IPP) and dimethylallyl diphosphate (DMAPP) are the C5 momomeric precursors of all terpenes. They derive from two distinct metabolic routes, namely the mevalonate (MVA) pathway, and the methylerythritol phosphate (MEP) pathway (Tholl 2015; Pazouki and Niinemets 2016; Abbas *et al.* 2017). After their synthesis, IPP and DMAPP are acted upon by prenyltransferases, which assemble them into dimer, trimer, tetramer or examer of the original C5 building blocks, yielding intermediates such as geranyl diphosphate (GPP, C10), farnesyl diphosphate (FPP, C15), geranylgeranyl diphosphate (GGPP, C20), and squalene (C30), respectively (Tholl 2015; Pazouki and Niinemets 2016). These intermediates are finally converted into terpenes by the action of terpene synthases (TPSs), which are named according to the length of their respective reaction products: hemiterpene- or isoprene- (C5), monoterpene- (C10), sesquiterpene- (C15), and diterpene- (C20) synthases (Chen *et al.* 2011; Tholl 2015; Pazouki and Niinemets 2016).

➢ *1.3.1. The first step: the terpenoid basic units, namely IPP and DMAPP, are synthetized in two compartmentally-separated metabolic pathways*

Both the IPP and DMAPP C5 precursors, which can be interconverted into each other, can derive from both the MVA and the MEP pathways; the former is localized into the cytosl, although it also operates in the endoplasmic reticulum and peroxisomes, whereas the latter operates in the plastids (Tholl 2015). Several studies indicate that the control of the metabolic fluxes into the two pathways is complex, being regulated at transcriptional, post-transcriptional and translational levels, as well as by feedback effects (reviewed in Hemmerlin 2013; Vranová *et al.* 2013; Tholl 2015).

The plant MVA pathway (Fig. 1.1, left) leads to the production of IPP through a sequence of six enzymatic steps. First, two molecules of acetyl-CoA are condensed to form acetoacetyl-CoA (AcAc-CoA) by the action of acetoacetyl-CoA thiolase (AACT). Then, AcAc-CoA is further condensed with a third molecule of acetyl-CoA by the action of 3-hydroxy-3-methylglutaryl (HMG) synthase (HMGS), yielding HMG-CoA, In the third, rate-limiting, reaction, HMG-CoA reductase (HMGR) reduces *(S)*-HMG-CoA to *(R)*-mevalonate at the expense of NADPH. The *(R)*-mevalonate is then phosphorylated twice at the expense of ATP, first by mevalonate kinase (MK), yielding mevalonate-5-phosphate, and then by phosphomevalonate kinase (PMK), converting mevalonate-5-phosphate into mevalonate-5-diphosphate. The last step in the MVA pathway is the

ATP-dependent decarboxylation of mevalonate-5-diphosphate to IPP, through the action of mevalonate diphosphate decarboxylase (MVD). The final product, namely IPP, can be isomerized to DMAPP by the action of IPP/DMAPP isomerase (IDI) (Tholl 2015; Abbas *et al.* 2017).

The MEP pathway is a series of seven reactions (Fig. 1.1, right). In the first step, 1-deoxy-D-xylulose 5-phosphate (DXP) synthase condensates (hydroxyethyl) thiamine diphosphate (derived from pyruvate) with glyceraldehyde-3-phosphate (GAP), to produce DXP. Next, DXP reducto-isomerase (DXR) catalyzes the rearrangement of the DXP molecule, which, after being reduced at the expense of NADPH, yields 2-C-methyl-D-erythritol 4-phosphate (MEP). In the third reaction, CTP donates its cytidyl moiety to MEP, and 4-diphosphocytidyl-2-C-methyl-D-erythritol (CDP-ME) is obtained, being CDP-ME synthase (MCT), the enzyme involved. Then CDP-ME is first phosphorylated by a kinase (CDP-ME kinase, CMK), to obtain CDP-ME 2-phosphate (CDP-ME2P), and then cyclized to 2-C-methyl-D-erythritol 2,4-cyclodiphosphate (MEcPP), after having lost its CMP (operated by MEcPP synthase, MDS). In the subsequent reaction, MEcPP is reduced to 4-hydroxy-3-methylbut-2-enyl diphosphate (HMBPP), by the action of HMBPP synthase (HDS). In the last step of the MEP pathway, HMBPP reductase (HDR) converts HMBPP into a combination of both IPP and DMAPP, with a stoichiometry of about 5:1 (Tholl 2015; Abbas *et al.* 2017).

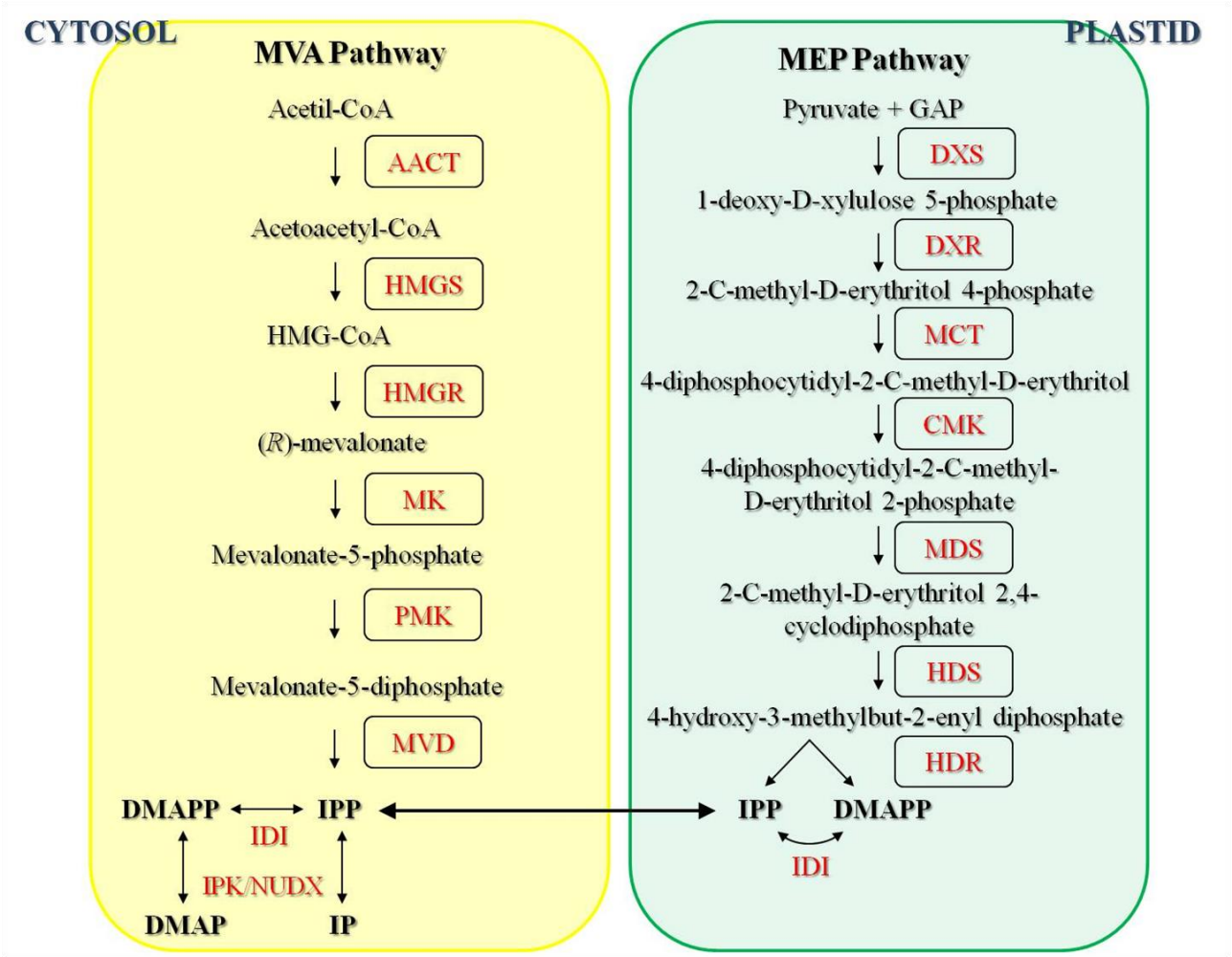


Fig. 1.1. *The two biochemical pathways leading to isopentenyl diphosphate (IPP), the basic unit for terpenoid biosynthesis in plants. Acronyms are in red for enzymes and in black for metabolites. Abbreviations: AACT, acetoacetyl-CoA thiolase; CMK, 4-diphosphocytidyl-2-C-methyl-D-erythritol kinase; DMAPP, dimethylallyl diphosphate; DXR, 1-deoxy-D-xylulose 5-phosphate reducto-isomerase; DXS, 1-deoxy-D-xylulose 5-phosphate synthase; GAP, glyceraldehyde-3-phosphate; HDR, 4-hydroxy-3-methylbut-2-enyl diphosphate reductase; HDS, 4-hydroxy-3-methylbut-2-enyl diphosphate synthase; HMG, 3-hydroxy-3-methylglutaryl-CoA; HMGR, HMG reductase; HMGS, HMG synthase; IDI, IPP/DMAPP isomerase; MCT, 4-diphosphocytidyl-2-C-methyl-D-erythritol synthase; MDS, 2-C-methyl-D-erythritol 2,4-cyclodiphosphate synthase; MEP, methylerythritol phosphate; MK, mevalonate kinase; MVA, mevalonic acid; MVD, mevalonate diphosphate decarboxylase; PMK, phosphomevalonate kinase. The double arrow denotes the cross-talk between cytosol and plastids based on the exchange of IPP. Adapted from Tholl 2015 and Abbas et al. 2017.*

The biosynthesis of terpenoids containing more than five carbon atoms (Fig. 1.2) requires an adequate supply of both IPP and of its more reactive isomer, i.e., DMAPP, being such isomerization accomplished by the intervention of IDI (see above). Therefore, IDI activity is of critical importance in the MVA pathway, since its final product is IPP only (Fig. 1.1, left), whereas in the MEP pathway, whose final products are both IPP and DMAPP (Fig. 1.1, right), IDI is thought to ensure an optimal IPP/DMAPP ratio for the assemblage of the C5 units leading to terpenoid precursors and/or to fuel export from plastids to cytosol. It is worth nothing (Fig. 1.2,

right) that DMAPP can be also used as a substrate for hemiterpene (C5) biosynthesis, by the activity of isoprene synthase (IPS in Fig. 1.2).

➢ *1.3.2. The intermediate step: IPP and DMAPP units are combined among each other to form prenyl diphosphates, the linear central precursors of all terpenoids*

Following their biosynthesis in the MVA and MEP pathways, IPP and DMAPP are used as building blocks for the assembling of the prenyl diphosphates (see below), from which all the terpenes derive. A large group of prenyltransferases (Fig. 1.2, see below), whose general name is also isoprenyl diphosphate synthases, are the enzymes in charge for producing prenyl diphosphates, whose biosynthesis always starts with the condensation of a single DMAPP with a single IPP, in a head-to-tail fashion. This allow the formation of a C10 prenyl diphosphate, to which one or more further IPP units can be added, again by head-to-tail condensation reactions, to produce short-chain (C15–C25), medium-chain (C30–C35), and long-chain (C40–Cn) prenyl diphospates.

Since the double bonds of the prenyl diphosphate to be formed can be either in *cis*- or in *trans* configuration, distinct families of *cis*-prenyltransferase or *trans*-prenyltransferase, respectively, have to come into play (Karel and Koyama 2003). Until recently, research on short-chain plant prenyltransferases was mostly concentrated on the *trans*-acting enzymes, because it was believed that the *cis*-acting ones were only involved in the synthesis of large prenyl diphosphates having more than 50 carbon atoms, such as the C70–C120 dehydrodolichol diphosphates (Takahashi and Koyama 2006; Surmacz and Swiezewska 2011) (Fig. 1.2). As a consequence, structural and catalytic features have been studied in details for *trans*-prenyltransferases, such as geranyl diphosphate synthase (GPS), farnesyl diphosphate synthase (FPS) and geranylgeranyl diphosphate synthase (GGPS), synthesizing the corresponding *trans*-prenyl diphosphates which play major roles in terpenoids biochemistry (Tholl 2015; Fig. 1.2). Recently, *cis*-prenyltransferase analogs of the *trans*-acting enzymes have been detected, among which neryldiphosphate synthase (NDPS), (*Z,Z*)-FPP synthase and nerylneryldiphosphate synthase (NNDPS), which are equally able to produce short-chain, metabolically versatile, prenylphosphates to be used in terpenes synthesis (Sallaud *et al.* 2009; Akhtar *et al.* 2013; Fig. 1.2).

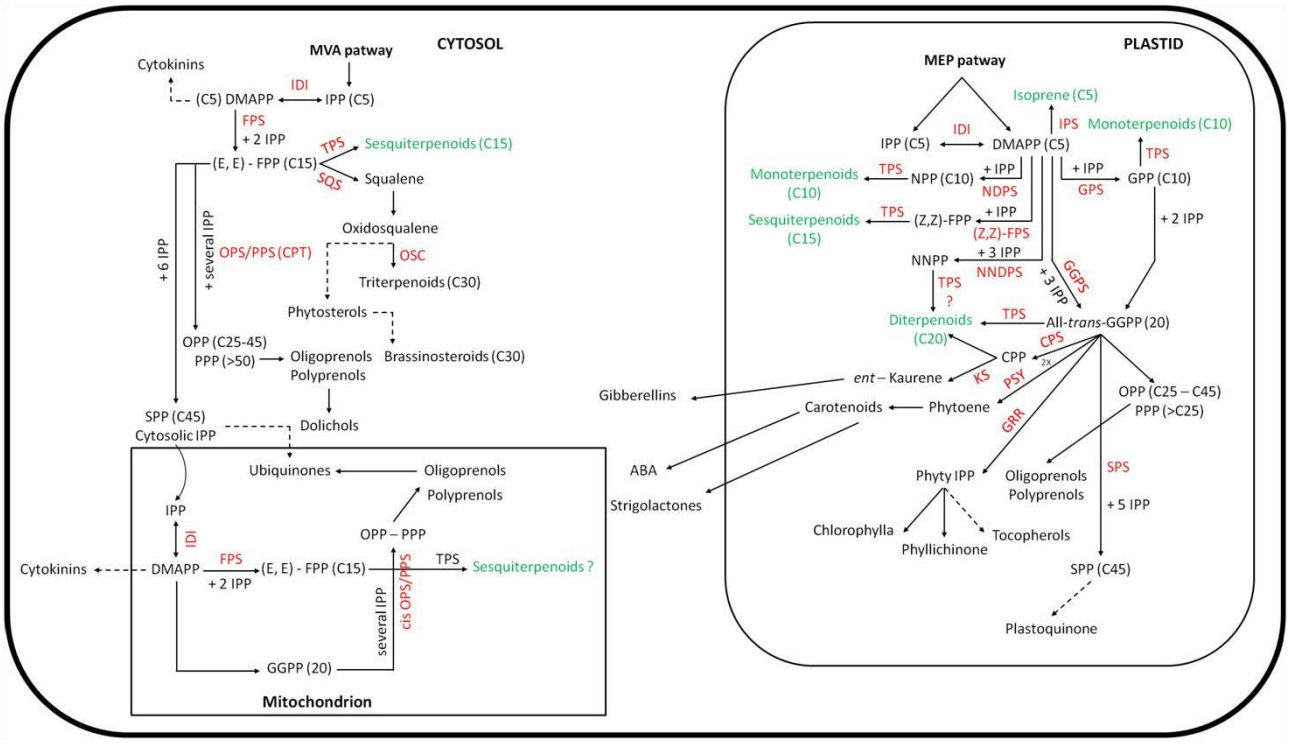


Fig. 1.2. *Terpenoid biosynthetic pathways and their subcellular localization. Enzymes are marked in red and specialized terpenoids are marked in green; all other intermediates and terpenoid end products are in black. Solid and dashed arrows indicate single and multiple enzymatic steps, respectively. Abbreviations: ABA, abscisic acid; CPS, ent-copalyl diphosphate synthase; CPT, cis-prenyltranferase; DMAPP, dimethylallyl diphosphate; FPP, farnesyl diphosphate; FPS, FPP synthase; GGPP, geranylgeranyl diphosphate; GGPS, GGPP synthase; GPP, geranyl diphosphate; GPS, GPP synthase; GRR, geranylgeranyl reductase; IPP, isopentenyl diphosphate; KS, ent-kaurene synthase; MEP, 2-C-methyl-D-erythritol 4-phosphate; NDPS, neryldiphosphate synthase; NNPP, nerylneryl diphosphate; NPP, neryl diphosphate; OPP, (all-E)-octaprenyl diphosphate; OSC, oxidosqualene cyclase; OPS, oligoprenyl diphosphate synthase; PPP, prenyl diphosphate; PPS, prenyl diphosphate synthase; PSY, phytoene synthase; SPP, solanesyl diphosphate; SQS, squalene synthase; TPS, terpene synthase. Adapted from Tholl et al. 2015.*

➢ *1.3.3. The final step: prenyl diphosphates are key precursors for the biosynthesis of both primary and specialized terpenoids*

The products of the catalytic action of prenyltransferases, namely *trans*- and *cis*-prenyl diphosphates, are then used in several plant cell compartments such as plastids, mitochondria, and the cytosol, for the formation of a myriad of terpenoids, destined to either the primary or the secondary metabolism (Fig. 1.2). For instance, *trans*- or *cis*-prenyl diphosphate pairs, such as GPP or NPP, (*E,E*)-FPP or (*Z,Z*)-FPP, and GGPP or NNPP, are acted upon by specific TPSs, to yield specialized metabolites such as monoterpenes, sesquiterpenes, and diterpenes, respectively (Fig. 1.2).

Moreover, the combined action of prenyltransferases and TPSs can lead to the formation of precursors used for the production of metabolites involved in primary metabolism. For example,

the condensation of two molecules of FPP in a head-to-head fashion and the consequent loss of both diphosphate groups, allow the production of squalene, the precursor of phytosterols (Fig. 1.2). A similar condensation reaction of two molecules of GGPP produces phytoene, fueling the downstream synthesis of carotenoids (Fig. 1.2). Geranylgeranyl diphosphate is also involved in the biosynthesis of *ent*-kaurene, from which all the plant gibberellins derive; this conversion is carried out in sequence by two structurally related TPSs, first *ent*-copalyl diphosphate (*ent*-CPP) synthase (CPS), which transform GGPP into *ent*-CPP, and then by kaurene synthase (KS), a lyase which removes diphosphate from *ent*-CPP and cyclizes it to *ent*-kaurene (Fig. 1.2).

➢ *1.3.4. Cross talk and interactions between the MVA and MEP pathways*

The MVA and the MEP pathways fuel distinct routes in terpenoids metabolism. The MVA pathway mostly feeds the cytosolic formation of sesquiterpenoids, polyprenols, phytosterols, brassinosteroids, and triterpenoids, and the mitochondrial production of ubiquinones and polyprenols. The MEP pathway, instead, acts mainly as a source of hemiterpenoids (isoprene), monoterpenoids, diterpenoids, carotenoids and their breakdown products, cytokinins, gibberellins, chlorophyll, tocopherols, and plastoquinones (Fig. 1.2). The metabolic cost of maintaining two IPP/DMAPP-metabolic pathways in plants has apparent benefits by allowing a wider capacity to evolve specialized terpenoid pathways and better control of compartment-specific isoprenoid pools.

The physical separation of the two pathways was supported by genome-wide co-expression analyses in *Arabidopsis*, which showed limited interaction between MVA and MEP genes (Vranova *et al.* 2013; Rodríguez-Concepción and Boronat 2015). However, there is now evidence that between the two compartmentally separated biosynthetic pathways metabolic “cross talk” does take place for substrate formation, via the exchange of IPP/DMAPP, and C10–20 prenyl diphosphate intermediates (GPP, FPP and GGPP) (Flügge and Gao 2005; Orlova *et al.* 2009; Dong *et al.* 2016). For instance, isotope-labelling experiments have shown the integration of MEP-derived IPP/DMAPP into both monoterpenoids and sesquiterpenoids in *Antirrhinum majus* and *Daucus carota* (Dudareva *et al.* 2005; Hampel *et al.* 2005). Analogously, the contribution of the MVA pathway to C10–C40 terpenoid biosynthesis was proved in *Gossypium hirsutum* (Opitz *et al.* 2014). Moreover, the application of MVA or MEP pathway-specific inhibitors cannot completely block terpenoid biosynthesis in cytoplasm or plastid, indicating that the common precursors of these two pathways can be freely transferred among different subcellular compartments (Kasahara *et al.* 2002; Bick and Lange 2003; Hemmerlin*et al.* 2003; Laule *et al.* 2003; Gutensohn *et al.* 2013).

A possible exchange of isoprenoid intermediates among plastids and cytosol has also been deduced from studies on mutant lines overexpressing the genes of the MVA or the MEP pathway. IPP and DMAPP are primarily derived from the MEP pathway in *Lavandula latifolia*. Nevertheless, a significant increase of monoterpenoids such as 1,8-eucalyptoland camphor was observed in mutant lines overexpressing HMGR, a MVA pathway enzyme (see above) (Mendoza-Poudereux *et al.* 2015). Likewise, overexpression of HMGR in *Salvia miltiorrhiza* also increased the production of tanshinones, a class of diterpenoid compounds that is assumed to be derived from the MEP pathway (Kai *et al.* 2011; Shi*et al.* 2014). One plausible explanation is that more IPP synthesized in the cytoplasm is transferred to the plastid to function as precursor to the MEP pathway. Reciprocally, *Artemisia annua* lines able to overexpress HDR, a plastidial enzyme involved in the MEP pathway (see above), showed an enhanced production of artemisinin and other sesquiterpenes, which are reputed to derive from the MVA pathway. Confocal microscopy and green florescence protein fusion showed that HDR was located in chloroplast and the transport of IPP from chloroplast to cell cytoplasm was observed after 13C labelling experiment, indicating that more IPP was accessible to the MVA pathway for terpenoid synthesis (Ma *et al.* 2017).

To date, no specific transporter mediating the flux of isoprenoid precursors among cellular compartments has been identified, although some transporter-assisted modes of inter change were suggested. The export of IPP from plastids to the cytosol was proposed to proceed by a plastidial proton symport system (Bick and Lange 2003), while the study of Flügge and Gao (2005) indicated that IPP is not transported by a plastidial phosphate translocator but it is instead dependent on the presence of phosphorylated counter-substrates. pH gradients may also be implicated in IPP translocation, as it appears to be the case for the movement of protonated abscisic acid among different compartments (Baier and Hartung 1988). Similarly, IPP protonation under acidic conditions may allow it to travel across the plastidial membrane without the intervention of a specific transporter (Cherian *et al.* 2019).

Molecular mechanisms regulating the cross talk among the cellular compartments remain also elusive. This might be due to the confounding effects of a variety of factors, including post-transcriptional, translational and post-translational processes that modulate the fluxes (Kumar *et al.* 2012; Lange *et al.* 2015; Tholl 2015). Moreover, it became clear that in addition to endogenous factors, exogenous stimuli could also affect the interaction between MVA and MEP pathways (Tholl 2015). For instance, sucrose supplementation induced the activity of SnRK1 (Sucrose non-fermenting-1-related protein Kinase), which reduced the activity of HMGR by phosphorylation (Polge and Thomas 2007), and therefore increased the substrates availability for the MEP pathway. As consequence, the production of chlorophyll was increased in *Arabidopsis* seedlings grown in a

medium supplemented with sucrose (Laby *et al.* 2000). Moreover, exposure to light down-regulates the expression of genes in the MVA pathway and decreases the level of sterols (Ghassemian *et al.* 2006; Rodríguez-Concepción 2006), but up regulates the expression of MEP pathway genes and genes for carotenoid and chlorophyll biosynthesis (Ghassemian *et al.* 2006; Rodríguez-Concepción 2006; Cordoba *et al.*2009; Meier *et al.* 2011). In addition, light-activated metabolism leads to a higher production of substrates for the MEP pathway, e.g., GAP from the Calvin cycle, which helps to increase the production of IPP and DMAPP in chloroplasts and leads to an increase in MEP-derived terpenoids. These results are supported by studies in which an increased carbon flux through the MEP pathway has been observed following plant exposure to increasing light intensity (Mongelard *et al.*2011). Conversely, the expression of MEP pathway genes is decreased during the transition from light to dark (Vranova*et al.* 2013) and exposure to dark can instead boost the activity of MVA enzymes, such as HMGRs, whose dark-induced up-regulation increased the biosynthesis of the triterpene ginsenoside in ginseng (Kim *et al.* 2014).

Although the exchange flux between compartments and pathways might be limited or negligible under non-stressed conditions, its importance might increase to compensate for stress conditions or developmental stages which specifically impair or suppress one of the two biosynthetic pathways (Dudareva *et al.* 2005; May *et al.* 2013; Rasulov *et al.* 2015). For instance, it has been proposed that, during the germination of *Arabidopsis* seedlings in the dark, prenyl diphosphates derived from the MVA pathway are transferred to etioplasts to fuel carotenoid and gibberellin synthesis prior to the induction of MEP pathway enzymes under illumination (Rodríguez-Concepción *et al.* 2004).

Late studies have shown that terpenoid biosynthesis by MVA and MEP pathways is not exclusively channelled via IPP and DMAPP, but may require a pool of the respective isopentenyl and dimethylallyl monophosphates, namely IP and DMAP, respectively (Henry *et al.* 2015, 2018). The IPP and IP pools are controlled by two classes of enzymes, the Nudix hydrolases (Nudxs), a large family of two-domain hydrolases/peptidases widely detected in bacteria, plants and animals, and the IP kinases (IPKs), which catalyse the hydrolysis and phosphorylation of IPP and IP, respectively (Fig. 1.1) (Henry *et al.* 2015, 2018). IPKs were first identified in archaebacteria as part of their modified MVA pathway for isoprenoid biosynthesis (Dellas *et al.* 2013). Lately, it has been shown that IPK homologs are extensively scattered in plant genomes, where they take place together with the complete set of MVA and MEP genes (Vannice *et al.* 2014). In *Arabidopsis*, it has been shown that IPK is localized in the cytosol and seems to control the production of both MVA- and MEP-derived terpenoids (Henry *et al.* 2015). Indeed, *Arabidopsis* T-DNA insertion knock outlines for IPK showed a significant decrease in the levels of

sesquiterpenes and sterols, whereas the overexpression of the *AtIPK* gene in transgenic *Nicotiana tabacum* led to significant increases in monoterpenes and sesquiterpenes. As far as the Nudxs genes are concerned, the *Arabidopsis* AtNudx1 and AtNudx3 proved their effectiveness in dephosphorylating IPP and DMAPP (Henry *et al.* 2018). *Arabidopsis* T-DNA insertion knockout lines for AtNudx1 and AtNudx3 showed an increased production of monoterpenes, sesquiterpenes, and sterols. Conversely, overexpression of these enzymes in *N. tabacum* resulted in a significantly decreased production of monoterpenes and sesquiterpenes (Henry *et al.* 2018). Although further studies are needed to understand the importance of IPK and Nudx genes in plant terpenoid metabolism, these findings highlight the potential of such pathway reactions to possibly operate as additional regulatory mechanisms for balancing IP/DMAP and IPP/DMAPP pools in terpenoid and other isoprenoid biosynthesis (Henry *et al.* 2015, 2018; Karunanithi and Zerbe 2019).

1.4. PLANT TERPENE SYNTHASES

As described above, TPSs transform acyclic *cis*- or *trans*-prenyl diphosphate intermediates bearing 5 to 20 carbon atoms into hemiterpenes (C5), such as isoprene, monoterpenoids (C10), sesquiterpenoids (C15), or diterpenoids (C20) (Fig. 1.2). These terpenoid compounds can be further modified to produce biologically active final products of greater structural diversity by means of secondary enzymatic reactions such as methylation, hydroxylation, glycosylation, peroxidation, acylation, or cleavage (Tholl *et al.* 2015). The enormous variety of terpenoids in specialized metabolism can be attributed mainly to the activity of TPS superfamily, which comprises a huge and still growing number of enzymes from nearly all plant species (Chen *et al.* 2011). TPS enzymes enable the adaptation of terpene metabolism to a changing environment due to their heterogeneous activity, which often leads to the production of more than one single compound; moreover, TPS proteins are able to easily acquire new catalytic properties as a consequence of minor structural changes (Tholl 2015; Pazouki and Niinemets 2016). As a result, TPS enzymes have attracted increasing consideration for *in planta* as well as heterologous metabolic engineering of terpenoid products, in view of their pharmaceutical and industrial uses (Bohlmann and Keeling 2008; Chen *et al.* 2011; Singh and Sharma 2015; Tholl 2015).

➢ *1.4.1. Structure of plant terpene synthases*

The functional diversity of TPSs is determined by their modular structure, based on a conserved α-helical folding pattern within which three common domains, denoted as γ, β, and α, are variably arranged (Cao *et al.* 2010) (Fig. 1.3). Based on the presence of either one, or two active sites, and their related catalytic motifs, TPSs are said to be monofunctional or bifunctional, respectively, and are categorized into class-I, class-II, or class-I/II enzymes (Chen *et al.* 2011; Tholl 2015).

In the class-I TPSs, which host only one active site and include all the MTPSs, all the STPSs and part of the DTPSs (Fig. 3), catalysis takes place in the C-terminal α-domain, where the ionization of the prenyl diphosphate substrate is mediated by a divalent cation. Electron abstraction operated by such metal cofactors increases the proneness of the enzymatic substrate to undergo cyclization, chemical shifts and molecular restructuring, to yield the final product (Tholl 2015). The α-domain of class-I TPSs contains two metal binding motifs, the highly conserved "DDXXD" motif and the less conserved "NSE/DTE" one, located on opposing helices near the entrance of the active site. Both motifs help to position the prenyl diphosphate substrate by binding of it to a trinuclear magnesium cluster, which triggers the ionization of the substrate and initiate the cyclization reaction.

The class-II TPSs, which also host just one active site and include only DTPSs (Fig. 1.3), contain a functional N-terminal β-domain together with a third "insertion" γ-domain. Within such β-domain, a conserved "DXDD" motif is present, which is responsible for the protonation-initiated cyclization of the substrate (Christianson 2017). The γ-domain has a highly acidic "EDXXD-like" motif, which further contributes to the activity of class-II TPSs (Cao *et al.* 2010).

Finally, a limited number of three-domain (γβα) TPSs contain all the three functional active sites, namely "EDXXD-like", "DDXXD" and "DDXD". These bifunctional TPSs, all of which are DTPSs, are said, therefore, to be class-I/II enzymes (Fig. 1.3). They include the diterpene synthases (CPS/KS) found in the mosses *Physcomitrella patens* and *Jungermannia subulata* and in the lycophyte *Selaginella moellendorffii,* which catalyze the formation of *ent*-kaurene (and 16-hydroxykaurene) via a CPP intermediate in the biosynthesis of kaurenoic acid (Hayashi *et al.* 2006; Mafu *et al.* 2011). Similar class-I/II DTPSs are also present in gymnosperms and are regarded as early DTPSs (Keeling and Bohlmann 2006; Hall *et al.* 2013a). These enzymes catalyze the formation of an enzyme-bound CPP from GGPP and then convert CPP to a diterpene (Peters *et al.* 2003).

The available protein crystal structures indicate that the majority of plant DTPSs and some STPSs have all the three domains, namely γ, β, and α (Cao *et al.* 2010). In general, however, just one of the domains is functional. For instance, DTPSs involved in gibberellin biosynthesis in both

angiosperms and gymnosperms, as well as those implicated in the biosynthesis of the large group of labdane-type specialized diterpenes in angiosperms, are class-II or class-I enzymes, with loss of activity in the α- or β-domains, respectively (Fig. 1.3). In the gibberellin biosynthetic pathway, the class-II enzymes *ent*-copalyl diphosphate (CPP) synthases (CPSs) protonates its linear substrate GGPP, so catalyzing its conversion into the two-rings cyclization product *ent*-CPP. A subsequent elimination reaction brought about by a class I *ent*-kaurene synthase (a lyase) converts *ent*-CPP into *ent*-kaur-16-ene plus diphosphate (Zi *et al.* 2014) (Fig. 1.2). Moreover, in both gymnosperms and angiosperms, some DTPSs involved in the secondary metabolism have been found to retain class-I activity only (without the loss of the β-domain) (Fig. 1.3) and use GGPP to directly produce a diterpene without a CPP intermediate (Chen *et al.* 2011; Köksal *et al.* 2011a; Hall *et al.* 2013a). Interestingly, the occurrence of three-domain (γβα) STPSs that retain only class-I TPS activity has also been reported (Fig. 1.3), as seen in α-bisabolene synthase (Ag BIS) from *Abies grandis* (McAndrew *et al.* 2011). Evolution by loss of function can also be envisaged for MTPSs (all belonging to class-I TPSs, see above), in most of which the β-domain is actually present, but is rendered inactive because of the absence of the conserved "DXDD" motif (Whittington *et al.* 2002) (Fig. 1.3).

Most MTPSs and DTPSs, unlike the cytosolic STPSs, have obvious N-terminal plastid transit peptides (Fig. 1.3). Transit peptides are removed from the mature TPS upstream of the "RR(X_8)W" motif, which is important for the catalysis of monoterpene cyclization (Whittington *et al.* 2002; Hyatt *et al.* 2007) and is also maintained with some differences in most STPSs and DTPSs (Chen *et al.* 2011) (Fig. 1.3).

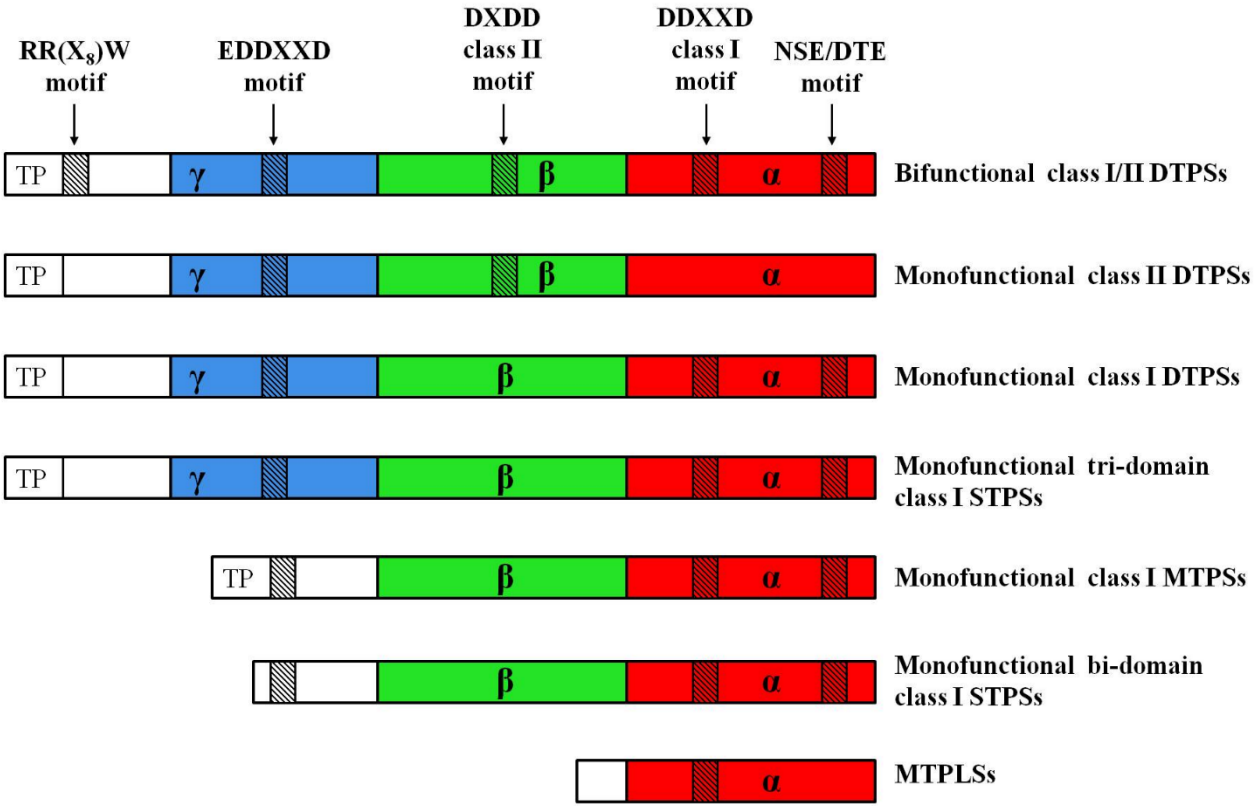


Fig. 1.3. *Structural features of plant monoterpene- (MTPSs), diterpene- (DTPSs), sesquiterpene (STPSs) synthases, and of microbial terpene synthase-like proteins (MTPSLs) based on the combination of the γ, β, and α domains and the presence of distinctive aminoacidic motifs. The two highly conserved aspartate-rich catalytic motifs "DDXXD" and "DXDD" responsible for class I and class II activities, and the less well conserved "NSE/DTE" and "EDDXXD"motifs, which also contribute to the activity of class-I and class-II TPSs, are indicated. The N-terminal plastid transit peptide (TP) and "RR(X8)W" motif are also shown.*

A new class of plant TPSs designated as microbial terpene synthase-like proteins (MTPLSs) has recently been recognized in the lycophyte *S. moellendorffii* (Li *et al.* 2012). This new type of TPSs, which have been found to be present in several cryptogamae, but not in seed plants, neither in green algae, are evolutionary much more strongly related to microbial TPSs rather than to the typical plant ones (Jia *et al.* 2018). According to their different origin (see below), MTPLSs and typical plant TPSs differ for numerous features. Firstly, at the gene level, because MTPSLs do not show a conserved intron–exon structure, unlike typical plant TPSs, which instead can be assigned to one out of three classes depending on the presence of 12–14 introns, nine introns, or six introns (Trapp and Croteau 2001). In the MTPSLs genes of *S. moellendorffi*i, for example, the number of introns can vary from none, or just one, in most cases, to as many as seven in certain others. Secondly, structural differences also emerged at the protein level. In fact, all known MTPSLs contain the α-domain only (Fig. 1.3), unlike the (βα)- or (γβα)-modular structures of typical plant TPSs (see above). As consequence, MTPSLs polypeptides are much shorter (about 350 amino acid residues), than those of typical plant TPSs (see above). Thirdly, although most MTPSLs, similarly

to typical plant TPSs, possess a canonical aspartate-rich "DDxxD" motif in their active sites, they can also host variants, such as "DDxxxD" and "DDxxx" (Jia *et al.* 2018).

➢ *1.4.2. Origin and evolution of plant terpene synthases*

Fig. 1.4 sketchily depicts current models and hypotheses illustrating the recombination among the γ, β, and α domains and key aminoacidic motifs during the evolution of terpene synthases in plants.

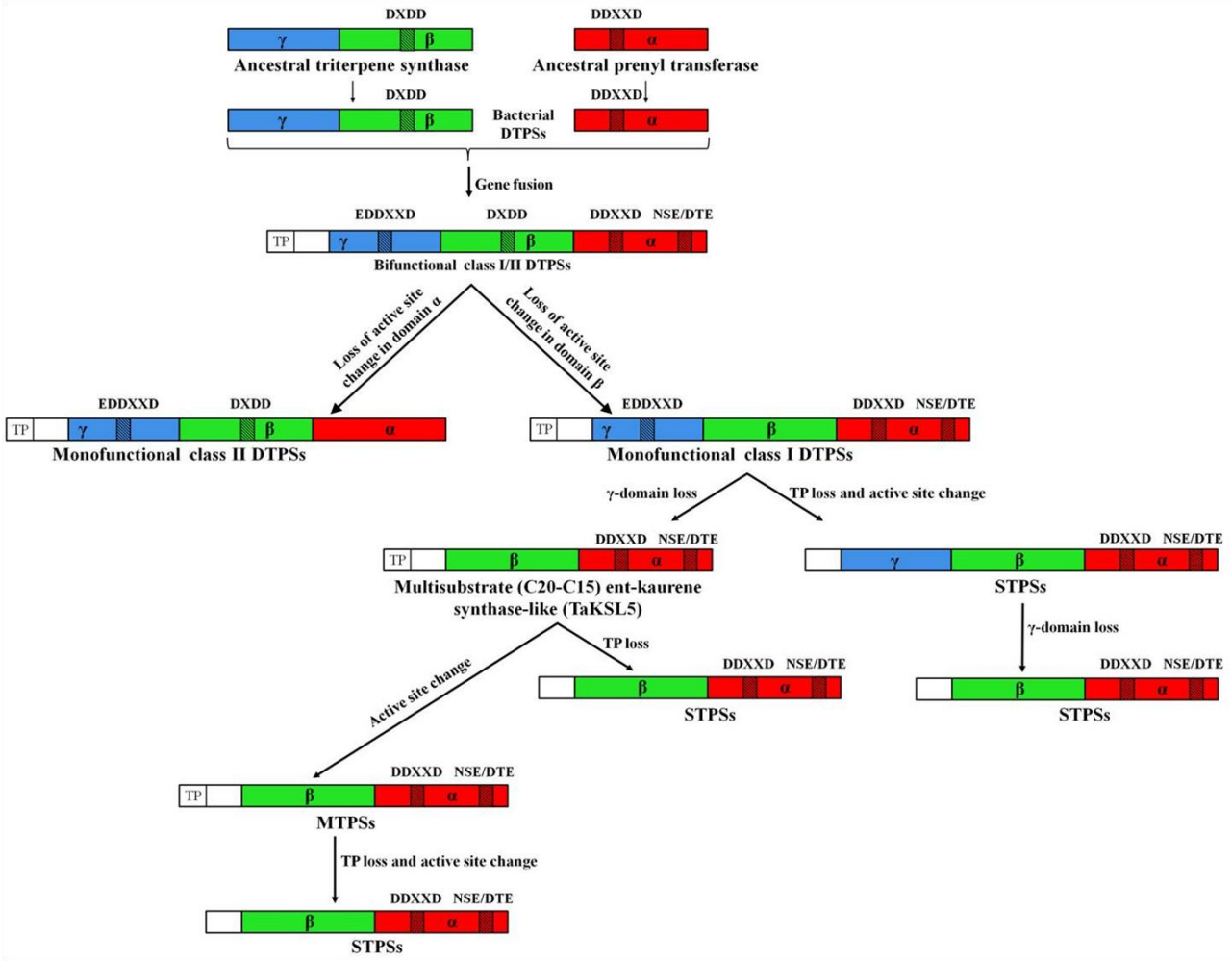


Fig. 1.4. *Proposed model for the structural evolution of diterpene synthases (DTPSs) (Cao et al. 2010; Gao et al. 2012) and hypothesis on evolution of sesquiterpene synthases (STPSs) and monoterpene synthases (MTPSs) according to two potential routes based on the analysis of multisubstrate enzymes (Pazouki and Niinemets 2016).*

According to the original evolutionary model proposed by Trapp and Croteau (2001), the plant TPS ancestor, which was similar to the present-day conifer DTPSs involved in gibberellin biosynthesis, originated prior to the divergence of gymnosperms and angiosperms. Afterwards, the class-I/II DTPS from *P. patens*, was assumed to be the common ancestor (Hayashi *et al.* 2010;

Keeling *et al.* 2010). Notably, two diterpene synthases has been found in the bacterium *Bradyrhizobium japonicum*, namely *ent*-copalyl diphosphate synthase and *ent*-kaurene synthase, which show structural relatedness with the β-domain (class-II activity) and the α-domain (class-I activity) of plant and fungal TPSs, respectively. This led Morrone *et al.* (2009) to hypothesize a common DTPS ancestor among plants, fungi, and bacteria.

As the crystal structures of several plant TPSs, such as those for IPS, MTPS, STPS, class-I, class-II and class-I/II DTPSs, became available (Kampranis *et al.* 2007; Gennadios *et al.* 2009; Köksal *et al.* 2011a, 2011b and 2014; Zhou *et al.* 2012), it became clearer that domain architecture is a highly conserved feature among all TPSs. This prompted Cao *et al.* (2010) to make an attempt to unify pre-existing models concerning TPS evolution. They hypothesized that a prototypical γβα-domain structure might have arisen from the fusion among two ancestral DTPS genes of bacterial origin, one bearing an γβ-domain and the other an α-domain, which, in turn, might have evolved from pre-existing ancestors, i.e., a γβ-domain triterpene synthase and an α-domain prenyltransferase, respectively (Fig. 1.4, upper part). In the course of evolution, the ancestral three-domain (γβα) class-I/II DTPS described above, with functional α- and β-domains and a transit peptide, resembling the present-day abietadiene synthase from gymnosperms and the CPS/KS from *P. patens*, acted as the progenitor of both class-II type DTPSs (γβα-assemblies in which the α-domain activity has been lost) and class-I type DTPSs (γβα-assemblies in which the β-domain activity has been lost) (Cao *et al.* 2010; Gao *et al.* 2012) (Fig. 1.4, central part). Beyond the milestone events outlined above, evolution might have exerted a strong pressure towards specialization and diversification of terpenes biochemistry, to the point that some of the structures which were present and functional in the ancestral TPS might have become redundant, if not disadvantageous, on an evolutionary perspective; this might explain why IPSs and MTPSs have lost their γ-domain, while STPSs lost the transit peptide and, in most cases, the γ-domain (Hillwig*et al.* 2011; Köksal*et al.* 2011a, 2011b; Rajabi*et al.* 2013) (Fig. 1.4).

Recently, the analysis of several proteins with mixed substrate specificity allowed to put forward new hypotheses about the pathways and the timing of the main evolutionary changes regarding the loss of γ-domain and transit peptide in TPSs (Pazouki and Niinemets 2016) (Fig. 1.4, bottom part). For instance, structural analysis of a bi-domain (α-β) DTPS from *Triticum aestivum* (TaKSL5), which can use either *ent*-CPP for the production of *ent*-kaurene (C20) or (*E*,*E*)-FDP for the production of (*E*)-nerolidol (C15) (Hillwig *et al.* 2011), suggested that evolution of STPSs probably took place first as a result of the loss of γ-domain, followed by changes in subcellular sorting (loss of transit peptide) and further diversification by the loss of capacity for using C20 substrates. This hypothesis is supported by the finding of multi-substrate (*E*)-nerolidol

(C15)/(*E,E*)-geranyllinalool (C20) synthases in *Vitis vinifera*, namely VvPNLNGl1 and VvCSENerGl, which are able to use both C15 and C20 substrates, but lack both the transit peptide and the γ-domain (Martin *et al.* 2010). The above mixed-substrate TPSs from *T. aestivum* and from *V. vinifera* represent three putative intermediates in the evolution of STPSs directly from DTPSs by loss of γ-domain, which is predicted to have preceded the loss of the transit peptide (Fig. 4). This evolutionary scenario was also confirmed by evidence concerning the evolution of STPSs from MTPSs (Fig. 1.4, bottom part). Indeed, mixed-substrate TPS have been found in a number of species, among which: *Antirrhinum majus,* which has two (*E*)-nerolidol (C15)/linalool (C10) synthases, i.e. AmNES/LIS-1 and AmNES/LIS-2 (Nagegowda *et al.* 2008); *Arabidopsis thaliana,* which has two (*E,E*)-α-farnesene (C15)/(*E*)-β-ocimene (C10) synthases, i.e. AtTPS02 and AtTPS03 (Huang *et al.* 2010); and *Fragaria ananassa,* which has two (*E*)-nerolidol (C15)/linalool (C10) synthases, i.e. FaNES1 and FaNES2 (Aharoni *et al.* 2004). In each of the above three species, one TPS in each pair, namely AmNES/LIS-1, AtTPS03, and FaNES1, lacks the N-terminal transit peptide, whereas the other, namely AmNES/LIS-2, AtTPS02, and FaNES2, does have it, which suggests that STPSs might have evolved from MTPSs by a loss of the transit peptide and a change in the active site (Fig. 1.4).

Taken together, the above findings suggest an evolutionary model in which, starting from a tri-domain (γβα) class-I DTPS, the loss of the γ-domain first led to the formation of a bi-domain (βα) DTPS (e.g., the TaKSL5 of *T. aestivum*) and ultimately to a STPS, by the loss of the transit peptide (e.g., VvPNLNGI1 and VvCSENerGI in *V. vinifera*) or to a MTPS by change of the active site using a different substrate. In this latter case, the loss of the transit peptide led to a cytosol-localized STPS (Fig. 1.4).

On the other hand, in gymnosperms, there is also evidence for a different evolutionary scenario for the formation of STPSs. According to this alternative hypothesis, a tri-domain (γβα) cytosol-active TPS was produced by the loss of the transit peptide from a class-I DTPS. This could be eventually followed by the loss of γ-domain, leading to the formation of a bi-domain (βα) cytosol-localized TPS. Evidence for such hypothesis came from the analysis of three *Abies grandis* C10/C15 multisubstrate TPSs, namely (*E*)-α-bisabolene, δ-selinene- and γ-humulene synthases (Bohlmann *et al.* 1998). All these three TPSs lack the transit peptide, but the first is a tri-domain (γβα) protein whereas the other two are bi-domain (βα) proteins (Bohlmann *et al.* 1998). This suggests that in the evolution of gymnosperm STPSs, the transit peptide could have been lost first, followed by the loss of the γ-domain (Fig. 1.4, bottom part).

New insights concerning the evolution of plant TPSs came from the identification in cryptogamae of the MTPSL genes, which encode proteins containing only the α-domain and

produce mainly monoterpenes and sesquiterpenes (Li *et al.* 2012; Jia *et al.* 2018). In accordance with the evolution of the two domains (βα) typical plant TPSs from the three-domain (γβα) proteins (Fig. 1.4), MTPSLs may have originated either from an ancestral γβα TPS, which has then lost its γ-β domains, or from a βα TPS progenitor after the loss of the β domain. All the evidence produced so far, however, negated the above evolutionary hypothesis, because it was found that MTPSLs are considerably more related to the microbial TPSs possessing only the α-domain, rather than to the α-domain of typical plant TPSs, thus suggesting that MTPSLs and typical plant TPSs descended from distinct progenitor proteins. Since both bacteria and fungi have evolved before land plants, then it is reasonable to assume that horizontal gene transfer played a role in passing TPS genes from the former to the latter groups of organisms (Jia *et al.* 2018). Indeed, phylogenetic analyses indicated that MTPSLs have different degrees of relationship to bacterial TPSs and fungal TPSs, and show lineage-specific features, to the point that two distinct MTPSLs families can be recognized, depending on their clustering with bacterial or with fungal TPSs. In the latter MTPSL family, in turn, three separate subgroups can be identified. On such basis, it is therefore firmly reputed that not only MTPSL genes in cryptogamae derived from bacteria and fungi but also that genes acquisition from fungi during evolution have occurred more than once.

Since cryptogamae harbor both MTPSLs and typical plant TPSs, whereas seed plants possess only the latter, an interesting evolutionary question may arise. By considering that in non-seed plants MTPSLs produce mainly monoterpenes and sesquiterpenes, whereas typical plant TPSs mainly diterpenes, then the development of seed plants must have been accompanied by the loss of MTPSLs and their replacement by TPSs able to produce monoterpenes and sesquiterpenes. Understanding when and why these processes may have taken place should provide new insights into the evolution of plant isoprenoids.

➢ *1.4.3. Size and phylogeny of typical plant TPS family*

Genome-wide analysis of several plant species indicated that typical TPSs are encoded by medium- to large gene families, whose size range from 20 to 170 members, counting both full-length genes and pseudogenes, with the only notable exception of the moss *P. patens*, whose genome contain a single functional TPS gene (Table 1.1). In particular, full-length genes coding for typical plant TPSs ranges from 14 in *S. moellendorffii* to 113, 106 and 69 in *Eucalyptus grandis*, *Eucalyptus globulus* and *V. vinifera*, respectively, with *Ocimum sanctum* (47), *Populus trichocarpa* (38), *Oryza sativa* (32), *A. thaliana* (32), *Solanum lycopersicum* (29) and *Sorghum bicolor* (24) possessing an intermediate but large number of putative functional TPS genes (Table 1.1). The recently draft genome assemblies of *Picea glauca* (Birol *et al.* 2013; Warren *et al.* 2015),

Picea abies (Nystedt *et al.* 2013), and *Pinus taeda* (Neale *et al.* 2014), provided the first opportunities for a genome-wide annotation of TPS genes in conifers. Indeed, Warren *et al.* (2015), using the *P. glauca* PG29 V3 genome assembly, identified 83 unique TPS genes having at least 400 amino acids of coding region, including 28 pseudogenes (Table 1.1). This confirmed previous results based on transcriptome analyses, which estimated that more than 70 distinct transcriptionally active TPS genes may be present in a single conifer species, (Keeling *et al.* 2011), comparable to the number of potentially active TPS genes found in the sequenced genomes of angiosperms (Table 1.1).

Table 1.1. *Size of typical plant TPS family and subfamilies in selected plant species*

Species	Genome size (Mb)	Chromosome number (1N)	Total TPS gene models	Putative full length TPSs	TPS subfamily (full length TPS genes)						
					a	b	c	d	e/f	g	h
Arabidopsis thaliana[1]	135	5	40	32	22	6	1	0	2	1	0
Eucalyptus globulus[2]	530	11	143	106	45	38	2	0	11	10	0
Eucalyptus grandis[2]	640	11	172	113	52	36	2	0	10	13	0
Ocimum sanetum[3]	386	8	81	47	18	16	5	0	3	5	0
Oryza sativa[4]	375	12	57	32	18	0	3	0	9	2	0
Physcomitrella patens[5]	480	27	4	1	0	0	1	0	0	0	0
Picea glauca[6]	20000	12	83	55	0	0	1	53	1	0	0
Populus trichocarpa[7]	430	18	57	38	16	14	2	0	3	3	0
Selaginella moellendorffii[8]	106	27	18	14	0	0	3	0	3	0	8
Solanum lycopersicum[9]	828	12	44	29	12	8	2	0	5	2	0
Sorghum bicolor[4]	730	10	48	24	15	2	1	0	3	3	0
Vitis vinifera[10]	486	19	152	69	30	19	2	0	1	17	0

Number of TPS genes retrieved from information based on literature data: [1]Auborg *et al.* 2002; [2]Kulheim *et al.* 2015; [3]Kumar *et al.* 2018; [4]Chen *et al.* 2011; [5]Hayashi *et al.* 2006; [6]Warren *et al.* 2015; [7]Irmisch *et al.* 2014; [8]Li *et al.* 2012; [9]Falara *et al.* 2011; [10]Martin *et al.* 2010.

The above data suggest that the evolution of typical plant TPS families in higher plants involved several gene duplication events, followed by sub- and neo-functionalization. This is also suggested by the fact that many TPS genes in the genomes of flowering plants are present in tandem arrays, each made of two or more genes, among which one or a few unrelated genes are sometimes interlayered. In *A. thaliana*, *E. grandis*, *P. trichocarpa*, *O. sativa*, *S. bicolor* and *V. vinifera*, 42, 54, 59, 64, 66 and 85% of TPS genes, respectively, occur in such twin arrangements (Chen *et al.* 2011; Külheim *et al.* 2015), which might result from local gene duplication by unequal crossover. Consistently, a high degree of homology is often observed among the members of each tandem array. For instance, two linked TPS genes in *Arabidopsis*, namely *AtTPS23* and *AtTPS27*, are identical to each other in terms of both coding region and intron sequences and are therefore thought to witness a very late gene duplication event (Chen *et al.* 2004). On this same vein, it was found that the amino acid similarities among TPS genes within tandem arrays in *E. grandis* were much higher than those observed in the corresponding whole gene sub-family (Külheim *et al.* 2015). In some instances, the tandem arrays of TPS genes are very broad, as occurs in grapevine, in which as many as 45 TPS genes are present in a highly compact gene cluster encompassing a stretch of 690 kb on chromosome 18 (Martin *et al.* 2010). Similarly, 17 TPS genes are stretched over a 317 kb region on the chromosome 6 of *E. grandis* (Külheim al. 2015), whereas in rice a 480 kb stretch on chromosome 4 hosts 14 TPS genes (Chen *et al.* 2011).

Based on phylogenetic analyses, typical plant TPSs can be divided into seven clades or subfamilies, namely a, b, c, d, g, e/f and h (Chen *et al.* 2011; Shalev *et al.* 2018; Kumar *et al.* 2018). Function and taxonomic distribution of the seven plants TPS subfamilies are summarized in Table 2, while the number of different members assigned to each subfamily in sequenced plant genomes is reported in Table 1.1.

The TPS-a, TPS-b and TPS-g gene subfamilies are angiosperm specific, while the TPS-d and TPS-h clades are gymnosperm-, and lycopod- (*S. moellendorffii)* specific, respectively (Tables 1 and 2). On the other hand, the TPS-c and TPS e/f gene subfamilies include mainly class-I and class-II DTPSs from both angiosperms and gymnosperms (Tables 1.1 and 1.2).

Phylogenetic analyses indicated that the angiosperm-specific TPS-a, TPS-b and TPS-g clades have substantially diverged from other TPS clades (Chen *et al.* 2011; Li *et al.* 2012; Shalev *et al.* 2018). Based on functional characterization of their members in model- and non-model plant systems, these three gene clades only include specialized MTPSs, STPSs or DTPSs, i.e., involved in abiotic and biotic plant interactions, rather than those involved in primary metabolism (Chen *et al.* 2011) (Table 1.2).

Table 1.2. *Function and taxonomic distribution of plant terpene synthases (TPSs) subfamilies.*

Subfamily	Groups	Functions	Taxonomy
TPS-a	TPS-a1	STPSs, some DTPSs	Dicots (Embryophyta>Tracheophyta> Spermatophyta>Magnoliophyta>Magnoliopsida)
	TPS-a2	STPSs	Monocots (Embryophyta>Tracheophyta> Spermatophyta>Magnoliophyta>Liliopsida)
TPS-b		MTPSs, IPSs	Angiosperms (Embryophyta>Tracheophyta> Spermatophyta>Magnoliophyta)
TPS-c		Bifunctional class I/II (CPS/KS), Monofunctional class II DTPSs (CPSs) and DTPSs involved in secondary metabolism	Land plants (Embryophyta)
TPS-d	TPS-d1	Primarily MTPSs, STPSs	Gymnosperms (Embryophyta>Tracheophyta> Spermatophyta>Gymnospermae)
	TPS-d2	STPSs	Gymnosperms
	TPS-d3	Primarily DTPSs, STPSs	Gymnosperms
TPS-e/f		Monofunctional class I DTPSs (KS), DTPSs, STPSs and MTPSs involved in secondary metabolism	Vascular plants (Embryophyta>Tracheophyta)
TPS-g		MTPSs, STPSs, DTPSs producing acyclic terpenoids	Angiosperms
TPS-h		Putative bifunctional DTPSs (class I/II)	*Selaginella moellendorffii* (Embryophyta>Tracheophyta>Lycopodiophyta)

Abbreviations: CPS, *ent*-copalyl diphosphate synthase; DTPS, diterpene synthase; IPS, isoprene synthase; KS, *ent*-kaurene synthase; MTPS, monoterpene synthase; STPS, sesquiterpene synthase

Analysis of several flowering plants whose genome has been sequenced indicates that genes belonging to the TPS-a subfamily represent more than half of the total, so constituting the main determinant of the size of TPS family of each species (Chen *et al.* 2011) (Table 1.1). Phylogenetic analyses suggest that the growth of the TPS-a gene subfamily took place after the separation of the monocot and dicot lineages, because the clade is clearly split into two groups, TPS-a1 and TPS-a2, with the first being dicot-specific and the second monocot-specific (Chen *et al.* 2011) (Table 1.2). Moreover, the positions of *Arabidopsis*, *P. trichocarpa*, *Eucaliptus* spp. and *V. vinifera* genes on the branches of the TPS-a1 clade suggest that many of them were due to gene duplication events that took place after the divergence of the four lineages, representing a clear example of species-specific

expansion of TPS genes (Chen *et al.* 2011; Irmisch *et al.* 2014; Külheim *et al.* 2015). Most of the functionally characterized proteins of the TPS-a subfamily are STPSs (Table 1.2), but the DTPS casbene synthases from *Ricinus communis* and *Euforbia esula* (Mau and West 1994; Kirby *et al.* 2010) can be also included in this clade, indicating that different substrate specificities may have evolved inside the TPS-a subfamily (Zerbe *et al.* 2013).

Another angiosperm-specific clade, i.e., TPS-b, contains either MTPSs or IPSs (Table 2). Although most of the TPS-b genes are from dicots, this group also includes two TPSs from sorghum, but none from rice (Table 1.1). Many of the enzymes of the TPS-b group form hemiterpenes and cyclic monoterpenes, which in gymnosperms derive from the activity of members of the gymnosperm-specific TPS-d1 clade (see below). This suggests that many specific MTPS functions might have evolved independently in the angiosperms and in the gymnosperms (Chen *et al.* 2011).

The third angiosperm-specific clade, i.e., TPS-g, is closely related to the TPS-b one, except that its genes encode MTPSs, STPSs and DTPSs that produce mainly acyclic terpenoids (Table 2). A common structural feature of the TPS-g subfamily is that its members lack the N-terminal "$RR(X)_8W$" motif, which is instead highly conserved in the MTPSs of the angiosperms TPS-b clade (mostly cyclases), as well as in the TPS-d1 clade of the gymnosperms (see below). Some members of the TPS-g gene subfamily, among which the mixed-substrate pairs AmNES/LIS-1-AmNES/LIS-2 from *A. majus* (Nagegowda *et al.* 2008), and AtTPS02- AtTPS03 from *Arabidopsis* (Huang *et al.* 2010), already discussed in Section 1.2 above, are seen as examples of evolution by means of neo-functionalization of duplicated TPS genes involving a change in subcellular localization.

The TPS-c clade includes the already mentioned CPS/KS (class-I/II DTPS) from *P. patens*, class-II CPSs involved in primary metabolism (gibberellin biosynthesis) in both angiosperms and gymnosperms, as well as three TPSs from *S. moellendorffii* that hold only the "DXDD" motif but not the "DDXXD" motif, indicating that they are mono-functional CPSs (Chen *et al.* 2011). This subfamily also contains class-II DTPSs involved in the specialized metabolism of angiosperms (Table 1.2). In rice, for instance, in addition to the *OsCPSsyn* gene involved in gibberellin biosynthesis, there are other two genes, namely *OsCPS1/2*, encoding for TPSs involved in antimicrobial defense (Peters 2006).

The TPS-e/f subfamily contains mainly class-I KSs from gymnosperms and angiosperms involved in the primary metabolism (gibberellin biosynthesis) (Table 1.2). Phylogenetic analyses showed that three TPSs from *S. moellendorffii* form a subclade located close to the bifurcation node of the TPS-c and the TPS-e/f clades (Chen *et al.* 2011; Li *et al.* 2012). In the proteins encoded by these *S. moellendorffii* genes, the presence of the "DDXXD" motif and the absence of the "DXDD" motif indicates that they function as class-I TPSs, likely KSs. Therefore, these three *S. moellendorffii*

TPSs are assigned to the TPS-e/f subfamily (Table 1.1). It is worth noting from Table 1 that a significant expansion within the TPS-e/f gene subfamily occurred in eucalypt and rice, compared to other plant species. Indeed, among the eleven analyzed plant species *E. globulus*, *E. grandis* and rice genomes contain eleven, ten and nine TPS-e/f genes, respectively, whereas the other eight species have 1-5 members (Table 1.1). It has been claimed (Xu *et al.* 2007) that the large number of TPS-e/f genes in rice might be aimed at the production of several labdane-type diterpenoids, which are involved in defense against pathogens, thus indicating that, in addition to class-I DTPSs (KSs) involved in gibberellin biosynthesis, TPS-e/f subfamily includes several TPSs involved in specialized metabolism (Table 1.2).

Additional lines of evidence support such evolutionary diversification within the TPS-e/f subfamily. For instance, functionally characterized genes assigned to the TPS-e/f clade include *AtTPS04* and *PtTPS10*, encoding for two DTPSs producing geranyl linalool in *Arabidopsis* (Herde *et al.* 2008) and *P. trichocarpa* (Irmisch *et al.* 2014), respectively, and *CbLIS*, encoding a MTPS producing linalool from *Clarkia breweri* flowers (Dudareva 1996). Moreover, the *Solanum lycopersicum* β-phellandrene synthase (PHS1), although belonging to the TPS-e/f subfamily, employs the uncommon substrate neryl diphosphate, the *cis*-isomer of GPP, to produce mainly β-phellandrene and a few other monoterpenes in the plastids of the glandular trichomes that cover the surfaces of the leaves and stems (Schilmiller *et al.* 2009). More interestingly, its analogous enzyme in *S. habrochaites* (89% identity) utilizes *Z,Z*-FPP to produce the two sesquiterpenes bergamotene and santalene in the plastids of the trichomes (Sallaud *et al.* 2009).

S. moellendorffii TPSs, which do not belong either to the TPS-c or to the TPS-e/f subfamilies, cluster into a new clade designated as TPS-h (Tables 1 and 2). In contrast to the *S. moellendorffii* TPS genes in other clades, seven out of the eight TPS-h proteins of such lycophyte contain both the "DXDD" and the "DDXXD" motifs. Interestingly, such feature has never been found so far in angiosperm TPSs, whereas it does occur in several class-I/II TPS-d DTPSs from gymnosperms, as well as in the class-I/II CPS/KS from *P. patens* (Table 2 and see below). As discussed earlier, gymnosperm class-I/II DTPSs are likely to have evolved from a CPS/KS ancestor probably prior to the divergence from angiosperms, since neither lineage appears to contain a CPS/KS similar to that of *P. patens* (Keeling *et al.* 2010). By the same manner, the putative class-I/II TPSs in the newly established TPS-h subfamily is likely to have evolved from a prototypic TPS similar to the CPS/KS of *P. patens*, and may be implicated in the specialized terpenoid metabolism of *S. moellendorffii*.

1.5. TERPENE SYNTHASE GENES IN GYMNOSPERMS

As mentioned before, conifer trees produce complex mixtures of terpenoids, prevalently in the form of oleoresin, which can play multiple roles in the physical and chemical defense against insects and pathogens (Keeling and Bohlmann 2006; Zulak and Bohlmann 2010). The diversity of conifer terpenoids consists mainly of monoterpenes, sesquiterpenes and diterpenes, which, besides the primary catalysis brought about by a large family of TPSs, arise from the action of other enzymes, among which cytochrome P450 monoxygenases and transferases. This, by introducing additional functionalities to the TPS products, could further increase the structural complexity of conifer terpenoids (Ro *et al.* 2005; Keeling and Bohlmann 2006; Nelson 2011).

In the preceding sections of the present review, it has been highlighted that, while the evolutionary diversification of angiosperms TPSs is thought to have arisen from DTPSs involved in gibberellin biosynthesis, the diversity amongst gymnosperm TPSs appears to have evolved from class-I/II DTPS ancestors sharing structural and functional similarities with the CPS/KS from *P. patens*. According to the different scenario in the evolution of TPSs from angiosperms and gymnosperms, the many MTPSs, STPSs and DTPSs of conifer specialized metabolism form the gymnosperm-specific TPS-d subfamily (Chen *et al.* 2011; Warren *et al.* 2015; Shalev *et al.* 2018). The functional variety of gymnosperm TPSs seems to have evolved via repeated gene duplication events and further sub- and neo-functionalization, contributing to the expansion of the TPS-d multigene family (Chen *et al.* 2011; Warren *et al.* 2015; Shalev *et al.* 2018), which represents the key player behind the chemical complexity of conifer specialized terpenes. In contrast, gymnosperm DTPSs involved in primary metabolism are members of the TPS-c and TPS-e/f subfamilies, which include also angiosperm orthologous genes, suggesting that DTPSs involved in gibberellin biosynthesis are conserved across the two phyla (Keeling *et al.* 2010; Chen *et al.* 2011).

Until about a decade ago, our knowledge concerning the number, structural and functional complexity and phylogeny of gymnosperm TPSs was based on targeted cDNA cloning and characterization in two conifer species only, namely *Abies grandis* and *P. abies*, together with a few TPSs in other gymnosperms (Keeling and Bohlmann 2006; Chen *et al.* 2011). In *A. grandis*, 11 distinct TPS genes were functionally characterized (Bohlmann *et al.* 1999). Martin *et al.* (2004) reported the isolation and characterization of nine different cDNAs coding for TPSs in *P. abies* and analyzed the phylogeny of 29 gymnosperm TPSs, all of which were included into the gymnosperm-specific TPS-d subfamily. More insights into the structural diversity and functional complexity of gymnosperm TPSs, have been gained from the analysis of transcriptomic and genomic resources recently obtained by using next generation sequencing platforms, not only in several members of the

Pinaceae family, such as spruce (*Picea ssp.*) (Keeling *et al.* 2011; Warren *et al.* 2015) and pine (*Pinus* ssp.) (Hall *et al.* 2013a, b), but also in the Cupressaceae, such as *Platycladus orientalis* (Hu *et al.* 2016) and *Thuja plicata* (Shalev *et al.* 2018).

Since most of the gymnosperm TPSs with recognized biochemical function and well-established association with TPS-subfamilies are from the Pinaceae, while the Cupressaceae TPSs have only sporadically been considered, in the present review 20 putative full-length TPS sequences from several Cupressaceae species, such as *Thuja plicata*, *Taiwania cryptomerioides*, *Chaemacyparis formosensis*, *Chaemacyparis obtusa* and *Callitropsis nootkatensis* were used together with a representative set (62 sequences) of functionally characterized TPSs from Pinaceae, Taxaceae, Ginkgoaceae and Cycadaceae (Table 1.3) to construct a maximum likelihood phylogeny (Fig. 1.5).

Table 1.3. *Full-length cDNA sequences retrieved from the NCBI database upon which the phylogenetic analysis of terpene synthases in gymnosperms was carried out (Fig. 1.5). The ent-kaurene synthase from the moss Physcomitrella patens was included as an outgroup.*

Species	Function	Abbreviation	Accession mRNA sequence	ORF (bp)	Accession protein sequence	Amino acid (aa)
Abies balsamea	Bifunctional cis-abienol synthase	Ab_CAS	JN254808	2604	H8ZM73	867
	Bifunctional isopimaradiene synthase	Ab_Iso	JN254806	2559	H8ZM71	852
	Bifunctional like-abietadiene synthase	Ab_LAS	JN254805	2535	H8ZM70	844
Abies grandis	β-phellandrene synthase	Ag_βPHEL	AF139205	1893	AAF61453	630
	Pinene synthase	Ag_(-)α/βPIN	U87909	1887	AAB71085	628
	Bifunctional abietadiene synthase	Ag_AS	U50768	2607	Q38710	868
	α-bisabolene synthase	Ag_BIS	AF006195	2454	O81086	817
	γ-humulene synthase	Ag_HUM	U92267	1782	O64405	593
	δ-selinene synthase	Ag_SEL	U92266	1746	O64404	581
Callitropsis nootkatensis	Terpene synthase	Cn_STPS	JX040471	1770	AFN21429	589
Chamaecyparis formosensis	β-cadinene synthase	Cf_STPS	JN715077	1812	AFJ23663	603
	α-pinene synthase	Cf_αPIN	EU099434	1887	ABW80964	628
Chamaecyparis obtusa	Limonene/borneol synthase	Co_LIM	AB120957	1818	BAC92722	605
Cycas taitungensis	Sesquiterpene synthase	Ct_STPS	AB154833	2541	BAF43701	846
Ginkgo biloba	(E,E) farnesol synthase	Gb_FAR	KM248383	2415	AIU94289	804
	Levopimaradiene synthase	Gb_AS	AF331704	2622	Q947C4	873
Picea abies	E-α-bisabolene synthase	Pa_BIS	AY473619	2424	AAS47689	807
	Longifolene synthase	Pa_LON	AY473625	1738	AAS47695	578
	2-methyl-3-buten-2-ol synthase	Pa_MBOS1	JN039264	1881	AFJ73582	626
	(-)-linalool synthase	Pa_LIN	AY473623	1872	AAS47693	623
	(+)-3-carene synthase	Pa_3CAR	AF461460	1884	AAO73863	627
	Myrcene synthase	Pa_MYR	AY473626	1902	AAS47696	633
	(-)-limonene synthase	Pa_LIM	AAS47694	1905	AAS47694	634
	(E,E)-α-farnesene synthase	Pa_FAR	AY473627	1743	AAS47697	580
	Levopimaradiene synthase	Pa_LAS	AY473621	2580	Q675L4	859
	Isopimaradiene synthase	Pa_ISO	AY473620	2604	Q675L5	867
Picea glauca	ent-copalyl diphosphate synthase	Pg_CPS	GU045755	2286	ADB55707	761
	(-)-ent-kaurene synthase	Pg_KS	GU045756	2274	ADB55708	757
	α-humulene synthase	Pg_HUM	HQ426155	1728	ADZ45513	575
	1,8-cineole synthase	Pg_1,8CIN	HQ426160	1839	ADZ45498	612
	3-carene synthase	Pg_3CAR	FJ609174	1884	ACM04452	627
	(-)-α/β-pinene synthase	Pg_(-)α/βPIN	HQ426153	1884	ADZ45507	627
Picea pungens	2-methyl-3-buten-2-ol synthase	Pp_MBOS1	JN039265	1881	AFJ73583	626
Picea sitchensis	ent-copalyl diphosphate synthase	Psi_CPS	GU045757	2286	ADB55709	761
	(-)-ent-kaurene synthase	Psi_KS	GU045758	2274	ADB55710	757
	Levopimaradiene/abietadiene synthase	Psi_LAS	HQ426170	2580	ADZ45517	859
	Isopimaradiene synthase	Psi_ISO	HQ426150	2625	ADZ45512	874
	α-longipinene synthase	Psi_LON	HQ426161	1740	ADZ45516	579
	1,8-cineole synthase	Psi_1,8CIN	HQ426165	1839	ADZ45499	612
	(-)-linalool synthase	Psi_LIN	HQ426164	1884	ADZ45501	627
	(+)-3-carene synthase	Psi_3CAR	HQ426167	1884	ADZ45511	627
	(+)-sabinene synthase	Psi_SAB	HQ336803	1884	ADU85929	627
	(-)-β-phellandrene synthase	Psi_βPHEL	HQ426159	1875	ADZ45503	624
	(-)-α/β-pinene synthase	Psi_(-)α/βPIN	HQ426166	1884	ADZ45509	627
Pinus banksiana	α terpineol synthase	Pb_MTPS5	JQ240308	1881	AFU73860	626
	(-)-β-pinene synthase	Pb_MTPS2	JQ240291	1887	AFU73843	628
	(+)-α pinene synthase	Pb_MTPS8	JQ240298	1887	AFU73850	628
Pinus contorta	Levopimaradiene/abietadiene synthase	Pc_DTPS LAS1	JQ240310	2574	AFU73862	857
	Diterpene synthase	Pc_MDTPS1	JQ240318	2559	AFU73870	852
	Isopimaradiene synthase	Pc_DTPS mISO1	JQ240314	2631	AFU73866	876
	Pimaradiene synthase	Pc_DTPS mPIM1	JQ240316	2607	AFU73868	868
	α-terpineol /1,8-cineole synthase	Pc_MTPS6	JQ240309	1851	AFU73861	616
	(+)-3-carene synthase	Pc_MTPS4	JQ240307	1881	AFU73859	626
	(-)-α pinene synthase	Pc_MTPS1	JQ240303	1890	AFU73855	629
	(-)-β-phellandrene synthase	Pc_MTPS8	JQ240301	1866	AFU73853	621
Pinus sabiniana	2-methyl-3-buten-2-ol synthase	Psab_MBOS1	JF719039	1845	AEB53064	614
Pinus sylvestris	e-β farnesene synthase	Ps_STPS4	GU248335	2436	ADH29869	811
Pinus taeda	Levopimaradiene synthase	Pt_DTPS LAS1	AY779541	2553	Q50EK2	850
	α-terpineol synthase	Pt_MTPS2	AF543529	1884	AAO61227	627
	α-farnesene synthase	Pt_STPS1	AF543528	1725	AAO61226	574
Pseudotsuga menziesii	(E)-β-farnesene synthase	Pme_FAR	HQ214483	2478	ADX42737	825
	Terpinolene synthase	Pme_TER	AY906866	1878	AAX07264	625
	(E)-γ-bisabolene synthase	Pme_BIS	AY906868	2448	Q4QSN4	815
Taiwania cryptomerioides	Diterpene synthase	Tcr_TPS1	KT588489	2151	AOG18235	716
	Diterpene synthase	Tcr_TPS2	KT588484	2499	AOG18230	832
	Pimara-8(14),15-diene synthase	Tcr_TPS4	GU575291	2556	ADL14246	851
Taxus brevifolia	Taxadiene synthase	Tb_TXS	U48796	2589	Q41594	862
Taxus cuspidata	Taxadiene synthase	Tc_TXS	DQ305407	2589	ABC25488	862
Taxus x media	Taxadiene synthase	Tm_TXS	AY461450	2589	AAS18603	862
Thuja plicata		Tp_TPS1		2502		833
		Tp_TPS25		2598		865
		Tp_TPS5		2598		865
		Tp_TPS3		2598		865
		Tp_TPS10		2424		807
		Tp_TPS7		1773		590
		Tp_TPS19		1743		580
		Tp_TPS30		1743		580
		Tp_TPS31		1752		583
		Tp_TPS27		1815		604
		Tp_TPS29		1827		608
		Tp_TPS13		1905		634
		Tp_TPS21		1833		610
Physcomitrella patens	ent-kaurene synthase	Pt_TPS-entKS	AB302933	2646	BAF61135	881

*Sequences retrived from Shalev *et al.* 2018

To such aim, the multiple alignment of protein sequences was performed by MUltiple Sequence Comparison by Log-Expectation (*MUSCLE*) algorithm and the phylogenetic tree was constructed by the Maximum Likelihood method using MEGAX software (Kumar *et al.* 2018).

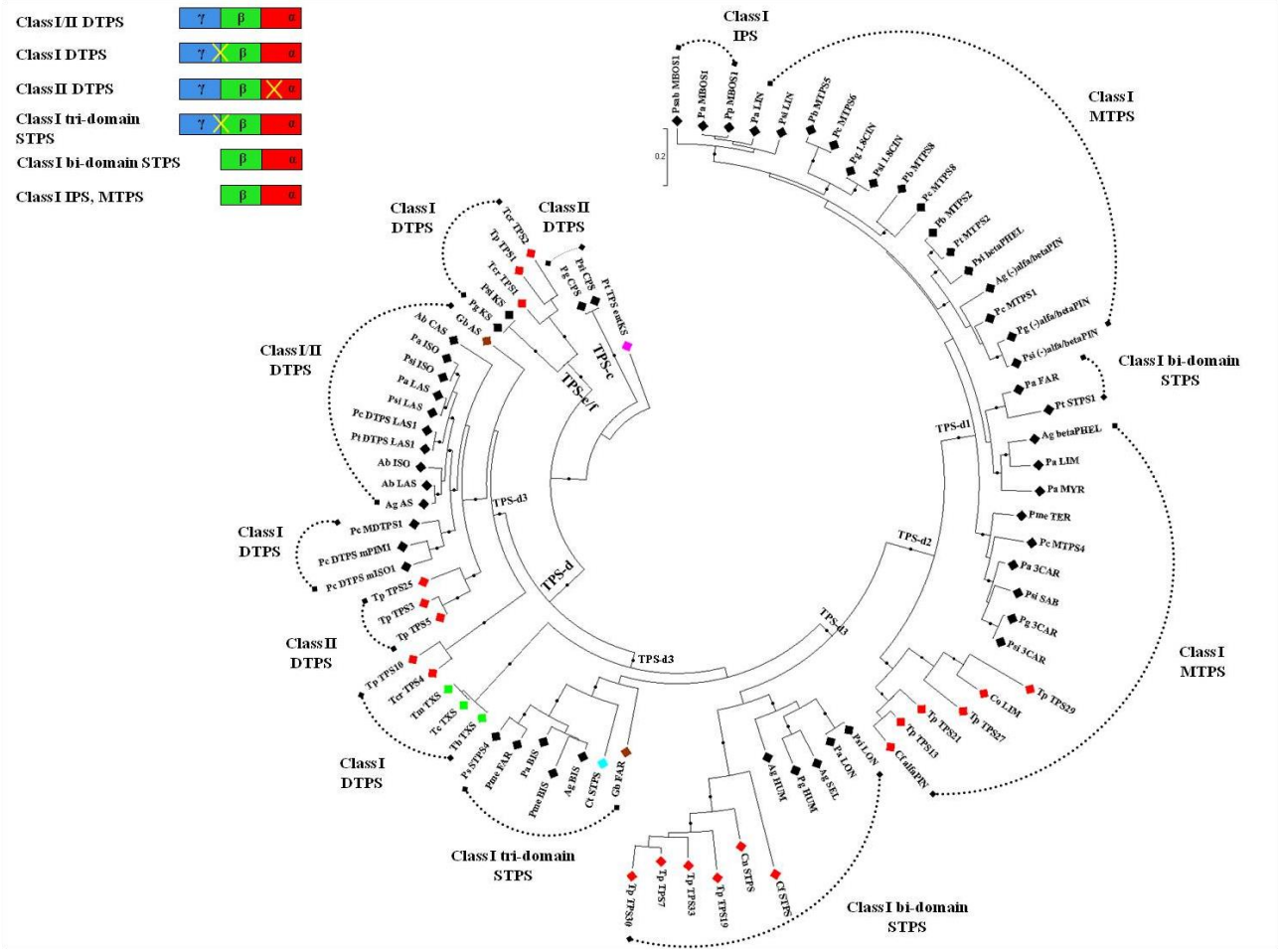


Fig. 1.5. *Phylogenetic tree of terpene synthases (TPSs) in gymnosperms: Cupressaceae (red diamonds), Pinaceae (black), Taxaceae (green), Ginkgoaceae (brown) and Cycadaceae (heavenly). The Physcomitrella patens ent-kaurene synthase (PtTPS-entKS; violet diamond) was used to root the tree. Branches indicated with dots represent bootstrap support more than 80% (100 repetitions). Modifications in the typical γβα-domain architecture of TPS and the presence of functional active sites (a yellow cross indicate loss of function) are illustrated corresponding to the different subfamilies of the TPS plant family and to the different group within the TPS-d3 subfamily. Abbreviations: DTPSs, diterpene synthases; MTPSs, monoterpene synthases; STPSs, sesquiterpene synthases.*

According to previous phylogenetic analyses (Chen *et al.* 2011; Shalev *et al.* 2018), the 82 gymnosperm TPSs considered here were divided into three major clades (Fig. 1.5), which correspond to the three TPS-c, TPS-e/f and TPS-d subfamilies of the plant TPS family.

The TPS-c subfamily included the two functionally characterized class-II CPSs from *P. glauca* (Pg CPS) and *P. sitchensis* (Psi CPS) (Fig. 1.5), whereas the TPS-e/f one, in addition to the two functionally characterized class-I KSs from the same *Picea* species (Pg KS and Psi KS, respectively),

also included putative KSs involved in primary metabolism from the two Cupressaceae species *T. plicata* (Tp TPS1) and *T. cryptomerioides* (Tcr TPS1, TcrTPS2) (Fig. 1.5).

The vast majority of the TPSs considered here fell into the TPS-d subfamily (Fig. 1.5), which, as already said, are reputed to be gymnosperm-specific and contains all the TPSs involved in the specialized metabolism. On the whole, Fig. 1.5 shows that proteins in the TPS-d subfamily appear to cluster together primarily on the basis of their general function, as indicated by the formation of distinct clades containing bifunctional and monofunctional DTPSs involved in secondary metabolism, three-domain (γβα) and two-domain (βα) STPSs, MTPSs and IPSs. Within each of such functional clades, TPSs of Cupressaceae mainly clustered apart from TPSs of Pinaceae, underlining the independent diversification and evolution of specific TPS functions in the two gymnosperm families (Fig. 1.5).

As previously discussed, evolution in plants seems to have maintained class-I/II DTPSs only in cryptogamae and gymnosperms. In the latter phylum, class-I/II DTPSs constitute the TPS-d3 group (Keeling *et al.* 2011; Hall *et al.* 2013a), which also includes class-I DTPSs from *Pinus* and *Taxus* species, as well as the ancestral γβα-domain STPSs (Fig. 1.5). Class I/II DTPSs include the levopimaradiene/abietadiene and isopimaradiene synthases (LASs and ISOs, respectively) of diterpene resin acid (DRA) metabolism isolated from several Pinaceae (Peters *et al.* 2000; Martin *et al.* 2004; Hall *et al.* 2013a), as well as the class-I/II *cis*-abienol synthase from balsam fir (Ab CAS; Zerbe *et al.* 2012). Interestingly, Ab CAS is phylogenetically equidistant from the LAS from *Ginkgo biloba* (Gb LAS) and ISO and LAS enzymes from other Pinaceae, indicating that this enzymatic activity evolved before the speciation of fir, pine and spruce, and was lost in other Pinaceae (Hall *et al.* 2011; Zerbe *et al.* 2012; Fig. 1.5). Class-I DTPSs from *P. contorta* (namely Pc MDTPS1, Pc DTPS mPIM1 and Pc DTPS mISO1) form a separate branch within the TPS-d3 group, located close to the class-I/II DTPSs of DRA biosynthesis, but distant from other class-I DTPSs, such as taxadiene synthases from *Taxus* spp (Fig. 1.5). The above three TPSs from *P. contorta* are closely related to three *T. plicata* sequences (Tp STS3, Tp STS5 and TPS25) (Fig. 1.5), which, however, contain the class-II "DXDD" motif but lack the functional class-I "DDXXD" motif. Interestingly, class-II DTPSs of specialized metabolism have not been previously reported in gymnosperms and appear to be exclusive of the Cupressaceae. Since the newly discovered class-II DTPSs identified in the transcriptome of *T. plicata* (Shalev *et al.* 2018) could add themselves to the array of catalytic diversity exhibited by conifer DTPSs involved in specialized metabolism, their functional characterization will be important to elucidate the role of this new class of TPSs in the secondary metabolism of the Cupressaceae. The other two proteins from Cupressaceae, namely Tp TPS10 from *T. plicata* and Tcr TPS4 from *T. cryptomerioides*, both of which clustered in the clade of the putative DTPSs (Fig. 1.5),

are class-I enzymes that lack the functional class-II "DXDD" motif, whereas class-I/II DTPSs have not been found so far in the Cupressaceae. Finally, the clade of ancestral gymnosperm (γβα-domain) STPSs, which is also included in the TPS-d3 group, contains members from the Pinaceae, as well as from *G. biloba* and *Cycas taitungensis*, but none from the Cupressaceae (Fig. 1.5).

In general, the phylogenetic analysis presented here suggests that events of mono-functionalization, i.e., a form of sub-functionalization from a duplicated bifunctional ancestor, took place independently many times during the evolution of gymnosperm DTPSs. Such events led to the appearance of class-II and class-I DTPSs of gibberellin metabolism, the *Taxus* spp. taxadiene synthases, the newly identified class-II DTPSs in *T. plicata*, and the class-I DTPSs found in the Cupressaceae and *Pinus* species. The pine class-I DTPSs appear to have evolved from loss of the class-II active site, which instead remained unchanged in the similar class-I/II LAS and ISO enzymes, while it remains to be determined whether the class-I/II DTPSs have been lost in the Cupressaceae. Apart from DTPSs, mono-functionalization ultimately also led to the ancestral tri-domain (γβα) STPSs, and to the large family of bi-domain gymnosperm MTPSs and STPSs constituting the TPS-d1 and TPS-d2 groups (Fig. 1.5). Within these two groups, it is evident that MTPSs and STPSs of the Cupressaceae clustered apart from those of the Pinaceae (Fig. 1.5), indicating that, as stated before, the diversification and the evolution of these two TPS functions occurred independently in the two gymnosperm families. While the TPS-d2 group included only the bidomain (βα) STPSs, TPS-d1 contained all the known gymnosperm MTPSs, in addition to Pinaceae TPSs that use DMAPP as a substrate to produce hemiterpenes, and two bidomain STPSs isolated from *P. taeda* and *P. abies* which produce the acyclic sesquiterpene *E,E*-α-farnesene (Phillips *et al.* 2003; Martin *et al.* 2004; Fig. 1.5).

1.6. IDENTIFICATION AND PHYLOGENY OF TPS GENE SEQUENCES IN *Pinus* SPECIES

Despite of its economic and ecological relevance, as well as of the importance of oleoresin terpenes in defense and as bioproducts, a comprehensive analysis of the functional diversity and evolution of TPSs in the *Pinus* genus still awaits to be carried out, to the best of our knowledge. Therefore, an extensive *in silico* search was performed here, to identify all the putative full-length TPSs for primary and specialized metabolisms in different *Pinus* species, and to describe their general characteristics, functional properties and phylogenetic relationships.

The identification of full-length cDNAs coding for putative pine TPSs was based on the BLAST search of the NCBI database, by using selected and functionally characterized TPSs from different conifer species (Table 1.4).

Table 1.4. *Full-length cDNA sequences of functionally characterized terpene synthases (TPSs) employed for the BLAST search in the NCBI database of the putative TPSs of Pinus spp.*

Species	Function	Type of TPS	Accession number	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-farnesene 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-farnesene 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)*

Abbreviations: CPS, ent-copalyl diphosphate synthase; DTPS, diterpene synthase; KS, *ent*-kaurene synthase; MTPS, monoterpene synthase; STPS, sesquiterpene synthase

Database search identified a total of 93 full-length cDNA sequences coding for putative TPSs from 28 different *Pinus* species (Table 1.5).

Table 1.5. *Full-length cDNA sequences of putative terpene synthases retrieved from the NCBI database upon which the phylogenetic analysis of terpene synthases in Pinus spp. was carried out (Fig. 1.7). The ent-kaurene synthase from the moss Physcomitrella patens was included as an outgroup (continue on the next page).*

Species	Function	Abbreviation	Accession mRNA sequence	ORF (bp)	Accession protein sequence	Amino acid (aa)
Pinus arizonica var. cooperi	2-methyl-3-buten-2-ol synthase	Par_MTPS1	JN039226	1845	AFJ73545	614
	2-methyl-3-buten-2-ol synthase	Par_MTPS2	JN039225	1845	AFJ73544	614
	2-methyl-3-buten-2-ol synthase	Par_MTPS3	JN039224	1845	AFJ73543	614
Pinus attenuata	2-methyl-3-buten-2-ol synthase	Pa_MTPS1	JN039215	1845	AFJ73535	614
Pinus banksiana	(-)-α pinene synthase	Pb_MTPS1	JQ240304	1890	AFU73856	629
	(-)-β-pinene synthase	Pb_MTPS2	JQ240291	1887	AFU73843	628
	(-)-β-pinene synthase (TPS-(-)Bpin2)	Pb_MTPS3	JQ240292	1884	AFU73844	627
	(-)-α/β-pinene synthase	Pb_MTPS4	JQ240290	1872	AFU73842	623
	α terpineol synthase	Pb_MTPS5	JQ240308	1881	AFU73860	626
	(+)-3-carene synthase	Pb_MTPS6	JQ240306	1881	AFU73858	626
	(+)-3-carene synthase	Pb_MTPS7	JQ240305	1881	AFU73857	626
	(+)-α pinene synthase	Pb_MTPS8	JQ240298	1887	AFU73850	628
	Monoterpene synthase	Pb_MTPS9	JQ240296	1887	AFU73848	628
	Monoterpene synthase	Pb_MTPS10	JQ240297	1887	AFU73849	628
	(-)-β-phellandrene synthase	Pb_MTPS11	JQ240302	1866	AFU73854	621
	Levopimaradiene/abietadiene synthase	Pb_DTPS_LAS1	JQ240312	2574	AFU73864	857
	Monofunctional diterpene synthase	Pb_MDTPS1	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	(-)-α pinene synthase	Pc_MTPS1	JQ240303	1890	AFU73855	629
	(-)-β-pinene synthase	Pc_MTPS2	JQ240293	1884	AFU73845	627
	Monoterpene synthase	Pc_MTPS3	JQ240294	1884	AFU73846	627
	(+)-3-carene synthase	Pc_MTPS4	JQ240307	1881	AFU73859	626
	(+)-α pinene synthase	Pc_MTPS5	JQ240295	1887	AFU73847	628
	α terpineol /1,8-cineole synthase	Pc_MTPS6	JQ240309	1851	AFU73861	616
	(-)-camphene / (+)-α-pinene synthase	Pc_MTPS7	JQ240299	1860	AFU73851	619
	(-)-β-phellandrene synthase	Pc_MTPS8	JQ240301	1866	AFU73853	621
	(-)-β-phellandrene synthase	Pc_MTPS9	JQ240300	1875	AFU73852	624
	Levopimaradiene/abietadiene synthase	Pc_DTPS_LAS1	JQ240310	2574	AFU73862	857
	Levopimaradiene/abietadiene synthase	Pc_DTPS_LAS2	JQ240311	2553	AFU73863	850
	Monofunctional diterpene synthase	Pc_MDTPS1	JQ240318	2559	AFU73870	852
	Monofunctional diterpene synthase	Pc_MDTPS2	JQ240319	2559	AFU73871	852
	Monofunctional diterpene synthase	Pc_MDTPS3	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 contorta var. murrayana	2-methyl-3-buten-2-ol synthase	Pmur_MTPS1	JN039217	1845	AFJ73537	614
	2-methyl-3-buten-2-ol synthase	Pmur_MTPS2	JN039216	1845	AFJ73536	614
	2-methyl-3-buten-2-ol synthase	Pmur_MTPS3	JN039221	1845	AFJ73541	614
	2-methyl-3-buten-2-ol synthase	Pmur_MTPS4	JN039218	1845	AFJ73538	614
	2-methyl-3-buten-2-ol synthase	Pmur_MTPS5	JN039219	1845	AFJ73539	614
	2-methyl-3-buten-2-ol synthase	Pmur_MTPS6	JN039220	1845	AFJ73540	614
Pinus coulteri	2-methyl-3-buten-2-ol synthase	Pcou_MTPS1	JN039227	1845	AFJ73546	614
	2-methyl-3-buten-2-ol synthase	Pcou_MTPS2	JN039229	1845	AFJ73548	614
	2-methyl-3-buten-2-ol synthase	Pcou_MTPS3	JN039228	1845	AFJ73547	614
Pinus densiflora	Abietadiene synthase	Pd_DTPS_ABS1	EU439295	2577	ACC54559	858
Pinus greggii	2-methyl-3-buten-2-ol synthase	Pg_MTPS1	JN039230	1845	AFJ73549	614
Pinus hartwegii	2-methyl-3-buten-2-ol synthase	Ph_MTPS1	JN039232	1845	AFJ73551	614
	2-methyl-3-buten-2-ol synthase	Ph_MTPS2	JN039231	1845	AFJ73550	614
Pinus jeffreyi	2-methyl-3-buten-2-ol synthase	Pj_MTPS1	JN039233	1845	AFJ73552	614
Pinus kesiya var. langbianensis		Pk_MTPS1	KX394684	1956	AQZ36562	651
	α-pinene synthase	Pk_MTPS2	KM382173	1875	AIY22674	624

Pinus massoniana	(-)-α pinene synthase	Pm_MTPS1	KF547035	1890	AGW25369	629
	α-terpineol synthase	Pm_MTPS2	KJ803197	1863	AIL88641	620
Pinus montezumae	2-methyl-3-buten-2-ol synthase	Pmon_MTPS1	JN039234	1845	AFJ73553	614
Pinus muricata	2-methyl-3-buten-2-ol synthase	Pmuri_MTPS1	JN039235	1845	AFJ73554	614
	2-methyl-3-buten-2-ol synthase	Pmuri_MTPS2	JN039236	1845	AFJ73555	614
Pinus patula	2-methyl-3-buten-2-ol synthase	Ppat_MTPS1	JN039245	1845	AFJ73563	614
	2-methyl-3-buten-2-ol synthase	Ppat_MTPS2	JN039243	1845	AFJ73562	614
Pinus pinaster	α-pinene synthase	Pp_MTPS1	KP780394	1890	ALB78130	629
	α-pinene synthase	Pp_MTPS2	KP780395	1890	ALB78131	629
Pinus pinea	α-pinene synthase	Ppinea_MTPS1	KR011842	1890	ALD18902	629
	α-pinene synthase	Ppinea_MTPS2	KR011841	1890	ALD18901	629
Pinus ponderosa var. scopulorum	2-methyl-3-buten-2-ol synthase	Ppon_MTPS1	JN039246	1845	AFJ73564	614
	2-methyl-3-buten-2-ol synthase	Ppon_MTPS2	JN039248	1845	AFJ73566	614
Pinus pseudostrobus	2-methyl-3-buten-2-ol synthase	Pps_MTPS1	JN039254	1845	AFJ73572	614
Pinus pseudostrobus var. apulcensis	2-methyl-3-buten-2-ol synthase	Papu_MTPS1	JN039240	1845	AFJ73559	614
	2-methyl-3-buten-2-ol synthase	Papu_MTPS2	JN039242	1845	AFJ73561	614
	2-methyl-3-buten-2-ol synthase	Papu_MTPS3	JN039241	1845	AFJ73560	614
	2-methyl-3-buten-2-ol synthase	Papu_MTPS4	JN039239	1845	AFJ73558	614
Pinus pseudostrobus var. estevezii	2-methyl-3-buten-2-ol synthase	Pest_MTPS1	JN039251	1845	AFJ73569	614
	2-methyl-3-buten-2-ol synthase	Pest_MTPS2	JN039252	1845	AFJ73570	614
	2-methyl-3-buten-2-ol synthase	Pest_MTPS3	JN039250	1845	AFJ73568	614
	2-methyl-3-buten-2-ol synthase	Pest_MTPS4	JN039249	1845	AFJ73567	614
Pinus radiata	2-methyl-3-buten-2-ol synthase	Prad_MTPS1	JN039257	1845	AFJ73575	614
	2-methyl-3-buten-2-ol synthase	Prad_MTPS2	JN039256	1845	AFJ73574	614
Pinus sabiniana	2-methyl-3-buten-2-ol synthase	Psab_MTPS1	JF719039	1845	AEB53064	614
Pinus sylvestris	Longifolene synthase	PS_STPS1	EF679332	1743	ABV44454	580
	1(10),5-germacradien-4-ol synthase	PS_STPS2	EF679331	1878	ABV44453	625
	Caryophyllene/humulene synthase	PS_STPS3	EF679330	1728	ABV44452	575
	e-β farnesene synthase	PS_STPS4	GU248335	2436	ADH29869	811
Pinus tabuliformis	ent-copalyl diphosphate-	Pta_CPS1	KJ158966	2391	AHW42450	796
	ent-kaurene synthases	Pta_KS1	KJ158985	2232	AHW42469	743
Pinus taeda	(-)-α-pinene synthase	Pt_MTPS1	AF543527	1890	AAO61225	629
	α-terpineol synthase	Pt_MTPS2	AF543529	1884	AAO61227	627
	(+)-α-pinene synthase	Pt_MTPS3	AF543530	1887	AAO61228	628
	α-farnesene synthase	Pt_STPS1	AF543528	1725	AAO61226	574
	Diterpene synthase	Pt_DTPS_LAS1	AY779541	2553	AAX07435	850
Pinus teocote	2-methyl-3-buten-2-ol synthase	Pteo_MTPS1	JN039258	1845	AFJ73576	614
	2-methyl-3-buten-2-ol synthase	Pteo_MTPS2	JN039260	1845	AFJ73578	614
	2-methyl-3-buten-2-ol synthase	Pteo_MTPS3	JN039259	1845	AFJ73577	614
Pinus torreyana	2-methyl-3-buten-2-ol synthase	Ptor_MTPS1	JN039263	1845	AFJ73581	614
	2-methyl-3-buten-2-ol synthase	Ptor_MTPS2	JN039262	1845	AFJ73580	614
	2-methyl-3-buten-2-ol synthase	Ptor_MTPS3	JN039261	1845	AFJ73579	614
Physcomitrella patens	ent-kaurene synthase	Pt_TPS-entKS	AB302933	2646	BAF61135	881

BLAST searches using as queries the CPS and KS from *P. glauca* (Table 1.5), assumed to represent DTPSs involved in primary metabolism (gibberellin biosynthesis, see above), detected orthologous full-length cDNA sequences only in *P. tabuliformis* (Pta CS1 and Pta KS1 in Table 1.5). It is worth noting that gymnosperm CPS and KS gene sequences have been previously isolated and characterized only in *P. glauca* and *P. sitchensis* (Keeling *et al.* 2010).

Five STPSs were identified for *Pinus* species in the NCBI database, of which four from *P. sylvestris* (Ps STPS1-4 in Table 1.5) and one from *P. taeda* (Pt STPS1 in Table 1.5). As described before, this latter STPS produces the acyclic sesquiterpene *(E,E)*-α-farnesene (Phillips *et al.* 2003)

and clusters into the TPS-d1 group (Fig. 1.5). Heterologous expression in *Escherichia coli* allowed to find out that the first three STPSs from *P. sylvestris* produce longifolene and α-longipinene (Ps STPS1), 1(10), 5-germacradiene-4-ol (Ps STPS2), and *(E)*-β-caryophyllene and α-humulene (Ps STPS3), as their main products (Köpke *et al.* 2008). When compared to each other, the deduced amino acid sequences of the three above *P. sylvestris* enzymes showed identities ranging from 60 to 65%, whereas a greatest amino acid identity (70–80%) was noticed with other conifer STPS, such as longifolene synthase from [*P. engelmannii* x *P. glauca*] and a α-humulene synthase from *P. glauca* (Keeling *et al.* 2011). This indicates the presence of putative orthologous genes coding for these STPSs in different Pinaceae species. The fourth STPS from *P. sylvestris*, namely Ps STPS4, which was reported to produce *(E)*-β-farnesene as unique terpenoid (Köpke *et al.* 2010), showed only a low amino acid identity (35-39%) with the other three STPS from the same species. In contrast, Ps STPS4 showed a 78–80% amino acid identity with several conifer tri-domain (γβα) STPSs, such as *(E)*-β-farnesene synthase from *Pseodotsuga menziesii* (Huber *et al.* 2005) and *(E)*-α-bisabolene synthases from *A. grandis* and *P. glauca* (Trapp and Croteau 2001; Martin *et al.* 2004).

BLAST searches using as queries the selected DTPSs from *P. contorta* and *P. abies* (Table S2) allowed to identify 13 DTPSs in *Pinus* species, of which seven and four in *P. contorta* and *P. banksiana,* respectively, and one each in *P. taeda* and *P. densiflora* (Table 1.5). Of these *Pinus* spp. DTPSs, five, namely Pc DTPS LAS1, Pc DTPS LAS2, Pb DTPS LAS1, Pt DTPS LAS1 and Pd DTPS ABS1, showed a high level (95-99%) of amino acid identity among each other, and were found to contain both the class-I and the class-II functional motifs, indicating that they are class-I/II DTPSs, similar to the already recognized conifer ISO and LAS enzymes of DRA biosynthesis (Peters *et al.* 2000; Martin *et al.* 2004; Keeling *et al.* 2011; Zerbe *et al.* 2012, and see above). Four of the five pine class-I/II DTPSs, namely Pt DTPS LAS1, Pb DTPS LAS1, Pc DTPS LAS1 and Pc DTPS LAS2, were functionally characterized, by expressing them as recombinant proteins in *E. coli* (Ro and Bohlmann 2006; Hall *et al.* 2013a). By using liquid chromatography-mass spectrometry, and supplying GGPP as the substrate, the major diterpene products of the three *P. banksiana* and *P. contorta* LAS enzymes were found to be stereoisomers of 13-hydroxy-8 (14)-abietene (Hall *et al.* 2013a). The replication of the same analysis by means of gas chromatography-mass spectrometry (GC-MS), which causes dehydration of 13-hydroxy-8(14)-abietene, led to the identification of abietadiene, levopimaradiene, and neoabietadiene, as the three major enzymatic products, consistent with the GC-MS results previously obtained for Pt DTPS LAS from *P. taeda* (Ro and Bohlmann 2006). The production of the unstable diterpene tertiary alcohol 13-hydroxy-8(14)-abietene by the three *P. banksiana* and *P. contorta* LAS enzymes, suggests that water capture by a carbocation at the

class-I active site is a conserved and typical reaction brought about by the LAS enzymes in many conifer species (Keeling *et al.* 2011; Zerbe *et al.* 2012).

The remaining eight putative DTPSs of *Pinus* spp. were found to be class-I enzymes, showing an intermediate amino acid identity (66-73%), both with the above five putative class-I/II DTPSs from *Pinus* species, and with the functionally characterized ISO and LAS enzymes of the DRA specialized metabolism in several conifers. Despite being putative monofunctional DTPSs, the above eight proteins showed only 33% to 34% sequence identity with the *P. glauca ent*-copalyl diphosphate synthase (Pg CPS) and *ent*-kaurene synthase (Pg KS) of gibberellin metabolism (Keeling *et al.* 2010), indicating their involvement in the specialized metabolism, rather than in the primary one. Showing 99% amino acid sequence identity to each other, Pc DTPS mISO1 and Pb DTPS mISO1, as well as Pc DTPS mPIM1 and Pb DTPS mPIM1, are probably two pairs of orthologous genes from *P. contorta* and *P. banksiana*, respectively. The other class-I DTPSs (namely Pc MDTPS1, Pc MDTPS2, PcmdiTPS3, and Pb MDTPS1), though very similar among each other's (97% to 98% protein sequence identity), exhibited a low identity (71-75%) with the other identified pine DTPSs. Functional characterization of four of the eight class-I DTPSs, identified the putative orthologous pair Pc DTPS mPIM1 and Pb DTPS mPIM1 as single-product pimaradiene synthases, whereas the orthologous pair of Pc DTPS mISO1 and Pb DTPS mISO1 were found to produce isopimaradiene as main product, whit small amounts of sandaracopimaradiene (Hall *et al.* 2013a). Interestingly, class-I DTPSs of specialized DRA metabolism have not been previously identified in Pinaceae. The only other known examples of class-I DTPSs of specialized metabolism in gymnosperms are the two TPSs identified in the Cupressaceae species *T. plicata* (Tp TPS10) and *T. cryptomerioides* (Tcr TPS4), whose functions remain to be determined, and taxadiene synthases in *Taxus* ssp. (Wildung and Croteau 1996), which specifically transforms GGPP into the macrocyclic taxadiene backbone without the need for a bicyclic diphosphate intermediate. The aforementioned class-I DTPS characterized by Hall *et al.* (2013a) are the first examples in gymnosperms of enzymes able to synthetize mainly pimaradiene; as such, they add themselves to the already known ISO and LAS conifers DTPS participating in the DRA specialized metabolism.

BLAST searches using as queries the seven selected MTPSs from *P. contorta, P. abies* and *P. glauca* (Table 1.4) detected 74 putative full-length cDNAs coding for MTPSs from 26 different *Pinus* species (Table 1.5). However, only 32 of them could be classified as true MTPSs. The deduced amino acid sequences of the remaining 42 cDNA sequences, from 18 different *Pinus* species, were predicted to synthetize 2-methyl-3-buten-2-ol (MBO), a C5 alcohol produced and emitted by several pine species (Lerdau and Gray 2003). MBO is related to isoprene by a structural and biosynthetic point of view and both derive from DMAPP (Gray *et al.* 2011). The gene for MBO synthase (MBOS) was

first isolated from *Pinus sabiniana* (towani pine, digger pine) and found to encode for a bifunctional enzyme able to produce both MBO and isoprene in a 90:1 ratio (Gray *et al.* 2011). Conifer MBOSs appear to have evolved independently from their homologous proteins in angiosperms, i.e., IPSs (Sharkey *et al.* 2013). Indeed, phylogenetic analysis showed that MBOSs fall into the TPS-d1 group, together with the gymnosperm MTPSs, and are most closely related to linalool synthases from *P. abies* (Martin *et al.* 2004) and *P. sitchensis* (Keeling *et al.* 2011) (Fig. 1.5). The 42 full-length MBOS sequences identified here showed a high level of homology among each other (93-99% amino acid sequence identity) as shown in the phylogenetic tree reported in Fig. 1.6.

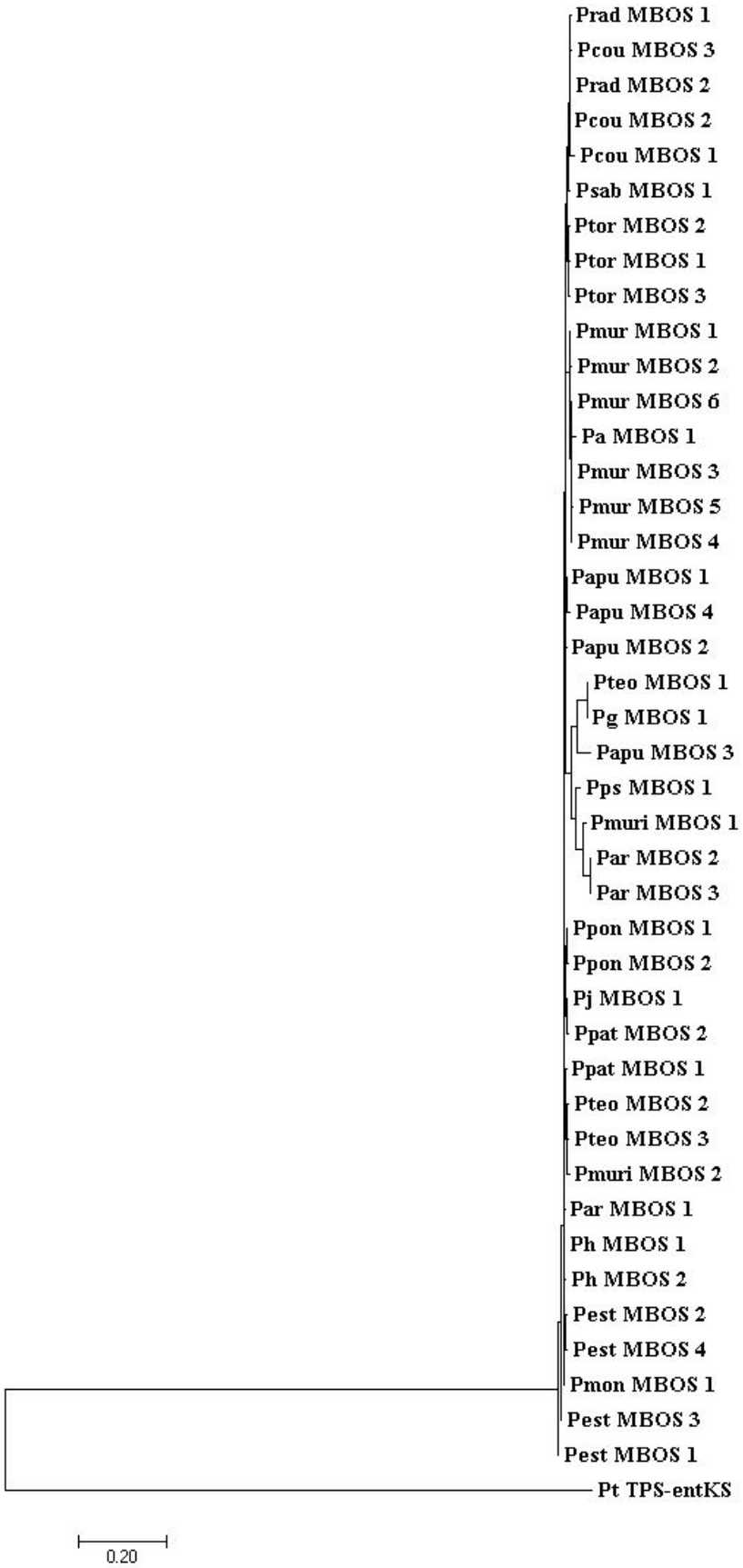


Fig. 1.6. *Phylogenetic tree for the deduced amino acid sequences of the 2-methyl-3-buten-2-ol synthases (MBOSs) from Pinus species identified in NCBI database (Table 1.5). Physcomitrella patens ent-kaurene synthase (Pt TPS-entKS) was used to root the tree.*

To gain a deeper understanding of the evolution of TPS family in the *Pinus* genus, we performed a phylogenetic analysis including all the identified pine MTPSs (32), DTPSs (13) and STPSs (5), the two *P. tabuliformis* class-I (KS) and class-II (CPS) of gibberellin biosynthesis, and, by considering the high level of homology among each other, five of the 42 MBOSs (see experimental procedures in the supplementary material). The corresponding phylogenetic tree is reported in Fig. 1.7.

All the pine TPSs involved in specialized metabolism, i.e., MTPSs, STPSs and DTPSs, were clearly separated from the two *P. tabuliformis* TPSs of primary (gibberellin) metabolism, i.e., Pta KS1 and Pta CPS1 (Fig. 1.7), consistent with their assignment to distinct TPS subfamilies, namely TPS-d for the formers, and TPS-c and TPS-e/f for the latter (Fig. 1.5). In turn, the pine TPSs involved in specialized metabolism can be divided into three major clades, corresponding to the three TPS-d1, TPS-d2 and TPS-d3 groups in which the gymnosperm-specific TPS-d subfamily has been subdivided (Martin *et al.* 2004; Keeling *et al.* 2011, Chen *et al.* 2011).

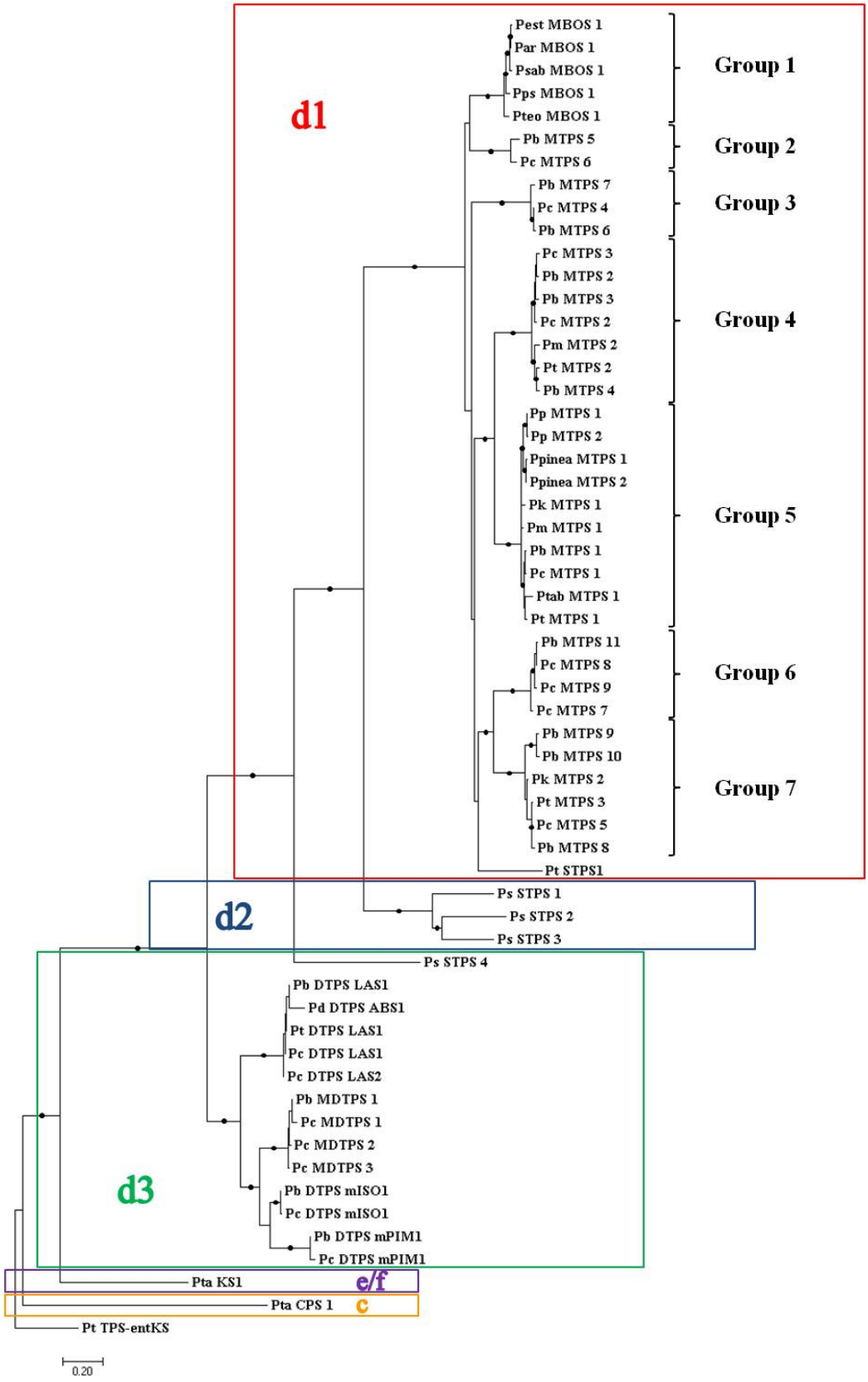


Fig. 1.7. *Phylogenetic tree of the deduced amino acid sequences of terpene synthases (TPSs) in Pinus spp. The Physcomitrella patens ent-kaurene synthase (Pt TPS-entKS) was used to root the tree. Branches indicated with dots represent bootstrap support more than 80% (100 repetitions). Colour rectangles denote TPS subfamilies and groups within the TPS-d subfamily. For acronyms denoting plants species, see Table 1.3. Abbreviations: CPS, ent-copalyl diphosphate synthase; DTPS, diterpene synthase; KS, ent-kaurene synthase; MBOS, 2-methyl-3-buten-2-ol synthase; MTPS, monoterpene synthase; STPS, sesquiterpene synthase.*

The present phylogenetic analysis showed that all the MTPSs identified in the *Pinus* species clustered with the pine MBOSs in the TPS-d1 group (Fig. 1.7), thus confirming the previous findings of Gray *et al.* (2011). The five identified STPSs in *Pinus* species were found to distribute in all of the three TPS-d clades (Fig. 1.7). Four of them are bi-domain (βα) STPSs, of which the three from *P. sylvestris* (Ps STPS1-3) were assigned to the TPS-d2 group, whereas the one from *P. taeda* (Pt STPS1) to the TPS-d1 group. The fifth STPS, identified in *P. sylvestris* (Ps STPS4), is a three-domain (γβα) enzyme, which clustered into the TPS-d3 group together with all the mono- and bi-functional DTPSs identified in the *Pinus* species (Fig. 1.7).

Consistent with previous phylogenetic analyses (Hall *et al.* 2013a), the five class-I/II DTPSs, namely Pc DTPS LAS1/LAS2, Pb DTPS LAS1, Pt DTPS LAS1 and Pd DTPS ABS1, formed a separate branch within the TPS-d3 group close to the eight monofunctional class-I DTPSs, namely Pb MDTPS1, Pc MDTPS1/MDTPS2/MDTPS3, Pc DTPS mISO1, Pb DTPS mISO1, Pc DTPS mPIM1 and Pb DTPS mPIM1 (Fig. 1.7). Furthermore, the putative orthologous pairs Pb DTPS mPIM1/Pc DTPS mPIM1 and Pb DTPS mISO1/Pc DTPS mISO1, for which Hall *et al.* (2013a) showed biochemical functions, clustered in a separate branch with respect to the four remaining monofunctional DTPSs, namely Pb MDTPS1, Pc MDTPS1/MDTPS2/MDTPS3, for which no activity was observed by the above authors. The topology of the phylogenetic tree in Fig. 1.7 indicates that the *P. contorta* and *P. banksiana* class-I DTPSs of specialized metabolism have evolved in relatively recent times through gene duplication of a class-I/II DTPS, accompanied by loss of the class-II activity and subsequent functional diversification. It is worth noting that while the class-I/II LAS enzymes of *P. contorta* and *P. banksiana* have orthologs in other conifers, within and outside of the *Pinus* genus, e.g., in *P. taeda* (Fig. 1.7), *P. abies*, *P. sitchensis*, *A. balsamea* and *A. grandis* (Fig. 1.5), class-I DTPSs of specialized metabolism have not yet been discovered outside of the *Pinus* genus. It is therefore conceivable that they constitute a lineage-specific clade of the TPS-d3 group arising from a common ancestor of the closely related species of *P. contorta* and *P. banksiana*, possibly after the split from *P. taeda*, and after that pine, spruce, and fir genera became separated from each other.

Within the major TPS-d1 clade, phylogenetic analysis showed that the 32 pine MTPSs and the five selected pine MBOSs clustered into seven distinct groups (Fig. 1.7). It is worth noting that some of the pine MTPSs, including the proteins responsible for hemiterpenes biosynthesis (MBOSs), grouped phylogenetically with functionally similar MTPSs from grand fir (*A. grandis*) and spruce (*P. glauca*, *P. abies* and *P. sitchensis*) (Fig. 1.5). This functional conservation across species indicates that significant gene duplication and functionalization took place before the speciation of pine, fir and spruce.

The Group 1 of TPS-d1 clade (Fig. 1.7) contained the five selected MBOSs that use DMAPP as a substrate to produce hemiterpenes (Gray *et al.* 2011, see above). Phylogenetic analysis at the gymnosperms level (Fig. 1.5) showed that the MBOSs from several Pinaceae species are closely related to linalool synthases from *P. abies* and *P. sitchensis*.

Group 2 (Fig. 1.7) included only two proteins from *P. contorta* (Pc MTPS6) and *P. banksiana* (Pb MTPS5), which were shown to form α-terpineol as the major product (Hall *et al.* 2013b). These two proteins have only a 62% sequence identity with Pt MTPS2, a *P. taeda* protein that also produces α-terpineol (Phillips *et al.* 2003) but was assigned to the Group 4 (Fig. 1.7). Indeed, Fig. 1.5 indicates that Pc MTPS6 and Pb MTPS5 were more closely related (77% identity) to 1,8-cineole synthases from *P. glauca* and *P. sitchensis* (Keeling *et al.* 2011).

Group 3 (Fig. 1.7) contained 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.* 2013b). As shown in the phylogenetic tree reported in Fig. 1.5, Pc MTPS4 grouped with functionally similar MTPSs from *P. abies, P. glauca, and P. sitchensis* (Keeling *et al.* 2011), indicating that the genes involved in the synthesis of (+)-3-carene originated before the speciation of pine and spruce.

Group 4 (Fig. 1.7) contained four MTPSs from *P. contorta* and *P. banksiana* (Pc MTPS2 and Pb MTPS 2-4), which were shown to produce (−)-β-pinene as their major product and also (−)-α-pinene, but in comparatively lower amounts (Hall *et al.* 2013b). These four MTPSs are closely related to a *P. taeda* protein (Pt MTPS2) (Fig. 1.7), which instead produces (−)-α-terpineol, but neither (−)-β-pinene nor (−)-α-pinene (Phillips *et al.* 2003). On one hand, this demonstrates that it is not always possible to predict the function of a putative MTPS only based on its sequence identity with homologous enzymes; indeed, it has been reported that few amino acid substitutions are sufficient to modify the product profiles of MTPSs from grand fir (Katoh *et al.* 2004; Hyatt *et al.* 2005). On the other hand, a high level of sequence identity as opposed to a clearly distinct catalytic competence provides a good example to illustrate the functional plasticity of MTPSs in conifers. The second member from *P. contorta* assigned to the Group 4, namely Pc MTPS3, did not show any activity with GPP, FPP or GGPP as substrates, either as full-length or as a truncated protein from which the putative plastid-targeting peptide had been removed (Hall *et al.* 2013b). Finally, another member of Group 4, namely Pm MTPS2 from *P. massoniana*, although reported to be a (-)-α-terpineol synthase in the NCBI database, most likely on the basis of the high sequence identity with Pt MTPS2 from *P. taeda*, was not functionally characterized yet, to the best of our knowledge.

Group 5 (Fig. 1.7) included ten putative α-pinene synthases, of which only those from *P. taeda* (Pt MTPS1), *P. contorta* (Pc MTPS1) and *P. banksiana* (Pb MTPS1) have been functionally characterized as forming (−)-α-pinene as their main product (Phillips *et al.* 2003; Hall *et al.* 2013b).

Phylogenetic analysis at the gymnosperms level (Fig. 1.5) showed that Pc MTPS1 groups most closely with spruce and fir enzymes that also produce (−)-α-pinene (Bohlmann *et al.* 1997; Keeling *et al.* 2011). This indicates that the genes involved in the synthesis of (−)-α-pinene originated before pine, fir and spruce became separated species, as also occurred for the genes encoding for (+)-3-carene synthases.

Group 6 (Fig. 1.7) contained three proteins from *P. contorta* and *P. banksiana* (Pc MTPS8-9 and Pb MTPS11) which form (−)-β-phellandrene as their major product (Hall *et al.* 2013b). Another member of this group, namely Pc MTPS7 from *P. contorta*, although showing a 95% identity with the above (−)-β-phellandrene synthases, forms predominantly (−)-camphene and (+)-α-pinene, along with other minor products (Hall *et al.* 2013b).

Group 7 (Fig. 1.7) included three MTPSs from *P. taeda* (Pt MTPS3), *P. contorta* (Pc MTPS5) and *P. banksiana* (Pb MTPS8) that were shown to form (+)-α-pinene as their dominant product (Phillips *et al.* 2003; Hall *et al.* 2013b). Two additional members of Group 7, namely Pb MTPS9-10 from *P. banksiana,* showed no activity with GPP, GGPP or FPP as the substrates (Hall *et al.* 2013b). Finally, the member of Group 7 from *P. kesiya*, although reported to be a α-pinene synthase in the NCBI database, was not functionally characterized so far.

Previous phylogenetic analyses (Hall *et al.* 2013b) showed that (+)-α-pinene synthases and (−)-β-phellandrene synthases from *P. contorta, P. banksiana* and *P. taeda* form a unique and apparently *Pinus*-specific subclade within the TPS-d1 group, as also shown in our phylogeny of gymnosperm TPSs, in which two representative members of Group 6 (Pc MTPS8) and Group 7 (Pb MTPS8) are clearly separated from the other conifer MTPSs (Fig. 1.5). In the phylogenetic tree of Fig. 1.5, moreover, the (−)-β-phellandrene synthases from *P. contorta* (Pc MTPS8), *A. grandis* (Ag betaPHEL) and *P. sitchensis* (Psi betaPHEL) clustered separately from each other, suggesting a multiple origin of (−)-β-phellandrene biosynthesis in Pinaceae. Finally, by comparing Fig. 1.7 with Fig. 1.5 it is worth noting that genes coding for MTPS producing (+)-α-pinene as their main product have not been identified so far in any other genus of the Pinaceae, except that in *Pinus*, indicating that this function may have evolved in the pine lineage after its separation from spruce and fir.

1.7. ISOLATION OF CDNA AND GENOMIC SEQUENCES CODING FOR MTPS IN *Pinus nigra* subsp. *laricio*

As a case study, we report here about the first attempt to isolate MTPS sequences in a non-model pine species by using a strategy based on the phylogeny of available MTPSs from different *Pinus* species (Fig. 1.7).

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 (Nicolaci *et al.* 2014). In Calabria, where it is considered an endemic species, it grows on the Sila and Aspromonte mountains, and represents an essential element of the forest landscape that plays an important role not only in soil conservation and watershed protection, but also in the local forest economy (Nicolaci *et al.* 2014).

To gain insights into the ecological and functional roles of terpenes in *P. laricio*, an insofar completely neglected species under such respect, we preliminarily determined, via a conventional GC-MS approach, the terpene profile of *P. laricio* needles, identifying several monoterpenes such as (−)-β-pinene, (−)-α-pinene, (+)-α-pinene, (+)-3-carene, and (−)-β-phellandrene, as the most abundant terpenoids in the above organs (data not shown; M. Badiani and A. Sorgonà, unpublished). Thus, we focused our attention in isolating cDNA sequences encoding MTPSs potentially involved in the synthesis of the aforementioned monoterpenes in *P. laricio*.

Deduced amino acid and nucleotide sequences of pine MTPSs belonging to each of the seven phylogenetic groups in the TPS-d1 clade (Fig. 1.7) 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 coding for MTPSs in *P. laricio* needles (see experimental procedures in the supplementary material). 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 the seven phylogenetic groups in which the TPS-d1 clade can be subdivided. Moreover, four partial *P. laricio* transcripts of Groups 1, 2, 5 and 7, were used as a template for isolating full-length MTPS cDNAs by 5' and 3' RACE extensions. The four full-length cDNAs contained ORFs of 1845, 1857, 1908 and 1890 bp encoding proteins of 614, 618, 635 and 629 aa, respectively (Fig. 1.8). The group 4 partial transcript of 1132 bp in length encoded an incomplete protein of 376 aa (Fig. 1.8).

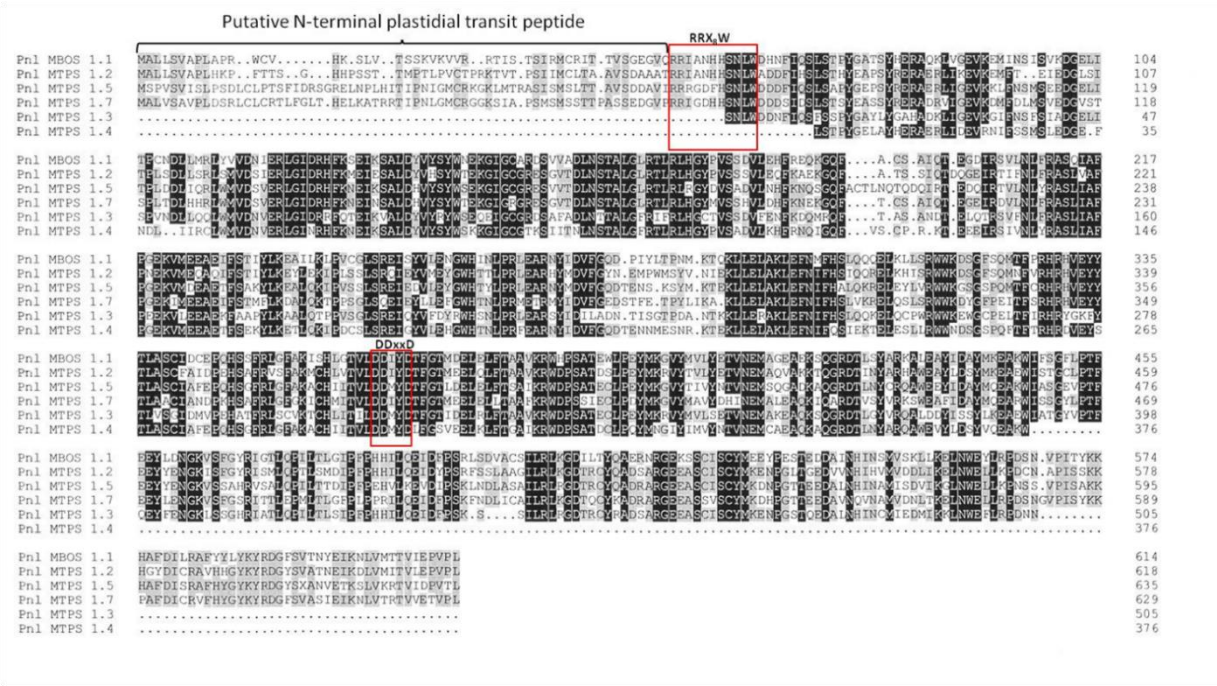


Fig. 1.8. *Alignment of deduced amino acid sequences of full-length cDNAs (Pnl MBOS-1.1, Pnl MTPS-1.2, -1.5 and -1.7) and partial genomic and cDNA sequences (Pnl MTPS-1.3 and-1.4) isolated from Pinus nigra subsp. laricio needles. The black-shaded residues are highly conserved ones; the gray-shaded residues are identical in at least three of the six sequences shown. The brace indicates the putative N-terminal transit peptide region. The "RRX8W" and the "DDxxD" motifs are indicated with red open rectangles. Abbreviations: MBOS, 2-methyl-3-buten-2-ol synthase; MTPS, monoterpene synthase.*

Putative orthologous genes for the phylogenetic TPS-d1 Group 3 were not found in the transcriptome of *P. laricio* needles, despite extensive efforts to amplify by PCR cDNA fragments of such genes, suggesting that they were not expressed in the source plant material (needles). Therefore, to verify 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 TPS-d1 Group 3 and the genomic DNA extracted from *P. laricio* needles as a template. Such strategy yielded a *P. laricio* genomic fragment of 2630 bp, extended from the 5' to the 3' ends of the coding region, which contained ten exons (with the first and the tenth incomplete) and nine introns (Fig. 1.9), being consistent with the previously characterized genomic sequences of conifer MTPSs (Trapp and Croteau 2001; Hamberger *et al.* 2009; Hall *et al.* 2011).

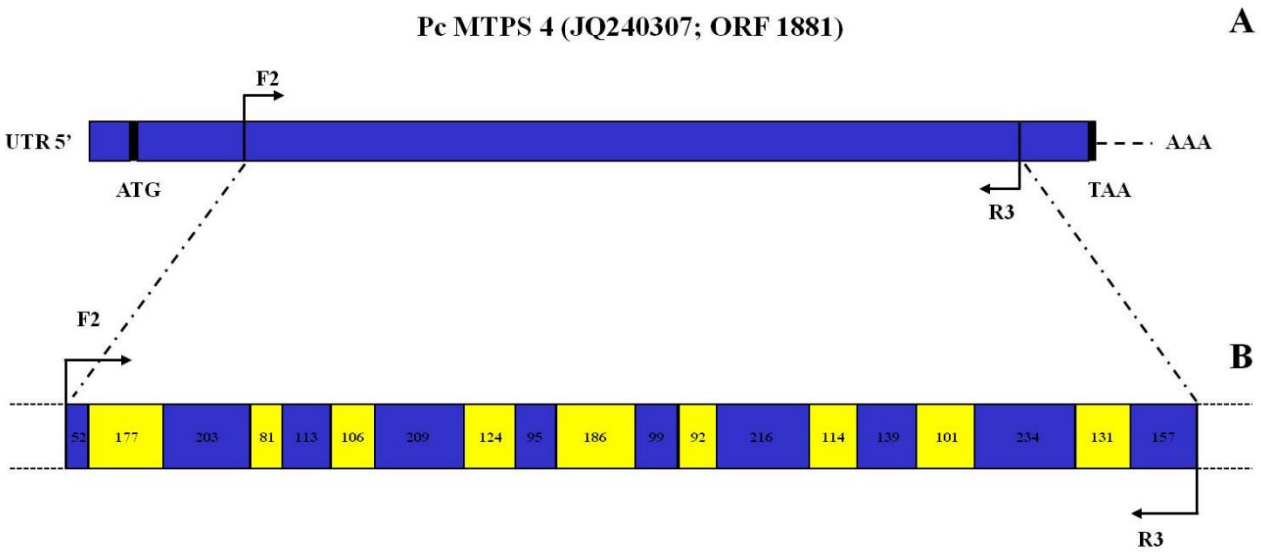


Fig. 1.9. *Strategy adopted for the genomic amplification of a putative Pinus nigra subsp. laricio gene belonging to the phylogenetic TPS-d1 Group 3. A) Schematic representation of the full-length cDNA of a representative member of the phylogenetic TPS-d1 Group 3 (PcMTPS4 from Pinus contorta, in the present case, see Fig. 1.7 and Table 1.5) in which the positions of the forward (F2) and the reverse (R3) primers used in the amplification of genomic DNA are shown. B) Intron (yellow)/exon (blue) structure of the amplified Pinus nigra subsp. laricio genomic sequence. The positions of the primers used to amplify the genomic fragment are also shown.*

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 the present study. Based on the determined intron/exon structure, the genomic fragment holds a partial nucleotide sequence potentially translated to having 1517 bp coding for an incomplete protein of 505 aa (Fig. 7).

A combined phylogenetic analysis of the six deduced aminoacid sequences from *P. laricio* (Pnl MBOS 1.1, Pnl MTPS 1.2, Pnl MTPS 1.5, Pnl MTPS 1.7, Pnl MTPS 1.4 and Pnl MTPS 1.3) with all the pine MTPSs (32) and the five selected MBOSs identified in the NCBI database (see Fig. 6), allowed to place the *P. laricio* predicted proteins in six out of the seven TPS-d1 phylogenetic groups (Fig. 1.10), 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 (Fig. 1.8). For instance, each of the four full-length predicted proteins incorporate sequences for a putative transit peptide ranging from 40 to 56 aa for import of mature proteins into plastids upstream of a conserved "RRX8W" domain. This, as reported before, is reputed to be essential for the catalysis of monoterpene cyclization (Whittington *et al.* 2002; Hyatt *et al.* 2007). Moreover, all the six *P. laricio* predicted proteins had a conserved aspartate-rich domain, i.e.,

"DDxxD", responsible for class-II activity that coordinates substrate binding via the formation of divalent-cation salt bridges (Tarshis *et al.* 1996; Lesburg *et al.* 1997).

Because a strategy based on the phylogeny of all the available MTPSs from different *Pinus* species was instrumental for isolating the full-length transcripts coding for MTPSs in *P. laricio*, this same approach could be promising for isolating from this non-model conifer species also the TPS-d members producing diterpenes and sesquiterpenes. 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.

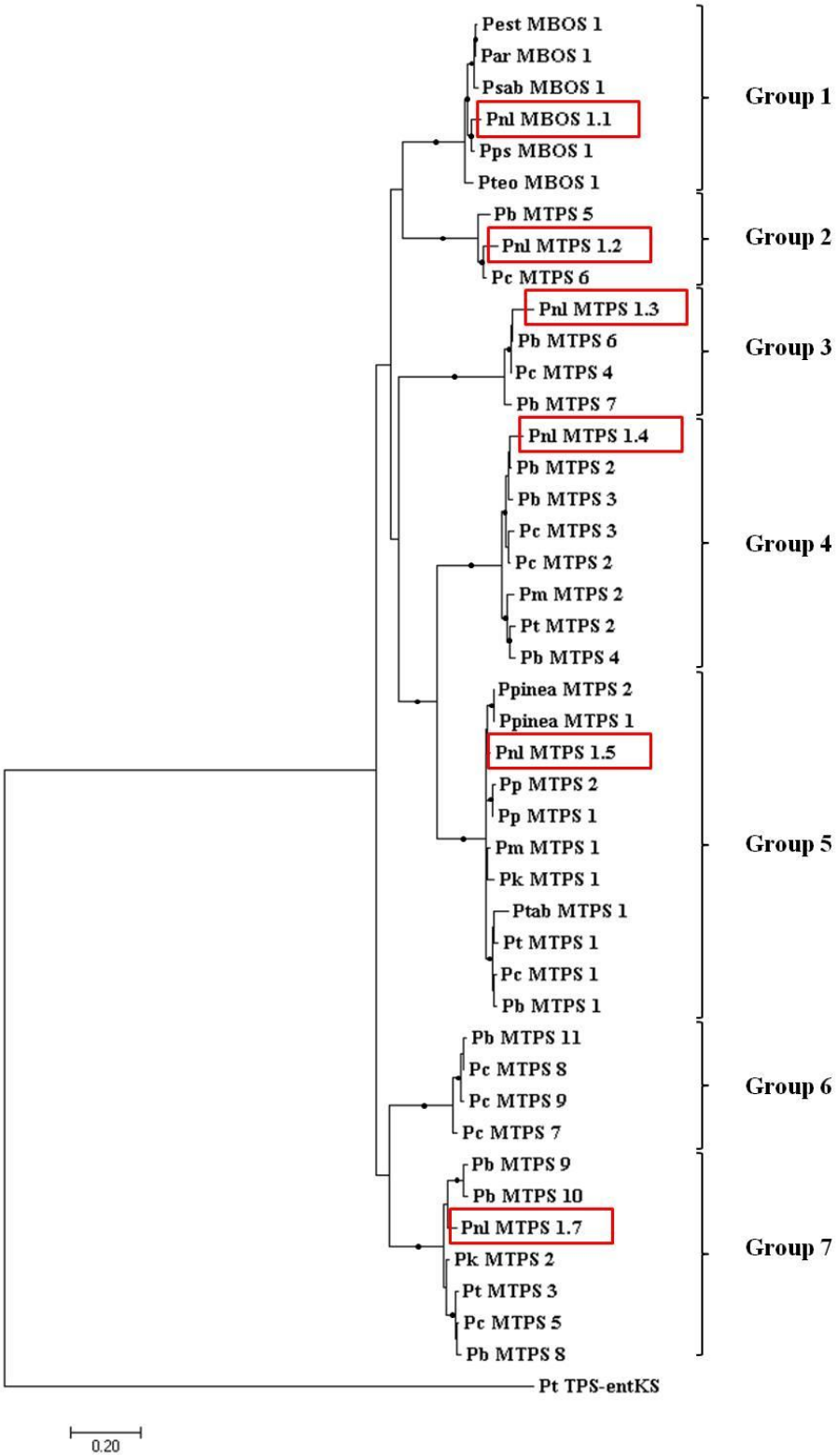


Fig. 1.10. *Phylogenetic tree of the deduced amino acid sequences obtained by combining monoterpene synthases (MTPSs) and the five selected 2-methyl-3-buten-2-ol synthases (MBOSs) identified in different Pinus species (Fig. 3.7 and Table 3.5) and the six sequences isolated from Pinus nigra subsp. laricio (outlined in red). Physcomitrella patens ent-kaurene synthase (Pt TPS-entKS) was used to root the tree. Branches indicated with dots represent bootstrap support more than 80% (100 repetitions). The seven phylogenetic groups identified in the pine members of TPS-d1 clade are highlighted with square brackets. For acronyms denoting plants species, see Table 3.5.*

1.8. CONCLUSIONS AND PROSPECTS

Conifers developed a variety of physical and chemical defenses against pathogens and herbivores, among which one of the most significant is the production of oleoresin, a complex blend of volatile mono- and sesqui-terpenes, along with non-volatile diterpene resin acids. In fact, the complex defense system of conifer oleoresin is considered one of the main factors that has enabled conifer trees to evolve and flourish as the dominant group of gymnosperms on the planet. Because of the ample physical and chemical diversity of oleoresin terpenoids and the resulting technological versatility, many of these compounds have also made their way in the food industry, as well as in the production of cosmetics, medicine, and chemicals.

The increasing interest in new terpenoid products for industrial uses makes dependence on natural resources alone not always sufficient or feasible. However, new possibilities exist nowadays for developing and improving the production of high-value terpenoid compounds on an industrial scale, via metabolic engineering of the natural biosynthetic pathway in bacterial (*E. coli*) and yeast (*Saccharomyces cerevisiae*) systems, and also in heterologous plant hosts. Therefore, novel and in-depth knowledge of the evolutionary diversification of members of conifer TPS family, their modular structure, and their putative functions appears to be important not only for a deeper understanding of their physiological and ecological roles, but also to foster metabolic engineering and synthetic biology tools for the production of high value terpenoid compounds.

Latest developments in conifer transcriptome and genome sequencing, together with metabolite analysis, have boosted the identification and annotation of terpenoid pathway genes. However, until now functional characterization has been achieved only for a subset of TPS members in each of the considered Pinaceae, while no functional analysis is reported for the TPSs recently identified in the Cupressaceae. It is worth nothing that precise computational annotation of TPS functions is significantly hindered by the high sequence identity of proteins with different enzymatic activities. Therefore, TPS characterization needs a laborious *in vitro* and *in vivo* evaluation of each candidate gene, often hampered by the absence of commercially available substrates and standards. Furthermore, comparatively few experiments have been dedicated so far to elucidate TPSs molecular architecture in gymnosperms.

Comprehensive structural and functional analyses of members of the TPS family in selected model Pinaceae species, such as *P. glauca*, *A. grandis*, *P. taeda* and *P. contorta*, for which large transcriptomic and genomic resources are available, as well as plants occupying key position in the conifer phylogeny, like species belonging to Cupressaceae and Taxaceae families, will provide new knowledge about the variety of terpenoid compounds that can be synthesized by a single conifer

species, and how its enzymes have developed the ability to do so. Simultaneously, wide transcriptome projects targeted at gymnosperm species producing interesting terpenoid metabolites, as recently reported for the Cupressaceae species *P. orientalis* and *T. plicata*, will boost the discovery and annotation of comprehensive sets of TPS genes in non-model conifer species. Alternatively, a comprehensive phylogenetic analysis of the identified TPS genes in a particular genus can help for isolating genomic and cDNA sequences from so-far neglected, but otherwise ecologically and economically relevant, conifer species, as shown in the case study reported here.

2. CHAPTER 2: DITERPENE RESIN ACIDS AND OLEFINS IN CALABRIAN PINE (*Pinus nigra* subsp. *laricio* (POIRET) MAIRE) OLEORESIN: GC-MS PROFILING OF MAJOR DITERPENOIDS IN DIFFERENT PLANT ORGANS, MOLECULAR IDENTIFICATION AND EXPRESSION ANALYSIS OF DITERPENE SYNTHASE GENES.

Adapted from: Alicandri et al. (2021) **Diterpene resin acids and olefins in Calabrian pine (*Pinus nigra* subsp. *laricio* (Poiret) Maire) oleoresin: GC-MS profiling of major diterpenoids in different plant organs, molecular identification and expression analysis of diterpene synthase genes.** Plants 2021, 10, 2391. https://doi.org/10.3390/plants10112391

Abstract:

A quali-quantitative analysis of diterpenoid composition in tissues obtained from different organs of *Pinus nigra* subsp. *laricio* (Poiret) Maire (Calabrian pine) was carried out. Diterpene resin acids were the most abundant diterpenoids across all the examined tissues. The same nine diterpene resin acids were always found, with the abietane type prevailing on the pimarane type, although their quantitative distribution was found to be remarkably tissue specific. The scrutiny of the available literature revealed species specificity as well. A phylogeny-based approach allowed us to isolate four cDNAs coding for diterpene synthases in Calabrian pine, each of which belonging to one of the four groups into which the d3 clade of the plants' terpene synthases family can be divided. The deduced amino acid sequences allowed predicting that both monofunctional and bifunctional diterpene synthases are involved in the biosynthesis of diterpene resin acids in Calabrian pine. Transcript profiling revealed differential expression across the different tissues and was found to be consistent with the corresponding diterpenoid profiles. The isolation of the complete genomic sequences and the determination of their exon/intron structures allowed us to place the diterpene synthase genes from Calabrian pine on the background of current ideas on the functional evolution of diterpene synthases in Gymnosperms.

Keywords: diterpenoids; diterpene synthase; *Pinus nigra* subsp. *laricio* (Poiret) Maire; Calabrian pine; pine oleoresin; genomic organization of gymnosperm diterpene synthases; diterpene resin acid

Key to plant species:

Abies balsamea (L.) Mill., 1768, balsam fir; *Abies grandis* (Douglas ex D. Don) Lindl., grand fir; *Physcomitrella patens* (Hedw.) Bruch & Schimp., spreading earthmoss; *Picea abies* (L.) H. Karst, Norway spruce; *Picea glauca* (Moench) Voss, white spruce; *Picea sitchensis* (Bongard) Carrière, 1855, Sitka spruce; *Pinus banksiana* Lamb., jack pine; *Pinus contorta* Douglas, lodgepole pine; *Pinus nigra* J.F.Arnold, Austrian pine or black pine; *Pinus nigra* subsp. *laricio* (Poiret) Maire, Calabrian pine; *Pinus pinaster* Aiton, maritime pine; *Pinus radiata* D.Don, Monterey pine; *Pinus taeda* L., loblolly pine; *Pseudolarix amabilis* (N. Nelson) Rehder, golden larch

2.1. INTRODUCTION

Gymnosperms developed a variety of physical and chemical defenses against pathogens and herbivores, among which one of the most significant is the production of terpenoid metabolites (Zulak and Bohlmann, 2020; Celedon and Bohlmann, 2019; Ma *et al.* 2019; Boncan *et al.* 2020. The complex terpenoid defense mechanisms have persisted throughout the long evolutionary history of gymnosperms and their decreasing geographical distribution during the Cenozoic era (Chou *et al.* 2011; Leslie *et al.* 2012), but diversified into often species-specific metabolite blends. For instance, structurally related labdane-type diterpenoids, such as ferruginol and derivative compounds, act as defense metabolites in many Cupressaceae species (Ma *et al.* 2019; Karchesy *et al.* 2018; Simoneit *et al.* 2018). On the other hand, diterpene resin acids (DRAs), together with mono- and sesquiterpenes, are the main components of the oleoresin defense system in the Pinaceae species (e.g., Conifers), and have been shown to provide an effective barrier against stem-boring weevils and associated pathogenic fungi (Robert *et al.* 2010, Lah *et al.* 2013, Wang *et al.* 2013, Kshatriya *et al.* 2018).

Diterpenoids from gymnosperms are also important for their technological uses, being employed in the production of solvents, flavours, fragrances, pharmaceuticals and a large selection of bioproducts (Zulak and Bohlmann, 2010; Zerbe *et al.* 2013): among the many other examples, the anticancer drugs pseudolaric acid B, obtained from the roots of the golden larch (*Pseudolarix amabilis*) (Mafu *et al.* 2017), and taxol, extracted from yew (*Taxus* spp.) (Croteau *et al.* 2006), as well as *cis*-abienol, produced by balsam fir (*Abies balsamea*), which is a molecule of interest for the fragrance industry (Zerbe *et al.* 2012).

The diterpenoids of conifer oleoresin are largely members of three structural groups, the abietanes, the pimaranes, and the dehydroabietanes, all of which are characterized by tricyclic parent skeletons

(Celedon and Bohlmann, 2019; Hamberger *et al.* 2011). These diterpenoids are structurally similar to the tetracyclic *ent*-kaurane diterpenes, which include the ubiquitous gibberellin (GA) phytohormones. Both the oleoresin diterpenoids of specialized metabolism and the GAs of general metabolism derive from the common non-cyclic diterpenoid precursor geranylgeranyl diphosphate (GGPP).

In conifers, among the other gymnosperms, the structural diversity of diterpenoids results from the combined actions of diterpene synthases (DTPSs) and cytochrome P450 monooxygenases (CP450s) (Celedon and Bohlmann, 2019). The former enzymes catalyze the cyclization and rearrangement of the precursor molecule GGPP into a range of diterpene olefins, often referred to as the neutral components of the oleoresins. Olefins are then functionalized at specific positions by the action of CP450s, through a sequential three-step oxidation first to the corresponding alcohols, then to aldehydes, and finally to DRAs (Celedon and Bohlmann, 2019), such as abietic, dehydroabietic, isopimaric, levopimaric, neoabietic, palustric, pimaric, and sandaracopimaric acids, which are the major constituents of conifers oleoresins (Celedon and Bohlmann, 2019; Hamberger *et al.* 2011; Celedon and Bohlmann, 2019)). The chemical structure of the most represented diterpenoids in *Pinus* spp. are reported in Figure 2.1.

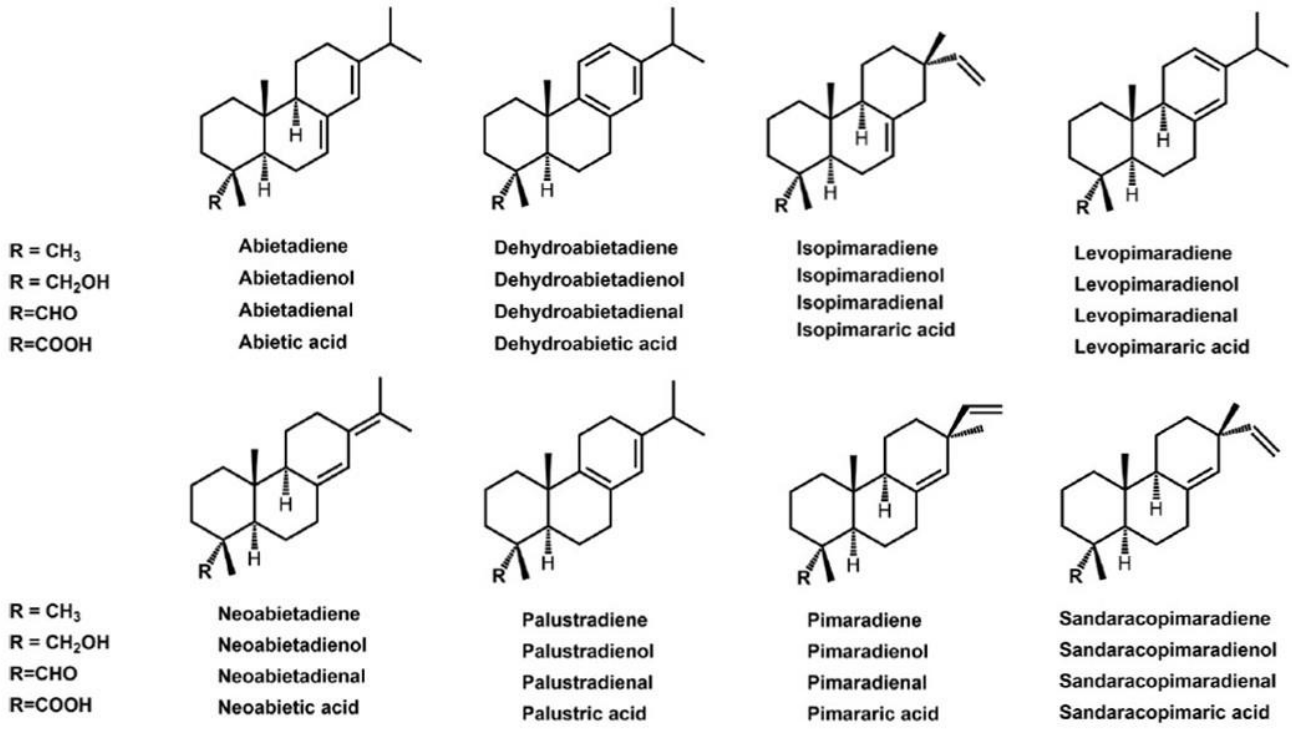


Figure 2.1. *Chemical structures of the most represented diterpenoids in Pinus spp. [R = CH3 olefins constituents; R = CH2OH alcoholic constituents; R = CHO aldehydic constituents; R = COOH diterpene resin acid (DRA) constituents]. (Adapted from Turner et al. 2019).*

Diterpene synthases in gymnosperms share a conserved α-helical fold with a common three-domains γβα architecture, and characteristic functional motifs (DxDD, DDxxD, NSE/DTE), which determine the catalytic activity of the enzymes (Celedon and Bohlmann, 2019; Gao *et al.* 2012). Indeed, depending on domain structure and presence/absence of signature active-site motifs, three major classes of DTPSs can be identified, namely monofunctional class I and class II DTPSs (mono-I-DTPS and mono-II-DTPS in the following, respectively) and bifunctional class I/II DTPSs (bi-I/II-DTPSs in the following) (Alicandri *et al.* 2020). Mono-II-DTPSs contain a conserved DxDD motif located at the interface of the γ and β domains, which is critical for facilitating the protonation-initiated cyclization of GGPP into bicyclic prenyl diphosphate intermediates (Peters, 2010), among which copalyl diphosphate (CPP) and labda-13-en-8-ol diphosphate (LPP) are the most common (Ma *et al.* 2019; Hall *et al.* 2013a; Ma *et al.* 2021). Mono-I-DTPSs then convert the above bicyclic intermediates into the tricyclic final structures, namely diterpene olefins, by ionization of the diphosphate group and rearrangement of the carbocation, which is facilitated by a Mg2+-cluster coordinated between the DDxxD and the NSE/DTE motifs in the C-terminal α-domain.

Bi-I/II-DTPSs, regarded as the major enzymes involved in the specialized diterpenoids metabolism in conifers, contain all the three functional active sites, namely DxDD (between γ and β domains), DDxxD and NSE/DTE (in the α-domain), and therefore are able to carry out in a single step the conversion of the linear precursor GGPP into the final tricyclic olefinic structures, which serve in turn as the precursors for the most abundant DRAs in each species (Keeling *et al.* 2011). In contrast, the synthesis of GA precursor ent-kaurene in gymnosperms involves two consecutively acting mono-I- and mono-II-DTPSs, namely *ent*-CPP synthase (ent-CPS) and *ent*-kaurene synthase (ent-KS), respectively, as has also been shown for both general and specialized diterpenoid metabolism in angiosperms (Zerbe and Bohlmann, 2015; Alicandri *et al.* 2020; Chen *et al.* 2011). Interestingly, class-I DTPSs involved in specialised diterpenoid metabolism were identified in *Pinus contorta* and *Pinus banksiana*, which can convert (+)-CPP produced by bifunctional DTPSs to form pimarane-type diterpenes (Hall *et al.* 2013a), while no (+)-CPP producing class-II DTPSs have been identified in other conifers.

Most of the existing knowledge concerning the genetics and metabolism of specialized diterpenes in gymnosperms was obtained from model Pinaceae species, such as *Picea glauca*, *Abies grandis*, *Pinus taeda* and *P. contorta* (Zulak and Bohlmann, 2020; Celedon and Bohlmann, 2019; Hall *et al.* 2013a), for which large transcriptomic and genomic resources are available, as well as, in recent times, from species occupying key position in the gymnosperm phylogeny, like those belonging to the Cupressaceae and the Taxaceae families (Ma *et al.* 2019; Ma *et al.* 2021).

In previous work of ours (Alicandri *et al.* 2020; Foti *et al.* 2020) we began to gain insight into the ecological and functional roles of the terpenes produced by the non-model conifer *Pinus nigra* subsp. *laricio* (Poiret) (Calabrian pine), one of the six subspecies of *P. nigra* (black pine) and an insofar completely neglected species under such respect. In terms of natural distribution, black pine is one of the most widely distributed conifers over the whole Mediterranean basin, and its *laricio* subspecies is considered endemic of southern Italy, especially of Calabria, where it is a basic component of the forest landscape, playing key roles not only in soil conservation and watershed protection, but also in the local forest economy (Nicolaci *et al.* 2015).

In the present study, we carried out for the first time, to the best of our knowledge, a quali-quantitative analysis of diterpenoids composition in different tissues of Calabrian pine via conventional gas chromatography-mass spectrometry (GC-MS). On this same subspecies, in addition, we report here about the isolation of full length (FL) cDNAs and the corresponding genomic sequences encoding for DTPSs involved in the specialized diterpenoid metabolism, obtained by using a strategy based on the phylogeny of available DTPSs from different *Pinus* species. The isolation of DTPS genes made it possible a tissue-specific gene expression analysis, to be confronted with the corresponding GC-MS diterpene profiles.

2.2. MATERIAL AND METHODS

➢ *2.2.1. Plant material*

Three-yr old Calabrian pine [*Pinus nigra* subsp. *laricio* (Poiret) Maire] saplings obtained from the Calabria Regional Forest nursery (Calabria Verde Agency, Catanzaro, Italy) were grown in the open within protective housings set up at the Calabria Regional Biodiversity Observatory, located at Cucullaro (38° 17’27” N, 15° 81’68” E; altitude 1010 MASL, exposed East), in the heart of the Aspromonte National Park, Southern Italy. In the course of four sampling campaigns from November 2019 to May 2020, different tissues/organs were collected, namely young needles (YN), mature needles (MN), bark and xylem combined from leader stem (LS), bark and xylem combined from interwhorl stem (IS), and roots (R). All collected tissues were immediately frozen in liquid nitrogen and stored at −80 °C until analysis.

➢ *2.2.2. Extraction and GC/MS analysis of diterpene metabolites*

After thawing, tissue samples were dried (48-72 hours, in the dark) at room temperature and then cut inTO fragments of about 1-2 mm by means of a scalpel. For all the tissue types, the extraction of the diterpenoid fraction was performed following the procedure described by López-Goldar *et al.*

(2020), with minor modifications. Briefly, approximately 250 mg of each of the five different tissue types were extracted twice with 2 mL of a n-hexane/dichloromethane mixture (1:1; v:v). During each extraction cycle, the extracts were kept in an ultrasonic bath at 25 °C for 20 min. After pooling together, the two aliquots obtained in a recovery glass vial. residual water was removed by passing the extracts onto a column containing 2 g of anhydrous Na_2SO_4, and the obtained eluates were kept in the dark and stored at -20 °C. For derivatization, first 200 µL of each extract were passed onto a column containing 15 mg of graphitized carbon, to remove non-terpenic impurities, and then 50 µL of each eluate were transferred into a conical vial and dried under gentle stream of N_2. After drying, 100µL of a 1:1 (*v/v*) mix of N,O-bis(trimethylsilyl)trifluoroacetamide, containing 1% (*v/v*) trimethylchlorosilane, plus pyridine were added to each sample, and the derivatization allowed to proceed for 30 min at 65 °C. Finally, the solution was brought to dryness under gentle stream of N_2, the residue was resuspended with 50 µL of n-hexane and finally stored in darkness at -20 °C until the GC-MS analysis. For each of the aforementioned tissue types, three biological replicates were processed and analysed, each of them in triplicate.

Qualitative and quantitative analysis of diterpenes from Calabrian pine tissues were carried out by means of a high-fast GC-MS approach an Agilent Technologies GC (model 7890A), equipped with a VF-5ms capillary column (Agilent Technologies; 15 m x 0.15 mm of inner diameter and a 0.15 µm film thickness) was used under the following thermal conditions: from 90 °C (2 min.) to 350 °C with a ramp of 44.7 °C min^{-1}, then in isothermal for 5 min. The He carrier gas constant flow was at 1.2 mL min^{-1}. The sample injection (0.5 µL) was performed under pulsed splitless technique (43 psi) and at 300°C. The coupled detector consisted of an Agilent mass selective detector (VL MSD-Triple-Axis Detector), mod. 5975C. The transfer line, the ion source and the analyzer were kept at 300 °C, 230 °C and 150 °C respectively. The acquisition was carried out under full scan mode (range m/z: 50-650).

The identification of the different diterpene metabolites was carried out by comparison of experimental mass spectra both with those in NIST08 and Wiley02 Libraries and those of the available reference literature (Hall *et al.* 2013; Ro and Bohlmann, 2006; Adams, 2007), as well as of their related retention indices (López-Goldar *et al.* 2020). As far as the Wiley and NIST mass spectra libraries are concerned, the spectral match scores obtained for the diterpenes analyzed in the present work were invariably higher than 850, consistently returning the correct identification of each metabolite as the "first hit". According to the NIST library guidelines, the above score value of mass spectra match is considered to be satisfactory and reliable for the correct identification of a given molecules.

The analytes concentrations, expressed as $\mu g \cdot g^{-1}$ dry weight (d.w.), were calculated by comparison with a calibration curve obtained by using a commercial standard of abietic acid (1*R*,4a*R*,4b*R*,10a*R*)-1,4a-dimethyl-7-(propan-2-yl)-1,2,3,4,4a,4b,5,6,10,10a-decahydrophenanthrene-1-carboxylic acid (Sigma-Aldrich catalog N. 00010). The GC/MS methods used in the present study for the extraction and analysis of plant metabolites were adequately validated for their selectivity, precision, and efficiency. Selectivity was verified by observing that no interfering peak was apparent at the elution time of each target analyte upon injecting three replicate blank samples. Precision was tested by measuring the inter- and intra-day variability in the chromatographic profiles of spiked samples, which ranged from 2 to 7% in terms of relative standard deviation. Finally, the intrinsic recovery of the extraction method was calculated as a mean of three replicate samples, in each of which the plant tissue was spiked with a known aliquot of abietic acid standard solution and then extracted, cleaned and derivatized prior to injection onto GC-MS. Regardless of the tissue extracted, the measured mean recovery always ranged from 80 to 90%.

➢ *2.2.3. RNA isolation and cDNA synthesis*

Total RNA was extracted from 250 mg of each of the five tissues considered according to Pavy *et al.* (2008). RNA concentration and integrity were checked, using a NanoDrop ND-1000 spectrophotometer (Labtech, East Sussex, UK). Only RNA samples with a 260/280 wavelength ratio between 1.9 and 2.1, and a 260/230 wavelength ratio greater than 2.0, were used for cDNA synthesis. First-strand cDNA was synthesized from 3 µg of total RNA of each of the five tissues using a Xpert cDNA Synthesis Kit (GRiSP Research Solution, Porto, Portugal) according to the manufacturer's instructions.

➢ *2.2.4. DNA extraction*

Genomic DNA was extracted from 100 mg of young and mature needles 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.

➢ *2.2.5. Isolation of partial and full-length cDNAs coding for diterpene synthases*

According to the methods reported in Alicandri *et al.* (2020), RT-Polymerase chain reaction (PCR) was used to amplify partial cDNAs coding for DTPSs in *P. nigra* subsp. *laricio*, by using

forward and reverse primers designed in conserved regions among DTPS sequences of *Pinus* species of the different groups identified by phylogenetic analysis. The complete list of the used forward and reverse primers is reported in Table 2.2.

Each PCR reaction was performed in a total volume of 50 μL containing 2 μL of RT reaction obtained from a pool of total RNA from the five different tissues (see section 2.3), 0.4 μM of each forward and reverse primer, and 25 μL of Xpert Taq Mastermix (2X) (GRiSP Research Solutions, Porto, Portugal) which includes pure Xpert Taq DNA Polymerase, dNTPs, $MgCl_2$ 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.

The partial DTPS cDNAs were used as templates for 5' and 3' RACE extensions using the 5′/3’ RACE System for Rapid Amplification of cDNA Ends Kit from INVITROGEN Life Technologies, following manufacturer's instructions and using 3 μg of a pool of total RNA from the five different tissues. The sequences of the RACE primers used are reported in Table 2.2.

➢ *2.2.6. Isolation of genomic sequences coding for diterpene synthases*

Genomic DNA was used to amplify *P.nigra* subsp. *laricio* DTPS genomic sequences by using specific forward and revers primers, designed, respectively, on the proximity of the initiation (ATG) and on the stop codons of each full-length isolated cDNA (Table 2.2). The PCR reactions and conditions were the same as described in section 2.5, with the exception of the extension step that was increased from 3 to 6 min at 72 °C.

➢ *2.2.7. Cloning and sequencing of RACE, cDNA and genomic amplification products*

Samples (5-10 μL) of the amplification products of RACE, partial cDNAs and genomic sequences were separated on 1.5% agarose gels and visualized under UV radiation after staining with ethidium bromide (0.001% *w/v*) 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, Mannheim, Germany) according to manufacturer's instructions, and cloned into the pGEM-T easy plasmid vector (Promega, Madison, WI, USA) following the manufacturer’s protocol. Three different clones for each cDNA, genomic and RACE amplicon, were sequenced. Plasmid DNA for sequencing reaction was prepared from 3

mL overnight cultures using a Wizard® Plus SV Minipreps DNA Purification Systems (Promega, Madison, WI, USA). A private company (MWG, Biotech AG, Germany) performed sequencing. 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 either vector or sequence specific primers. The sequences of the genomic clones were obtained by sequencing them with internal primers complementary to the cDNA sequences, and designed near the predicted exon/intron junctions, so to amplify each exon and nearby intron on both strands to fill gaps and resolve uncertainties (primers are available upon request).

The sequence data from this article can be found in the GeneBank database under accession numbers OK245418 to OK245421 for *Pnl DTPS1-4* FL cDNAs, and OK245422 to OK245425 for *Pnl DTPS1-4* genomic sequences.

➢ *2.2.8. Analysis of the nucleotide and of the deduced amino acid sequences*

All the nucleotide sequences obtained were analysed by DNAMAN Sequence Analysis Software (Version 3, Lynnon Biosoft) and their homologies were scored using the BLASTX program through the National Center for Biotechnology Information (NCBI) database. The software developed by NetGene (Hebsgaard, 1996) was used for the prediction of intron splice sites within the genomic sequences. The predicted protein sequences were analysed by searching for conserved motifs in CDD (Conserved Domain Database in the NCBI) and SMART (Simple Modular Architecture Research Tool, European Molecular Biology Laboratory) databases; their subcellular locations were predicted by ChloroP (Emanuelsson *et al.* 1999), Predotar (Small *et al.* 2004) and WoLF PSORT (Horton *et al.* 2007).

➢ *2.2.9. Phylogenetic analysis*

A multiple sequence alignment of pine DTPS deduced proteins was performed by ClustalX version 1.83 (Thompson, 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 MEGA X software (Kumar *et al.* 2018). 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 modelled with a gamma distribution (shape parameter = 1). Reliability of the tree obtained was tested using bootstrapping with 1000 replicates.

➢ *2.2.10. Gene expression analysis*

The expression patterns of the isolated *P. nigra* subsp. *laricio* DTPS sequences were analyzed in the five tissue types considered by quantitative real time (qRT-PCR).

As the reference genes for expression analysis, we looked at those showing stable expression in different pine tissues in the presence of stress conditions of different origin (Chen *et al.* 2016; Mo *et al.* 2019). The reference genes which were selected encode the following proteins: Actin 1 (*ACT1*, NCBI accession no KM496527), Cyclophilin (*CYP*, KM496534), Tubulin alpha (*TUBα*, KM496535), Polyubiquitin 4 (*UBI4*, KM496539), and uncharacterized protein LOC103705956 (*upLOC*, MN172175).

Quantitative RT-PCR analysis was performed using the AriaMX real-time PCR system, with the Fast Q-PCR Master Mix (SMOBIO, Hsinchu, Taiwan), according to the manufacturer's protocol. Each reaction was run in a 20 μL final volume containing 1 μL of cDNA, and 150 nM forward and reverse primers. No template and RT-minus controls were run to detect contamination, dimer formation or the presence of genomic DNA. Specific primer pairs were designed both for the target and the selected reference genes, using the Beacon Designer 6 software (Stratagene, La Jolla, CA), and the following stringency criteria: Tm of 55 °C ± 2 °C; PCR amplicon length between 60 and 200 bp; primer length of 21 ± 3 nt; and 40% to 60% guanine-cytosine content. Primers were also designed at the 3′ end of each sequence, to encompass all potential splice variants and ensure equal RT efficiencies (Table 2.1) Only primer pairs generating a sharp peak by melting curve analysis (without unspecific products or primer-dimer artifacts) and showing efficiencies between 90 and 110% and R^2 values (coefficient of determination) calculated for standard curves higher than 0.995 were selected for expression analysis of the target and references genes.

Table 2.1. *List of primer pairs of Calabrian pine DTPS and reference (CYP and upLOC) genes used in qRT-PCR analyses*

Gene	Forward primer (5'-3')	Reverse primer (5'-3')
DTPS1	CGACACTTGCAGGAGATTG	TATGTGAAAGTGTCAAACCGTC
DTPS2	ACAGTACGTGCCGAACAA	AACCGGTTCGAAGAGGAC
DTPS3	AAAGATGTGCCGGAGAAGT	AGGATCTTCTTGACGTGTTGT
DTPS4	ATGCGAGATCAATGCAGC	ACGCAACTGGTTCGAACA
CYP	TGTAGAGGGCTTGGAGGTC	CAAGCGAGCTGTCCAGAGT
upLOC	GGTTTGCTTTGGAGGATATG	GTCCAATGTGCACCTCGT

Standard curves based on five-points, corresponding to a fivefold dilution series (1:1–1:243) from pooled cDNA, were used to compute the PCR efficiency of each primer pair. The PCR efficiency (E) was derived by the eq. $E = (10[-1/m] - 1) \times 100$, where m is the slope of the linear regression model fitted over log-transformed data of the input cDNA concentration versus C_t values, according to the

linear equation y = m × log(x) + b. The thermal profile comprised three segments: 95 °C for 2 min; 40 cycles of 15 s denaturation at 95 °C, 1 min annealing at 56 °C and the dissociation curve, consisting of 1 min incubation at 95 °C, 30 s incubation at 60 °C and a ramp up to 95 °C. Three biological replicates, resulting from three different RNA extractions, were used in the quantification analysis. Three technical replicates were analysed for each biological replicate.

Raw C_t values were transformed to relative quantities by using the delta-C_t formula $Q = E^{\Delta Ct}$, where E is the efficiency of the primer pair used in the amplification of a specific gene (100% = 2), and ΔC_t is the difference between the sample with the lowest C_t (highest expression) from the dataset and the C_t value of the sample in question. For the the comparison of the relative expression levels of the isolated DTPSs in all tissue types the formula used to convert C_t values into relative quantity was $Q=2^{\Delta Ct}$. This assumption was justified by the fact that the amplification efficiencies of the considered genes were approximately the same, ranging from 96 to 100%.

The expression stability of the candidate reference genes was evaluated by the software program NormFinder (Andresen *et al.* 2004), as described in Paolacci *et al.* (2017). The best two-genes combination proposed by NormFinder was *CYP* + *upLOC*, with a stability score considerably lower than that of the most stable single reference gene, namely *CYP*, among those considered. Therefore, the expression levels of the genes of interest were normalized against the geometric mean of the two reference genes, and their normalized relative values reported as mean value ±SD. Standard deviation values for normalized expression levels were calculated according to the geNorm user manual (geNorm manual, updated 8 July 2008).

➢ *2.2.11. Statistical analysis*

Each reported value for metabolites and gene expression levels represents the mean of a total of nine values, obtained from three biological replicates and three technical replicates for each biological replicate. The statistical significance of the differences was evaluated by one-way ANOVA followed by the Tukey's test.

2.3. RESULTS AND DISCUSSION

➢ *2.3.1. In the Pinaceae, the diterpene metabolites profiles are tissue-specific and species-specific*

The diversity of oleoresin diterpenoids and the extent of diterpenes oxidation were quali-quantitatively evaluated in five different Calabrian pine tissues, namely young- (YN) and mature (MN) needles, bark and xylem combined from leader- (LS) and interwhorl (IS) stems, and roots (R).

GC-MS analysis showed that diterpene resin acids (DRAs) are the most abundant diterpenoids across all the examined tissue types, together with remarkably lower amounts of the corresponding aldehydes and olefins (Fig. 2.2).

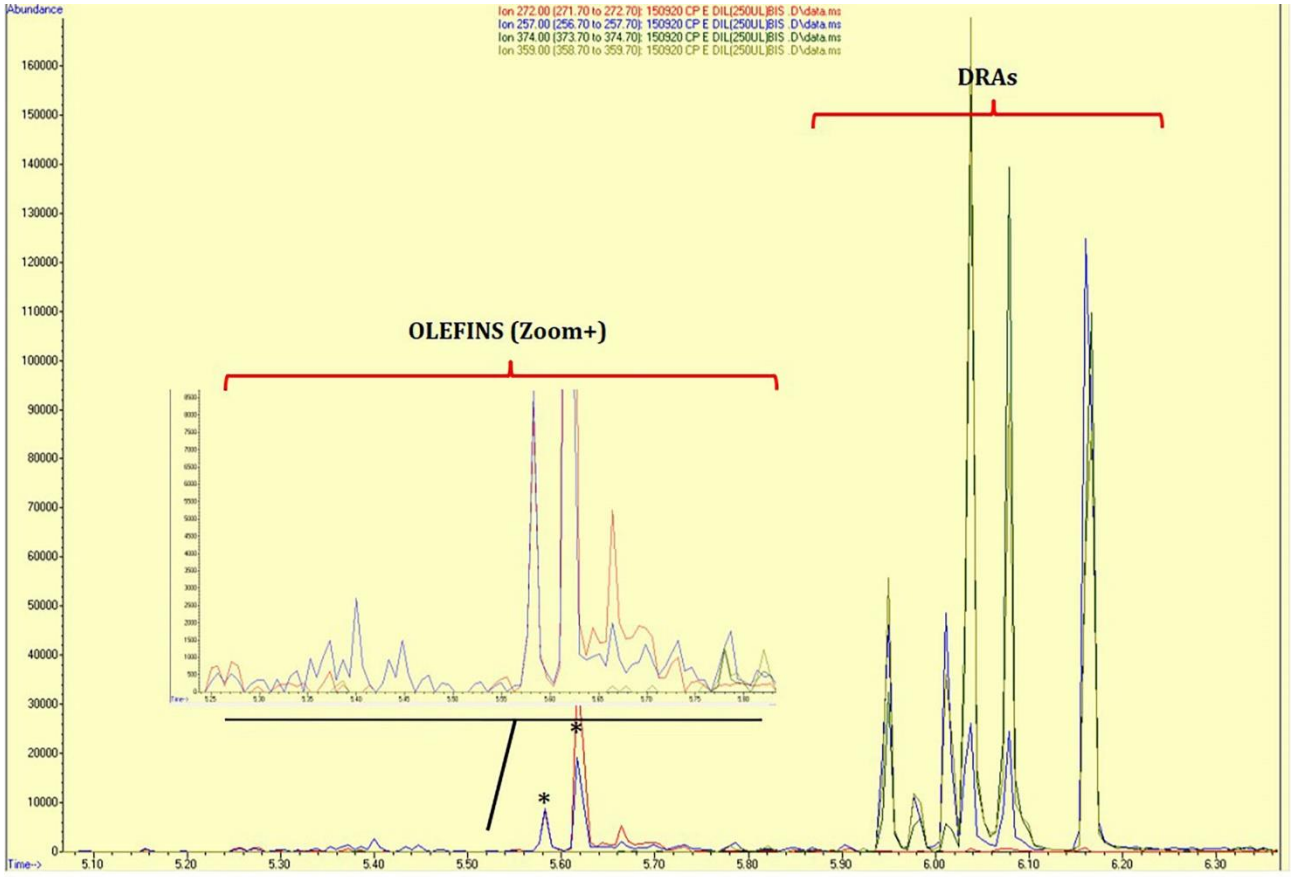


Fig. 2.2. *A representative example of the quantitative relationships among acidic (diterpene resin acids, DRAs) and neutral (olefins) components of the diterpenes extracted from Pinus nigra subsp. laricio (Calabrian pine) tissues, visualized by overlapping GC-MS ion chromatograms at selected m/z, i.e., 374/359 for DRA and 272/257 for olefins (magnified inset on the bottom left side of the item).*

Similar quantitative relationships among acidic and neutral diterpenoids were previously observed in various tissue types of other *Pinus* species, such as *P. banksiana* and *P. contorta* (Hall *et al.* 2013a), as well as *P. pinaster* and *P. radiata* (López-Goldar, 2020). Likewise, in Sitka spruce (*Picea sitchensis*) the DRA fraction in stem tissues accounted for more than 92% of the total diterpenoids (Hamberger *et al.* 2011). Because of their very low concentrations in all the tissues of Calabrian pine examined, olefins and aldehydes are described here only qualitatively, whereas the corresponding DRAs are quantitatively compared among each other in the different tissues (see below).

All the Calabrian pine tissues examined here showed the presence of the same nine DRAs, seven of which being non-dehydrogenated species - namely pimaric acid, sandaracopimaric acid, isopimaric acid, palustric acid, levopimaric acid, abietic acid, and neoabietic acid - and two being dehydrogenated ones, namely dehydroabietic acid and a non-identified putative dehydroisomer. This is exemplified by Fig. 2.3, showing the DRAs elution profile obtained from the LS tissue and by Figure 2.4 illustrating their mass spectra.

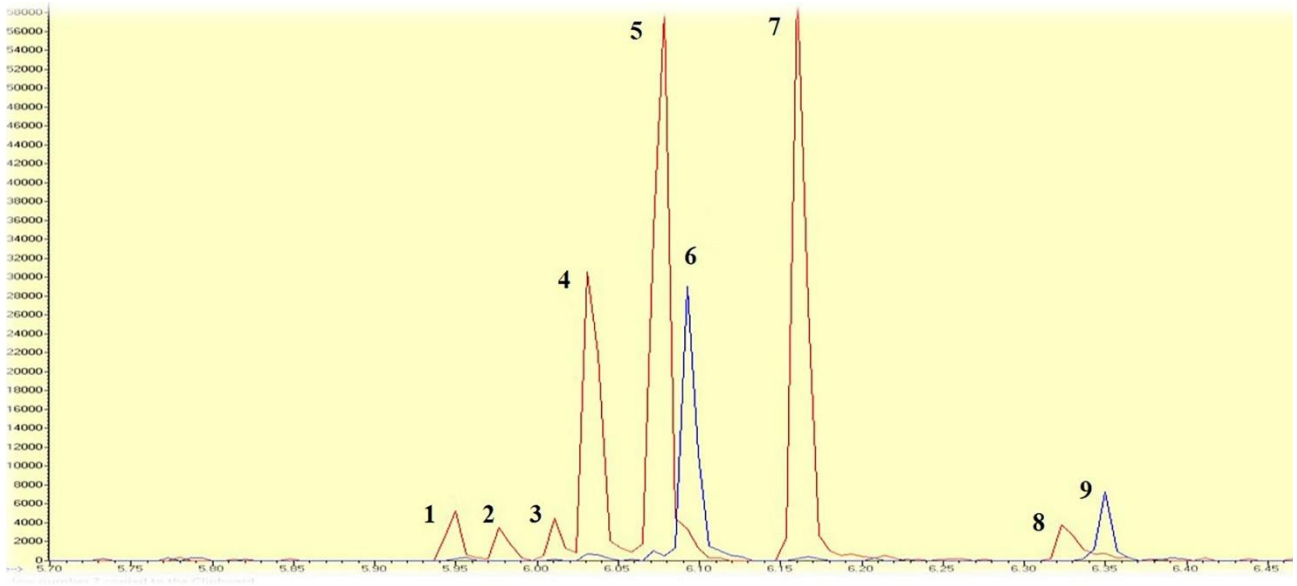


Fig. 2.3. *A representative GC-MS profile of the diterpene resin acids extracted from the leader stem of Calabrian pine. The single ion monitoring at m/z 374 (red line, non-dehydrogenated species) was overlapped with the single ion monitoring at m/z 372 (blue line, dehydrogenated species). (1), pimaric acid; (2), sandaracopimaric acid; (3), isopimaric acid; (4), palustric acid; (5), levopimaric acid; (6), dehydroabietic acid; (7), abietic acid; (8), neoabietic acid; (9), non-identified dehydroisomer.*

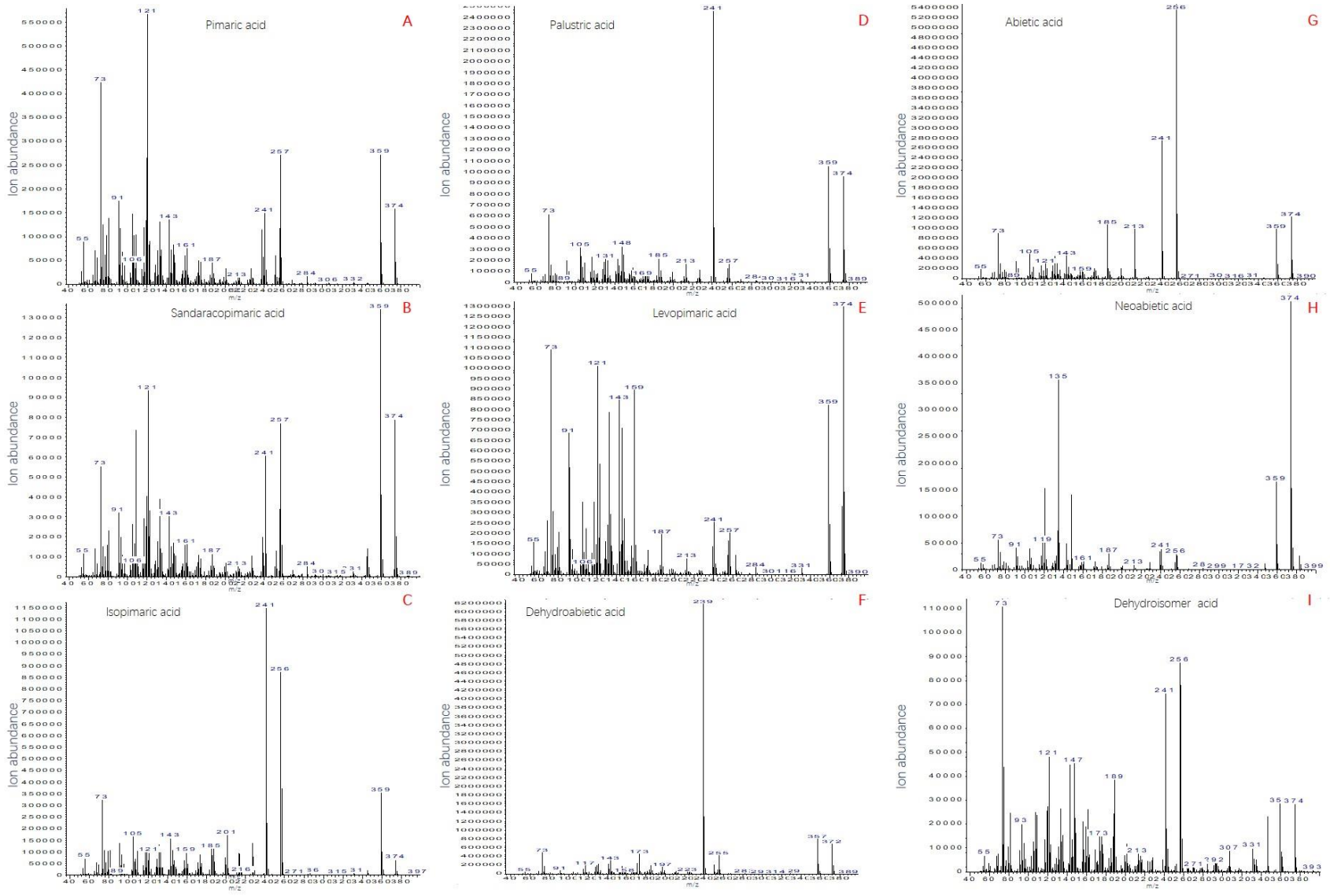


Fig 2.4. *Mass spectra of the nine diterpene resin acids identified in Pinus nigra subsp. laricio. Pimaric acid (A), Sandaracopimaric acid (B), Isopimaric acid (C), Palustric acid (D), Levopimaric acid (E), Dehydroabietic acid (F), Abietic acid (G), Neoabietic acid (H) and Dehydroisomer acid (I).*

Quantitatively speaking, Fig. 2.5 shows that the highest contents of total DRAs were found in the LS and IS tissues, with decreasing concentrations being observed in the R, MN and YN ones. Fig. 2.5 also shows the quantitative distribution of the nine DRAs in the different tissue examined: in both MN and YN, dehydroabietic-, isopimaric- and abietic acids were found to be the main components, while the other DRAs were detected at lower concentrations (1-6% of the total). This confirms the results obtained by López-Goldar *et al.* (2020) on the same tissues of *P. radiata* and *P. pinaster,* but not those reported by Hall *et al.* (2013a), who instead observed a prevalence of levopimaric and neoabietic acids in both young and mature needles from *P. contorta* and *P. banksiana.* In the LS tissue, abietic acid was the dominant DRA component (about the 33% of the total), followed by dehydroabietic- and palustric acids. On the other side, the IS tissue showed a prevalence of dehydroabietic- and palustric acids, each contributing about 30% of the total DRAs, followed by abietic acid. In both the stem tissues, namely LS and IS, comparatively lower abundances were observed for levopimaric-, isopimaric-, pimaric-, sandaracopimaric-, and neoabietic acids, as well as for the non-identified dehydroisomer. These results significantly differ from those reported by Hall *et al.* (2013a), who instead observed that levopimaric acid is the most abundant DRA in the LS and IS tissues from *P. contorta* and *P. banksiana*. Finally, dehydroabietic-, palustric- and abietic acids, although with significant differences in their amounts, were found to be the predominant DRAs of the R tissue, in which, compared to the aerial tissues, intermediate abundances of isopimaric- and levopimaric acids, as well as lower amounts of pimaric-, sandaracopimaric-, neoabietic acids, and of the non-identified dehydroisomer, were measured.

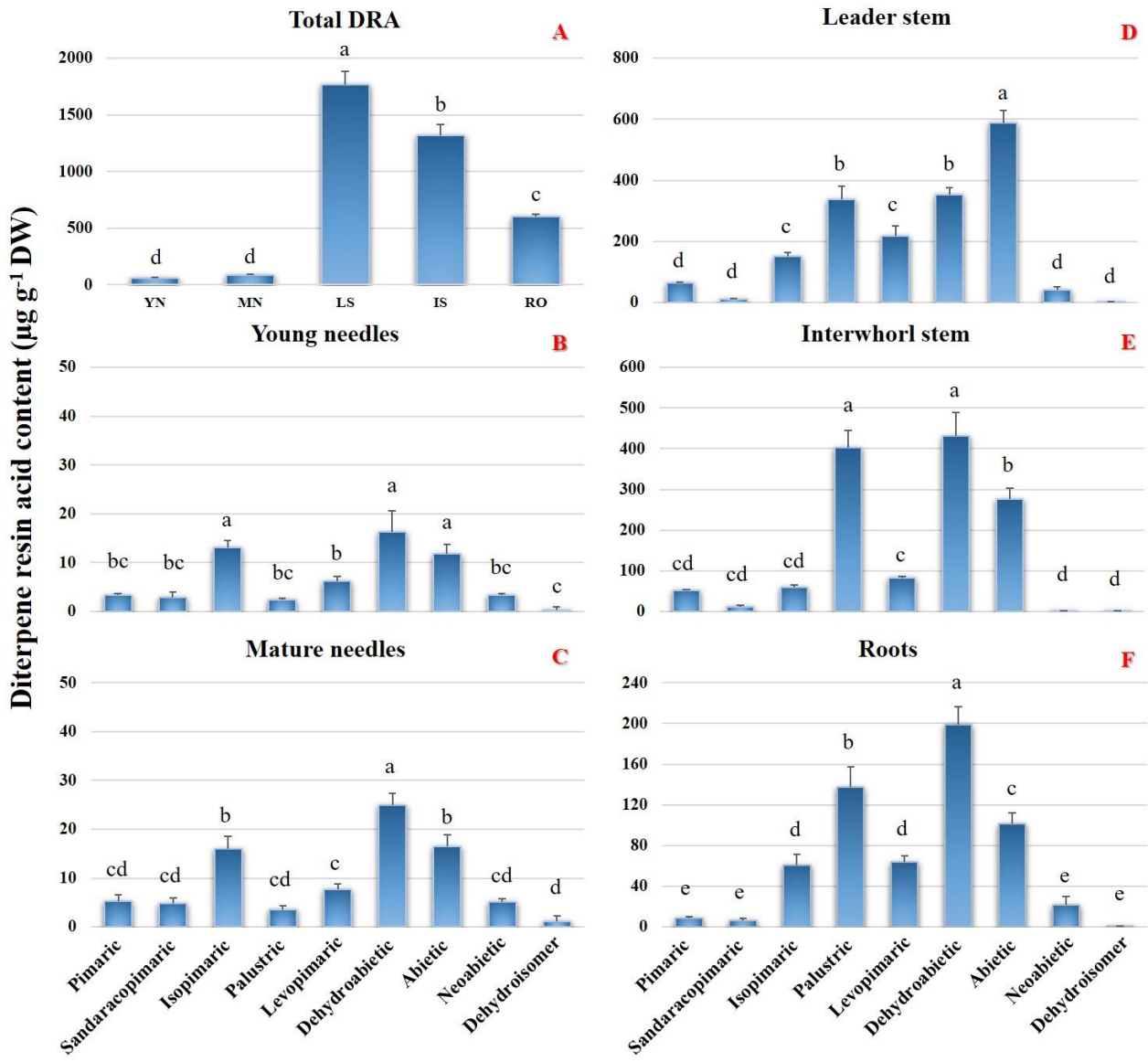


Fig. 2.5. *Total diterpene resin acids (DRAs, panel at the top) and levels of individual DRAs in different tissues of 3-year-old Pinus nigra subsp. laricio (Calabrian pine) saplings. Error bars indicate the standard deviation of the mean. The statistical significance of the differences was evaluated by one-way ANOVA, followed by the Tukey's test. Different letters denote statistical significance of the difference at $p < 0.01$. YN, young needle; MN, mature needle; LS, leader stem (LS); IS, interwhorl stem; R, root; DW, dry weight.*

Again, at variance with our results, Hall *et al.* (2013a) reported comparatively higher concentrations of palustric- and levopimaric acids in the roots of both *P. contorta* and *P. banksiana*. Taken together, the reported results could suggest that the DRAs fingerprint in *Pinus* spp. is not only tissue-specific, but also species-specific.

In conifers oleoresins, both due to their nature of precursors, and because of their higher volatility and tendency to undergo UV-induced photooxidation, olefins are normally found in lower concentration with respect to their oxygen-containing counterparts, i.e., DRAs. In agreement with such view, we detected in the Calabrian pine tissues only trace amounts of the neutral components of oleoresin, of which five olefins, namely sandaracopimaradiene, levopimaradiene, palustradiene,

abietadiene, and neoabietadiene, and five aldehydic derivatives, namely sandaracopimaradienal, palustradienal, isopimaradienal, abietadienal, and neoabietadienal (Fig. 2.6). Qualitatively speaking, the olefins and the corresponding aldehydes found in Calabrian pine tissues were the same found by Hall *et al.* (2013a) in the homologous tissues of *P. contorta* and *P. banksiana*, although at different relative concentrations.

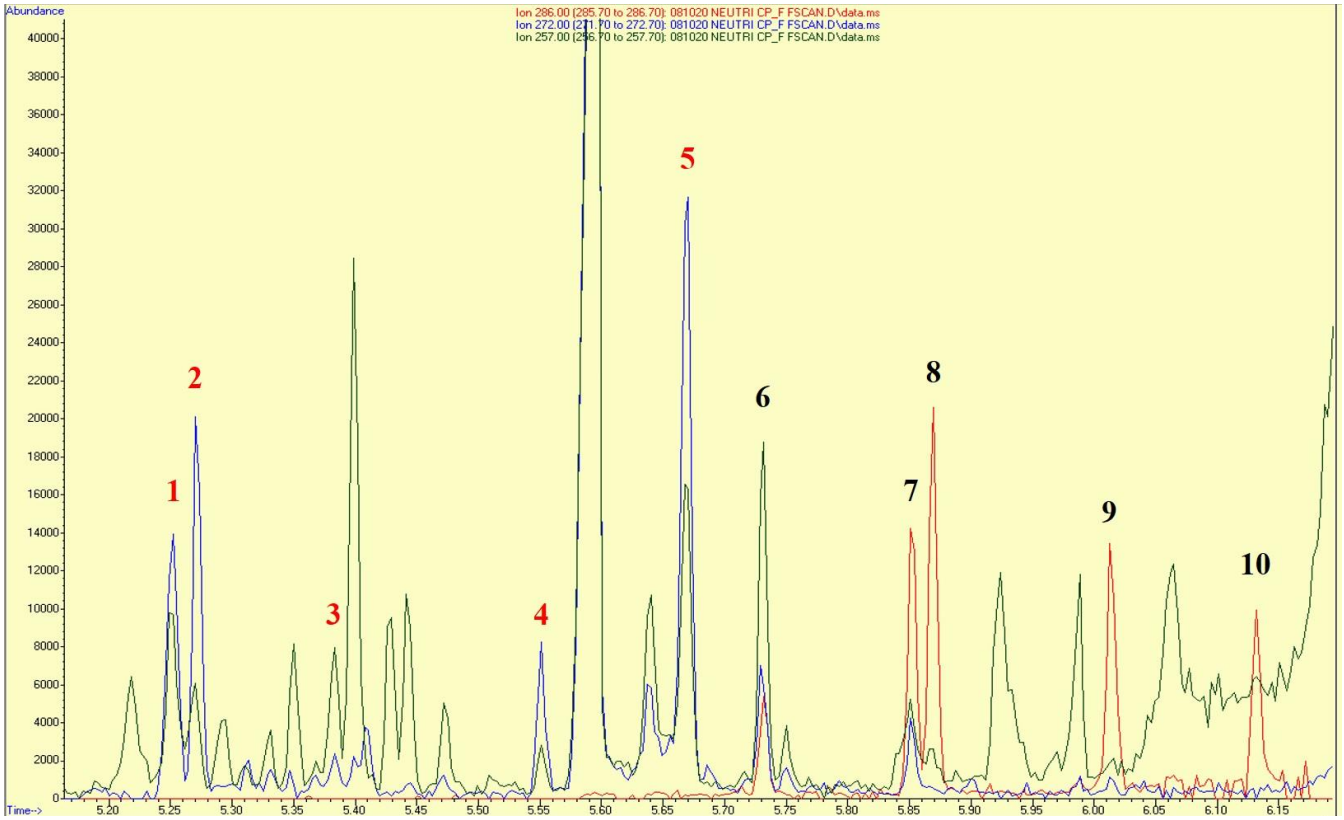


Fig. 2.6. *A representative GC-MS profile of the diterpene olefins and aldehydes in the extract obtained from leader stem tissue of Calabrian pine. The selected m/z 272 and 257 for diterpene olefins (red number above peaks) were overlapped with the selected m/z 286 for aldehydes (black number above peaks). (1) sandaracopimaradiene; (2) levopimaradiene; (3) palustradiene; (4) abietadiene; (5) neoabietadiene; (6) sandaracopimaradienal; (7) palustradienal; (8) isopimaradienal; (9) abieta-dienal; (10) neoabietadienal.*

➢ *2.3.2. A phylogeny-based approach for isolating partial and full-length cDNAs coding for diterpene synthases in Calabrian pine*

To gain insights into the structural diversity of diterpenoids in Calabrian pine, we isolated cDNA sequences encoding DTPSs potentially involved in the synthesis of the specialised diterpenes acting as DRAs precursors in such species. The strategy adopted was based on the PCR amplification of cDNA sequences by using specific primers designed on conserved regions of pine DTPSs belonging

to distinct phylogenetic groups, an approach we successfully used previously for the isolation of genes encoding monoterpene synthases in the same non-model conifer species (Alicandri *et al.* 2020).

In such previous work of ours (Alicandri *et al.* 2020), we carried out an extensive *in silico* search to identify all the putative full-length TPSs for primary and specialized metabolisms in different *Pinus* species, and to analyse their phylogenetic relationships. As far as DTPS are concerned, such database search allowed to identify 13 FL sequences involved in secondary diterpenoid metabolism in the *Pinus* species (Table 2.2).

Table 2.2. *Full length cDNA sequences identified in the National Center for Biotechnology Information (NCBI) database coding for putative diterpene synthases (DTPS) in the Pinus species. ORF, open reading frame; bp, base pair.*

Species	Function	Abbreviation	Accession mRNA sequence	ORF (bp)	Accession protein sequence	Predicted number of amino acids
Pinus banksiana	Levopimaradiene/Abietadiene synthase	Pb DTPS LAS1	JQ240312	2574	AFU73864	857
	Monofunctional diterpene synthase	Pb MDTPS1	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	Levopimaradiene/Abietadiene synthase	Pc DTPS LAS1	JQ240310	2574	AFU73862	857
	Levopimaradiene/Abietadiene synthase	Pc DTPS LAS2	JQ240311	2553	AFU73863	850
	Monofunctional diterpene synthase	Pc MDTPS1	JQ240318	2559	AFU73870	852
	Monofunctional diterpene synthase	Pc MDTPS2	JQ240319	2559	AFU73871	852
	Monofunctional diterpene synthase	Pc MDTPS3	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 densiflora	Abietadiene synthase	Pd DTPS ABS1	EU439295	2577	ACC54559	858
Pinus taeda	Diterpene synthase	Pt DTPS LAS1	AY779541	2553	AAX07435	850

Phylogenetic analysis clustered all the 13 pine DTPSs sequences into the TPS-d3 clade, which includes four well-supported major groups, denoted as 1-4. Each of these groups contains DTPS proteins from different pine species thought to be functionally related among each other (Alicandri *et al.* 2020).

Based on our previous phylogenetic analysis, in the present work the deduced amino acid- and the nucleotide sequences of pine DTPSs belonging to each of the groups 1-4 (Table 2.2) were aligned, to identify highly conserved regions among members of each of the four groups. Such conserved regions within each group were then used to design specific primers for the isolation by RT-PCR of partial transcripts of orthologous genes in Calabrian pine. Fig. 2.7 schematically outlines the FL cDNAs for representative members of the four phylogenetic DTPS groups, and the positions of the specific primers used, of which a complete list is reported in Table 2.3.

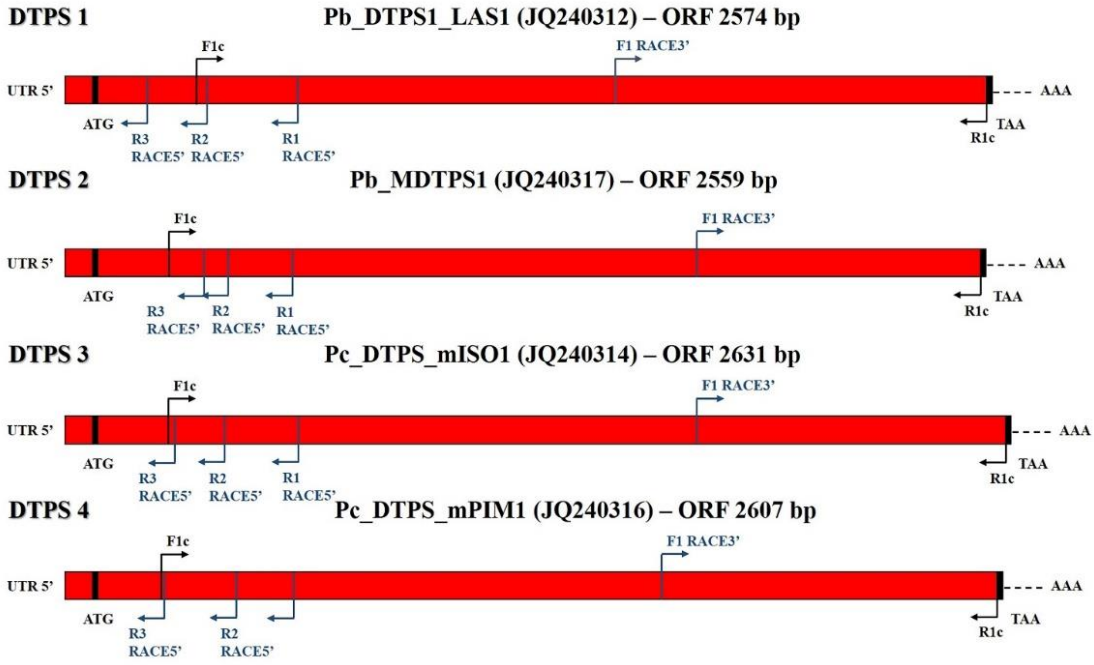


Fig. 2.7. *Schematic representation of the full length cDNAs of four representative diterpene synthase (DTPS) members of the phylogenetic d3 clade of genes, two from Pinus banksiana (Pb) and two from Pinus contorta (Pc) and the positions of their Forward and Reverse primers used in the present study for the isolation of the partial transcripts coding for the orthologous genes (F1c/R1c) in Calabrian pine. The position of the specific primers used for 5' and 3' RACE (Rapid Amplification of cDNA Ends) extensions of partial transcripts of Calabrian pine DTPS (groups 1-4) are indicated in blue. UTR, untranslated region.*

Table 2.3. *Forward and Reverse primers used for the isolation of cDNAs and genomic diterpene synthase sequences in Pinus nigra subsp. laricio. RACE, Rapid Amplification of cDNA Ends.*

	Phylogentic group	**Forward primers 5'- 3'**	**Reverse primers 5'- 3'**
Partial cDNA sequences	Group 1	F1c: CAAGGTTGCACCATCAGACG	R1c: GCCTCTTCCAGCCAGTTGC
	Group 2	F1c: ACGAAACGAGAATTCCCTGA	R1c: AACCGGTTCGAAGAGGAC
	Group 3	F1c: TTGGAGAAGGCGCAACCT	R1c: AGGGACTGGTTCGAATAGG
	Group 4	F1c: GTTCAAGCAACATCGTTGCC	R1c: ACGCAACTGGTTCGAACA
		RACE 5'	**RACE 3'**
Full length cDNA sequences by RACE	Group 1	R1: CTTCHCCCCAAGAGCCATC R2: CGTCTGATGGTGCAACCTTG R3: TGCCTGGTTATCTTTACCC	F1Race3': ATAGAAGTATGCCAAGGCTGG
	Group 2	R1: CTTCHCCCCAAGAGCCATC R2: AACAGCCGGAATCCTGG R3: TGACTCCACACGCTTCTC	F1Race3': AAGGTGGTGGAAATCATCG
	Group 3	R1: CTTCHCCCCAAGAGCCATC R2: GCTGCTGCCACCTTATTAGA R3: CCTTCTCCAACACCAGCA	F1Race3': AAGGTGGTGGAAATCATCG
	Group 4	R1: CTTCHCCCCAAGAGCCATC R2: CAGGCTTGTGAGATGACATT R3: TTGCTTGAACCTGATCGT	F1Race3': GACCGCAAATACTTGGAACTG
		Forward primers 5'- 3'	**Reverse primers 5'- 3'**
Genomic sequences	Group 1	F1g: ATGCCTTCCTCTTCATTG	R1g: GCCTCTTCCAGCCAGTTGC
	Group 2	F1g: ATGGCCATGCCCTCCTCT	R1g: AACCGGTTCGAAGAGGAC
	Group 3	F1g: ATGGCCATGCCTTCGTAC	R1g: AGGGACTGGTTCGAATAGG
	Group 4	F1g: ATGGCCATGCCTTTGTG	R1g: ACGCAACTGGTTCGAACA

By using such strategy, we were able to isolate and sequence partial DTPS transcripts of putative orthologous genes in Calabrian pine belonging to each of the groups 1-4, which confirms the validity of the phylogenetic approach used. These partial DTPS transcripts were then used as templates for isolating the corresponding FL cDNA sequences by means of 5' and 3' RACE (Rapid Amplification of cDNA Ends) extensions; the primer sequences of 5’ and 3’ RACE are reported in Table 2.3 and their positions indicated in Fig. 2.7.

In the cases of the partial DTPS transcripts belonging to groups 1 and 2, two slightly different sequences were recognized among the three clones analysed for each cDNA fragments, due to nucleotide substitutions, most of them synonymous, on a background otherwise showing high levels of sequence identity among each other (over 97%). These slightly different DTPS transcripts might derive from alleles of the same gene and/or from duplicated copies of the same gene, and this would imply that we might have as many more DTPS closely related genes belonging to each phylogenetic group in Calabrian pine, as observed in other *Pinus* species (Hall *et al.* 2013a). This possibility will be tested in future studies. However, among the three sequenced clones for the corresponding 3’ and 5’ RACE products we identified the same sequences that were identical to the 3’ and 5’ ends of two of the three sequenced cDNA products, indicating that they are part of the same FL transcript. Therefore, the assembled four unique FL cDNAs isolated from Calabrian pine, denoted as *Pnl DTPS1*, *Pnl DTPS2*, *Pnl DTPS3*, and *Pnl DTPS4*, each of them belonging to one of the four groups of the TPS-d3 clade, contained open reading frames (ORFs) of 2574, 2559, 2631 and 2607 bp, respectively, and were predicted to encode proteins of 857, 852, 876 and 868 aa, respectively (Fig. 2.8).

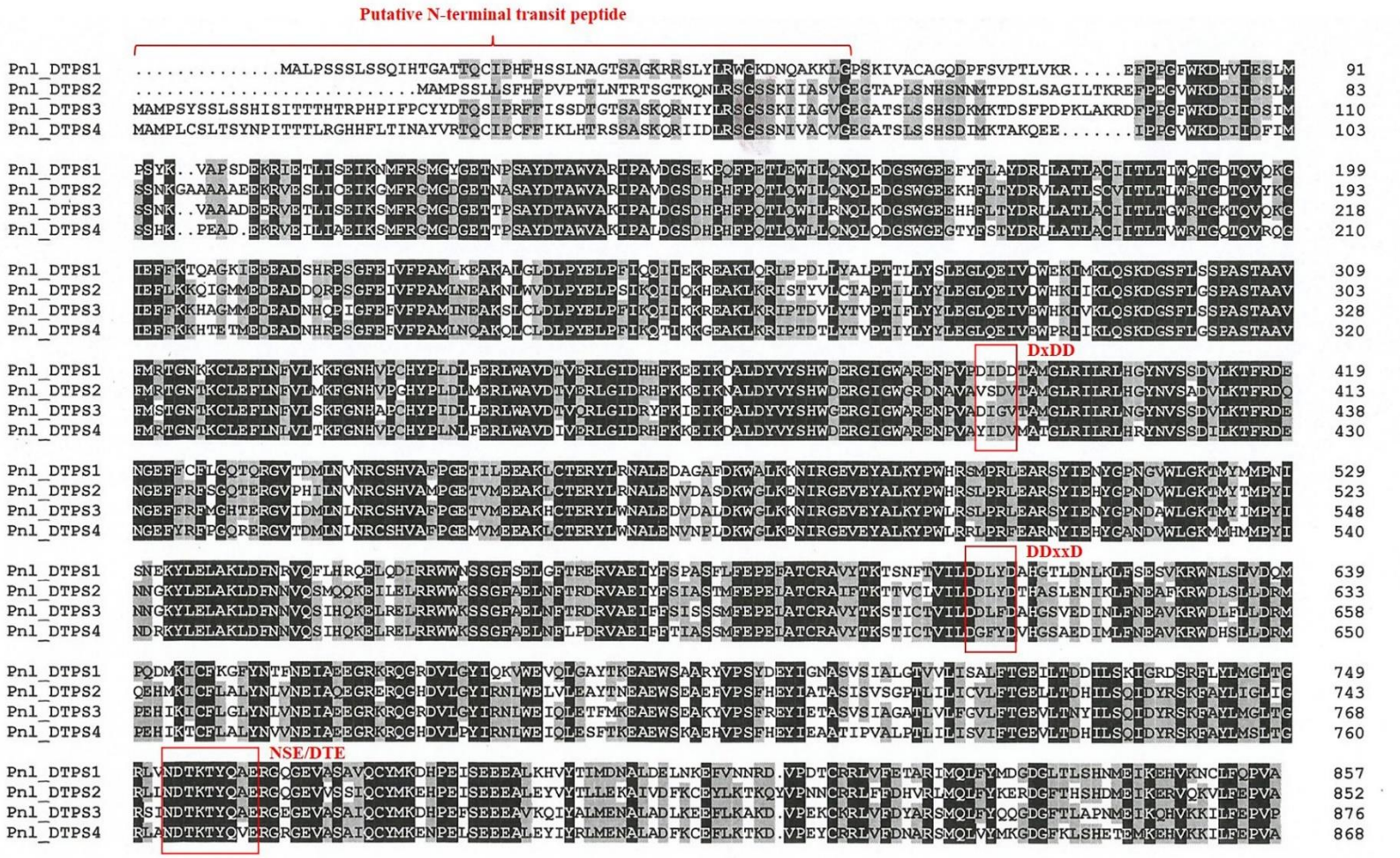


Fig. 2.8. *Alignment of deduced amino acid sequences of the four putative diterpene synthases from Calabrian pine (Pnl DTPS 1-4) isolated in the present study. Amino acid residues with black background indicate highly conserved regions, while amino acid residues which are identical in more than 50% of the proteins are in grey background. The DTPS class II (DxDD) and class I (DDxxD, NSE/DTE) signature motifs are indicated.*

➢ *2.3.3. Sequence-based analysis predicts that both monofunctional and bifunctional diterpene synthases are involved in the biosynthesis of diterpene resin acids in Calabrian pine*

The deduced amino acid sequences of the four full-length cDNAs isolated from Calabrian pine (see above) were found to contain highly conserved and characteristics regions of plant DTPSs (Fig. 2.8). First, a putative transit peptide, ranging from 33 (Pnl DTPS1) to 68 aa (Pnl DTPS4) in length, probably needed for the import of the mature DTPS proteins into plastids. Secondly, DTPS active-site signature motifs (Fig. 2.8): Pnl DTPS1 was found to contain both class-II and class-I motifs, suggesting its nature of proper bi-I/II DTPS, alike the already known bifunctional DTPSs involved in DRAs biosynthesis in conifers, namely the isopimaradiene synthase-type (ISO) and levopimaradiene/abietadiene synthase-type (LAS) enzymes from grand fir (*A. grandis*) and balsam fir (*A. balsamea*) (Zerbe *et al.* 2012; Stofer Vogel *et al.* 2004), Norway spruce (*Picea abies*) and Sitka spruce (*P. sitchensis*) (Keeling *et al.* 2011; Martin *et al.* 2004), loblolly pine (*P. taeda*), lodgepole pine (*P. contorta)* and jack pine (*P. banksiana*) (Hall *et al.* 2013a; Ro and Bohlmann, 2006). All the three-remaining putative DTPS isolated from Calabrian pine, instead, were found to contain only the class-I signature motifs, plus incomplete versions of the class-II one, lacking D residues known to be critical for class-II catalysis (Peters and Croteau, 2002) either in the middle (Pnl DTPS3) or in the first and last positions (Pnl DTPS2 and Pnl DTPS4). Although representing putative monofunctional DTPSs, the three sequences only showed 33% to 34% protein sequence identity to the conifer monofunctional class II *ent*-copalyl diphosphates synthases and class I *ent*-kaurene synthases involved in GA metabolism (data not shown), suggesting their roles in specialized as opposed to general metabolism.

A phylogenetic analysis including the four deduced amino acid sequences from Calabrian pine and all the pine DTPSs identified in the NCBI database (Fig. 2.9), allowed to locate the isolated predicted proteins in the four phylogenetic groups in which the *Pinus* members of the TPS-d3 clade can be divided (Alicandri *et al.* 2020).

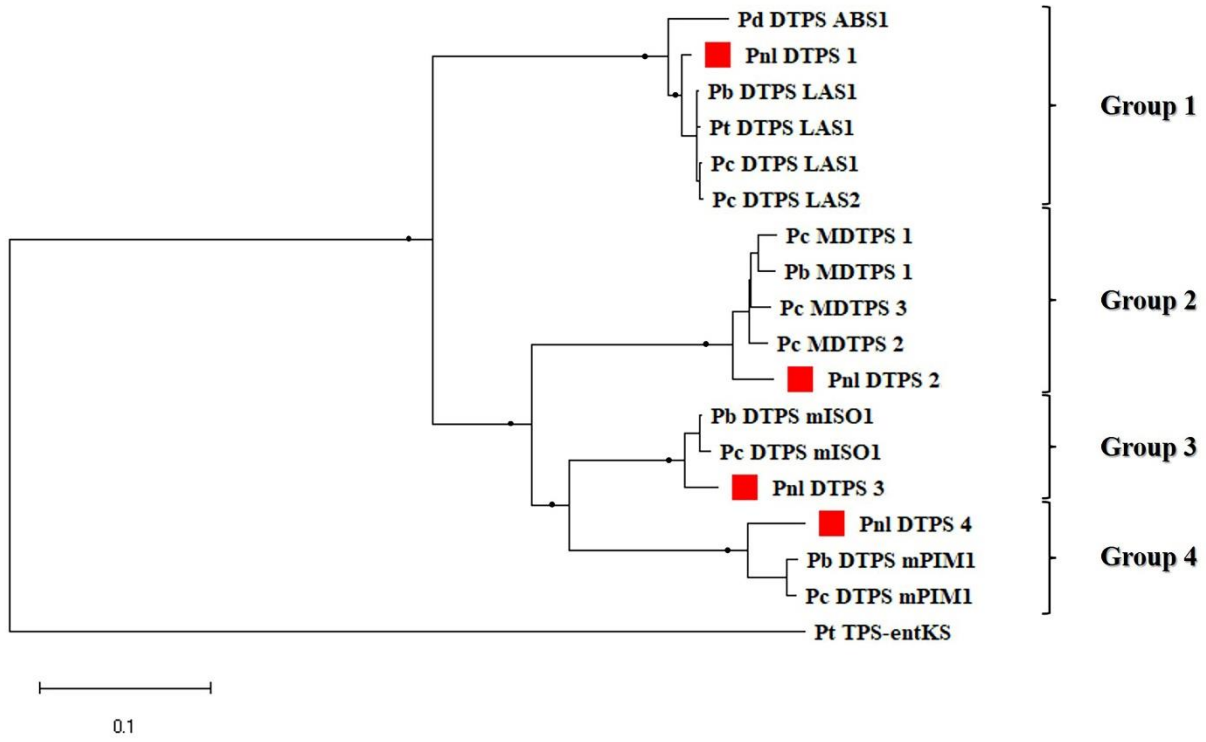


Fig. 2.9. *Phylogenetic tree of the deduced amino acid sequences of 13 diterpene synthases (DTPSs) identified in different Pinus species (Table 2.2) and the four DTPSs from Calabrian pine isolated in the present study (red squares). The ent-kaurene synthase from Physcomitrella patens (Pt TPS-entKS, BAF61135) was used to root the tree. Branches marked with dots represent bootstrap support more than 80% (1000 repetitions). The four phylogenetic groups identified in the pine members of the d3 clade of terpene synthases are indicated by square brackets.*

Based on the sequence relatedness with the previously characterized pine DTPSs, it was possible to predict the potential functions of three out of four DTPSs isolated from Calabrian pine. Pnl DTPS1 was found to cluster in group 1 with other five bi-I/II-DTPSs, showing 98-99% aa sequence identity with the four of them which have been functionally characterized so far, namely Pc DTPS LAS1, Pc DTPS LAS2, Pb DTPS LAS1 and Pt DTPS LAS1. Of these, the three bifunctional DTPSs from *P. banksiana* and *P. contorta* were shown to produce the diterpene alcohol 13-hydroxy-8 (14)-abietene (Hall *et al.* 2013a). This unstable allylic alcohol can undergo dehydration resulting in the formation of abietadiene, neoabietadiene, palustradiene and levopimaradiene, consistent with the GC–MS results previously obtained for Pt DTPS LAS from *P. taeda* (Ro and Bohlmann, 2006). Based on such sequence similarity, Pnl DTPS1 could be predicted to be involved in the synthesis of abietane-type diterpene olefins. Interestingly, however, when aligned with the other group-1 DTPSs (Fig. 2.10), Pnl DTPS1 from Calabrian pine revealed distinctive amino acids substitutions, namely D/G-515, G/E-565, and D/N-632, which could lead to a change in the protein structure and hence in its product(s) profile.

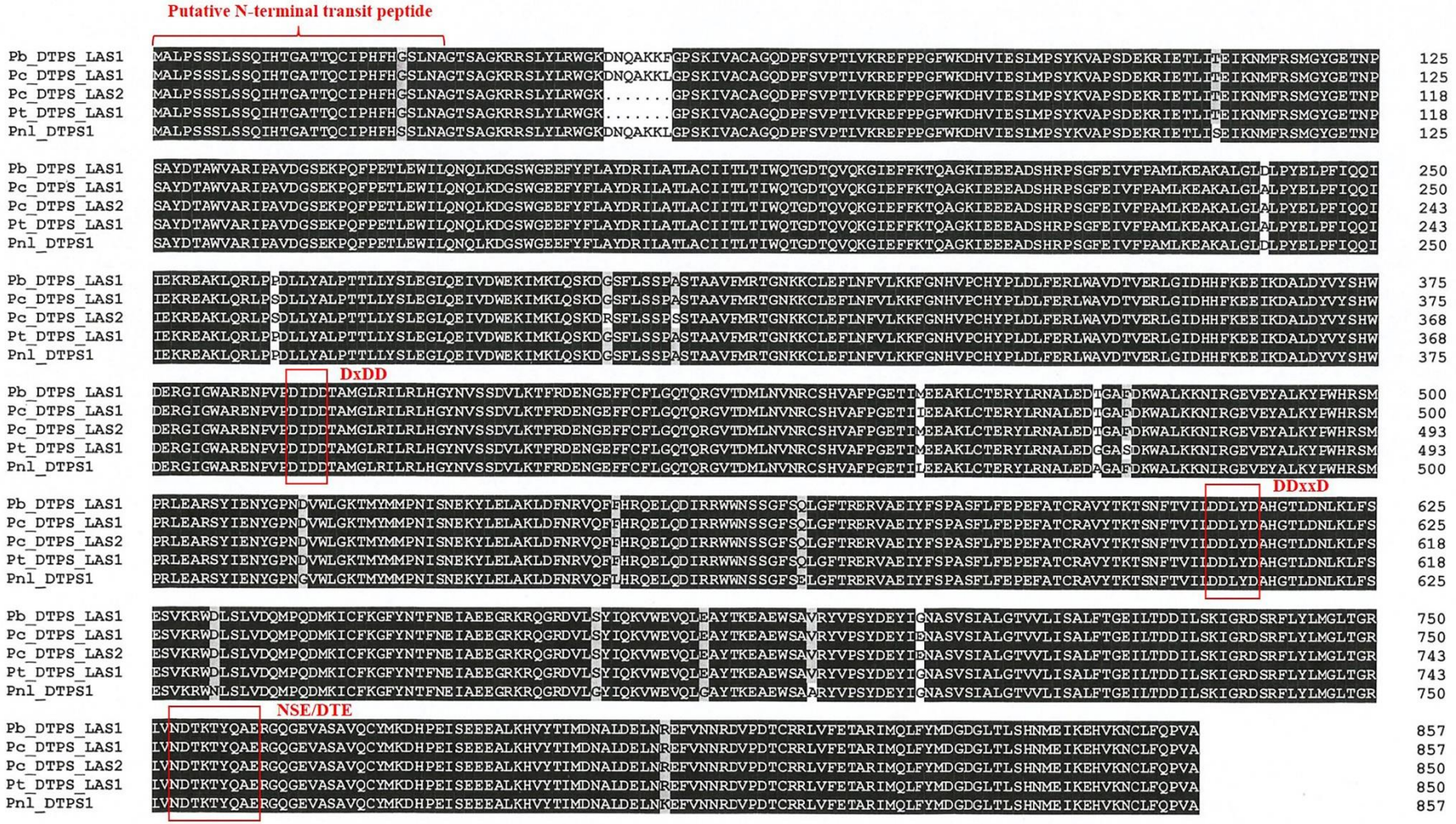


Fig. 2.10. *Alignment of deduced amino acid sequences of diterpene synthase (DTPS) belonging to the phylogenetic group 1. Amino acid residues with black background indicate highly conserved regions, while amino acid residues which are identical in more than 50% of the proteins are in grey background. The DTPS class II (DxDD) and class I (DDxxD, NSE/DTE) signature motifs are indicated. Pb, Pinus banksiana; Pc, Pinus contorta; Pt, Pinus taeda; Pnl, Pinus nigra subsp. laricio.*

The Pnl DTPS2 was found to be closely related to four mono-I DTPSs belonging to the phylogenetic group 2 (Fig. 2.9), for which Hall *et al.* (2013a) observed no biochemical activity. All of these proteins, though very similar among each other (95% to 98% protein sequence identity), show a low identity both with the above five putative bi-I/II DTPSs from *Pinus* species (64-65%), and with the other identified pine mono-I DTPSs (73-76%) (Table 2.4). Although the four mono-DTPS from *P. contorta* and *P. banksiana* contain the class-I signature motif, and their homology modelling (Koksal *et al.* 2011) predicts that they possess a conserved γβα-domain folding pattern (Hall *et al.* 2013a), the presence of unique structural features near their active sites, conserved also in the Pnl DTPS2 from Calabrian pine (Fig. 2.11), could explain their absence of function. In such respect, it was proposed that, in these group-2 DTPSs, the side chains of F-592, located upstream of the class I motif, and likewise those of F-814 and H-817, can protrude into the active site cavity and may cause a steric hindrance, possibly impeding catalytic activity (Hall *et al.* 2013a). It has been therefore speculated that these enzymes may have evolved from functional DTPSs into a trough of no function, from where they may evolve toward new DTPS activities or simply represent dead-end mutations of functional DTPSs (Hall *et al.* 2013a).

Table 2.4. *Amino acid sequence identity matrix of DTPS candidate genes from P. laricio (in red) with previously characterized DTPSs from P. taeda (Pt), P. contorta (Pc) and P. banksiana (Pb).*

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1) Pnl_DTPS1	100															
2) Pb_DTPS_LAS1	98.5	100														
3) Pc_DTPS_LAS1	98.2	99.4	100													
4) Pc_DTPS_LAS2	98.0	99.4	99.6	100												
5) Pt_DTPS_LAS1	98.4	99.6	99.4	99.3	100											
6) Pnl_DTPS2	64.3	64.5	64.2	64.6	65.0	100										
7) Pc_MDTPS_1	64.2	64.3	64.1	64.5	64.8	95.2	100									
8) Pc_MDTPS_2	64.6	64.7	64.5	64.9	65.1	95.4	96.8	100								
9) Pc_MDTPS_3	64.7	64.8	64.6	65.0	65.2	95.3	96.8	97.5	100							
10) Pb_MDTPS_1	64.1	64.2	64.0	64.4	64.6	95.4	97.8	97.7	97.7	100						
11) Pnl_DTPS3	66.6	66.6	66.5	66.9	67.0	75.0	74.9	74.8	74.7	74.3	100					
12) Pb_DTPS_mISO1	67.4	67.4	67.3	67.7	67.8	75.7	75.9	75.9	75.5	75.4	97.0	100				
13) Pc_DTPS_mISO1	67.1	67.1	66.9	67.4	67.5	75.3	75.3	75.3	74.9	74.8	96.5	99.2	100			
14) Pnl_DTPS4	63.0	63.2	63.1	63.4	63.5	74.2	74.0	74.3	73.6	73.7	79.2	79.7	79.2	100		
15) Pb_DTPS_mPIM1	63.5	63.8	63.7	64.0	64.1	73.4	73.4	73.7	73.1	73.0	79.8	80.0	79.5	93.8	100	
16) Pc_DTPS_mPIM1	63.5	63.7	63.6	63.9	64.0	73.8	73.8	74.2	73.5	73.5	80.0	80.3	79.7	93.8	98.6	100

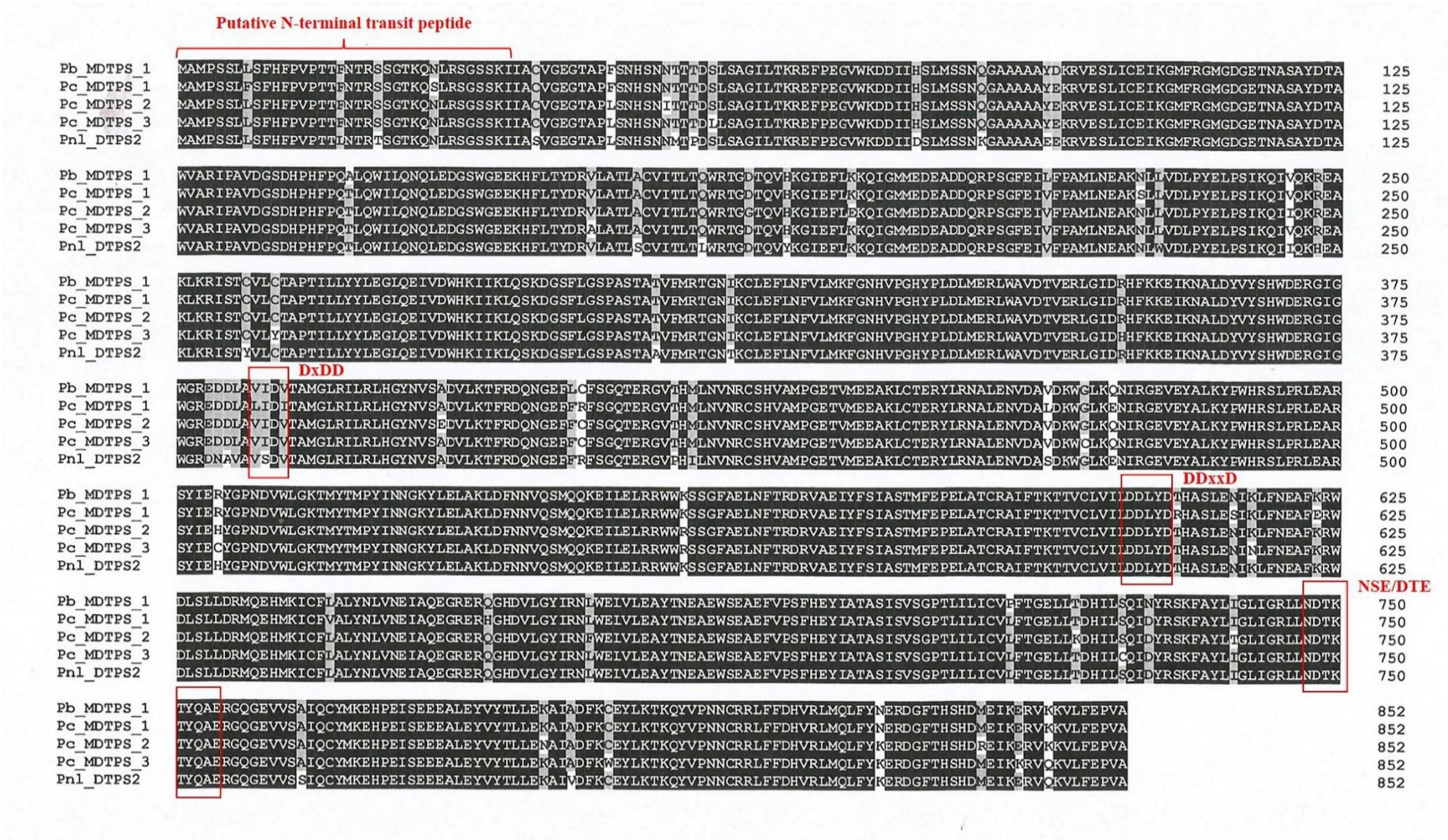


Fig. 2.11. *Alignment of deduced amino acid sequences of diterpene synthase (DTPS) belonging to the phylogenetic group 2. Amino acid residues with black background indicate highly conserved regions, while amino acid residues which are identical in more than 50% of the proteins are in grey background. The DTPS class II (DxDD) and class I (DDxxD, NSE/DTE) signature motifs are indicated.*

Based on sequence similarity (Fig. 2.9), and diverging from Pnl DTPS1, Pnl DTPS3 and Pnl DTPS4 were predicted to produce pimarane-type olefins, namely pimaradiene, sandaracopimaradiene, and isopimaradiene. In particular, Pnl DTPS3 was found to cluster in the phylogenetic group 3, together with one protein from *P. contorta* (Pc DTPS mISO1) and one from *P. banksiana* (Pb DTPS mISO1) (Fig. 2.9), both of which were found to produce isopimaradiene as the main product, whit small amounts of sandaracopimaradiene (Hall *et al.* 2013a). These members of such group, showing 96% to 99% protein sequence identity among each other, were found to be more similar to the mono-I DTPSs from the phylogenetic group 4 (79-80%) than to those of phylogenetic group 2 (74-76%; Table 2.4). Also for the group-3 DTPS, as noticed above for the group-1 ones, sequences alignment revealed amino acid substitutions exclusively present in the Pnl DTPS3 from Calabrian pine, namely K/N-642, D/N-748, and H/Y-749 (Fig. 2.12), which could lead to a change in the protein structure and hence in its product(s) profile. Likewise, Pnl DTPS4 was found to cluster in the phylogenetic group 4 (Fig. 2.9), together with two previously described mono-I DTPS, one from *P. banksiana* (Pb DTPS mPIM1) and one from *P. contorta* (Pc DTPS mPIM1), both of which were functionally characterised as forming pimaradiene as their major product (Hall *et al.* 2013a). Despite of the pronounced sequence identity among the group-4 predicted proteins (about 94%; Table 2.4), the high number of amino acid substitutions found in the Pnl DTPS4, compared to the other two DTPSs (Fig. 2.13), suggests that only its functional characterization might elucidate its specific catalytic competence.

Although we tried to predict the potential functions of Calabrian pine DTPSs based on sequence relatedness, it has to be mentioned that examples of apparent lack of structure-function correlation have been observed in the plants' TPS family. Hall *et al.* (Hall *et al.* 2013b), for instance, reported that conifer monoterpene synthases sharing 80-90 % aa identity among each other can catalyse biochemically distinct reactions, while others sharing only 50-60% protein identity among each other can form the same product. For this reason, a functional characterization, consisting of heterologous expression in bacterial systems and testing of the recombinant enzymes with their potential terpenoids substrates, would be essential to elucidate the actual functions of Calabrian pine DTPSs.

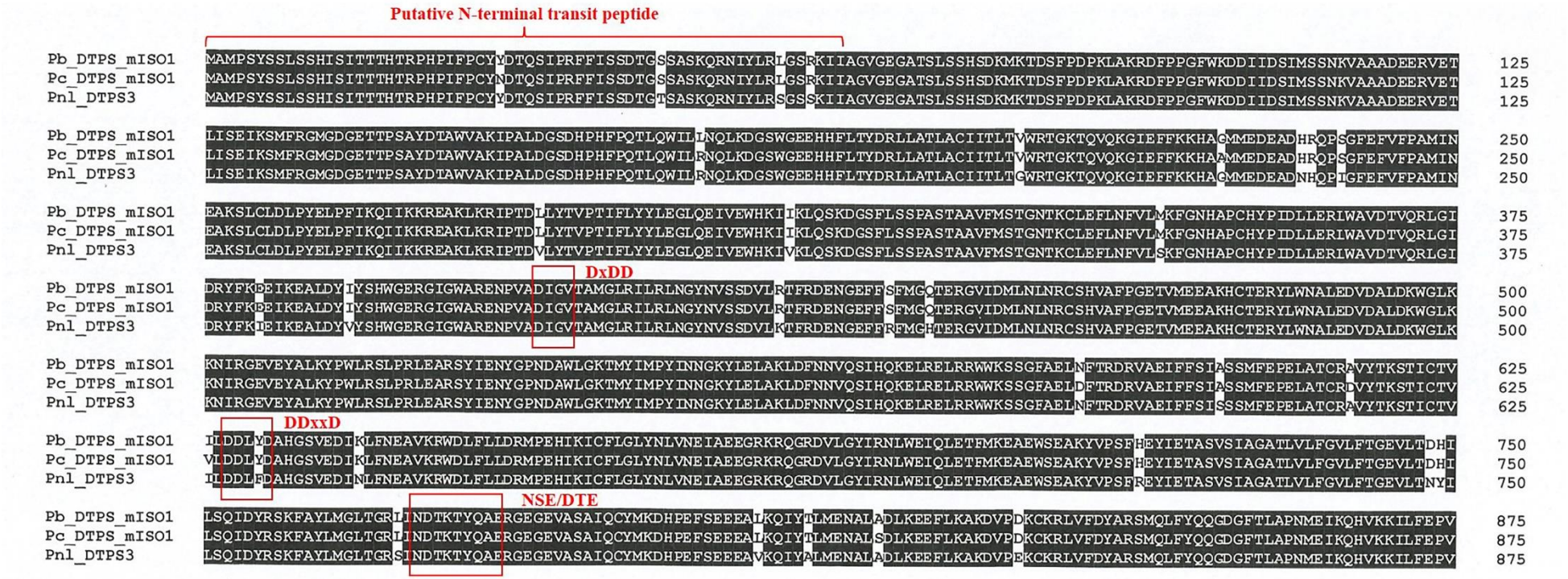


Fig. 2.12. *Alignment of deduced amino acid sequences of diterpene synthase (DTPS) belonging to the phylogenetic group 3. Amino acid residues with black background indicate highly conserved regions, while amino acid residues which are identical in more than 50% of the proteins are in grey background. The DTPS class II (DxDD) and class I (DDxxD, NSE/DTE) signature motifs are indicated.*

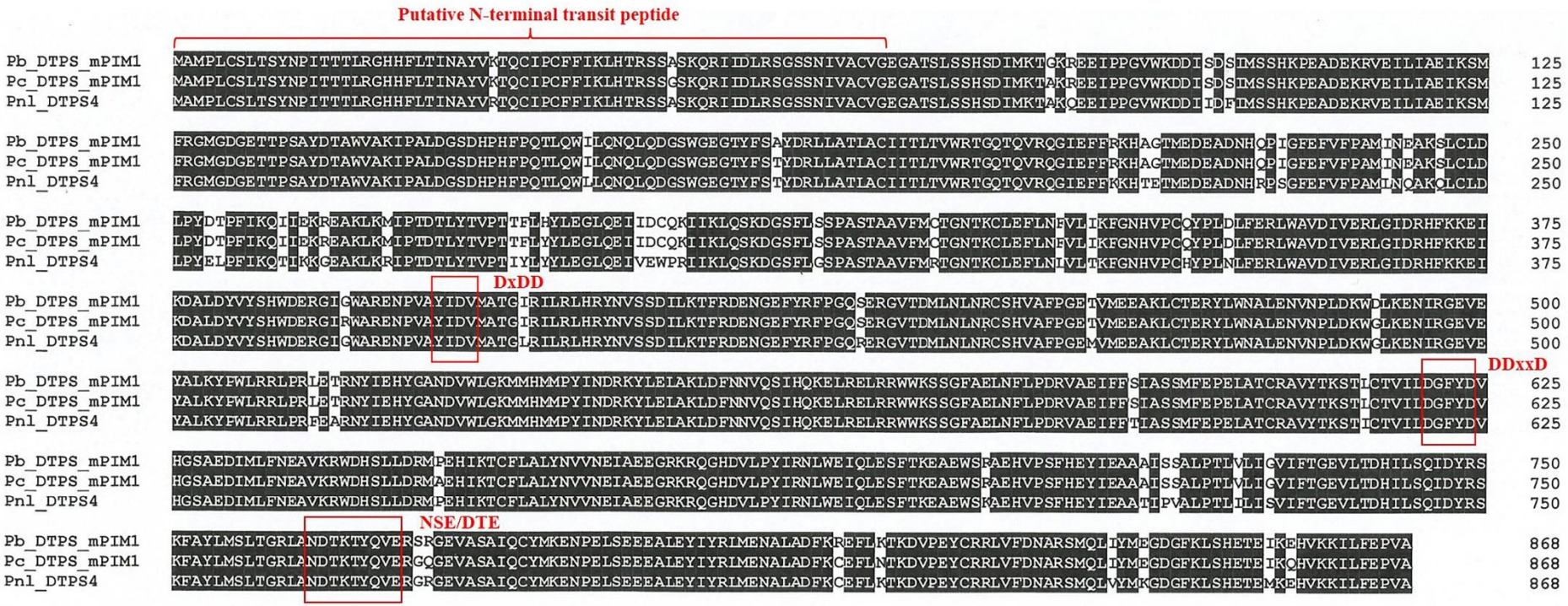


Fig. 2.13. *Alignment of deduced amino acid sequences of diterpene synthase (DTPS) belonging to the phylogenetic group 4. Amino acid residues with black background indicate highly conserved regions, while amino acid residues which are identical in more than 50% of the proteins are in grey background. The DTPS class II (DxDD) and class I (DDxxD, NSE/DTE) signature motifs are indicated.*

➢ *2.3.4. Genomic organization of diterpene synthases in Calabrian pine on the background of DTPS functional evolution*

The genomic sequences encompassing the ORFs of the four Calabrian pine DTPS genes isolated in the present study are schematically shown in Fig. 2.14. The alignment of each genomic sequence with its corresponding cDNA revealed an almost perfect matching among the latter and the exonic regions of the former, thus allowing a reliable determination the exon/intron structure of each DTPS gene. *Pnl DTPS1* and *Pnl DTPS2* were found to contain 16 exons and 15 introns, whereas 15 exons and 14 introns were found in the *Pnl DTPS3* and *Pnl DTPS4* sequences (Fig. 2.14). Apart from the 5' end, which showed considerable variability in terms of gene structure and sequences, the four DTPS genes from Calabrian pine were found to exhibit a high level of conservation of their genomic structural features, in terms of introns location, exons number and size, and position of the class-I active site functional motif (Fig. 2.14). Obvious patterns of intron sizes and sequences were not detected, although there was a strong conservation of their position along the genomic sequences (introns IV to XV in *Pnl DTPS1* and *Pnl DTPS2* and introns III to XIV in *Pnl DTPS3* and *Pnl DTPS4*; Fig. 2.14). The intron sizes were found to be generally small (about 50-200 nt), although some large introns (more than 300 nt) were also detected (Fig. 2.14). In addition, these introns were AT-rich, with repetitive sequences rich in T (3-10 mers; data not shown). All the four Calabrian pine DTPS genes were found to contain intron-exon junctions, which, with few exceptions, followed the GT/AG boundary rules (data not shown) (Brown and Simpson, 1998). Moreover, the phasing of the introns insertion, defined as the placement of intron before the first, second or third nucleotide position of the adjacent codon and referred to as phase 0, 1 and 2, respectively (Li, 1997), appeared to be equally well conserved (Fig. 2.14).

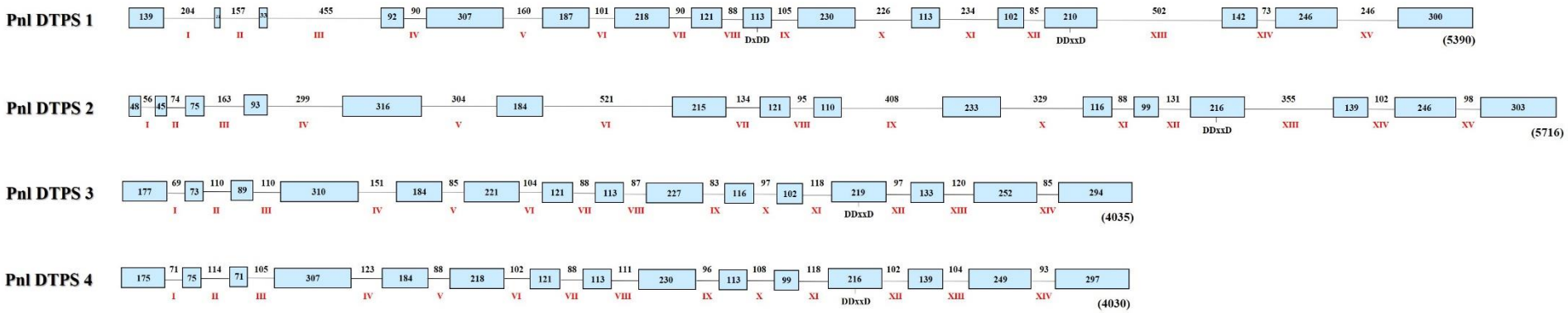


Fig. 2.14. *Schematic representation of the exon/intron structures of the four diterpene synthase (DTPS) genes isolated from Calabrian pine (Pnl) in the present study. For both exons (blue boxes) and introns (black lines) the lengths in bp are indicated. Introns were numbered (Roman numerals) starting from the 5' end of each genomic sequences.*

To gain insights into the functional evolution of terpene synthases genes in plants, Trapp and Croteau (2001) divided them into three classes, namely I, II, and III, which might have evolved sequentially by introns loss mechanisms. According to such classification, the four Calabrian pine *DTPSs* isolated in the present study belong to the class I, formed mainly by both mono- and bi-DTPS genes containing 12-14 introns, present in both gymnosperms (secondary metabolism) and angiosperms (primary metabolism). Indeed, the authors (Trapp and Croteau, 2001) showed a strong conservation of the genomic structure between the genes encoding monofunctional CPS and KS enzymes of angiosperm gibberellin metabolism, on one side, and a gene coding for the bifunctional DTPS abietadiene synthase from *Abies grandis* (AgAS), involved in specialized metabolism, on the other side. This led to above authors to propose that AgAS might be reminiscent of a putative ancestral bifunctional DTPS from which the monofunctional CPS and KS were derived through gene duplication and subsequent specialization of each of the duplicated genes for only one of the two ancestral activities. This model of an ancestral bifunctional DTPS was validated later on by the discovery of a bifunctional CPS/KS from the moss model species *Physcomitrella patens*, showing a similarly conserved gene structure (Keeling *et al.* 2010).

In the present work, the isolation of the complete genomic sequences of Calabrian pine *DTPSs* made it possible to further and complete the analysis of Trapp and Croteau (2001), by comparing them with the *DTPSs* already assigned to class-I (Fig. 2.15). Such comparison confirms that, as already noticed among the four *DTPSs* from Calabrian pine (see above), number, position, and phase of the introns III-XIV are highly conserved in all the class-I DTPS genes, among which *AgAS*, regarded as descending from a putative ancestral bifunctional DTPS gene (see above). In contrast, number, placement and phase of introns preceding the intron III, on the 5'terminus side, were not conserved among the compared DTPS genes, and an additional, equally not conserved, intron was also found in this region in the genomic sequences of *Pnl DTPS1* and *Pnl DTPS2* (Fig. 2.15).

Even though conifer bifunctional DTPSs of specialized metabolism, and monofunctional DTPSs of specialized metabolism and GA biosynthesis represent three separate branches of DTPS evolution (Hall *et al.* 2013a; Alicandri *et al.* 2020), their conserved gene structure provides strong evidence for a common ancestry of DTPS of general and specialized metabolism. In agreement with the phylogenetic analysis (Fig. 2.9), the high conserved genomic organization detected among the four *P. laricio* genes confirmed also that the monofunctional class-I DTPSs of specialized metabolism in *Pinus* species have evolved in relatively recent times by gene duplication of a bifunctional class-I/II DTPS, accompanied by loss of the class-II activity and subsequent functional diversification. It is worth noting that while the bifunctional class-I/II DPTS of *P. laricio*, and the putative homologous proteins from *P. taeda*, *P. contorta* and *P. banksiana* have orthologs in other conifers, e.g., in *P. abies*, *P. sitchensis*, *Abies balsamea* and *A. grandis*, class-I DTPSs of specialized metabolism have

not yet been discovered in other conifers outside of the *Pinus* genus. It is therefore conceivable that they constitute a lineage specific clade of the TPS-d3 group arising from a common ancestor of the closely related species of *P. laricio*, *P. contorta* and *P. banksiana*, and possibly of all the *Pinus* species, after that pine, spruce, and fir genera became separated from each other.

Fig. 2.15. *Genomic organization of plant DTPS genes. Black vertical bars represent introns (indicated by Roman numerals) and are separated by colored boxes with specified lengths in aa (amino acids) representing exons. The number above the intron Roman numeral represents the intron phase number and demonstrates conservation throughout the plant DTPS genes. Introns are classified into three phase types according to Li [33]. Schematic, intron numbers, and exon coloring scheme are based upon Trapp and Croteau (2001). Genomic DNA sequences compared are as follows: AgAS, A. grandis abietadiene synthase (NCBI accession no. AF326516); GbLS, G. biloba levopimaradiene synthase (AY574248); TbTXS, Taxus brevifolia taxadiene synthase (AF326519); PgKS, P. glauca ent-kaurene synthase (GU059905); AtCPS, A. thaliana copalyl diphosphate synthase (AT4G02780); Pnl DTPS 1-4 isolated in this study.*

➢ *2.3.5. Transcripts profiling of Calabrian pine DTPSs genes reveal differential expression across different tissues and suggest their putative roles in the biosynthesis of diterpene resin acids.*

The four DTPS genes isolated from Calabrian pine were found to be constitutively expressed in all the five tissues analysed, although their transcription levels were highly variable (Fig. 2.16).

Compared to the other three DTPS genes, *Pnl DTPS1* was highly expressed in LS and IS. The expression levels of such gene were also comparatively high in R, respect to very low amount of transcripts detected in YN and MN (Fig. 2.16). Overall, the expression pattern of *Pnl DTPS1* in each of the different Calabrian pine tissues was consistent with the corresponding diterpenoids profiles: first, by comparing Fig. 2.16 and Fig. 2.5A, it can be seen that *Pnl DTPS1* transcripts abundances and the total amounts of DRAs in the different tissues are essentially correlated. Secondly, by considering its high predicted protein sequence identity with other bi-I/II DTPS from *Pinus* spp. known to produce abietane-type diterpene olefins, namely abietadiene, neoabietadiene, palustradiene, and levopimaradiene, the expression levels of *Pnl DTPS1* were comparatively higher in those same tissues, namely LS, IS and R, in which abietic and palustric acids were found to be amongst the predominant DRAs (see Fig. 2.5 D-E).

Although significantly lower than those of the other two DTPS genes, the expression levels of *Pnl DTPS3* and *Pnl DTPS4* were similar in LS, IS and R, with comparatively lower amount detected in YN and MN (Fig. 2.16). Again, tissue-specific genes expression levels were found to be consistent with the corresponding DRAs profiles: indeed, the predicted protein sequences of Pnl DTPS3 and Pnl DTPS4 were found to be highly homologous with the ISO and the PIM DTPSs from *P. contorta* and *P. banksiana*, respectively, known to produce pimarane-type olefins, namely pimaradiene, sandaracopimaradiene and isopimaradiene, acting as precursors of the corresponding DRAs. As a matter of fact, pimarane-type DRAs were found to accumulate in considerably lower amounts than the abietane-type DRAs in most of the tested Calabrian pine tissues (see Fig. 2.5 D-E).

Among the pimarane-type DRAs, isopimaric acid was significantly more abundant than pimaric acid in most of the tissues tested (Fig. 2.5), although no significant differences were detected in the amount of transcripts of the two genes potentially involved in their synthesis, namely *Pnl DTPS3* and *Pnl DTPS4*, respectively (Fig. 2.16 and see above). These findings suggest that other TPSs might be involved in the production of isopimaric acid in Calabrian pine. Indeed, bifunctional enzymes producing isopimaric acid have been previously identified from *P. abies* (Martin *et al.* 2004), *P. sitchensis* (Keeling *et al.* 2011), and *A. balsamea* (Zerbe *et al.* 2012), although no obvious ISO candidate has been identified so far in the *Pinus* species (Celedon and Bohlmann, 2019; Hall *et al.* 2013b) It would be conceivable that an orthologous bifunctional ISO enzyme is present in Calabrian

pine, which would account for the discrepancy between the transcript abundances and metabolite levels in the analysed tissues.

Finally, transcript levels of *Pnl DTPS2* were the highest in LS and IS, although significantly lower than those of *Pnl DTPS1*, and moderate in R, MN and YN (Fig. 2.16). It is worth noting that in both types of needles the expression levels of *Pnl DTPS2* were remarkably higher than those of the other three genes (Fig. 2.16). Because previous attempts to functionally characterize orthologous genes in other pine species were unsuccessful, it is not possible at present to make correlative hypotheses on the possible role of *Pnl DTPS2* in DRAs biosynthesis. Nonetheless, its sustained and tissue-specific expression levels observed here, which appears to be correlated with the accumulation of dehydroabietic acid (compare Fig. 2.5 and Fig. 2.16), warrant further and deeper studies to elucidate the true function of *Pnl DTPS2* and orthologous genes from *Pinus* species in conifer DRA biosynthesis.

In summary, the diterpenoid profiles determined in the different tissues of Calabrian pine appear to be consistent with the potential roles of three of the four DTPSs genes isolated in the present study. It should be noted, however, that none of the DTPSs genes isolated here can be associated to the synthesis of dehydroabietic acid, despite this was one of the most abundant DRA detected across all the Calabrian pine tissues (Fig. 2.5 C-F). As a matter of fact, the biosynthesis of dehydroabietadiene has not been resolved yet in any plant species (Hall *et al.* 2013a), while one member of the CP450 family in *P. sitchensis* (PsCYP720B4) was found to be able to interact with the dehydroabietadienate group of substrates (dehydroabietadiene, dehydroabietadienol, and dehydroabietadienal) to produce dehydroabietic acid (Hamberger *et al.* 2011).

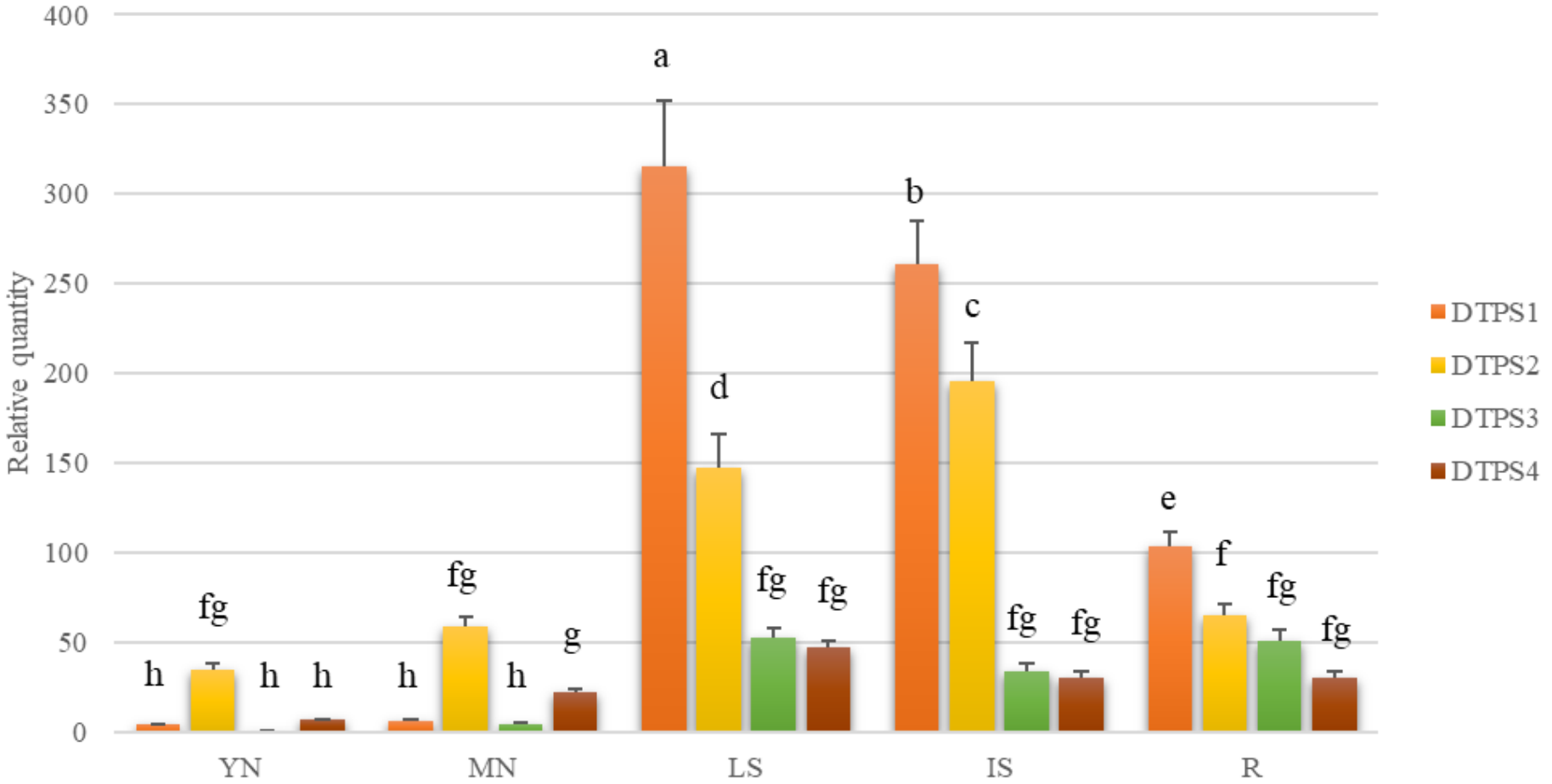


Fig. 2.16. *Relative expression levels of four diterpene synthase genes (DTPS1-4) in five different tissues of Calabrian pine. The expression data of each gene were normalized using the geometric average of the two reference genes CYP and upLOC. Relative expression levels of the different DTPS genes were referred to a calibrator, set to the value 1, which was represented by the gene in the five tissues with the lowest expression (DTPS3 in YN). YN, young needles; MN, mature needles; LS, bark and xylem combined from the leader stem; IS, bark and xylem combined from the interwhorl stems; R, roots. Different letters denote significant differences according to the Tukey's test ($p < 0.01$).*

2.4. CONCLUSIONS

In the present study, we carried out for the first time, to the best of our knowledge, a quali-quantitative analysis of diterpenoids composition in tissues obtained from different organs of *Pinus nigra* subsp. *laricio* (Poiret) Maire (Calabrian pine), namely young and mature needles, leader stem, interwhorl stem, and roots. In these same tissues, we carried out the isolation and sequencing of full length cDNAs and of the corresponding genomic sequences encoding for diterpene synthases involved in the specialized diterpenoid metabolism.

It was shown that diterpene resin acids are the most abundant diterpenoids across all the examined tissue types, together with remarkably lower amounts of the corresponding aldehydes and olefins. All the Calabrian pine tissues examined showed the presence of the same nine diterpene resin acids, whose quantitative distribution was nonetheless found to be remarkably different in the different tissues examined. In agreement with other studies on model *Pinus* species, our analyses showed that abietane-type DRAs were more abundant than pimarane-type DRAs in all the Calabrian pine tissues, even though with considerable differences in their amounts among the different tissues. Taken together, the above results seem to suggest a remarkable tissue-specificity, as well as species-specificity, of terpenoids composition in conifers, whose functional significance in terms of plant's biological performance, as well as in terms of possible exploitation for a variety of applications, awaits and deserves further studies.

A phylogeny-based approach allowed the isolation and sequencing of four partial and full length cDNAs coding for diterpene synthases in Calabrian pine, denoted as *Pnl DTPS1*, *Pnl DTPS2*, *Pnl DTPS3*, and *Pnl DTPS4*, each of which was found to belong to one of the four group into which the d3 clade of the plants' terpene synthases family can be divided. The subsequent analysis of the deduced amino acid sequences allowed to predict that both monofunctional, such as Pnl DTPS 2-4, and bifunctional, such as Pnl DTPS 1, diterpene synthases are involved in the biosynthesis of diterpene resin acids in Calabrian pine. Transcripts profiling of the Calabrian pine DTPSs genes revealed differential expression across the different tissues and were found to be consistent with the corresponding diterpenoids profiles, suggesting potential roles for three of the four DTPSs genes in the biosynthesis of diterpene resin acids.

Finally, the obtained full-length DTPS cDNAs were also used to isolate the corresponding complete genomic sequences, for each of which the exon/intron structure was determined. This allowed to place the DTPSs genes isolated from Calabrian pine on the background of current ideas on the functional evolution of diterpene synthases in plants and, in particular, on the functional diversification accompanying genera and species evolutionary segregation within the gymnosperms.

Beyond their roles in conifer defense, because of their ample physical and chemical diversity, and their resulting technological versatility, diterpene resin acids provide a large-volume, renewable resource for industrial and pharmaceutical bioproducts. Therefore, novel and in-depth knowledge of the evolutionary diversification of members of conifer DTPS family, their modular structure, and their putative functions appears to be important not only for a deeper understanding of their physiological and ecological roles, but also to foster metabolic engineering and synthetic biology tools for the production of high value terpenoid compounds.

3. CHAPTER 3: MONOTERPENE SYNTHASE GENES AND MONOTERPENE PROFILES IN CALABRIAN PINE [*Pinus nigra* subsp. *laricio* (POIRET) MAIRE].

Adapted from: Alicandri et al. **Monoterpene synthase genes and monoterpene profiles in Calabrian pine [*Pinus nigra* subsp. *laricio* (Poiret) Maire].** Submitted to *Plants* Basel

Abstract

In conifers, the constitutive or induced production of volatile and non-volatile terpenoids play key roles in tolerance and defence against biotic and biotic stresses. The range of action of monoterpenes (mainly volatile terpenoids) is not restricted to plants protection, but also encompasses conspecific as well as allospecific plant-to-plant communication. In the present study, we carried out a quantitative analysis of the monoterpenes composition in different tissues of the non-model conifer *Pinus nigra* subsp. *laricio* (Calabrian pine), an endemic and widespread species in Southern Italy. All the Calabrian pine tissues examined showed the presence of the same fourteen monoterpenes, among which the most abundant were β-phellandrene, α-pinene, and β-pinene, whose distribution was markedly tissue-specific. Using a strategy based on the phylogenesis of monoterpene synthases within the *Pinus* genus made it possible to isolate, for the first time, seven full length orthologous cDNA transcripts in Calabrian pine, each of which was found to be attributable to one of the seven phylogenetic groups in which the d1-clade of the canonical classification of plants' terpene synthases can be subdivided. The amino acid sequences deduced from the above cDNA transcripts allowed to predict their involvement in the biosynthesis of five of the monoterpenes identified. Transcripts profiling revealed differential gene expression across the different tissues examined and was found to be consistent with the corresponding metabolites profiles, suggesting potential roles for the isolated monoterpene synthase genes in the biosynthesis of monoterpene compounds. The genomic organization of the seven isolated monoterpene synthase genes was also determined.

Keywords: monoterpenes; monoterpene synthase genes; *Pinus nigra* subsp. *laricio*; Calabrian pine, genomic organization of plant terpene synthase genes; pine oleoresin; terpene volatile compounds, β-phellandrene; α-pinene; β-pinene.

Key to plant species: *Abies grandis* (Douglas ex D. Don) Lindl., grand fir; *Picea abies* (L.) H. Karst, Norway spruce; *Picea glauca* (Moench) Voss, white spruce; *Picea sitchensis* (Bongard) Carrière, 1855, Sitka spruce; *Pinus banksiana* Lamb., jack pine; *Pinus contorta* Douglas, lodgepole pine; *Pinus jeffreyi*

Balf., Jeffreyi pine; *Pinus nigra* subsp. *laricio* (Poiret) Maire, Calabrian pine; *Pinus ponderosa* Douglas ex. C. Lawson, ponderosa pine; *Pinus sabiniana* Douglas ex D. Don, grey pine; *Pinus taeda* L., loblolly pine.

3.1. INTRODUCTION

The production of terpenoids in conifer species, both in the form of oleoresin and emitted as volatile compounds, play a significant role in the physical and chemical defence responses against herbivores and pathogens (Celedon *et al.* 2019; Alicandri *et al.* 2020). Oleoresin, whose main components are mono- (C_{10}) and di-terpenes (C_{20}) (including diterpene resin acids, DRAs), with lower quantities of sesquiterpenes (C_{15}), accumulates in specialized anatomical structures, mainly resin ducts, which are abundantly present in different tissues of conifer trees, such as bark, wood and needles (Celedon *et al.* 2019). In case of wounding, the resin spreads out from the ducts to reach the damaged area and acts as a physical and chemical weapon against invading organisms (Celedon *et al.* 2019; Zulak and Bohlmann 2010). Conifer foliage also emits terpenes originating from two sources: those stored in the needle resin ducts and those directly emitted from mesophyll cells (Celedon *et al.* 2019). These volatile terpenes, mainly mono- and sesqui-terpenes, can play a significant function in plant-to-environment interactions, acting as signals to stimulate defence responses in nearby plants or in healthy tissues of the same plant or in attracting natural enemies of pathogens and herbivores (Celedon *et al.* 2019; Raffa *et al.* 2014; Boncan *et al.* 2020).

In addition to chemical and physical defence, constitutive and induced conifer terpenes, mainly hemi- and mono- terpenes, can perform a physiological and ecological role in the protection against abiotic stresses, such as drought, salinity, high and low temperatures and air and soil pollution (reviewed in Kopaczyk *et al.* 2020). Moreover, their emission from conifer trees plays a considerable role in global carbon cycle and atmospheric processes (e.g., in the cycle of photochemical production/destruction of ozone, particle formation and aerosols) (Loreto *et al.* 2010; Sharkey *et al.* 2013). Many stress conditions cause a rapid rise in reactive oxygen species, resulting in oxidative stress and the activation of signalling pathways that lead to metabolic reprogramming (Akula and Ravishankar 2011). Several studies showed that some terpenes exhibit considerable antioxidant activity (Dahham *et al.* 2015; Porres-Martìnez *et al.* 2016), suggesting that they may have a role in preventing oxidative stress induced by abiotic stressors (Kopaczyk *et al.* 2020).

While the constitutive and induced terpene-related defence mechanisms are thought to have contributed to the evolutionary success of conifers in forest ecosystems worldwide, the increased pressure from the combined and related effects of climate change and host expansion of conifer pests and disease may challenge the effectiveness of these defence systems (Celedon *et al.* 2019).

Conifer's monoterpenes as major components of the oleoresin and of the emitted volatile terpene fraction, are largely involved in the constitutive and induced defence responses of conifer trees to

insects and fungal diseases (Celedon *et al.* 2019; Kopaczyk *et al.* 2020; Nicole *et al.* 2006; Hall *et al.* 2011; Pollastrini *et al.* 2015; Hall *et al.* 2013b; Boone *et al.* 2011).

Because many monoterpenes are volatile molecules, some of them, among which ocimene, myrcene, limonene, linalool, camphene and pinene, just to mention a few of those found in conifers, are well-known fragrants found in the essential oils, and are used in perfumes, cosmetics, and cleaning products. Moreover, several conifer monoterpenes, among which bornyl acetate, camphene and limonene are used as food flavors and food additives. Finally, yet importantly, a few conifer monoterpenes have been claimed to exhibit antimicrobial, antiviral or antitumor activities (Cristiani *et al.* 2007; Sobral *et al.* 2014; Allenspach *et al.* 2021).

The biosynthesis of monoterpenes results from the activity of a specialized group of terpene synthases (TPSs), namely monoterpene synthases (MTPSs), which use geranyl diphosphate as a substrate to form, typically in a stereo specific fashion, acyclic, monocyclic or bicyclic compounds. All the known MTPSs belong to the so called "class-I" TPSs, hosting only one active site located in the C-terminal α-domain, which contains two metal binding amino acidic motifs, the highly conserved "DDXXD" and the less conserved "NSE/DTE". Most MTPSs have an obvious N-terminal plastid transit peptide, which is removed from the mature MTPS upstream of the "RR(X_8)W" motif, which in turn is important for the catalysis of monoterpene cyclization (see Alicandri *et al.* 2020 and references therein).

The abundance and diversity of the different monoterpenes in conifers (Kopaczyk *et al.* 2020), and their roles in conifer-herbivore interactions, prompted the cloning and characterization of MTPSs in several conifer species (reviewed in Celedon *et al.* 2019; Alicandri *et al.* 2020). For instance, the white spruce (*Picea glauca*) genome annotation identified 24 MTPS genes (Warren *et al.* 2015), but only four of them have been functionally characterised, i.e., heterologously expressed in bacterial/yeast systems and tested *in vitro* with their potential terpenoids substrates, including (–)-α-pinene synthase, (–)-β-pinene synthase, (–)-linalool synthase and 1,8-cineole synthase (Keeling *et al.* 2011). Another study allowed cloning nine different jack pine (*Pinus banksiana*) and eight different lodgepole pine (*Pinus contorta*) MTPS full-length cDNAs (Hall *et al.* 2015). Functional characterization of the 17 different MTPSs and monoterpene profiles in different tissues of lodgepole pine and jack pine identified orthologous sets of enzymes that contribute to the biosynthesis of oleoresin monoterpenes such as (-)-β phellandrene, (-)-β-pinene, (+)-α-pinene, (-)-α-pinene and (+)-3-carene (Hall *et al.* 2015).

It is apparent from the above that most of the existing knowledge concerning the genetics and metabolism of monoterpenes in conifers was obtained from model Pinaceae species, for which large transcriptomic and genomic resources are available. In previous work of ours (Alicandri *et al.* 2020; Foti *et al.* 2020; Alicandri *et al.* 2021) we began to gain insight into the ecological and functional roles

of the terpenes produced by the non-model conifer *Pinus nigra* subsp. *laricio* (Poiret), vernacular name Calabrian pine, one of the six subspecies of *P. nigra* (black pine) and an insofar completely neglected species under such respect. In terms of natural distribution, black pine is one of the most widely distributed conifers over the whole Mediterranean basin, and its *laricio* subspecies is considered endemic of Southern Italy, especially of Calabria, where it is a basic component of the forest landscape, playing key roles not only in soil conservation and watershed protection, but also in the local forest economy (Nicolaci *et al.* 2015).

In one of our previous works on Calabrian pine (Alicandri *et al.* 2021), five different plant tissues were analyzed in terms of oleoresin diterpenoids composition, and, in parallel, the isolation, characterization and expression analysis of diterpene synthase (DTPS) genes were carried out in the same plant material. To broaden the characterization of the terpene oleoresins in this non-model conifer species, the same combined biochemical and genetic approach was adopted in the present work, in which, again for the first time to the best of our knowledge, we used gas chromatography-mass spectrometry (GC-MS) to determine the monoterpenes profiles in five different tissues of Calabrian pine. In this same subspecies, in addition, we report here about the isolation of full length (FL) cDNAs and the corresponding genomic sequences encoding for six MTPSs and one hemi-TPS, obtained by using a strategy based on the phylogeny of available MTPSs from different *Pinus* species. The isolation of Calabrian pine *MTPS* genes allowed a tissue-specific gene expression analysis, to be confronted with the corresponding GC-MS monoterpenes profiles.

3.2. MATERIAL AND METHODS

➢ *3.2.1 Plant material*

The plant material, as well as the sampling procedure, were the same adopted in our previous work concerning the metabolic and genetic profiling of diterpenoids in different tissues of three years old seedlings of Calabrian pine [*Pinus nigra* subsp. *laricio* (Poiret) Maire] (Alicandri *et al.* 2021). Briefly, the Calabrian pine seedlings were grown in the open inside protective housings set up within the premises of the Aspromonte National Park, Southern Italy (38° 17’27” N, 15° 81’68” E; altitude 1010 MASL, exposed East). Five tissue types were collected for analysis: young needles (YN); mature needles (MN); bark and xylem combined from leader stem (LS); bark and xylem combined from interwhorl stem (IS), and roots (R). All the collected tissues were immediately frozen in liquid nitrogen and stored at −80 °C until use for nucleic acids isolation and for metabolite profiling analysis.

➢ *3.2.2 Extraction and GC/MS analysis of monoterpene metabolites*

Approximately 500 mg of the five different tissues were extracted by sonication in 1 mL of n-pentane at room temperature for 15 min. One hundred µL of each extract were transferred into a conical vial and 0,5µL were analysed by high-fast GC-MS techniques. For all the five tissue types, three biological replicates were processed, and then each of them analysed in triplicate.

Qualitative and quantitative GC/MS analysis of monoterpenes from Calabrian pine tissues were carried out essentially as reported in our previous work on diterpenes from the same plant source [23], in which full details are provided concerning the analytical equipment, the GC capillary column, the nature and flow rate of the carrier gas, the sample volume and the injection technique, the couped MS detector and the temperatures selected for the transfer line, the ion source and the analyser. For the present GC analysis of monoterpenes, the following thermal conditions were adopted: from 45°C (1 min.) to 250°C (0.5 min) at 30 °C min^{-1}, then to 325 at 45°C min^{-1}, then isothermal for 5 min. The MS acquisition was carried out under full scan (m/z 37-250) or selective ion monitoring mode at m/z: 136, 121 and 93, for qualitative and quantitative analysis, respectively.

Monoterpenes identification was achieved by comparing the experimental mass spectra both with those in the NIST08 and Wiley02 Libraries and with those reported in the available literature (Hall *et al.* 2015; Adams, 2007). As far as the Wiley and NIST mass spectra libraries are concerned, the spectral match scores obtained for the analysed monoterpenes in the pine tissues were invariably higher than 850, consistently returning the correct metabolite identification as the "first hit". According to the NIST library guidelines, this match value is considered satisfactory and reliable for a correct identification of a given molecule.

The analyte concentrations, expressed as µg (g dry weight)$^{-1}$, were calculated by calibration curves obtained by using with commercial standards of α-pinene [(1*S*,5*S*)-2,6,6-trimethylbicyclo(3.1.1)hept-2-ene, 98% purity, Sigma-Aldrich catalogue # 147524] and of (R)-(+)-limonene [1-methyl-4-(prop-1-en-2-yl)cyclohex-1-ene, 97% purity, Sigma-Aldrich catalogue # 183164].

The GC/MS methods used in the present study for the extraction and analysis of plant metabolites were adequately validated for their selectivity, precision, and efficiency. Selectivity was verified by observing that no interfering peak was apparent at the elution time of each target analyte upon injecting three replicate blank samples. Precision was tested by measuring the inter- and intra-day variability in the chromatographic profiles of spiked samples, which ranged from 2 to 7% in terms of relative standard deviation. Finally, the extraction method recovery was computed as the average of three replicate samples of the plant tissue spiked with a known aliquot of α-pinene and (R)-(+)-limonene standard mix

and then analysed by GC-MS. Regardless of the tissue extracted, the measured mean recovery always ranged from 80 to 90%.

➢ *3.2.3 Isolation, characterization, and expression analysis of monoterpene synthases genes in Calabrian pine*

As the first step, a multiple sequence alignment of the MTPS deduced proteins from the genus Pinus was carried out from which the corresponding phylogenetic tree was generated. The adopted in silico approach and the criteria used for phylogenetic analysis are described in detail in previous work of ours (Alicandri *et al.* 2020; Alicandri *et al.* 2021).

The genomic DNA and the total RNA were extracted from the five different tissue types of Calabrian pine as reported in Alicandri *et al.* (Alicandri *et al.* 2021). All the DNA and cDNA samples were stored at -80 °C until used.

RT-polymerase chain reaction (PCR) was used to amplify partial cDNA coding for MTPSs by using forward and reverse primers designed in conserved regions among the *Pinus* MTPS sequences of the different groups identified by the phylogenetic analysis (Alicandri *et al.* 2020). The complete list of the forward and reverse primers used is reported in Table 3.1.

The partial cDNAs sequences obtained were then used as templates to isolate the corresponding full-length *MTPS* cDNAs by means of 5' and 3' RACE (Rapid Amplification of cDNA Ends) extensions, as detailed in our previous work (Alicandri *et al.* 2021). The sequence of RACE primers used are reported in Table 3.1.

To isolate the genomic MTPS sequences, the genomic DNA was amplified by using specific forward and revers primers, designed in the close proximity of the initiation (ATG) or stop codons, respectively, of each full-length cDNA (Table 3.1) as described in Alicandri *et al.* (2021).

Cloning and sequencing of partial cDNAs, RACE and genomic amplification products were conducted as described in Alicandri *et al.* (2021). In the same previous work of ours, full details can be found concerning the analysis of the nucleotide sequences obtained and of the corresponding deduced amino acid sequences.

Table 3.1. *Forward and Reverse primers used for the isolation of cDNAs and genomic sequences coding for MBOS (Group 1) and MTPSs (Groups 2-7) in Pinus nigra subsp. laricio. RACE, Rapid Amplification of cDNA Ends*

	Phylogentic group	Forward primers 5'→3'	Reverse primers 5'→3'
cDNA sequences	Group 1	F1c: TCATCATTCCAACCTCTGGGA (165)	R1c: AGGCACAGGCTCAATGAC (1839)
	Group 2	F1c: CACCATGTGTTTGACAGCCC (147)	R1c: TTTATTTCATTGGTGGCGACG (1838)
	Group 3	F1c: TTCTAACCTGTGGGACGACAA (207)	R1c: TACATTAGCACGGGTTCG (1880)
	Group 4	F1c: TTCTGTCAACGCCTTATGGG (242)	R1c: TATAAAGGCACAGGTTCAAGGAG (1883)
	Group 5	F1c: GGACCGCCGTATCTGATGATG (167)	R1c: GTGACAGGATCAATGACGGT (1883)
	Group 6	F1c: CGCAACGTCTTATGAGGCAC (240)	R1c: CACAGGTTCAAGGACGGT (1851)
	Group 7	F1c: GGGAGGGAAATCCATAGCAC (120)	R1c: GCACAGTTTCAACGACGG (1879)
		Race 5'	**Race 3'**
Race 5'-3' ends	Group 1	R1: AACACATCTGAAGACACCGGG (545) R2: CCATAAGGCGTTGAGAGGGAC (221) R3: GGTTATCCGACACATGCGG (126)	F1Race3': AATAGACTTTCCTTCGAGGC (1306)
	Group 2	R1: AAGGCCCAGTGCAGTTGAGT (534) R2: GGCGCCTCATAATGTGTGG (236) R3: GCATCTGACACAGCGGTTGTT (149)	F1Race3': CAGATAGTGCCCGTGGAGAA (1212)
	Group 3	R1: CGCCAGAGCAGTTGTGTTGAG (534) R2: CAGCATGCTCACCGTACGA (283) R3: GTGGCCGAACTCATCCTGAT (161)	F1Race3': GCATTGACATGGTGCCTGAAC (1055)
	Group 4	R1: GAAGAGTTCGAAAGCCCAAGG (553) R2: GCACGTTCACGGTAAGCGA (296) R3: ACTGATTTCCGTGGCCTCG (140)	F1Race3': GCAATTAAGAGATGGGATCCG (1189)
	Group 5	R1: GGTTCGAAGACCCAAGGCAG (552) R2: TCTCAGCACGTTCCCGATAAG (295) R3: GCTCATGGAAGCACGTGTCA (156)	F1Race3': AAGAGATGGGATCCGTCG (1198)
	Group 6	R1: CAGCACTTTGTCTCCCACGTC (508) R2: TCGGCACGTTTGAGGTAGG (287) R3: ACTCACGCGAATGGAAGGTC (153)	F1Race3': CCGTCGGTTGTAGATTGTC (1186)
	Group 7	R1: TCTCTGCCACGTCCAATG (503) R2: AGGCTATGGAATCATCGTCC (244) R3: AGGTGCTATGGATTTCCCTC (141)	F1Race3': CTTCGGAACAATGGAGGAGC (1152)
		Forward primers 5'→3'	**Reverse primers 5'→3'**
Genomic sequences	Group 1	F1g: ATGTCTCTGCTCTCTGTCGC (1)	R1g: AGGCACAGGCTCAATGAC (1839)
	Group 2	F1g: ATGGCTCTACTTTCTGTCGC (1)	R1g: TTTATTTCATTGGTGGCGACG (1838)
	Group 3	F1g: ATGTCTCTTATTTCCGCTGTG (1)	R1g: TACATTAGCACGGGTTCG (1880)
	Group 4	F1g: ATGGATTTAATATCTGTC (1)	R1g: TATAAAGGCACAGGTTCAAGGAG (1883)
	Group 5	F1g: ATGTCTCCTGTTTCTGTGATC (1)	R1g: GTGACAGGATCAATGACGGT (1883)
	Group 6	F1g: ATGGCTCTGGCTCTGGTT (1)	R1g: CACAGGTTCAAGGACGGT (1851)
	Group 7	F1g: ATGGCTCTGGTTTCTGCTG (1)	R1g: GCACAGTTTCAACGACGG (1879)

As far as the expression analysis of the isolated MTPS genes is concerned, full methodological details can be found in previous work of ours (Alicandri *et al.* 2021; Ciaffi *et al.* 2019). These include samples replication, quantitative real time (qRT-PCR) conditions, the selection of the most appropriate and stable reference genes for normalisation, the designing of primer pairs for both target and reference genes (Table 3.2), the evaluation of primers specificity and amplification efficiency, and the criteria used to calculate normalized relative values of gene expression and their standard deviation.

Table 3.2. *List of primer pairs of Calabrian pine MBOS, MTPS and reference (CYP and upLOC) genes used in qRT-PCR analyses*

Gene	Forward primer (5'-3')	Reverse primer (5'-3')
MBOS1	GCAATGTTCCAATCACTTCCAAG	AGGCACAGGCTCAATGACG
MTPS2	GCTTCTCAAACCAGACTGC	TTGGTGGCGACGCTGTAG
MTPS3	TCTCGGGTCAACACAGGAAG	CCATCTCGGTAGTTGTAGAAGTG
MTPS4	CTGGAGCAACAGAGGAAGATG	GTATTTGTAACCGTAATGGAAAGC
MTPS5	GCCCGTGGAGAAGAAGC	TGGGAACGCTGCTGTTTG
MTPS6	TTAAACCCGACAGCAATGTTCCC	GTGGCATCACCGTAACCATCTC
MTPS7	TGCTGTCAATCAAGTCAATGC	GCAACACTGAAGCCATCTCTG
CYP	TGTAGAGGGCTTGGAGGTC	CAAGCGAGCTGTCCAGAGT
upLOC	GGTTTGCTTTGGAGGATATG	GTCCAATGTGCACCTCGT

➢ *3.2.4 Statistical analysis*

Each reported value for metabolites and gene expression levels represents the mean of a total of nine replicates, obtained from three biological replicates and three technical replicates for each biological replicate. The statistical significance of the differences observed was evaluated by one-way ANOVA, followed by the Tukey's test. All statistical analyses were performed using JMP PRO 15 (Trial Version ©SAS Institute Inc.).

3.3. RESULTS AND DISCUSSION

➢ *3.3.1 In the Pinaceae, the monoterpene metabolites profiles are tissue-specific and species-specific*

The diversity of monoterpenes composition was evaluated in five different tissues, namely young (YN) and mature (MN) needles, leader and interwhorl stems (LS and IS, respectively), and roots (R) obtained from three-year old Calabrian pine saplings. All the tissues examined contained the same set of fourteen monoterpene compounds, namely, bornyl acetate, camphene, δ-3-carene, α-fenchene, limonene, myrcene, β-phellandrene, α- and β-pinene, sabinene, α-terpineol, terpinolene, α-thujene, and tricyclene, whose typical elution order under the adopted GC conditions is reported in Figure 3.1, and whose chemical structures are shown in Figure 3.2.

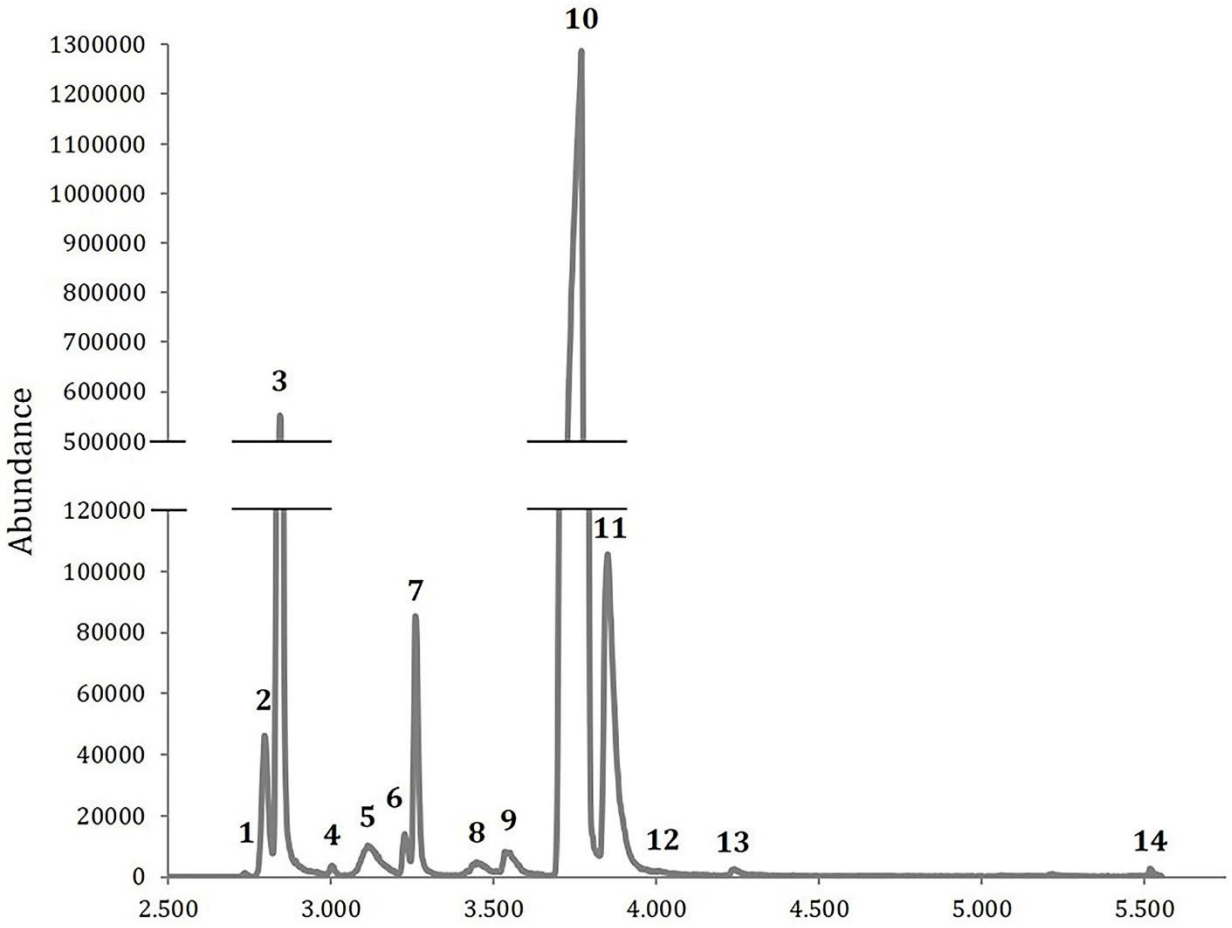


Figure 3.1. *A representative GC-MS profile of the monoterpenes extracted from the young needles of Calabrian pine. Single ion monitoring at m/z 136, 121 and 93. (1) tricyclene, (2) α-thujene, (3) α-pinene, (4) α-fenchene, (5) camphene, (6) sabinene, (7) β-pinene, (8) myrcene, (9) δ-3-carene, (10) β-phellandrene, (11) limonene, (12) terpinolene, (13) α-terpineol and (14) bornyl acetate.*

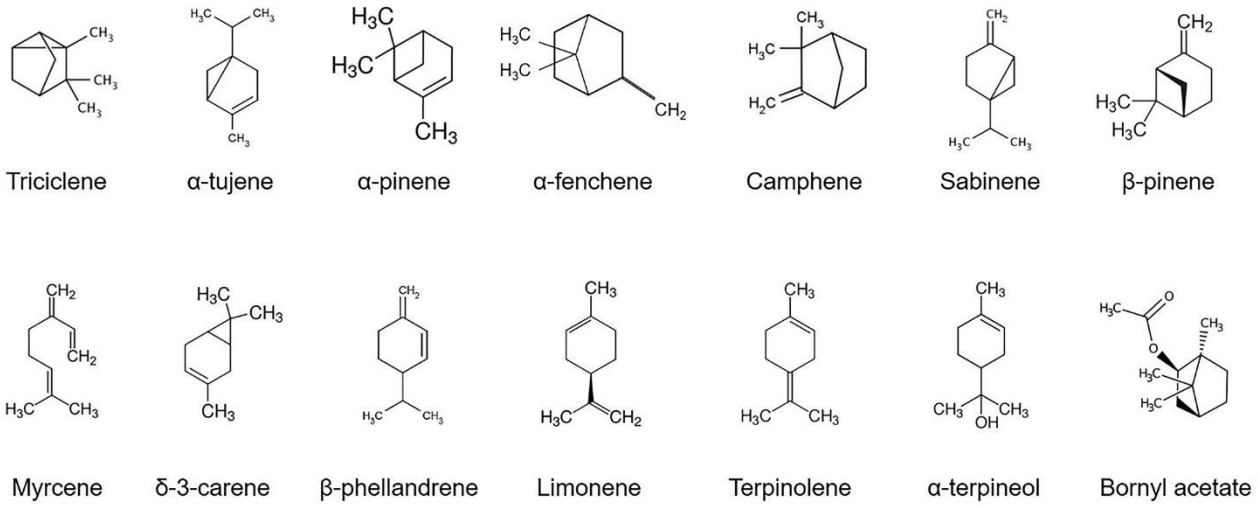


Figure 3.2. *Chemical structures of the monoterpenes found in different tissues of Calabrian pine.*

Comparison with homologous tissues obtained from coeval saplings of jack pine and lodgepole pine (Hall *et al.* 2013b) showed a remarkably similar monoterpenes profiling in Calabrian pine, albeit not identical, in that the latter contained α-thujene and α-fenchene in addition, but was devoid of α-terpinene, γ-terpinene, linalool and terpin-4-ol.

Quantitatively speaking, Figure 3.3A shows that the highest content of total monoterpenes was found in the LS tissue, while decreasing amounts were observed in R, IS and MN tissues, with the lowest concentration detected in YN.

Three of the five tissues, namely YN, MN, and IS, contained 31-44% β-phellandrene as the most abundant monoterpene (Figure 3.3 B-C-E), whereas α-pinene (approximately 44% of the total) and β-pinene (about 51% of the total) were found to be the most abundant monoterpenes in LS (Figure 3.3D) and R (Figure 3.3F), respectively. These findings confirm the results obtained by Hall *et al.* (2013) on the same tissues of three-year old *P. contorta* seedlings, with the exception that in the leader stem of lodgepole pine the predominant monoterpene was β-phellandrene (more than the 50% of the total), with very low amount of α-pinene (approximately 3%).

Besides β-phellandrene, the most abundant monoterpenes found in the needles (both YN and MN) were α-pinene (26-27% of the total), β-pinene (7-17%) and limonene (6-14%), with lower amounts (less than 3% each) of the remaining ten compounds (Figure 3.3 B-C). Conversely, the root and the stem (both LS and IS) revealed a more fragmented monoterpenes composition (Figure 3.3 D-F). For instance, the root, in addition to β-pinene (more than half of the total monoterpene fraction), contained approximately 12% each of β-phellandrene and myrcene, 6% each of α-pinene and camphene, 4% each of α-thujene and limonene, and less than 1% each of the remaining seven monoterpenes (Figure 3.3 F). In this context, it is worth nothing that δ-3-carene, which was previously found to be one of the most abundant monoterpenes in *P. contorta* roots (more than the 30% of the total) (Hall *et al.* 2013b), was found to be present at very low levels in the same tissue of Calabrian pine (0.036% of the total). On the other hand, monoterpene profiles of Calabrian pine LS and IS tissues contained 13-44% α-pinene, 11-31% β-phellandrene, 13-17% α-thujene, 9-13% β-pinene, 7-8% δ-3-carene, 3-8% myrcene, and less than 3% of the other eight identified monoterpene compounds (Figure 3.3 E-F). In these last two tissues, remarkable was the amount of α-thujene, which, as stated before, was not identified in *P. banksiana* and *P. contorta* (Hall *et al.* 2013b). This bicyclic compound was found to represent 1-10% of the total monoterpene fraction in several widespread American conifers, such as *Pinus strobus* (eastern white pine), *Tsuga canadensis* (eastern hemlock), *Thuja occidentalis* (eastern white cedar), and *Juniperus* spp. (juniper) (Geron *et al.* 2000). At the same time, in these same tissues, it is worth nothing the relatively high amount of δ-3-carene, which, contrary to what was observed for the roots, is comparable

to that determined in leader and interwhorl stem tissues from *P. contorta* and *P. banksiana* (Hall *et al.* 2013b).

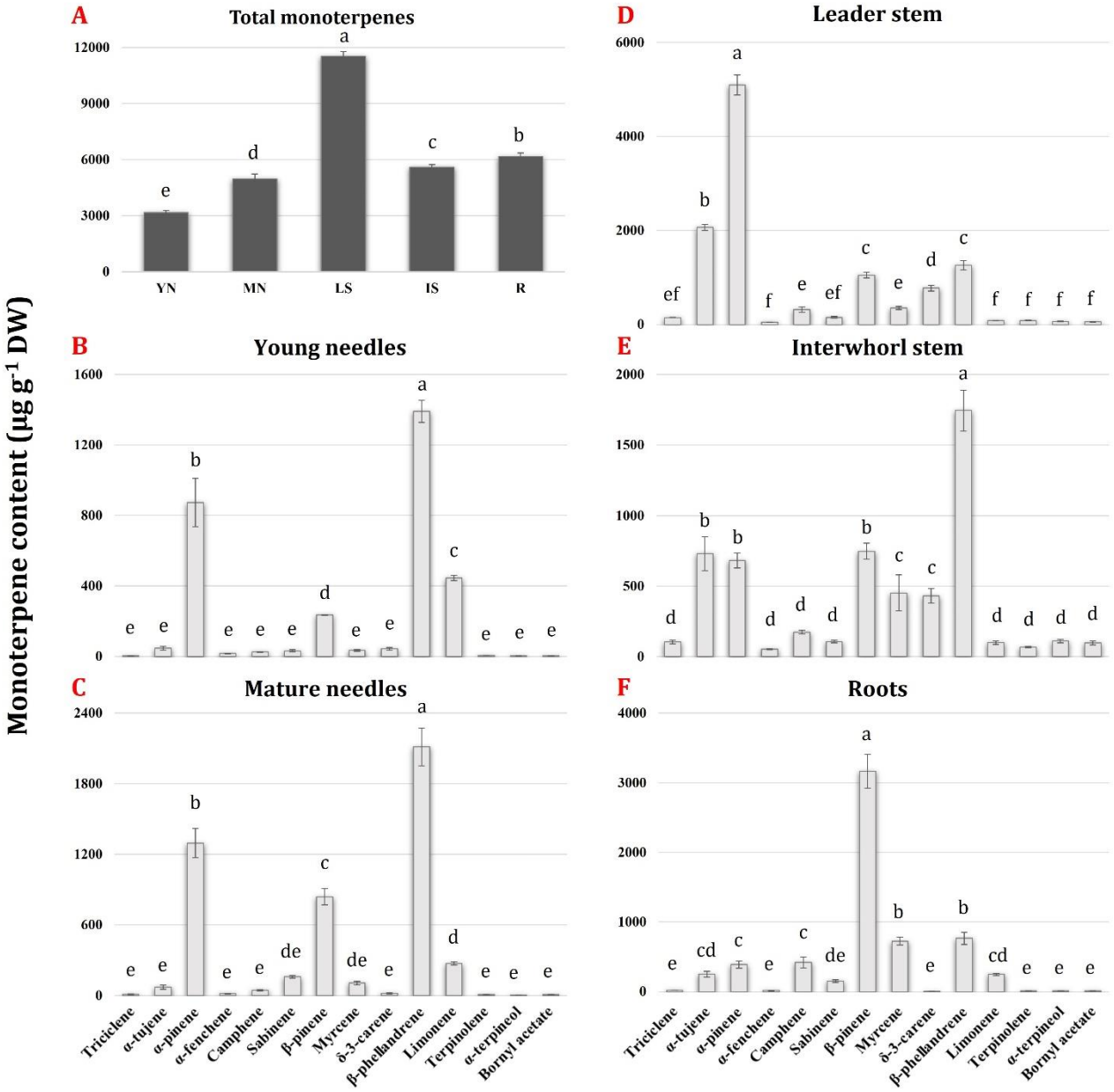


Figure 3.3. *Total monoterpene contents (top left panel A) and levels of individual monoterpenes (panels B-F) in different tissues of three-years old Calabrian pine saplings. Error bars indicate the standard deviation of the mean. The statistical significance of the differences was evaluated by one-way ANOVA, followed by the Tukey's test. Different letters denote statistical significance of the difference at $p < 0.01$.*

As Figure 3.3 indicates that the needles, and in particular the young ones, are the tissues in which, comparatively speaking, the lowest accumulation of monoterpenes occurs, it could be of interest to understand the possible functional and ecological basis of such circumstance. One possible explanation could be that, in the needles, monoterpenes are not only components of the oleoresin, but also actively

emitted as volatile compounds, to play the well-known array of physiological and ecological functions in plant-to-environment interactions (see Introduction). As a matter of fact, in a previous study of ours conducted on adult individuals of Calabrian pine thriving in the same natural context of the present study (Foti *et al.* 2020) we adopted the "headspace" GC/MS approach to collect volatile terpenoids emitted from vegetation and we found that, together with a tenth of different sesquiterpenes, at least a half of the same monoterpenes reported in Figure 3.3, namely bornyl acetate, limonene, myrcene, β-phellandrene, α- and β-pinene, and α-terpineol, are persistently and consistently present in the blend of volatiles released by the young needles throughout their growing season, with β-phellandrene, α- and β-pinene being the major components, on a comparative basis.

Taken together, the above results suggest that the quali-quantitatively profile of monoterpenes in Calabrian pine is remarkably tissue-specific, and the perusal of the available literature indicate species-specificity as well, thus confirming our previous results on the diterpenoids from the same conifer species (Alicandri *et al.* 2021). As it was shown by several studies (Kopaczyk *et al.* 2020; Geron *et al.* 2000; Pokorska *et al.* 2012), there is a notable variation in monoterpenes composition between different genera and species of conifer trees. Moreover, even closely related species may differ significantly with regard the quality and quantity of monoterpene compounds they produce, as observed in the *Pinus* genus (Kopaczyk *et al.* 2020; Hall *et al.* 2015; Blanch *et al.* 2009; Mukrimin *et al.* 2019).

Another aspect of potential interest in the non-model conifer species studied here, in view of its possible physiological, ecological and technological implications, concerns the reciprocal quantitative relationships among the major terpenoids components of oleoresin, as well as the developmental and environmental factors driving the differential distribution of them in plant tissues. In the aforementioned twin study of ours (Alicandri *et al.* 2021), we analyzed quantitatively, in the identical plant material studied here and by adopting an essentially similar GC/MS approach, the diterpenoid components of oleoresin, in the form of diterpene resin acids (DRAs), i.e., the stable and functional conformers deriving from the combined action of DTPSs and of cytochrome P450 monooxygenases (Celedon and Bohlmann, 2019). By comparing the results obtained in the two experiments (Figure 3.3 and Alicandri *et al.* 2021), on a DW basis, total monoterpenes as a whole, in Calabrian pine, are about six times more abundant than total DRAs. Upon dissecting such cumulative data at the level of each specific tissue, however, it can be found that while such approx. 6:1 ratio among monoterpenes and DRAs is maintained in the leader stem and in the root, it is otherwise substantially lower, i.e., approx. 2-3:1, in the interwhorl stem, but it is much higher, and as much as 60:1 and 120:1, in the mature and young needles, respectively. Although a much more rigorous analysis would be needed, and albeit the third macro-component, namely sesquiterpenes, is not considered here, the above considerations confirm that monoterpenes are the most abundant terpenoids in conifer oleoresin (Alicandri *et al.* 2021) and seem to

suggest that the relative abundance of the different terpenoids categories might be tissue specific. While the possible functional significance, if any, of such tissue-specificity awaits and deserves further studies, it can be hypothesized that the massive prevalence of monoterpenes over DRAs in needles, and especially in the young ones, might be associated with the ability of the former terpenoids, but not of the latter, to act as volatile signals for the chemical communication with the plant's environment (see above), by using the most obvious and pervious communication channel between the plant and the surrounding atmosphere, i.e. stomata.

➢ *3.3.2 A phylogeny-based approach for isolating partial and full-length cDNAs coding for monoterpene synthases in Calabrian pine*

Based on the same strategy we previously used for isolating DTPS genes in the same non-model conifer species (Alicandri *et al.* 2021), a phylogeny-based approach was adopted to isolate *MTPS* cDNA sequences potentially involved in the synthesis of the monoterpenes identified in the different tissue types of Calabrian pine. In practice, PCR amplification of cDNA sequences was conducted by utilizing specific primer pairs designed on conserved regions of available *MTPS*s from *Pinus* species belonging to different phylogenetic groups.

Phylogenetic relationships concerning conifer *MTPS* were drawn from a broader study (Alicandri *et al.* 2020) in which we conducted an exhaustive *in silico* search to discover all possible FL *TPS*s for primary and specialized metabolisms in several *Pinus* species. As far as the present work is concerned, BLAST searches using as queries selected gymnosperm *MTPS*s enabled us to identify 74 FL sequences implicated in the synthesis of hemi- and mono- terpenes in pine species (Alicandri *et al.* 2021). Indeed, the deduced amino acid sequences from 42 cDNA sequences out of 74, belonging to 18 *Pinus* species, were predicted to synthetize 2-methyl-3-buten-2-ol (MBO), a C_5 alcohol related to isoprene by a structural and biosynthetic point of view, since both derive from the common precursor dimethylallyl diphosphate (Gray *et al.* 2011) (referred to as MBO synthases, MBOSs, in the following). Because the predicted amino acid sequences of the 42 MBOSs showed a high level of homology among each other (93-99% identity), only five of them were used in the phylogenetic analysis, together with the remaining 32 FL cDNA retrieved, whose deduced amino acids sequences were predicted to belong to proper MTPSs (Alicandri *et al.* 2021). The resulting phylogenetic tree showed that all the 37 pine MBOS/MTPS sequences clustered together into the d1 clade of TPS, which includes seven well-supported major groups, denoted as 1-7. The first group contained the five selected MBOSs, while each

of the remaining six groups contained MTPS proteins thought to be functionally related among each other (Alicandri *et al.* 2021).

In each of the seven groups, we aligned the deduced amino acid and nucleotide sequences contained therein (Table 3.3) to reveal group-specific highly conserved regions, which were then used to design group-specific primers, shown in the upper panel of Table 3.1, for the isolation of partial transcripts of orthologous genes in Calabrian pine.

By using the above strategy, we isolated and sequenced partial *MBOS/MTPS* transcripts of putative orthologous genes in Calabrian pine, one for each of the groups 1-7, which were then utilised as templates for the isolation of as many full length cDNAs by carrying out 5' and 3' RACE (Rapid Amplification of cDNA Ends) extensions. The primer sequences used for 5' and 3' RACE are listed in the middle panel of Table 3.1.

In the case of the partial *MTPS* transcripts from groups 2, 4, and 7, two slightly different sequences were identified among the three clones analysed for each cDNA fragment, due to nucleotide substitutions, the majority of which being synonymous, i.e., not changing the identity of the amino acid encoded by the nucleotides triplet, on a background of otherwise high level of sequence identity among each other (over 96%). These slightly different *MTPS* transcripts, already observed in other *Pinus* species (Hall *et al.* 2013b), and reminiscent of analogous observation we previously made during the isolation of *DTPS* transcripts (Alicandri *et al.* 2021), might derive from alleles of the same gene or from duplicated copies of the same gene, implying that Calabrian pine may contain as many more *MTPS* closely related genes belonging to each phylogenetic group, a possibility which deserve further studies. However, across the three sequenced clones for the corresponding 5' and 3' RACE products, we found the same sequences that were identical to the 5' and 3' ends of two of the three sequenced cDNA products, demonstrating that they are part of the same FL transcript. Therefore, the assembled seven unique FL cDNAs isolated from Calabrian pine and denoted as *Pnl MBOS1/Pnl MTPS2-7*, each of them belonging to one of the seven TPS-d1 groups, contained open reading frames (ORFs) of 1845, 1881, 1866, 1887, 1911, 1866 and 1890 bp, respectively, and were predicted to encode proteins of 614, 626, 624, 628, 636, 622 and 629 aa (Figure 3.4).

The FL cDNA sequences of the *Pnl MBOS1/Pnl MTPS2-7* genes have been deposited in the GeneBank database under the accession numbers from OL689404 to OL689410.

Table 3.3. *Full-length cDNA sequences of putative MBOSs and MTPSs in Pinus spp. retrieved from the NCBI database*

Species	Function	Abbreviation	Accession mRNA sequence	ORF (bp)	Accession protein sequence	Amino acids (aa)
Pinus arizonica var. cooperi	2-methyl-3-buten-2-ol synthase	Par MBOS1	JN039226	1845	AFJ73545	614
Pinus banksiana	(-)-α pinene synthase	Pb MTPS1	JQ240304	1890	AFU73856	629
	(-)-β-pinene synthase	Pb MTPS2	JQ240291	1887	AFU73843	628
	(-)-β-pinene synthase (TPS-(-)Bpin2)	Pb MTPS3	JQ240292	1884	AFU73844	627
	(-)-α/β-pinene synthase	Pb MTPS4	JQ240290	1872	AFU73842	623
	α terpineol synthase	Pb MTPS5	JQ240308	1881	AFU73860	626
	(+)-3-carene synthase	Pb MTPS6	JQ240306	1881	AFU73858	626
	(+)-3-carene synthase	Pb MTPS7	JQ240305	1881	AFU73857	626
	(+)-α pinene synthase	Pb MTPS8	JQ240298	1887	AFU73850	628
	Monoterpene synthase	Pb MTPS9	JQ240296	1887	AFU73848	628
	Monoterpene synthase	Pb MTPS10	JQ240297	1887	AFU73849	628
	(-)-β-phellandrene synthase	Pb MTPS11	JQ240302	1866	AFU73854	621
Pinus contorta	(-)-α pinene synthase	Pc MTPS1	JQ240303	1890	AFU73855	629
	(-)-β-pinene synthase	Pc MTPS2	JQ240293	1884	AFU73845	627
	Monoterpene synthase	Pc MTPS3	JQ240294	1884	AFU73846	627
	(+)-3-carene synthase	Pc MTPS4	JQ240307	1881	AFU73859	626
	(+)-α pinene synthase	Pc MTPS5	JQ240295	1887	AFU73847	628
	α terpineol /1,8-cineole synthase	Pc MTPS6	JQ240309	1851	AFU73861	616
	(-)-camphene / (+)-α-pinene synthase	Pc MTPS7	JQ240299	1860	AFU73851	619
	(-)-β-phellandrene synthase	Pc MTPS8	JQ240301	1866	AFU73853	621
	(-)-β-phellandrene synthase	Pc MTPS9	JQ240300	1875	AFU73852	624
Pinus kesiya var. langbianensis	Monoterpene synthase	Pk MTPS1	KX394684	1956	AQZ36562	651
	α-pinene synthase	Pk MTPS2	KM382173	1875	AIY22674	624
Pinus massoniana	(-)-α pinene synthase	Pm MTPS1	KF547035	1890	AGW25369	629
	α-terpineol synthase	Pm MTPS2	KJ803197	1863	AIL88641	620
Pinus pinaster	α-pinene synthase	Pp MTPS1	KP780394	1890	ALB78130	629
	α-pinene synthase	Pp MTPS2	KP780395	1890	ALB78131	629
Pinus pinea	α-pinene synthase	Ppinea MTPS1	KR011842	1890	ALD18902	629
	α-pinene synthase	Ppinea MTPS2	KR011841	1890	ALD18901	629
Pinus pseudostrobus	2-methyl-3-buten-2-ol synthase	Pps MBOS1	JN039254	1845	AFJ73572	614
Pinus pseudostrobus var. estevezii	2-methyl-3-buten-2-ol synthase	Pest MBOS1	JN039251	1845	AFJ73569	614
Pinus sabiniana	2-methyl-3-buten-2-ol synthase	Psab MBOS1	JF719039	1845	AEB53064	614
Pinus tabuliformis	α-pinene synthase	Ptab MTPS1	EF608499	1890	ABY65904	629
Pinus taeda	(-)-α-pinene synthase	Pt MTPS1	AF543527	1890	AAO61225	629
	α-terpineol synthase	Pt MTPS2	AF543529	1884	AAO61227	627
	(+)-α-pinene synthase	Pt MTPS3	AF543530	1887	AAO61228	628
Pinus teocote	2-methyl-3-buten-2-ol synthase	Pteo MBOS1	JN039258	1845	AFJ73576	614
Physcomitrella patens	*ent*-kaurene synthase	Pt TPS-entKS	AB302933	2646	BAF61135	881

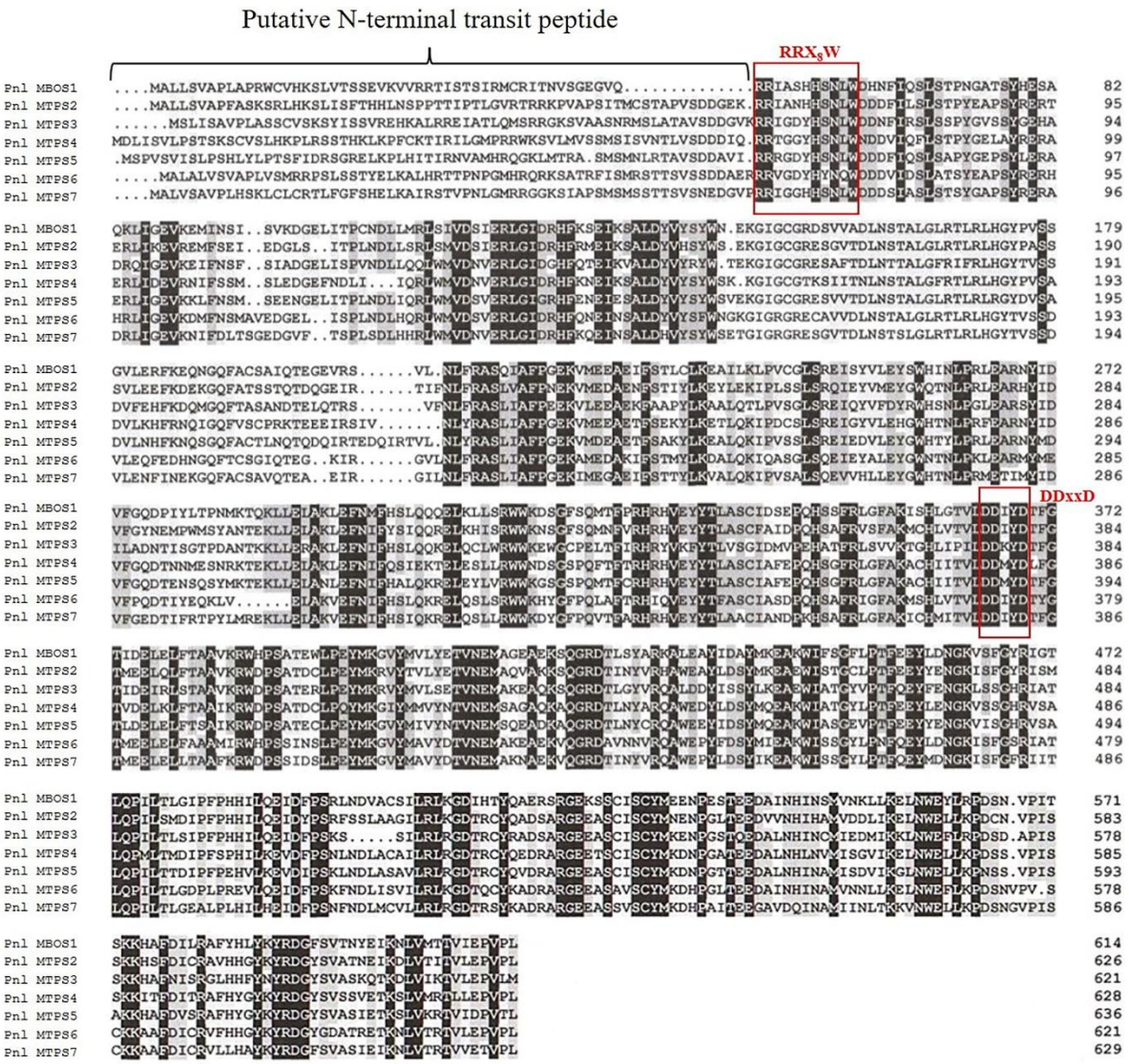


Figure 3.4. *Alignment of deduced amino acid sequences of the seven putative hemi- and mono-terpene synthases from Calabrian pine (Pnl MBOS1/Pnl MTPS 2-7) isolated in the present study. Amino acid residues with black background indicate highly conserved regions, while amino acid residues which are identical in more than 50% of the proteins are in grey background. The horizontal square bracket indicates the putative N-terminal transit peptide region. The "RRX8W" and class-I "DDxxD" signature motifs are indicated with red open rectangles.*

➢ *3.3.3 Sequence-based analysis of the predicted MBOS and MTPS proteins in Calabrian pine*

The deduced amino acid sequences of the seven FL cDNAs obtained from Calabrian pine were found to possess highly conserved and distinctive regions of plant MTPSs (Figure 3.4). First, each of the seven predicted proteins featured a potential transit peptide, ranging in length from 50 (Pnl MBOS1) to 67 amino acids (Pnl MTPS4) (Figure 3.4), which is predicted to facilitate the import of mature proteins into plastids. Such putative transit peptides are located just before a conserved RR(X_8)W motif

(Figure 3.4), which is thought to be required for the catalysis of monoterpene cyclization (Alicandri *et al.* 2020; Whittington *et al.* 2002; Hyatt *et al.* 2007). Furthermore, all the seven predicted proteins showed a conserved Asp-rich domain, namely DDxxD, which is thought to be responsible for class I activity and coordinates substrate binding through the formation of divalent cation salt bridges (Tarshis *et al.* 1996; Lesburg *et al.* 1997).

As shown in Figure 3, phylogenetic relationships were found among the 37 pine MBOS/MTPS sequences retrieved from the NCBI database (Table 3.3) and the seven MBOS/MTPS sequences obtained from Calabrian pine, thus confirming the validity of the approach used for their isolation. Pnl MBOS1 clustered with the five selected pine MBOSs in the phylogenetic group 1 (Figure 3.5), of which only that from *Pinus sabiniana* (Psab MBOS1) was functionally characterized as producing both MBO and isoprene in a 90:1 ratio (Gray *et al.* 2011). The high amino acid sequence identity detected among Pnl MBOS1 and Psab MBOS1 (about 95%) suggested that we isolated a FL transcript from a putative MBOS orthologous gene in Calabrian pine. Conifer MBOSs seem to have evolved independently from their homologous proteins in angiosperms, i.e., isoprene synthases (Sharkey *et al.* 2013). Indeed, phylogenetic analysis revealed that MBOSs fall into the TPS-d1 clade, together with the gymnosperm MTPSs, and are most closely related to linalool synthases from *P. abies* and *P. sitchensis* (Alicandri *et al.* 2020; Gray *et al.* 2011).

Pnl MTPS2 was found to cluster in the phylogenetic group 2 (Figure 3.5), together with two proteins, one from *P. contorta* (Pc MTPS6) and the other from *P. banksiana* (Pb MTPS5), both of which were functionally characterised as producing α-terpineol as their major product (Hall *et al.* 2013b). It is worth nothing that these two proteins showed only 60 to 62% sequence identity with Pt MTPS2 from *P. taeda*, which also was found to produce α-terpineol (Phillips *et al.* 2003) but was assigned instead to the phylogenetic group 4 (Figure 3.5).

As shown in Figure 3.5, Pnl MTPS3 was closely related (92-93 % protein sequence identity) to two proteins from *P. banksiana* (Pb MTPS 6-7) and one from *P. contorta* (Pc MTPS4), all of them assigned to the phylogenetic group 3, which were previously shown to produce (+)-3-carene as their major product (Hall *et al.* 2013b). The most remarkable difference among the four proteins was a deletion of 15 bp in the nucleotide sequence of Pnl MTPS 3, which determines the loss of five amino acids in its C-terminal region (Figure 3.6).

Pnl MTPS4 was found to be closely related (93-96% sequence identity) to seven MTPSs from four *Pinus* species belonging to phylogenetic group 4 (Figure 3.5), of which those from *P. contorta* (Pc MTPS 2), *P. banksiana* (Pb MTPS 2-4) and *Pinus taeda* (Pt MTPS2) were functionally characterized. Pc MTPS 2 and Pb MTPS 2-4 were reported to produce (−)-β-pinene as their major product, and (−)-

α-pinene as well, although in comparatively lower amounts (Hall *et al.* 2013b), whereas Pt MTPS2 was shown to form (−)-α-terpineol, but neither (−)-β-pinene nor (−)-α-pinene (Phillips *et al.* 2003).

Pnl MTPS5 clustered in the phylogenetic group 5, together with ten putative α-pinene synthases (Figure 3.5), among which only three, namely Pt MTPS1 from *P. taeda*, Pc MTPS1 from *P. contorta,* and Pb MTPS1 from *P. banksiana*, have been functionally characterized as producing (−)-α-pinene as their dominant product (Hall *et al.* 2013b; Phillips *et al.* 2003). Pnl MTPS5, although highly like the three functionally characterised α-pinene synthases (94-95% protein sequence identity), showed an insertion of six aa in its N-terminal region, not present in any members of the phylogenetic group 5 (Figure 3.7).

Pnl MTPS6 was found to cluster in the phylogenetic group 6 (Figure 3.5), together with four MTPSs, three from *P. contorta* (Pc MTPS7-9) and one from *P. banksiana* (Pb MTPS11), which were functionally characterized by Hall and Co-workers (Hall *et al.* 2013b). These Authors showed that Pc MTPS8, Pc MTPS9 and Pb MTPS11 form (-)-β-phellandrene as their major product, whereas Pc MTPS7, although showing a 95% identity with the aforementioned proteins, predominantly produces (-)-camphene and (+)-α-pinene, together with other monoterpene products to a lesser extent.

Finally, Pnl MTPS7 clustered in the phylogenetic group 7 (Figure 3.5), together with six MTPSs from four pine species, of which those from *P. taeda* (Pt MTPS3), *P. banksiana* (Pb MTPS8) and *P. contorta* (Pc MTPS5) were functionally characterized as forming (+)-α-pinene as their dominant product (Hall *et al.* 2013b; Phillips *et al.* 2003).

In summary, placing the phylogenetic analysis reported in Figure 3 on the background of the available literature leads us to hypothesize that, on the basis of their predicted protein sequences, the seven FL *MBOS/MTPS* cDNAs isolated from Calabrian pine could be involved in the biosynthesis of 2-methyl-3-buten-2-ol (Pnl MBOS1), α-terpineol (Pnl MTPS2), (+)-3-carene (Pnl MTPS3), (−)-β-pinene/(−)-α-pinene (Pnl MTPS4 and Pnl MTPS5), and (-)-β-phellandrene/(-)-camphene/(+)-α-pinene (Pnl MTPS6 and Pnl MTPS7). As a matter of fact, except for MBO, which probably escaped the extraction methodology used here probably because of its volatile nature [35], all of the aforementioned monoterpenes were actually found to be present in all the tissues of Calabrian pine examined, some of which, such as β-phellandrene, α-pinene, and β-pinene, even in comparatively conspicuous amounts (Figure 3.3; see also section 3.3.1). It is important to recall, however, that predicting potential MTPSs functions solely based on sequence homology could not be always reliable, as proven by the apparent lack of structure-function correlation previously reported for some members of the TPS-d1 clade (see above). For this reason, a functional characterization of the isolated FL transcripts by expression of recombinant proteins in bacterial or yeast systems and *in vitro* enzyme assays would be crucial to

decipher the actual functional roles of Calabrian pine MTPSs, regardless of their sequence homology with MTPSs from other conifers.

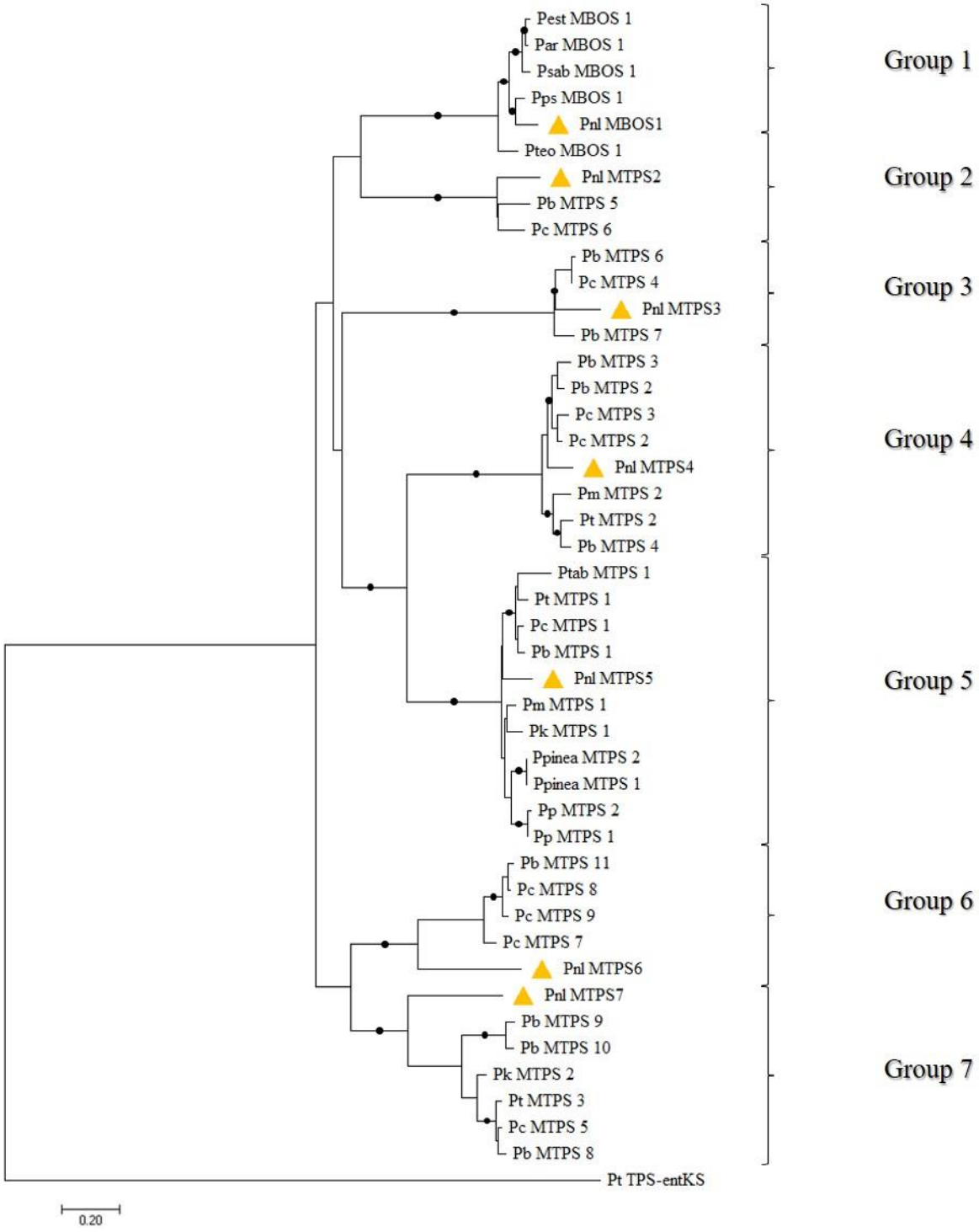


Fig. 3.5. *Phylogenetic tree of the deduced amino acid sequences of MTPS and MBOS genes identified in different Pinus species (Table S1) and those from the seven Calabrian pine MBOS and MTPS genes isolated in the present study (yellow triangles). The ent-kaurene synthase from Physcomitrella patens (Pt TPS-entKS, BAF61135) was used to root the tree. Branches marked with dots represent bootstrap support more than 80% (1000 repetitions). The seven phylogenetic groups identified in the pine members of the TPS- d1- clade are indicated by square brackets.*

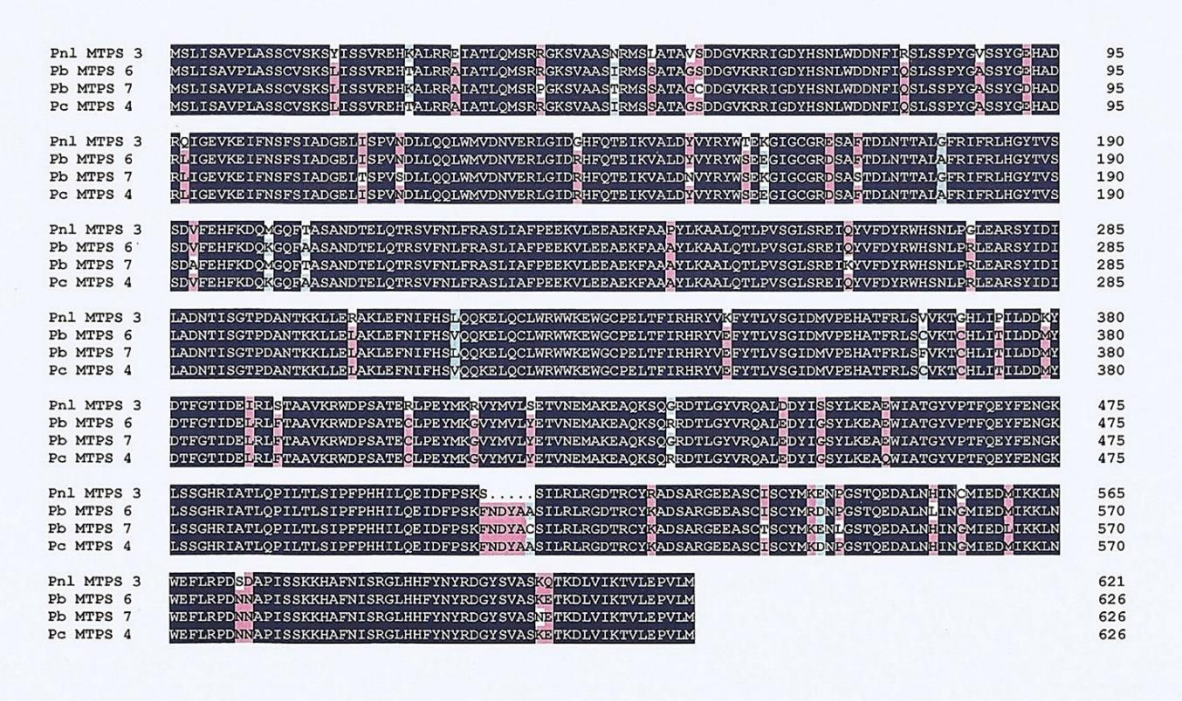


Figure 3.6. *Alignment of deduced amino acid sequences of MTPSs belonging to the phylogenetic group 3. Amino acid residues with blue background indicate highly conserved regions, while amino acid residues which are identical in more than 50% of the proteins are in pink background. Pb, Pinus banksiana; Pc, Pinus contorta; Pnl, Pinus nigra subsp. laricio (Calabrian pine).*

Figure 3.7. *Alignment of deduced amino acid sequences of MTPSs belonging to the phylogenetic group 5. Amino acid residues with blue background indicate highly conserved regions, while amino acid residues which are identical in more than 50% of the proteins are in pink background. Pb, Pinus banksiana; Pc, Pinus contorta; Pt, Pinus taeda; Pnl, Pinus nigra subsp. laricio (Calabrian pine).*

➢ *3.3.4 Genomic organization of MBO/monoterpene synthases in Calabrian pine on the background of MTPS functional evolution*

The genomic sequences encoding the seven MBOS and MTPS proteins in Calabrian pine were amplified by using primers designed to the 5'and 3' termini of the coding region of the corresponding cDNAs (Table 3.1) and the genomic DNA extracted from Calabrian pine needles as template. In all cases, a single PCR product larger than the corresponding cDNA was obtained. Three clones for each of the seven genomic fragments were partially sequenced at their ends using universal primers; once the correspondence of each cDNA with its parent genomic fragment was confirmed, a single clone for each gene was chosen and both strands were completely sequenced using internal primers designed based on the corresponding cDNA.

The lengths of the genomic sequences encompassing the ORFs of the seven genes were: *Pnl MTPS1*: 2988 bp; *Pnl MTPS2*: 3313bp; *Pnl MTPS3*: 2978 bp; *Pnl MTPS4*: 3386 bp; *Pnl MTPS5*: 2854 bp; *Pnl MTPS6*: 2893 bp; *Pnl MTPS7*: 3132 bp. These genomic sequences have been deposited in the GeneBank database under the accession numbers from OL689411 to OL689417.

Sequence alignment showed an almost perfect matching among the cDNAs and the corresponding exonic regions of the genomic sequences, allowing a reliable determination of the exon/intron structure of each gene. All the seven genomic sequences were found to contain 10 exons and 9 introns (Table 3.4), being consistent with the previously characterized genomic sequences of conifer MTPSs ((Hall *et al.* 2011; Trapp and Croteau, 2001; Hamberger *et al.* 2009). Moreover, the genomic structural characteristics of the seven *MBOS/MTPS* genes from Calabrian pine were found to be highly conserved, in terms of intron placements, exon number and sizes, and the location of RR(X_8)W and class-I DDxxD amino acids motifs (Figure 3.8). The intron sizes were generally small (about 70-200 nt), except for the last intron of *Pnl MTPS4*, encompassing 469 bp (Table 3.4). In addition, the introns of the seven *MBOS/MTPS* genes were AT rich, with repetitive sequences rich in T (3-10 mers). With few exceptions, all the seven genes had intron-exon junctions that matched the GT/AG boundary rules (Brown and Simpson, 1998). Furthermore, the phasing of the intron insertion, identified as the location of intron before the first, second, or third nucleotide position of the neighbouring codon and referred to as phase 0, 1, and 2, respectively (Li, 1997), seemed to be equally well conserved (Table 3.4). The high conserved genomic organization detected among the Calabrian pine *MBOS/MTPS* genes provides strong evidence of their common origin in conifers, confirming previous phylogenetic and protein structural studies, which demonstrated that the genes involved in hemiterpenes synthesis evolved independently in angiosperms and gymnosperms (Alicandri *et al.* 2020; Sharkey *et al.* 2013; Gray *et al.* 2011).

Table 3.4. *Comparison of the introns for the hemi-terpene synthase (MBOS) and of the monoterpene synthases (MTPS) genes isolated in the present study: position (the letters are referred to the last coded amino acid of the exon), size (in parentheses), total number and phase. No asterisk indicates no interruption between codons; single asterisk indicates intron inserted between the first and the second nucleotide; double asterisk indicates intron inserted between the second and third nucleotide; hyphen indicates intron not present. The introns are numbered according to Trapp and Croteau (2001).*

Intron	***Pnl MBOS1***	***Pnl MTPS2***	***Pnl MTPS3***	***Pnl MTPS4***	***Pnl MTPS5***	***Pnl MTPS6***	***Pnl MTPS7***
I	-	-	-	-	-	-	-
II	-	-	-	-	-	-	-
III	G74 (97)	E87 (318)	G86 (177)	Y91 (102)	G89 (123)	E87 (91)	G88 (183)
IV	-	-	-	-	-	-	-
V	-	-	-	-	-	-	-
VI	-	-	-	-	-	-	-
VII	Y141** (159)	S153** (79)	R154** (81)	S156** (124)	S157** (90)	S155** (90)	S156** (89)
VIII	S179* (115)	S190* (101)	S191* (106)	A193* (148)	A195* (80)	S192* (111)	S193* (104)
IX	E249 (94)	Q261 (225)	E261 (124)	E263 (263)	E271 (74)	E262 (97)	E263 (86)
X	L281** (269)	E290** (289)	S292** (186)	T292** (69)	E301** (181)	Q295** (221)	I293** (297)
XI	R314** (105)	R326** (108)	R326** (91)	R328** (125)	R336** (93)	R321** (94)	R328** (98)
XII	R386** (111)	R398** (111)	R398** (114)	R400** (91)	R408** (92)	R393** (98)	R400** (208)
XIII	A432 (99)	A444 (89)	A444 (100)	A446 (108)	A454 (111)	W440 (113)	A446 (89)
XIV	Q515 (94)	Q527 (112)	G543 (133)	Q529 (469)	Q537 (99)	A523 (112)	K529 (88)
Genomic size (bp)	2988	3313	2978	3386	2854	2893	3132
Protein length (aa)	614	626	621	628	636	621	629

In an attempt to gain insights into the functional evolution of terpene synthase genes in plants, Trapp and Croteau [36] divided them into three classes, namely I, II, and III, which might have evolved sequentially by introns loss mechanisms. According to such classification, the seven Calabrian pine *MBOS/MTPS* genes isolated in the present study belong to class II, formed by gymnosperm monoterpene and sesquiterpene synthase genes containing nine introns. According to Trapp and Croteau (2001), the gene coding for the bifunctional DTPS abietadiene synthase from *A. grandis* (*AgAs*)

could represent the most obvious modern candidate that resembles the ancestral progenitor gene in gymnosperms.

To further and complete the analysis of Trapp and Croteau (2001), the genomic sequences of Calabrian pine MTPSs isolated in the present work were compared with those of the MTPSs already assigned to class II and with representative members of class I and III genes (Figure 4). Such comparison indicates that, as already noticed among the seven *MBOS/MTP*S genes from Calabrian pine (see above), number, position, and phase of the nine introns, numbered according to Trapp and Croteau (Trapp and Croteau, 2001) as III and VII-XIV, are highly conserved in all the class II *MTPS* genes. Moreover, the comparison of the isolated Calabrian pine genes and the available class II genes, with the class-I *AgAs* gene, regarded as descending from a putative ancestral progenitor of all the TPS genes (see above), confirms that class I genes gave rise to class II genes by a complex intron loss mechanism. In particular, class II genes have lost the introns I and II and the entire N-terminal domain, named conifer diterpene internal sequence (CDIS), spanning the regions of exons 4, 5, and 6, and a small portion of exon 7, which includes introns IV, V and VI of all class I genes (Figure 3.8). In addition, by considering the comparison with the gene encoding for the limonene synthase in *Arabidopsis* (*AtLS*), it was confirmed that the class III genes, which include mono-, sesqui- and di-terpene synthase genes involved in secondary metabolism in angiosperms, derive from class II types by a further loss of intron VII and sequential loss of introns IX and X (Figure 3.8). According to Trapp and Croteau (2001), the class III genes contain the six conserved introns (III, VIII, XI-XIV) that are found in all *TPS* genes (Figure 3.8).

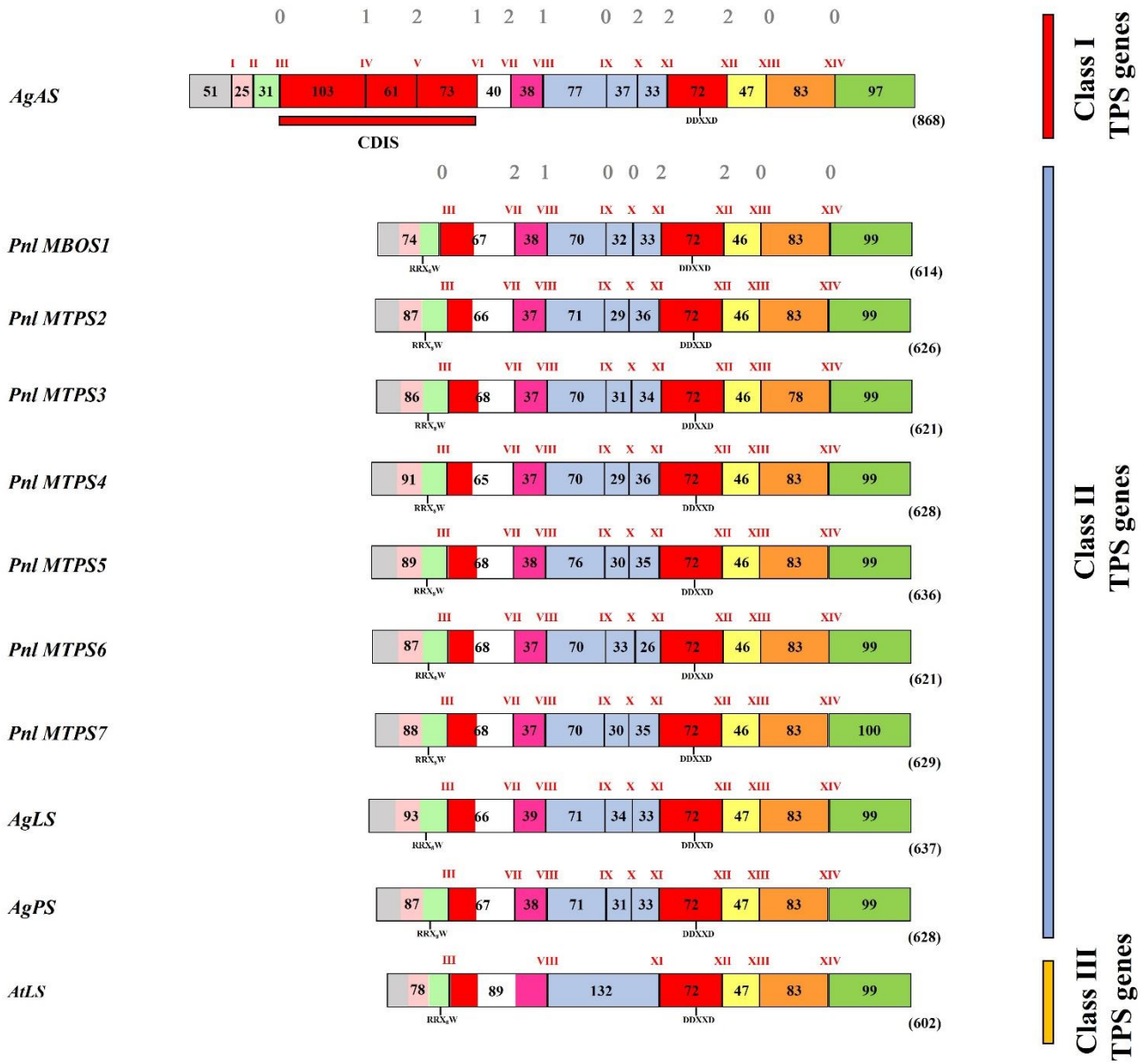


Figure 3.8. *Genomic organization of class I, class II and III terpene synthase genes sensu Trapp and Croteau (2001). Black vertical slashes represent introns (indicated by Roman numerals) and are separated among each other by colored boxes with indicated lengths in amino acids, representing exons. The numbers above the introns of the first row from the top represent the intron phase type classification according to Li (1997) and indicate conservation throughout the plants' TPS genes. Schematization, intron numbers, and exon coloring scheme are based upon Trapp and Croteau (2001). Genomic DNA sequences compared are as follows: AgAS, Abies grandis abietadiene synthase (NCBI accession no. AF326516), AgLS, Abies grandis limonene synthase (AF326518); AgLS, Abies grandis pinene synthase (AF326517); AtLS, Arabidopsis thaliana putative limonene synthase (Z97341); Pnl MBOS1/Pnl MTPS2-7 denote the MBOS and MTPS genes isolated from Calabrian pine in the present study.*

➢ *3.3.5 Transcripts profiling of Calabrian pine monoterpene synthase genes reveal differential expression across different tissues and suggest their putative roles in the biosynthesis of monoterpenes*

All the seven Calabrian pine *MBOS/MTPS* genes were shown to be constitutively expressed in all the five tissues studied, but their transcription levels were highly variable (Figure 3.9 and Table 3.5).

Compared to the six MTPS genes, *Pnl MBOS1* was highly expressed in all the five tissues studied, with the highest transcript levels detected in LS. The expression levels of such gene were also comparatively high in YN, MN and R, while the lower amount of transcripts were detected in IS (Figure 3.9 and Table 3.5). Based on protein sequence homology (see above, Figure 3.4), the protein encoded by *Pnl MBOS1* may be involved in the synthesis of hemiterpenes, such as isoprene and/or its isomer MBO. Isoprene is known to be produced by many plant groups, especially Angiosperms trees, while in gymnosperms this hemiterpene is known to be emitted from species belonging to the genus *Picea*, including *P. abies* and *P. glauca*, but not from species belonging to the genus *Pinus* (Lerdau *et al.* 2003; Loreto, 2015). Instead, the related hemiterpene MBO is emitted by *Pinus* species native of Northern America, e.g., *P. ponderosa*, *P. contorta* and *P. jeffreyi* (Gray *et al.* 2011; Harley *et al.* 1998). A quali/quantitative analysis of the hemiterpenes was not carried out in the five Calabrian pine tissues examined, because the identification and quantification of these compounds requires dedicated sampling equipment and analytical procedures, such as selected-ion flow-tube (SIFT)-MS (Lehnert *et al.* 2020), respect to those used for the identification and analysis of the C_{10} monoterpenes. However, the comparatively high expression level detected in the present study for *Pnl MBOS1*, suggests that the hemiterpenes could be produced in, and conceivably emitted by, different tissues of Calabrian pine. Based on these findings, further studies will be planned to elucidate the molecular and biochemical basis of hemiterpene formation in Calabrian pine.

Among the five tissues analysed, the relative amount of transcripts for each of the six MTPS genes was the highest in LS (Figure 3.9 and Table 3.5), which is also the tissue that accumulated the largest amount of monoterpenes (Figure 3.3). By considering all the five tissues, the highest level of expression was detected for *Pnl MTPS6* and *Pnl MTP5*, compared to the remaining four *MTPS* genes (Figure 3.9 and Table 3.5).

In addition to LS, the expression levels of *Pnl MTPS6* were also comparatively high in MN and IS, and moderate in R and YN (Figure 3.9 and Table 3.5). The tissue-specific expression pattern of *Pnl MTPS6*, found to be highly homologous to *P. contorta* and *P. banksiana MTPS* genes encoding β-phellandrene synthase, appeared to be indeed consistent with the tissue-specific levels of β-

phellandrene, which was found to be the predominant monoterpene in three of the five tissues studied, namely MN, IS and YN (Figure 3.3 B-F).

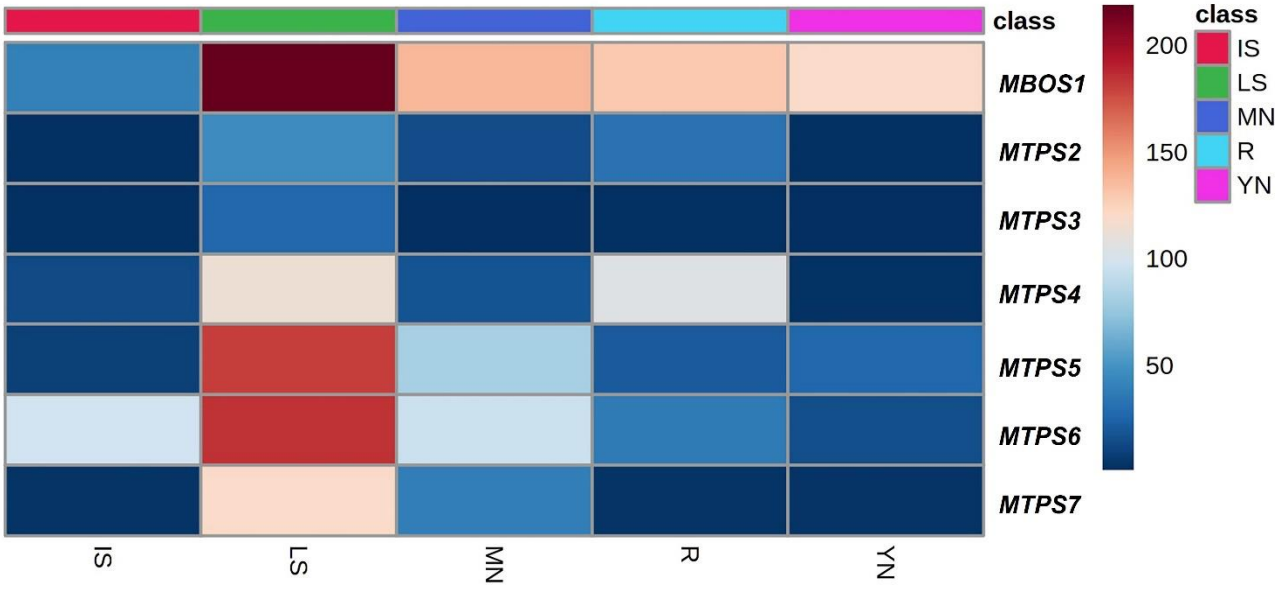


Figure 3.9. *Heat maps of the relative expression levels of the hemi-terpene synthase (MBOS) and of the monoterpene synthases (MTPS) genes isolated from five different tissues of Calabrian pine. The expression data of each gene were normalized using the geometric average of the two reference genes CYP and upLOC. Relative expression levels were calculated by setting a value of 1 for the lowest value among the seven genes in the five tissues considered (MTPS3 in YN). For each gene, the differences in the relative expression levels were shown in colour according to the scale and statistical evaluation of the differences among the seven genes in the five tissues is reported in Table 3.5. IS: bark and xylem combined from the interwhorl stems; LS: bark and xylem combined from the leader stem; MN: mature needles; R: roots; YN: young needles.*

Pnl MTPS5, which based on sequence homology was predicted to produce α-pinene (see section 2.3, above), was highly expressed in LS and MN tissues, followed by moderate levels of transcripts detected in YN and R, and low levels in IS (Figure 3.9 and Table 3.5). A similar expression pattern was observed for *Pnl MTPS7*, whose predicted protein could also be involved in the synthesis of α-pinene, although its transcription levels were significantly lower than those of *Pnl MTPS5*. By matching the expression of the two genes with the corresponding monoterpene profiles, a remarkable level of correlation was found. Indeed, the highest amount of α-pinene was detected in LS, followed by MN, YN, R and IS tissues (Figure 3.3 B-F).

The expression levels of *Pnl MTPS4* were comparatively high in LS and R, respect to the very low amount of transcripts detected in YN, MN and IS (Figure 3.9 and Table 3.5). Again, tissue-specific gene expression levels appeared to be consistent with the corresponding monoterpene profiles; indeed, the predicted protein sequence of *Pnl MTPS4* exhibited high sequence identity with four proteins from *P. contorta* and *P. banksiana*, which were shown to produce β-pinene as their major product. As matter of fact, compared to the other *MTPS* genes, *Pnl MTPS4* showed the highest level of expression in roots (R), in which the β-pinene was the predominant monoterpene (Figure 3.3 F).

Finally, *Pnl MTPS2* and *Pnl MTPS3* exhibited the lowest transcript levels among the six isolated *MTPS* genes in all the five tissues studied (Figure 3.9 and Table 3.5). Based on sequence homology, the encoded proteins from *Pnl MTPS2* and *Pnl MTPS3* were predicted to produce α-terpineol and δ-3-carene, which were found to accumulate in very low amount in the Calabrian pine tissues studied here (Figure 3.3 B-F).

In summary, transcript profiling of the Calabrian pine *MTPS* genes revealed differential expression across the different tissues and were found to be consistent with the corresponding monoterpene profiles, suggesting potential roles for the six *MTPS* genes in the biosynthesis of monoterpene compounds in this non-model species. However, because it is not always possible to predict the catalytic capability of the TPSs based only on the sequence similarity, the specific functions of the isolated genes need to be further verified. Moreover, the genes coding for enzymes potentially involved in the synthesis of some monoterpenes that accumulate in comparatively low/moderate amounts in some Calabrian pine tissues, such as limonene, bornyl acetate and α-tujene (see Figure 3.3 B-F), have not been identified. The isolation of FL cDNAs encoding for less represented monoterpene components of conifer oleoresin is fraught with difficulties arising first from the scarcity of retrievable sequences in public databases: as a matter of fact, while monoterpene synthases producing limonene have been isolated from a few conifers, including Norway spruce (Martin *et al.* 2004), Sitka spruce (Byun-McKay *et al.* 2006) and grand fir (Bohlmann *et al.* 1997), those responsible for the biosynthesis of (+)-bornyl diphosphate, the likely precursor of bornyl acetate, and of α-tujene have been isolated and characterized in *Salvia officinalis* (Wise *et al.* 1998) and *Litsea cubeba* (Chang and Chu, 2011), respectively, but never in any conifer species. Secondly, those less represented monoterpenic components might conceivably result from the transcription of low-expression/under-represented genes, whose characterization is difficult on the background of a very complex genetic family as MTPS is in conifers.

This notwithstanding, studying those less represented monoterpenes at the molecular levels could be worth, because some of them could undergo inductive dynamics in response to a wealth of environmental cues: as a matter of fact, in a previous work of ours conducted on adult individuals of Calabria pine thriving in the same environment of the present study (Foti *et al.* 2020) we found that the foliar emissions of bornyl acetate and, to a lesser extent, of β-ocimene, differentiated plants infested by the caterpillars of the pine processionary moth (*Thaumetopoea pityocampa*) from their respective non infested controls, being higher in the former during the period of maximal trophic activity of the larvae. On such basis, we are currently pursuing the isolation of FL cDNAs coding for the enzymes responsible to produce limonene, bornyl acetate and α-tujene in Calabrian pine, to decipher their possible functional and ecological roles in plant-environment biotic and abiotic interactions.

Table. 3.5 *Statistical evaluation of the differences (one-way ANOVA followed by Tukey test) among the relative expression levels of MBOS and MTPS genes in five different tissues of Calabrian pine. The expression data of each gene were normalized using the geometric average of the two reference genes CYP and upLOC. Relative expression levels of the different MBOS and MTPS genes were referred to a calibrator, set to the value 1, which was represented by the gene in the five tissues with the lowest expression (MTPS3 in YN). Different letters denote significant differences according to the Tukey's test ($p < 0.01$). YN: young needles; MN: mature needles; LS: bark and xylem combined from the leader stem; IS: bark and xylem combined from the interwhorl stems; R: roots.*

MTPS 1 _LS	218,873	[a]	± 28.42
MTPS 6_LS	185,546	[a]	± 21.82
MTPS 5_LS	180,853	[a]	± 22.09
MTPS 1_MN	137,811	[b]	± 14.98
MTPS 1_R	129,661	[b]	± 13.33
MTPS 1_YN	120,568	[bc]	± 14.55
MTPS 7_LS	120,355	[bc]	± 14.52
MTPS 4_LS	113,445	[bc]	± 12.97
MTPS 6_R	105,215	[bc]	± 11.61
MTPS 6_IS	98,0454	[c]	± 10.89
MTPS 6_MN	95,839	[c]	± 9.06
MTPS 5_MN	82,6652	[c]	± 10.08
MTPS 2_LS	45,5297	[d]	± 5.83
MTPS 1_IS	40,0133	[d]	± 5.27
MTPS 7_MN	38,7512	[d]	± 4.74
MTPS 4_R	36,3261	[de]	± 4.23
MTPS 2_R	32,1126	[de]	± 4.86
MTPS 3_LS	27,2339	[de]	± 3.80
MTPS 5_YN	27,1918	[de]	± 3.43
MTPS 5_R	21,3925	[de]	± 3.00
MTPS 4_MN	17,3541	[ef]	± 2.80
MTPS 6_YN	15,1923	[ef]	± 2.90
MTPS 2_MN	14,0932	[f]	± 2.76
MTPS 4_IS	13,318	[f]	± 2.56
MTPS 5_IS	9,31126	[f]	± 2.20
MTPS 7_R	4,37264	[g]	± 1.18
MTPS 7_YN	4,23804	[g]	± 1.37
MTPS 7_IS	3,66427	[g]	± 1.26
MTPS 4_YN	3,40483	[g]	± 1.16
MTPS 2_IS	2,39381	[g]	± 0.89
MTPS 3_IS	2,31444	[g]	± 0.77
MTPS 3_R	2,25211	[g]	± 0.66
MTPS 2_YN	1,95116	[g]	± 0.43
MTPS 3_MN	1,2958	[g]	± 0.38
MTPS 3_YN	1	[g]	± 0.30

3.4. CONCLUSIONS

The importance of terpenes in the physiological and ecological processes in plants, and in conifers, as well as their current and potential practical uses, deserve adequate consideration. In the present study we carried out for the first time, to the best of our knowledge, a quantitative analysis of monoterpene composition in different tissues of the non-model conifer species *Pinus nigra* subsp. *laricio* (Calabrian pine), namely young needles, mature needles, leader stem, interwhorl stem and roots. In a bid to understand the molecular mechanism regulating terpene synthesis in the studied species, we have also faithfully isolated and characterized monoterpene synthase genes, in an attempt to predict their involvement in the specialised monoterpenoid metabolism.

All the Calabrian pine tissues examined indicated the presence of the same fourteen monoterpenes. It was shown that β-phellandrene is the most abundant monoterpene in the young and mature needles and in interwhorl stem, while leader stem and roots show α-pinene and β-pinene, respectively, as the most abundant compounds. However, the leader stem revealed to contain the major quantity of monoterpenes among all the tested tissues. Taken together, these results indicate a tissue-specificity in the quantitative composition of C_{10} terpenoids.

Phylogenetic analysis of the *Pinus* members of the terpene synthases d1-clade allowed the recognition of seven distinct groups. By examining the members of each phylogenetic group for their conserved regions, it was possible to design specific primers to be used for isolating from Calabrian pine seven FL transcripts, denoted as *Pnl MBOS1/Pnl MTPS2-7*, each belonging to one of the above phylogenetic groups. The subsequent analysis of the deduced amino acid sequences allowed to predict the potential roles of each of the Calabrian pine MTPSs in the synthesis of the monoterpenoids identified in this same species. Gene expression analysis revealed that transcripts profiling of the Calabrian pine MTPSs genes have differential abundances across the different tissues. Such tissue-specific expression profiles were found to be consistent with the corresponding monoterpene profiles, suggesting the potential involvement of the isolated *MTPS* genes in the biosynthesis of monoterpenoids.

Finally, the FL *MBOS/MTPS* cDNAs from Calabrian pine were the basis for isolating the corresponding complete genomic sequences, for each of which the exon/intron structure was determined. This filled a knowledge gap in the genomics of the *MTPSs* genes in *Pinus* spp., since no complete genomic sequence has been characterized so far in the non-model conifer species studied here.

The study of monoterpene synthase genes and of their putative functions in conifers appears to be relevant because of the multiple functions that monoterpenes can provide, not only in terms of physiological and ecological roles in plant fitness, but also for their proneness to be employed in a vast array of bio-based technological fields, including the biological control of plant pests and pathogens.

4. CHAPTER 4: PROFILING VOLATILE TERPENOIDS DURING THE INFESTATION BROUGHT ABOUT BY THE PINE PROCESSIONARY MOTH ON CALABRIAN PINE STANDS WITHIN THE ASPROMONTE NATIONAL PARK (SOUTHERN ITALY)

Adapted from: Foti et al. (2020) **Profiling volatile terpenoids during the infestation brought about by the pine processionary moth on Calabrian pine stands within the Aspromonte National Park (Southern Italy).** Plants, 9, 1362. https://doi.org/10.3390/plants9101362

Abstract: Terpenoids make up the biggest and most diversified class of chemical substances discovered in plants, encompassing over 40,000 individual compounds. In conifers, the production of terpenoids, either as oleoresin or emitted as volatile compounds, play an important role in the physical and chemical defense responses against pathogens and herbivores. In the present work, we examined, for the first time to the best of our knowledge, the terpenic defensive relations of Calabrian pine [*Pinus nigra* subsp. *laricio* (Poiret) Maire], facing the attack of the pine processionary moth [*Thaumetopoea pityocampa* (Denis and Schiffermüller, 1775)], brought about in the open on adult plant individuals growing at two distinct forest sites. Among the volatile terpenoids emitted from pine needles, bornyl acetate [(4,7,7-trimethyl-3-bicyclo[2.2.1]heptanyl) acetate] was the most frequently and selectively associated with the infestation, increasing during the period of most intense trophic activity of the caterpillars (defoliation), and decreasing thereafter. Although further work is needed to clarify whether the observed response reflects defense reactions and/or they are involved in communication among the infested plants and their biotic environment, the present results boost the currently growing interest in the isolation and characterization of plant secondary metabolites that can be used to control pests, pathogens and weeds.

Keywords: Calabrian pine; *Pinus nigra* subsp. *laricio* (Poiret) Maire; pine processionary moth; *Thaumetopoea pityocampa* (Denis and Schiffermüller, 1775); terpenoids; bornyl acetate; green leaf volatiles; foraging behavior; headspace analysis.

4.1. INTRODUCTION

Terpenoids, also referred to as terpenes or isoprenoids, make up the biggest and most diversified class of chemical substances discovered in plants, encompassing over 40,000 individual compounds (Tholl, 2015; Singh *et al.* 2014, Abbas *et al.* 2017; Alicandri *et al.* 2020). The evolutionary success of the terpenoid metabolites largely depends on the flexibility of building molecules of various sizes. Indeed, terpenoids, arising from the two basic five-carbon (C5) isoprenoid units, namely isopentenyl diphosphate and its isomer, dimethylallyl diphosphate, can be categorized as hemiterpenoids (C5), monoterpenoids (C10), sesquiterpenoids (C15), diterpenoids (C20), triterpenoid (C30), tetraterpenoid (C40) or polyterpenoids (C5n), based on the number of C5 units they contain (Alicandri *et al.* 2020; Tholl and Lee, 2011).

While terpenoids are known to play essential primary functions as precursors of phytohormones and growth regulators, photosynthetic pigments, electron carriers, and key components of membrane structures, "secondary" terpenoid metabolites have been identified as having a range of specialized roles in plant/environment and plant/plant interactions (Tholl, 2015; Abbas *et al.* 2017; Zhou *et al.* 2012). Low-molecular-weight terpenoids such as isoprene, monoterpenoids, sesquiterpenoids, and diterpenoids, which are volatile, semi-volatile or non-volatile at ambient temperature, respectively, are involved in plant defense from abiotic stress and in many above- and below-ground biotic interactions (Tholl, 2015; Abbas *et al.* 2017; Loreto *et al.* 2014).

The involvement of induced volatile terpenoid compounds in attracting natural enemies of pathogens and herbivores is also well documented (reviewed in Gols, 2014; Pierik *et al.* 2014). Such indirect defense strategy is used by plants to protect their photosynthetic tissues from pathogens and herbivores, as well as to limit insect oviposition (Buchel *et al.* 2011).

Besides their role in the interaction with herbivores and their antagonists, constitutive and induced volatile terpenoids can act as interspecific, intraspecific, and intraplant signals to promote defense responses in nearby plants or in healthy tissues of the same plant (Tholl, 2015; Heil, 2014). Monoterpenes and sesquiterpenes, are particularly suited as long-distance chemical messengers, because of their low-molecular-weight, high vapor pressure at ordinary temperatures, and lipophilic nature, which facilitates their interactions with membrane systems (Tholl, 2015; Abbas *et al.* 2017; Alicandri *et al.* 2020). However, there is still a poor understanding of the molecular mechanisms involved in plant-to-environment communication mediated by volatile compounds, and especially so as far as non-model plant species are concerned.

In conifers, the production of terpenoids, either as oleoresin or emitted as volatile compounds, play an important role in the physical and chemical defense responses against pathogens and herbivores (Tholl, 2015; Keeling and Bohlmann, 2006; Zulak and Bohlmann, 2010; Hall *et al.* 2011).

The objective of the present work was to study, for the first time, to the best of our knowledge, the terpenic defensive relations of an endemic conifer of the Calabria territory, namely Calabrian black pine [*Pinus nigra* subsp. *laricio* (Poiret) Maire], facing the attack of the pine processionary moth (PPM) [*Thaumetopoea pityocampa* (Denis and Schiffermüller, 1775)] brought about on adult forest stands in the open.

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, 2016). Outside Europe, it has become naturalized in the midwestern states of the U.S, normally south of the normal 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, 2016). *Pinus nigra* subsp. *laricio* (Poiret) is one of the six subspecies of black pine; it is found in Corsica, and in southern Italy, with a natural range extending from Calabria to Sicily (Nicolaci *et al.* 2015). In Calabria, where it is considered an endemic species, it grows on the Sila and Aspromonte mountains, and represents an essential element of the forest landscape, playing an important role not only in soil conservation and watershed protection, but also in the local forest economy (Nicolaci *et al.* 2015).

There are about 40 different species of processionary moths, the most widespread in Italy being *Thaumetopoea pityocampa*. This is one of the most important defoliators of pine trees in the Mediterranean region, and its eruptive dynamics inflict serious economic and ecological losses, with desiccation and defoliation which can even lead to the death of the attacked tree, as well as severe allergic reactions in humans and other mammals (Werno and Lamy, 1990). This insect is limited by the low winter temperatures, and the current climate warming is consequently expanding the limits of its distribution to more northerly territories and to higher altitudes, where it was not previously common (Battisti *et al.* 2005). *Pinus nigra* is among the preferred species of such insect.

To shed light on the chemical interactions among Calabrian pine and PPM, we aimed mainly to determine whether the concentrations of volatile terpenes differ between infested individuals (I plants in the following) and unattached ones (not infested plants; NI, in the following), from both the qualitative and semi-quantitative point of view, to evaluate which substance(s) might be involved in plant protection. To do this, we measured the volatile terpenoids emitted from pine needles by means of gas chromatography/mass spectrometry (GC/MS) analysis.

To increase its informative and predictive value, the study of plant-host interactions was matched with the biological cycle of the insect, by collecting plant material for analysis throughout almost all of its phenological phases, to understand in which of these the plant might show the need to modulate the emission of its volatile terpenes to possibly implement its defensive strategies. To such aim, samples of Calabrian pine needles were periodically collected at each of two different forest locations, namely Bova and Canolo, within the premises of the Aspromonte National Park, Southern Italy. The first sampling of pine needles was carried out when the young PPM larvae spend most of their time inside their nests and come out sporadically to start feeding (NEST stage, in the following). The second and third samplings were coincident with the maximal trophic activity of the insect, occasionally leading to massive defoliation (denoted as DEFO-1 and DEFO-2, in the following). The fourth and fifth samplings took place at the beginning of the "subsidence" of the insect underground, preceded by the well-known “procession” of the PPM caterpillars (denoted as PROCE-1 and PROCE-2, in the following). Finally, the sixth and final sampling was carried out during the stage in which adult PPM females lay their eggs (oviposition; OVI, in the following; see Materials and Methods for further details).

An additional aim of the present study was to characterize, for the first time, the blend of volatile terpenes emitted by the studied forest stands under the peculiar environmental conditions occurring in the Aspromonte territory, to learn lessons concerning forest management practices and the possible technological exploitation of these substances.

4.2.MATERIALS AND METHODS

➢ *4.2.1. Plants sampling sites*

The Calabrian pine-PPM interaction which is the object of the present work was studied in two artificial and pure pine plantations located near Bova Superiore (Bova, in short) and Canolo Nuova (Canolo, in short), which are located at the southern and the northern limits, respectively, of the Aspromonte National Park, about 40 Km apart from each other as the crow flies, in the south-most part of continental Italy. Access to the above study areas and sampling of plant material was approved and authorized by the Aspromonte National Park Authority within the framework of an ad hoc research agreement with the Department of Agriculture of the Mediterranean University of Reggio Calabria, Italy.

The sampling area of Canolo (38° 33'33" N, 16° 15'63" E; altitude 904 MASL, exposed South), at the northern limit of the Park, hosts a main arboreal composition of Calabrian pine and a secondary composition of brushwood with the bracken fern [*Pteridium aquilinum* (L.) Kuhn], with signs of plant renewal of *Quercus* spp. These are mostly adult monoplane high forests, derived from artificial afforestation carried out in the Seventies of the previous century. Two Calabrian pine sampling plots were chosen there, one with plants showing visible signs and symptoms of PPM infestation, and another representing the not infested control. Three infested plants, each showing several and characteristic PPM nests also on branches close to the ground, were chosen. The mean percentage of defoliation due to the PPM caterpillars was around 40%. As many control plants, not showing any visible sign or symptom of PPM infestation, were selected from a nearby forest plot (38 ° 33'92" N; 16 ° 14'99" E; altitude 950 MASL, exposed South-East), about 100 m apart from the infested plot as the crow flies. Once selected, the Calabrian pine individuals in the two plots were marked with cuttings of white-red signal tape around the trunk bases, in order to facilitate their retrieval. Always these same marked plants were used for needles collection throughout the entire sampling campaign (see below).

The sampling area of Bova is located on the eastern slopes of the Aspromonte massif, at the southern limit of the park. Similarly, to the Canolo sampling site (see above), also Bova hosts a main arboreal composition of Calabrian pine, which forms mostly adult monoplane high forests derived from artificial afforestation carried out in the Fifties of the previous century. As for Canolo (see above), two Calabrian pine sampling plots were chosen at Bova, one with plants showing visible signs and symptoms of PPM infestation, and another acting as the not infested control. Three infested plants, each showing several and characteristic PPM nests also on branches close to the ground, were chosen from a plot located at 38 ° 1'58"N, 15 ° 26'46 "E, 1194 MASL, exposed South-East. The mean percentage of defoliation due to the PPM caterpillars was around 50%. As many control plants, not showing any visible sign or

symptom of PPM infestation, were selected from a plot located at 38 ° 1'58"N, 15 ° 26'46 "E, 1194 MASL, exposed North-West, about 70 m apart from the infested plot as the crow flies. Once selected, the Calabrian pine individuals in the two plots were marked as described above. Always these same marked plants were used for needles collection throughout the entire sampling campaign (see below).

➢ *4.2.2. Sampling of pine needles*

From each of the two sampling areas previously described, and from each infested or not infested plot within each of them, three samples of needles were collected from as many individual plants. Samples were identified as I1, I2, and I3 for PPM-infested plants, and as NI1, NI2, and NI3 for not infested (control) plants. Needles were collected by means of a pruner from branches located at 3-4 m from the ground, at 15-25 cm from their tips, taking care to avoid branches showing visible symptoms of defoliation and/or damage, whatever the cause. Once excised from the plant, each collected twig was cut into portions of about 10 cm in length, each bearing one or two tufts of needles, placed inside a transparent plastic bag and stored in a thermal bag at 4 °C. Once back into the laboratory, needles samples were stored (24-48 h) in the fridge at 4 °C until GC/MS analysis.

The number and frequency of pine needles samplings from the aforementioned forest plots was planned *a priori* by keeping in mind the progression of the PPM biological cycle. A total of six samplings were carried out in each of the two sampling areas: The first sampling (half of February) coincided with the 3rd phenological stage of the PPM, that is when the young larvae spend most of their time inside the nests and come out sporadically to start feeding on pine needles (denoted as the NEST stage). The second and third samplings (bimonthly during March) were coincident with maximal trophic activity of the insect (4th phenological stage), occasionally leading to massive defoliation (denoted here as the DEFO-1 and the DEFO-2 stages). The fourth and fifth samplings (bimonthly during April) coincided with the 5th phonological stage of the insect, with reduced or no defoliation activity, which marks the beginning of the "subsidence" of the insect underground, preceded by the well-known “procession” of the PPM caterpillars (denoted here as the PROCE-1 and PROCE-2 stages). Finally, the sixth and final sampling (late August) was carried at the end of the pupal stage, upon which, after emerging and mating, adult females lay their eggs on the nearest pines (6th phonological stage, oviposition; denoted here as the OVI stage).

Upon each sampling date, plant material was always collected from both the Bova and Canolo experimental sites during the same day, within a total time span of about three hours.

A voucher specimen of the collected plant material has been deposited in the Department of Agriculture of the Mediterranean University of Reggio Calabria, Italy. A total of 2580 g of Calabrian pine needles were collected during the six sampling campaigns.

➢ *4.2.3. Head space GC/MS analysis of volatiles from pine needles*

Volatiles from pine needles were chemically characterized by means of a Thermo Fisher gas chromatograph apparatus (Trace 1310) equipped with a single quadrupole mass spectrometer (ISQ LT). The capillary column was a TG-5MS 30 m × 0.25 mm × 0.25 μm; the carrier gas was helium with a flow rate of 1 mL min^{-1}. Before the GC/MS analysis, the pine needles samples were taken out from the fridge and left at room temperature for one hour, to allow the emission of volatiles, otherwise inhibited by the low temperatures. Subsequently, needles were selected from the twigs among those not showing evident signs of deterioration or yellowing.

For each sample, one gram of the selected needles was placed inside a SPME incubation vial, the vial closed, and left in a well-balanced position for 90 min. After this time, a SPME device holding a DVB/CAR/PDMS (gray) fiber (Supelco, Bellefonte, PA, USA] was inserted into the vial through its rubber screw cap and left in incubation for 30 min, to allow the sample volatiles to be adsorbed. The fiber was subsequently inserted into the GC injection port, in splitless mode and the volatile substances desorbed.

The gas chromatographic conditions were as follows: isocratic for 3 min at 60 °C, from 60 °C to 240 °C at 6 °C min^{-1}, then isocratic for 4 min at 240 °C. The mass spectra were recorded in EI mode at 70 eV, with scanning at 30-300 m/z. The pine volatilome constituents were identified from their retention indices (KI), calculated as relative to the homologous series of (C5–C36) alkanes analyzed under the same GC/MS conditions, and by comparison with the built-in mass spectra database of the GC/MS apparatus (NIST 2005 and Wiley 7.0).

➢ *4.2.4. Statistical analysis*

The statistical analysis of the volatiles from the Calabrian pine needles was carried out by means of the MetaboAnalyst software, taking into consideration the percentage area of each GC/MS peak. Metabolite concentrations were checked for integrity, and missing values were replaced with a small positive value (the half of the minimum positive number detected in the data). Data were then normalized by a reference sample, by creating a pooled average sample from control groups, transformed through log normalization, to make the metabolite concentration values more comparable

among different compounds, and scaled through the Pareto-Scaling, i.e., mean-centered and divided by the square root of standard deviation of each variable (Araniti *et al.* 2017). Data were then classified through Principal Component Analysis (PCA). If separation was not achieved, data were further analyzed through Partial Least Squares - Discriminant Analysis (PLS-DA), built by using the first two Latent Variables (or components) allowing samples separation. In order to identify the metabolites responsible for the discrimination among the volatiles profiles, the Variable Importance for Prediction (VIP) score was used to select those with the most significant contribution in a PLS-DA model. VIPs are a weighted sum of PLS weights for each variable and measure the contribution of each predictor variable to the model (Hodgson *et al.* 2008). Further, the VIP statistic summarizes the importance of the metabolites in differentiating the study groups (i.e., not infested vs infested plants, in the present case) in a multivariate space. In the present experiment, therefore, the volatiles exhibiting the higher VIPS (≥ 1.4) were assumed to be the most influent variables.

4.3. RESULTS

Headspace GC-MS analysis revealed twenty-one volatile compounds, all of which monoterpenes or sesquiterpenes, released by the Calabrian pine needles (Table 4.1).

When a more restrictive criterion was applied, in which only compounds present in at least two replicates out of three were considered, Tables 2.2 and 2.3 were obtained, in which volatiles from NI and I plants are listed in decreasing order on the basis of % area of each peak over the total area of the chromatogram.

Fig. 4.1 is intended to exemplify that the needles of I plants in Bova seemed to have a more massive terpenes emission than their NI counterparts (Fig. 4.1a), but not in Canolo (Fig. 4.1b).

Fig. 4.2 shows the results obtained by applying the Principal Component Analysis (PCA) to the volatiles dataset from the two sampling locations and to the six sampling times. It is apparent from the results shown that PCA was not able to separate I plants from NI ones, because of the high variability among samples. Therefore, data were further analyzed by using Partial Least Squares - Discriminant Analysis (PLS-DA)

PLS-DA applied to the NEST sampling in Bova (Fig. 4.3) showed that components 1 and 2 alone explained 65.2% of the variability. For NI and I plants, the first two components (1 and 2) helped explain 47.4% and 17.8% of the variability, respectively. In order to identify the volatiles responsible for the discrimination among infested plants and not infested ones, the Variable Importance for Prediction (VIP) scores were calculated. Such analysis revealed that the metabolites with highest VIPS

in the NEST samples from Bova were β-ocimene, β-myrcene, bornyl acetate and terpinolene, all of which were up-accumulated in the PPM-infested plants (Fig. 4.3)**.**

Table 4.1. *Volatile organic compounds identified by GC/MS analysis in the head space of Calabrian pine needles. RT, retention time; KI, retention index.*

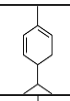
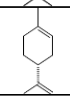
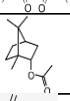
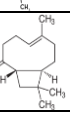
N°	Common name	IUPAC name	Type of terpene	RT	KI	Structural formula
1	Pinene alpha*	(1*S*,5*S*)-2,6,6-Trimethylbicyclo[3.1.1]hept-2-ene ((−)-α-Pinene)	monoterpenoid	7,38	939	
2	Pinene beta*	6,6-Dimethyl-2-methylidenebicyclo[3.1.1]heptane	monoterpenoid	8,35	982	
3	Myrcene beta	7-Methyl-3-methylene-1,6-octadiene	monoterpenoid	8,75	1000	
4	Phellandrene alpha	2-Methyl-5-(1-methylethyl)-1,3-cyclohexadiene	monoterpenoid	9,61	1035	
5	Limonene*	1-Methyl-4-(prop-1-en-2-yl)cyclohex-1-ene	monoterpenoid	9,8	1043	
6	Ocimene beta*	(*Z*)-3,7-Dimethyl-1,3,6-octatriene	monoterpenoid	10,06	1053	
7	Terpinolene*	4-Methyl-1-(1-methylethyl)-1,3-cyclohexadiene	monoterpenoid	11,09	1095	
8	Thymol methyl ethere	2-methoxy-4-methyl-1-propan-2-ylbenzene	monoterpenoid	11,98	1131	
9	Camphor*	1,7,7-Trimethylbicyclo[2.2.1]heptan-2-one	monoterpenoid	12,48	1152	
10	Bornyl acetate	4,7,7-Trimethyl-3-bicyclo[2.2.1]heptanyl) acetate	monoterpenoid	15,82	1292	
11	Gurjunene gamma	(1R,3aR,4R,7R)-1,4-dimethyl-7-prop-1-en-2-yl-1,2,3,3a,4,5,6,7-octahydroazulene	Sesquiterpenes	16,4	1318	
12	Elemene delta	(3*R* , 4*R*) -1-Isopropil-4-metil-3- (prop-1-en- 2-il) -4-vinylcyclohex-1-ene	Sesquiterpenes	16,99	1344	
13	Cubebene alpha*	(1R,5S,6R,7S,10R)-10-methyl-4-methylidene-7-(propan-2-yl)tricyclo[4.4.0.0^{1,5}]decane	Sesquiterpenes	17,27	1357	
14	Copaene alpha*	(1S,6S,7S,8S)-1,3-dimethyl-8-(propan-2-yl)tricyclo[4.4.0.0^{2,7}]dec-3-ene	Sesquiterpenes	17,89	1384	
15	Bourbonene beta	1-methyl-5-methylidene-8-(propan-2-yl)tricyclo[5.3.0.0^{2,6}]decane	Sesquiterpenes	18,82	1427	
16	Caryophyllene*	(1*R* ,4*E* ,9*S*)-4,11,11-Trimethyl-8-methylidenebicyclo[7.2.0]undec-4-ene	Sesquiterpenes	18,88	1430	
17	Bisabolene alpha	(*E*)-1-Methyl-4-(6-methylhepta-2,5-dien-2-yl)cyclohex-1-ene	Sesquiterpenes	19,63	1466	
18	Humulene*	2,6,6,9-Tetramethyl-1,4-8-cycloundecatriene	Sesquiterpenes	19,93	1480	
19	Murolene gamma*	(1S,4aS,8aR)-7-methyl-4-methylidene-1-propan-2-yl-2,3,4a,5,6,8a-hexahydro-1H-naphthalene	Sesquiterpenes	20,4	1492	
20	Germacrene D*	(1*E* ,5*E* ,8*S*)-1,5-dimethyl-8-(prop-1-en-2-yl)cyclodeca-1,5-diene	Sesquiterpenes	20,18	1502	
21	Candinene gamma*	(1S,4aR,8aR)-7-methyl-4-methylidene-1-propan-2-yl-2,3,4a,5,6,8a-hexahydro-1H-naphthalene	Sesquiterpenes	20,89	1526	

Table 4.2. *Volatile organic compounds from the needles of not infested Calabrian pines detected in at least two replicates out of three at each sampling time in the two sampling sites of Canolo and Bova.*

Common name	Type of terpene	Sampling time											
		1°		2°		3°		4°		5°		6°	
		Area %		Area %		Area %		Area %		Area %		Area %	
		Canolo	Bova	Canolo	Bova	Canolo	Bova	Canolo	Bova	Canolo	Bova	Canolo	Bova
Pinene alpha	monoterpenoid	73,043	82,013	89,573	29,99	91,603	75,727	83,95	78,38	76,787	76,63	73,77	77,38
Pinene beta	monoterpenoid	3,1067	2,32	4,1167	-	2,5267	-	3,0633	2,4033	2,76	2,26	3,98	2,84
Limonene	monoterpenoid	5,285	4,07	2,41	10,79	2,0567	5,68	3,475	4,245	3,935	6,945	4,91	4,99
Ocimene beta	monoterpenoid	4,7867	-	-			-	-	-	3,855	-	6,99	2,62
Terpinolene	monoterpenoid	0,205	-	0,145	0,28	0,14	0,2633	0,22	0,1933	0,36	0,35	0,81	0,16
Camphor	monoterpenoid	-	-	0,05	-	0,1	-	-	-	-	0,2	-	-
Bornyl acetate	monoterpenoid	-	-	-	0,295	-	-	0,035	0,08	-	0,235	0,13	0,04
Cubebene alpha	Sesquiterpenes	0,12	0,195	-	0,265	-	-	-	0,19	-	-	-	0,36
Copaene alpha	Sesquiterpenes	0,145	0,135	-	0,445	-	0,19	-	0,18	0,08	-	0,04	0,28
Caryophyllene	Sesquiterpenes	2,78	2,59	-	1,4833	0,1467	1,9033	0,5	1,2167	2,8667	1,53	2,18	2,70
Humulene	Sesquiterpenes	0,485	-	-	-	-	-	0,055	-	0,33	-	0,16	0,07
Murolene gamma	Sesquiterpenes	-	-	-	0,715	-	0,54	-	0,585	-	-	-	0,31
Germacrene D	Sesquiterpenes	0,51	0,83	-	1,69	-	2,12	-	1,55	1,0433	2,6933	0,1	2,48
Candinene gamma	Sesquiterpenes	0,5	0,565	-	1,23	-	0,6967	-	0,8	0,2033	-	0,03	0,67

Table 4.3. *Volatile organic compounds from the needles of pine processionary moth-infested Calabrian pines detected in at least two replicates out of three at each sampling time in the two sampling sites of Canolo and Bova.*

Common name	Type of terpene	Sampling time											
		1°		2°		3°		4°		5°		6°	
		Area %		Area %		Area %		Area %		Area %		Area %	
		Canolo	Bova	Canolo	Bova	Canolo	Bova	Canolo	Bova	Canolo	Bova	Canolo	Bova
Pinene alpha	monoterpenoid	1,91	44,673	86,83	66,403	89,867	55,06	72,28	60,04	2,04	65,31	81,517	63,003
Pinene beta	monoterpenoid	74,383	7,505	3,4433	4,64	2,6067	16,485	3,59	12,93	2,435	6,565	6,6433	2,94
Myrcene beta	monoterpenoid	5,06	11,33	-	4,3833	-	9,8	8,92	5,04	5,225	-	0,5133	3,23
Phellandrene beta	monoterpenoid	-	17,105	-		-	11,89	-	9,96	-	8,785	-	3,46
Limonene	monoterpenoid	10,11	-	-	8,16	3,8033	-	9,9	-	-	-	7,53	3,27
Ocimene beta	monoterpenoid	2,08	6,24	-	4,545	-	7,485	-	3,94	-	3,88	0,85	0,19
Terpinolene	monoterpenoid	0,4467	0,6067	0,205	0,36	0,1533	0,47	0,485	0,27	0,2833	0,28	0,24	0,31
Bornyl acetate	monoterpenoid	-	0,2867	0,27	1,8033	0,065	0,1333	0,24	0,0767	0,1	0,53	0,2	0,38
Copaene alpha	Sesquiterpenes	0,06	0,22	-	0,21	-	0,0967	-	0,13	-	0,2033	0,1	0,22
Caryophyllene	Sesquiterpenes	1,4867	4,9867	0,4667	3,61	0,27	3,35	1,1	4,39	1,05	6,6033	1,21	5,61
Humulene	Sesquiterpenes	0,225	-	-	0,3867	0,0333	-	0,135	-	0,13	0,965	0,10	0,77
Murolene gamma	Sesquiterpenes	-	0,3867	-	0,93	-	0,275	-	0,48	-	0,265	-	0,33
Germacrene D	Sesquiterpenes	0,145	1,1367	-		-	1,11	-	1,57	-	4,0567	0,25	0,41
Candinene gamma	Sesquiterpenes	-	0,38	-	0,375	-	0,42	-	0,3	-	0,5033	-	0,24

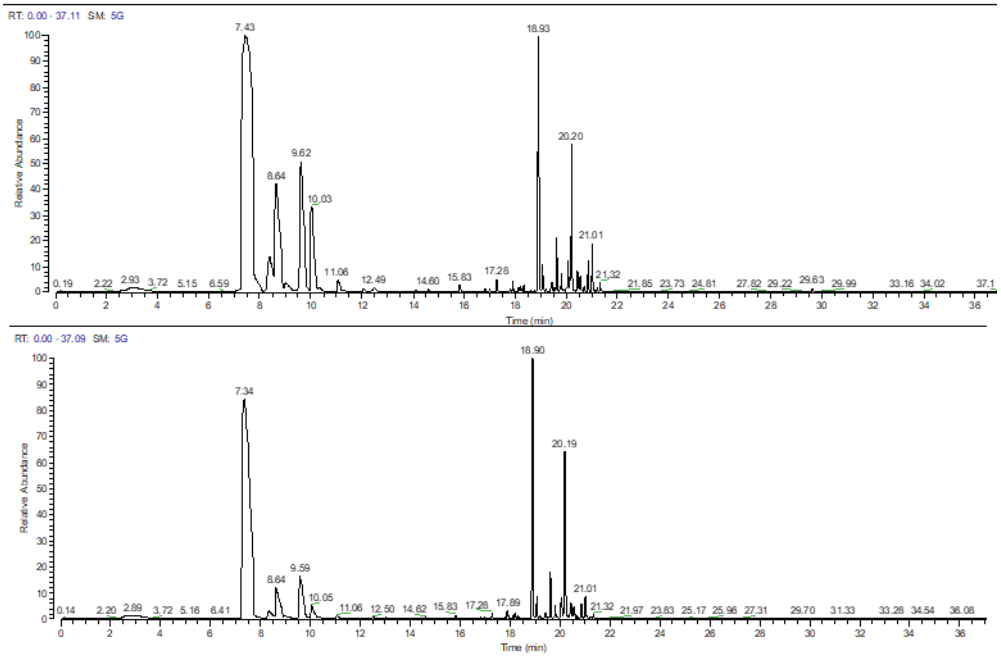


Fig. 4.1a. *A typical headspace-GC/MS chromatogram obtained from Calabrian pine needles collected at Bova. Upper panel, plant infested by the pine processionary moth; lower panel, not infested (control) plant.*

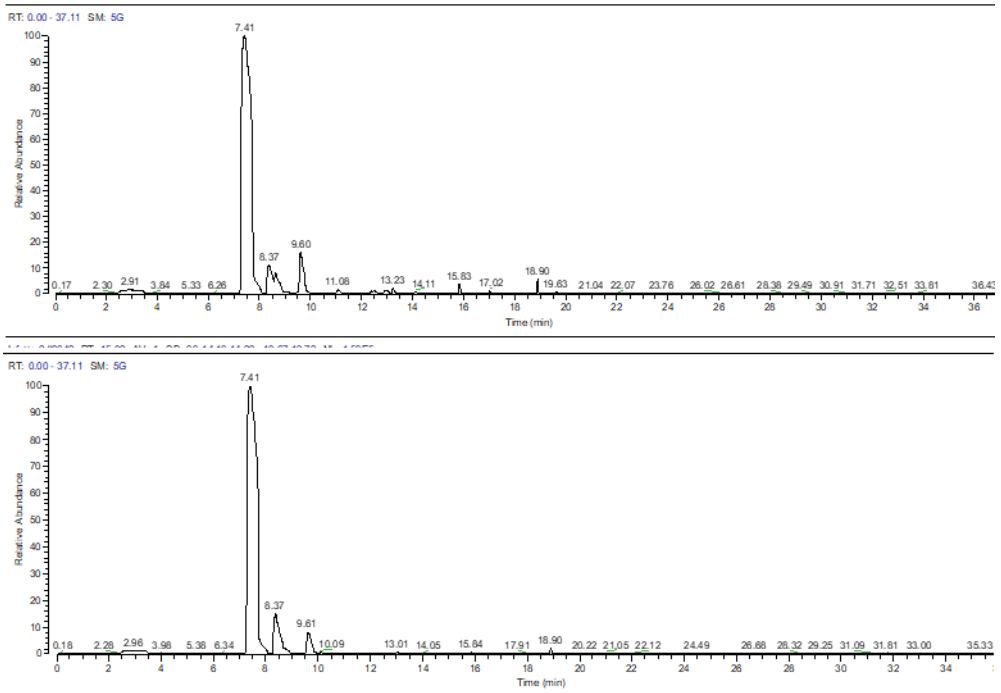


Fig. 4.1b. *A typical headspace-GC/MS chromatogram obtained from Calabrian pine needles collected at Canolo Nuova. Upper panel, plant infested by the pine processionary moth; lower panel, not infested (control) plant.*

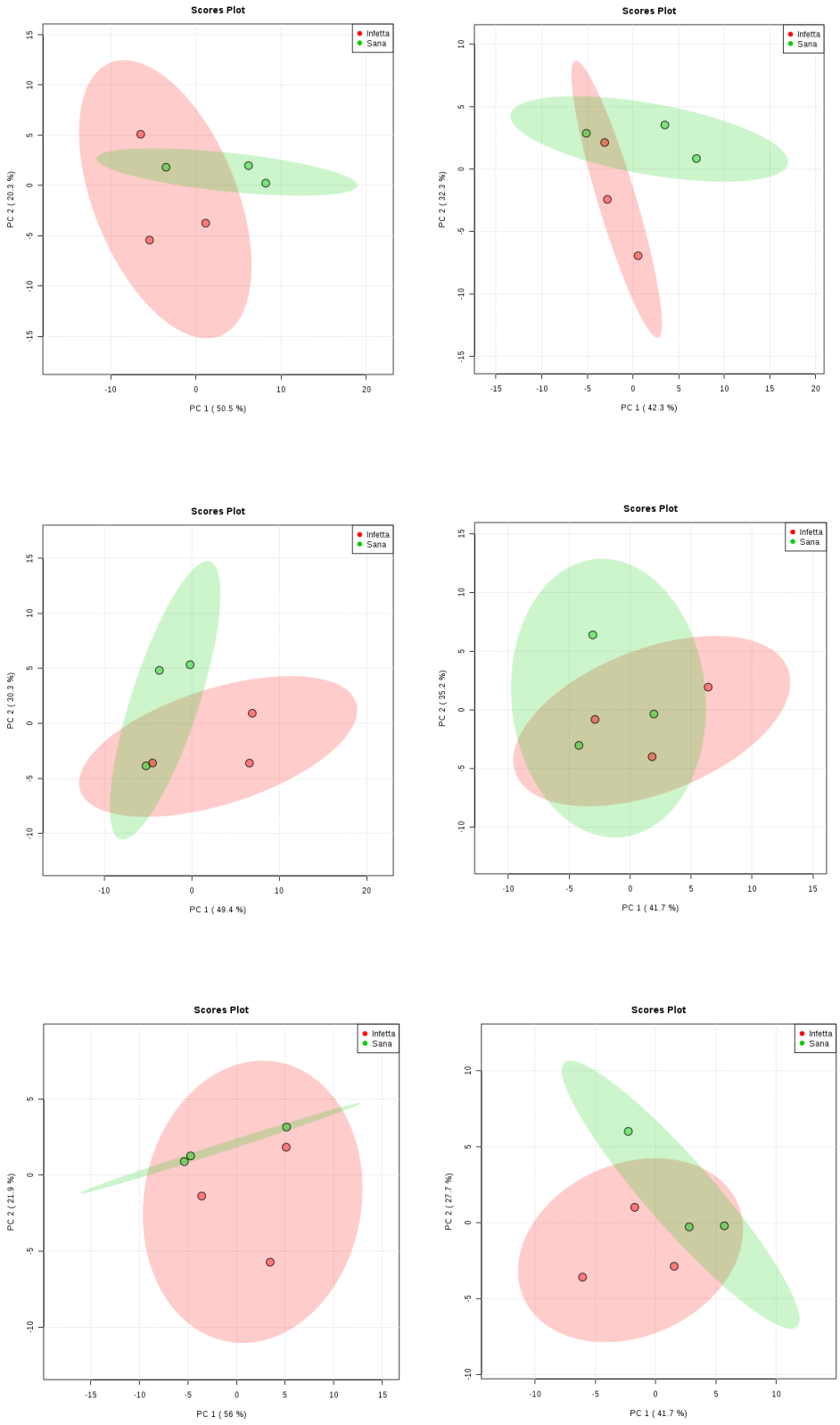


Fig. 4.2. *Principal Component Analysis applied to volatiles emission data obtained from pine processionary moth-infested- (red circles) or not infested (green circles) pine needles collected in Bova (left) or in Canolo (right) during six samplings along the insect biological cycle. (Continues).*

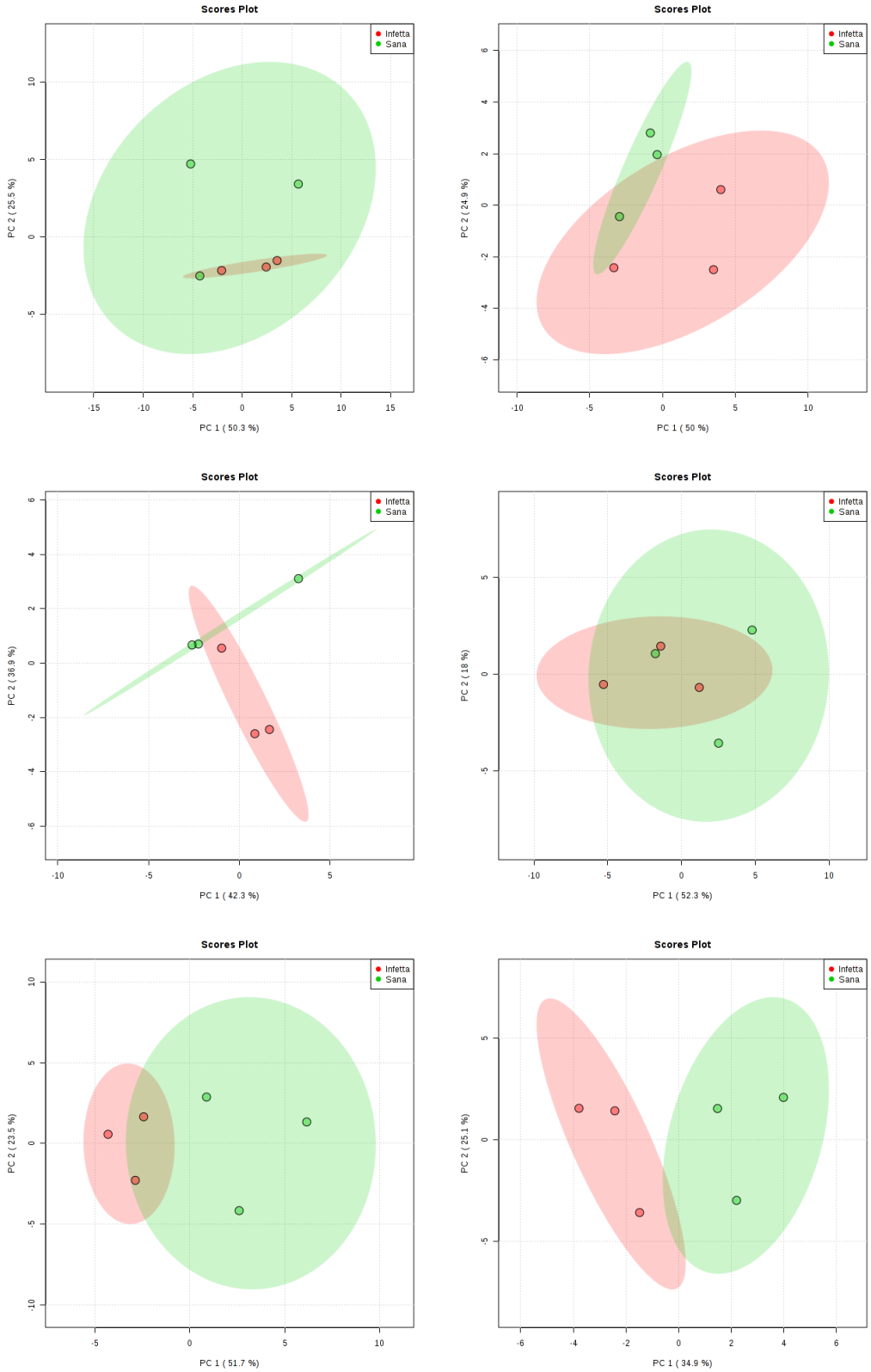


Fig. 4.2. *Continued from the previous page.*

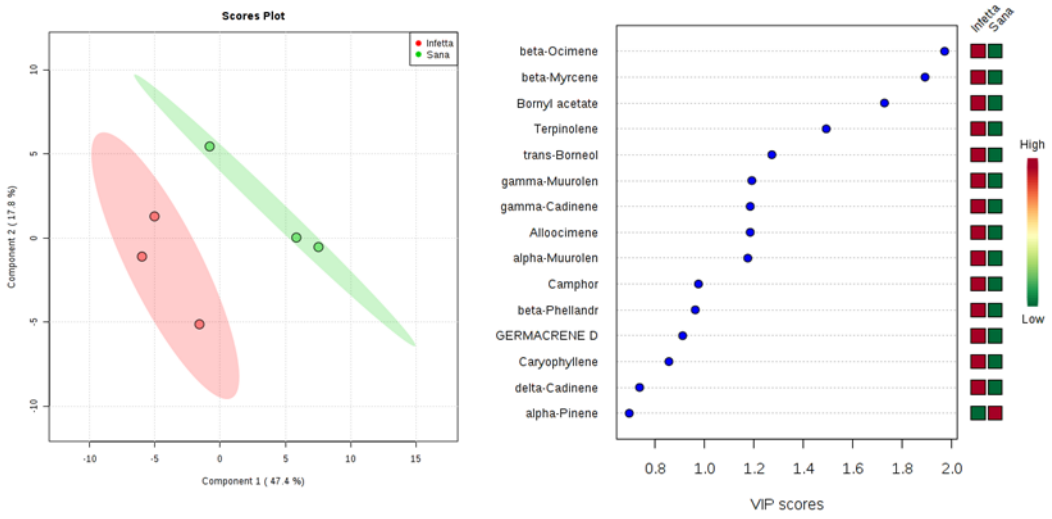


Fig. 4.3. *Partial Least Squares - Discriminant Analysis applied to the volatiles emitted from pine needles collected upon the NEST sampling in Bova. Symbols as in Fig. 4.2.*

In the DEFO-1 sampling in Bova (Fig. 4.4), the components (latent variables) 1 and 2 explained 63.2% of the variability. In particular the PC1 explained 35.6%, whereas the PC2 explained 27.6%. VIPS > 1.4 were found for α-humulene, β-ocimene, bornyl acetate, germacrene D, β-myrcene, α-muurolene and thymol methyl. The first three metabolites, the fifth, and the seventh were up-accumulated in the I plants, whereas the fourth and the sixth in the NI ones.

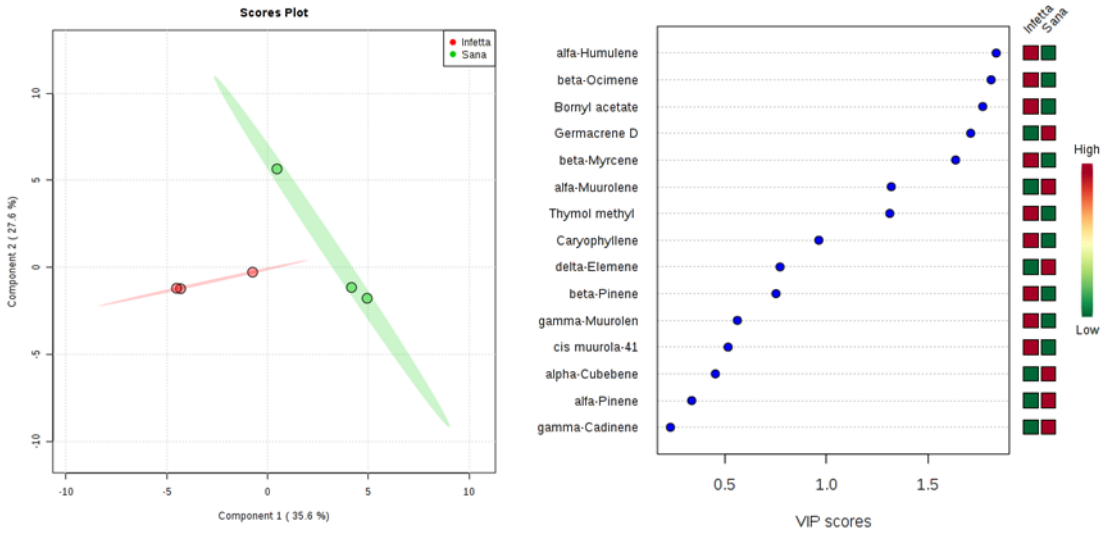


Fig. 4.4. *Partial Least Squares - Discriminant Analysis applied to the DEFO-1 sampling in Bova. Parameters and symbols as in Fig. 4.3.*

In the DEFO-2 sampling in Bova (Fig. 4.5), the components 1 and 2 explained 72.2% of the variability. In particular the PC1 explained 44%, whereas the PC2 explained 28.2%. VIPS > 1.4 were found for thymol methyl, β-ocimene, camphor, germacrene D, bornyl acetate, all of which were up-accumulated in the PPM-infested plants**.**

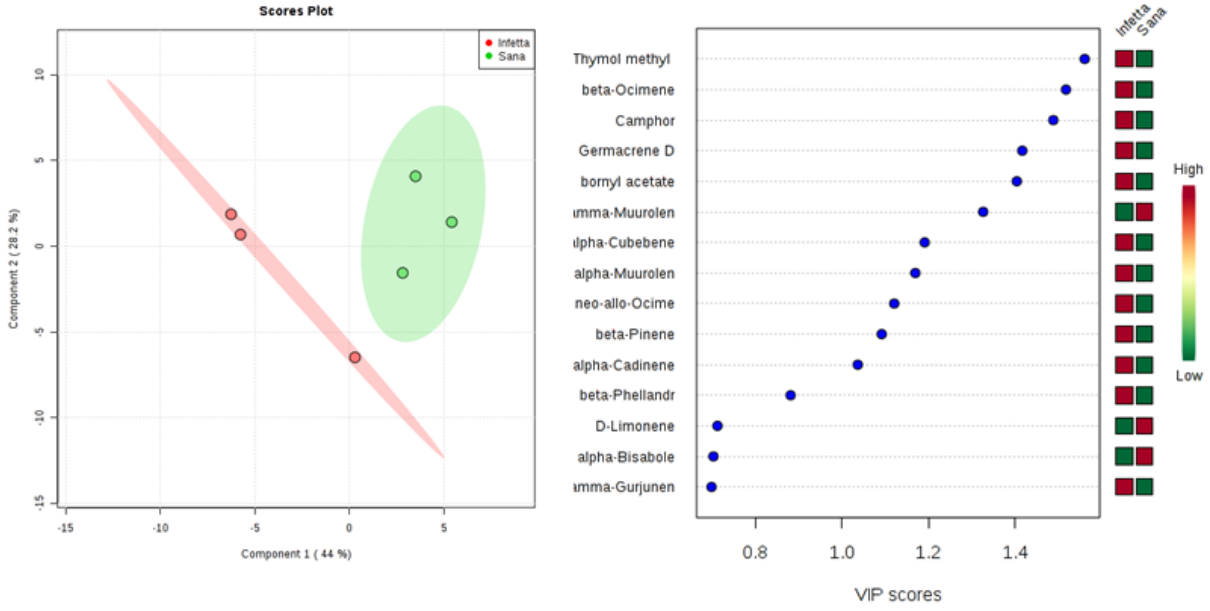


Fig. 4.5. *Partial Least Squares - Discriminant Analysis applied to the DEFO-2 sampling in Bova. Parameters and symbols as in Fig. 4.3.*

In the PROCE-1 sampling in Bova (Fig. 4.6), the components 1 and 2 explained 52.8% of the variability. In particular the PC1 explained 36.8%, whereas the PC2 explained 16%. VIPS > 1.4 were found for β-ocimene, α-bisabolene and caryophyllene. The first and the third metabolite were up-accumulated in the PPM-infested plants, whereas the second in the NI plants.

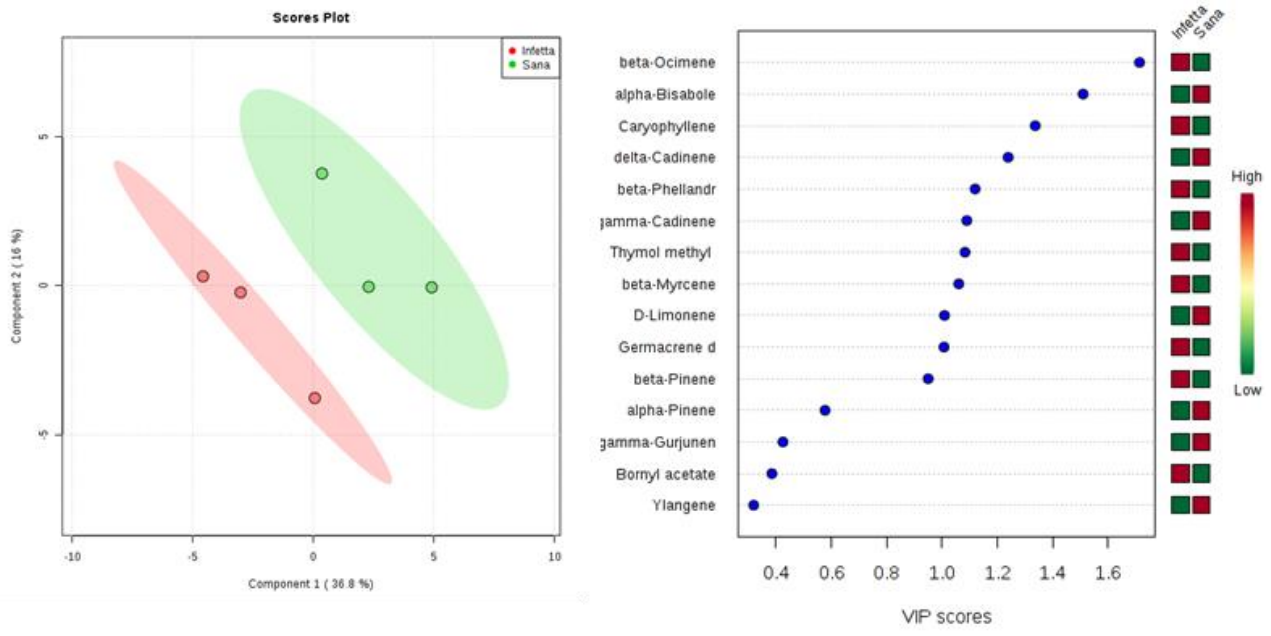


Fig. 4.6. *Partial Least Squares - Discriminant Analysis applied to the PROCE-1 sampling in Bova. Parameters and symbols as in Fig. 4.3.*

In the PROCE-2 sampling in Bova (Fig. 4.7), the components 1 and 2 explained 73.8% of the variability. In particular PC1 explained 42.5%, whereas the PC2 explained 31.3%. VIPS > 1.4 were found for thymol methyl, *trans*-caryophyllene and α-humulene, all of which were up-accumulated in the PPM-infested plants.

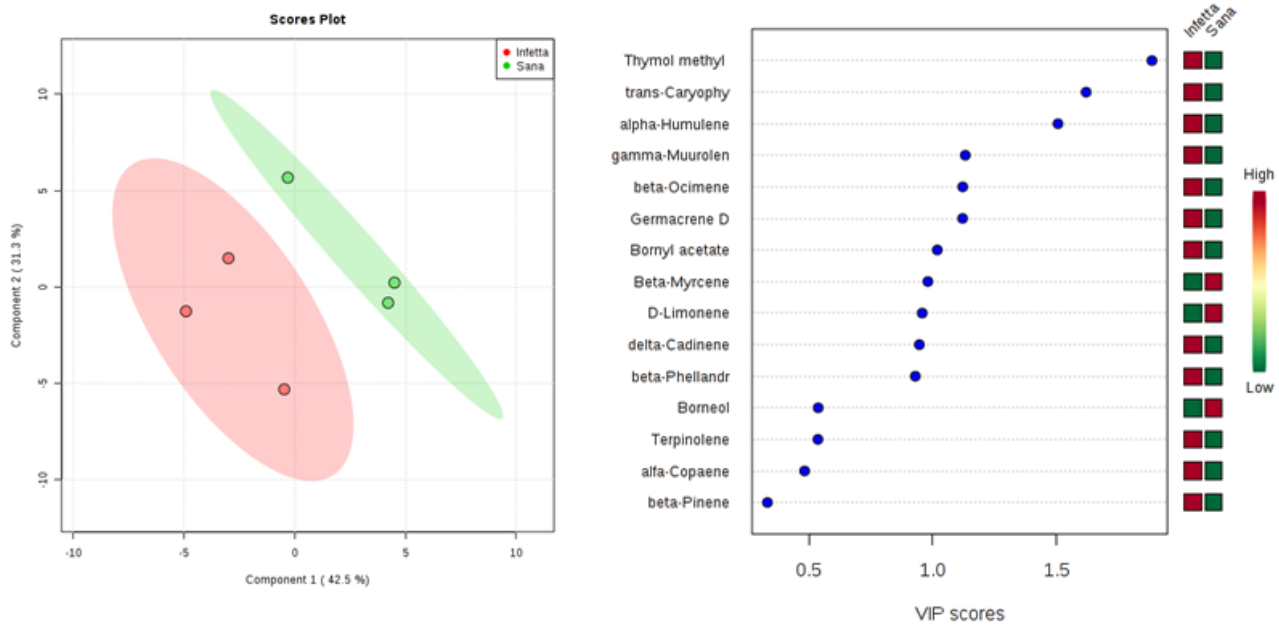


Fig. 4.7. *Partial Least Squares - Discriminant Analysis applied to the PROCE-2 sampling in Bova. Parameters and symbols as in Fig. 4.3.*

In the OVI sampling in Bova (Fig. 4.8), the components 1 and 2 explained 62.7% of the variability. In particular, the PC1 explained 35.6%, whereas the PC2 explained 27.1%. VIPS > 1.4 were found for β-phellandrene, β-ocimene, δ- cadinene, humulene and bornyl acetate. The first, third, fourth, and fifth metabolites were up-accumulated in PPM-infested plants, whereas the second in the NI plants.

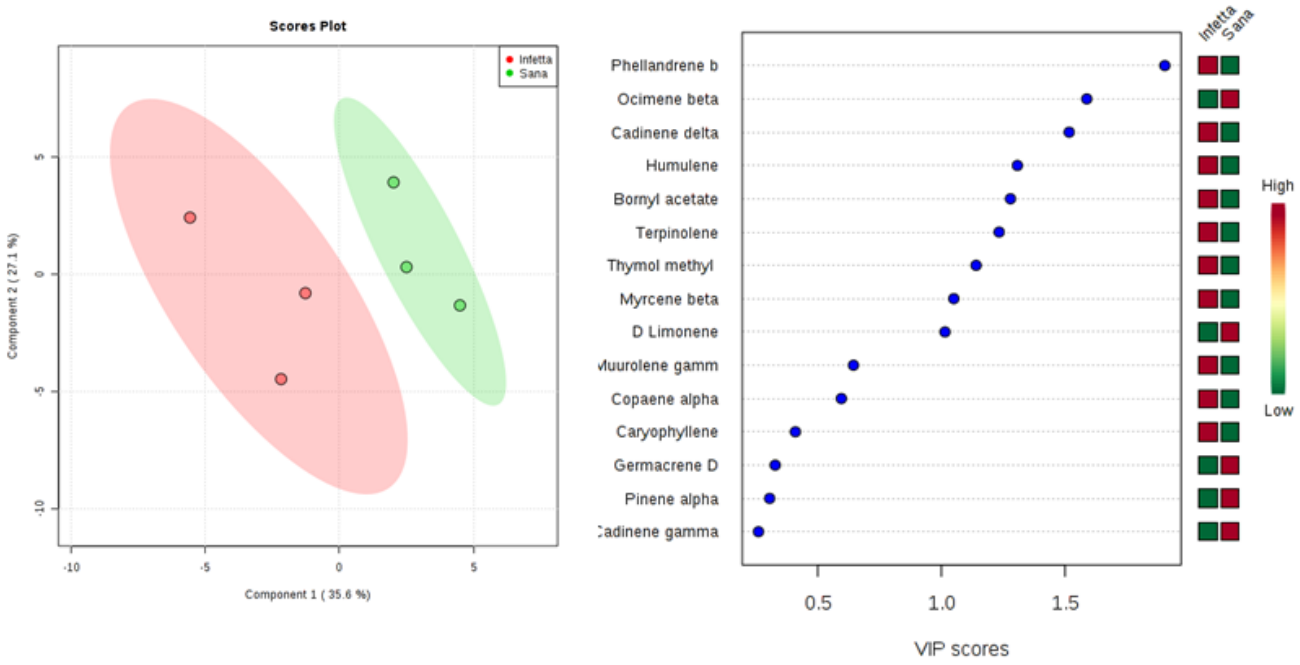


Fig. 4.8. *Partial Least Squares - Discriminant Analysis applied to the OVI sampling in Bova. Parameters and symbols as in Fig. 4.3.*

In the NEST sampling in Canolo (Fig. 4.9), the components 1 and 2 explained 69.7% of the variability. In particular, the PC1 explained 33.8%, whereas the PC2 explained 35.9%. VIPS > 1.4 were found for β-phellandrene, β-myrcene, D-limonene and α-terpinolene. The first metabolite was up-accumulated in the NI plants, whereas the remaining three in the PPM-infested plants.

In the DEFO-1 sampling in Canolo (Fig. 4.10), the components 1 and 2 explained 73.8% of the variability. In particular, the PC1 explained 41%, whereas the PC2 explained 32.8%. VIPS > 1.4 were found for caryophyllene, δ-elemene, and bornyl acetate, all of which were up-accumulated in the PPM-infested plants.

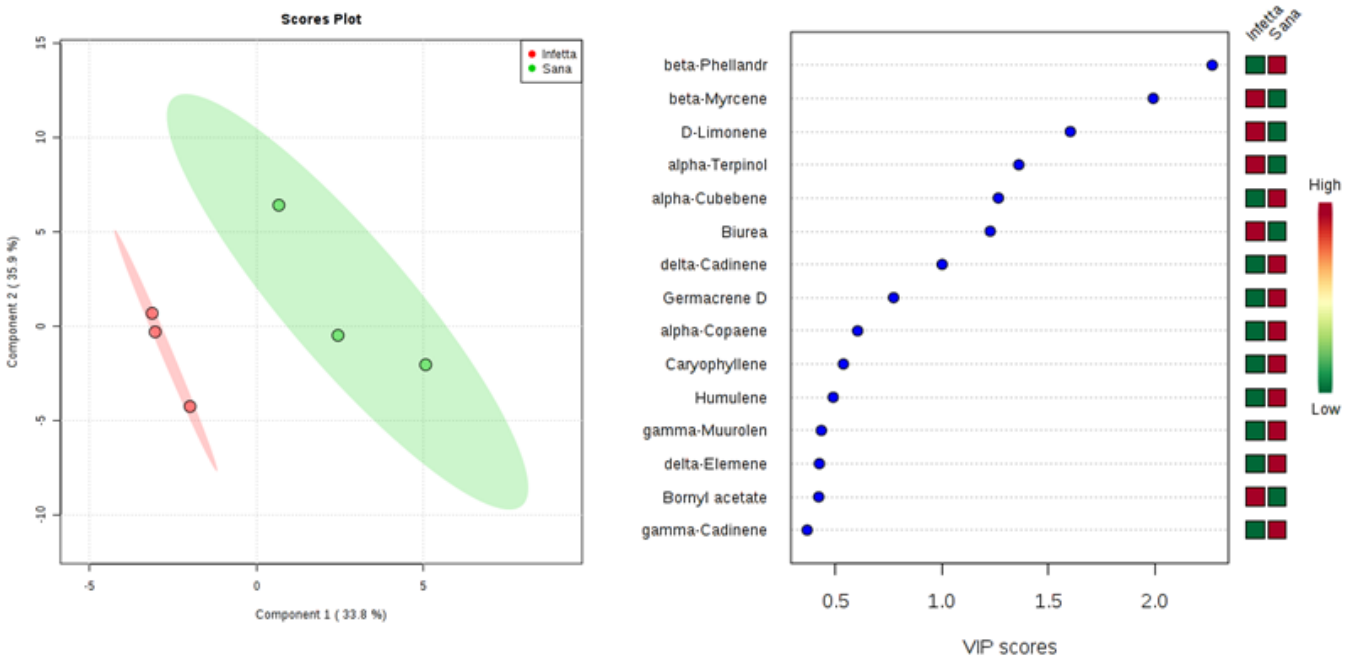


Fig. 4.9. *Partial Least Squares - Discriminant Analysis applied to the NEST sampling in Canolo. Parameters and symbols as in Fig. 4.3.*

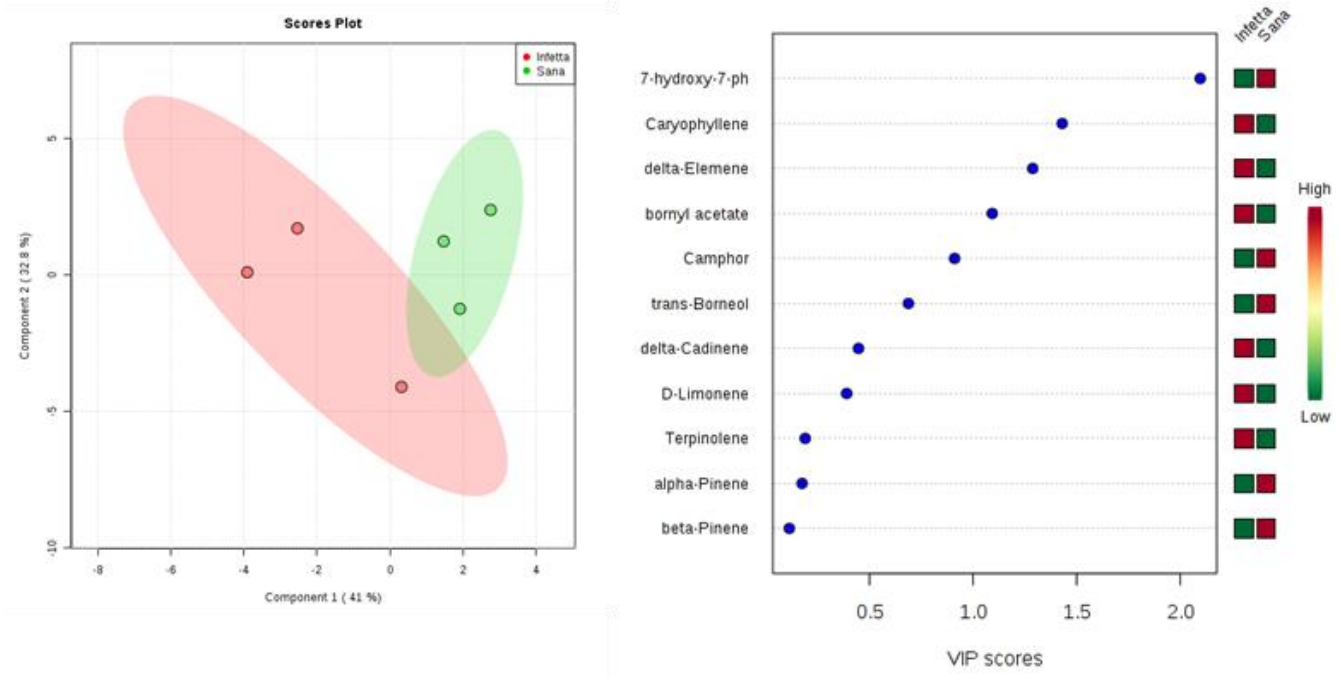


Fig. 4.10. *Partial Least Squares - Discriminant Analysis applied to the DEFO-1 sampling in Canolo. Parameters and symbols as in Fig. 4.3.*

In the DEFO-2 sampling in Canolo (Fig. 4.11), the components 1 and 2 explained 72.2% of the variability. In particular the PC1 explained 44%, whereas the PC2 explained 28.2%. VIPS > 1.4 were found for thymol methyl, β-ocimene, camphor, germacrene D, and bornyl acetate, all of which were up-accumulated in the PPM-infested plants.

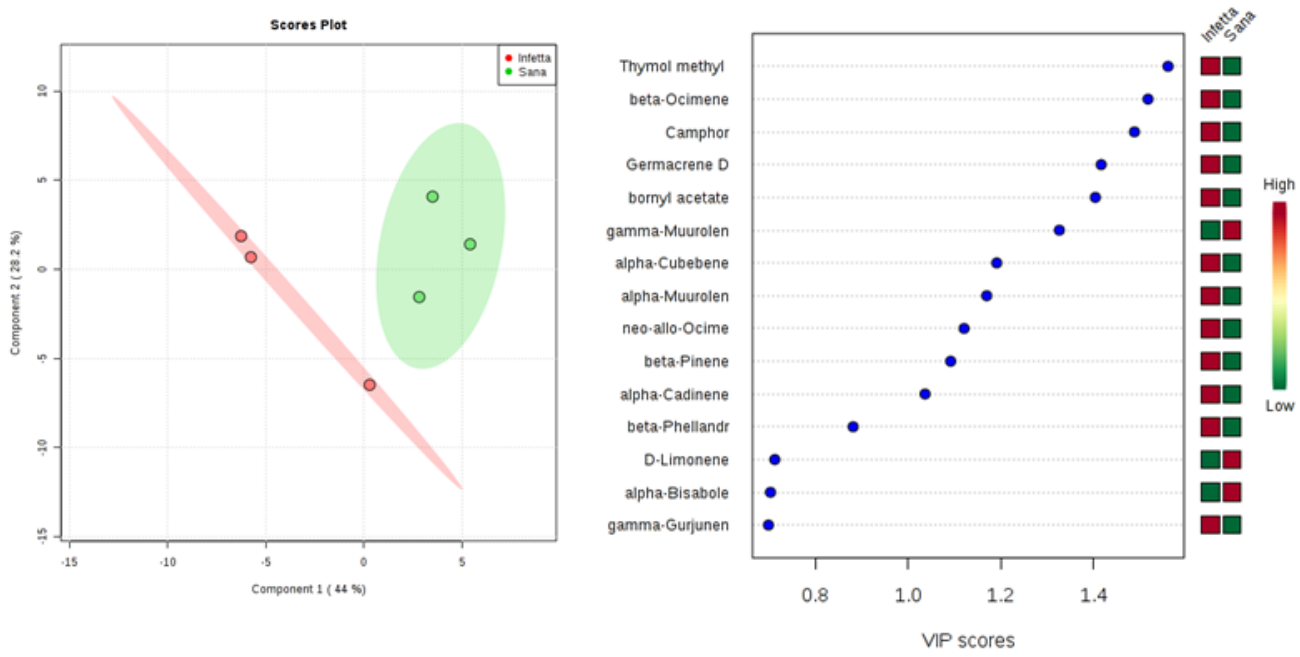


Fig. 4.11. *Partial Least Squares - Discriminant Analysis applied to the DEFO-2 sampling in Canolo. Parameters and symbols as in Fig. 4.3.*

In the PROCE-1 sampling in Canolo (Fig. 4.12), the components 1 and 2 explained 60.7% of the variability. In particular, the PC1 explained 50.5%, whereas the PC2 explained 10.2%. VIPS > 1.4 were found for D-limonene, β-myrcene and terpinolene, all of which were up-accumulated in the PPM-infested plants.

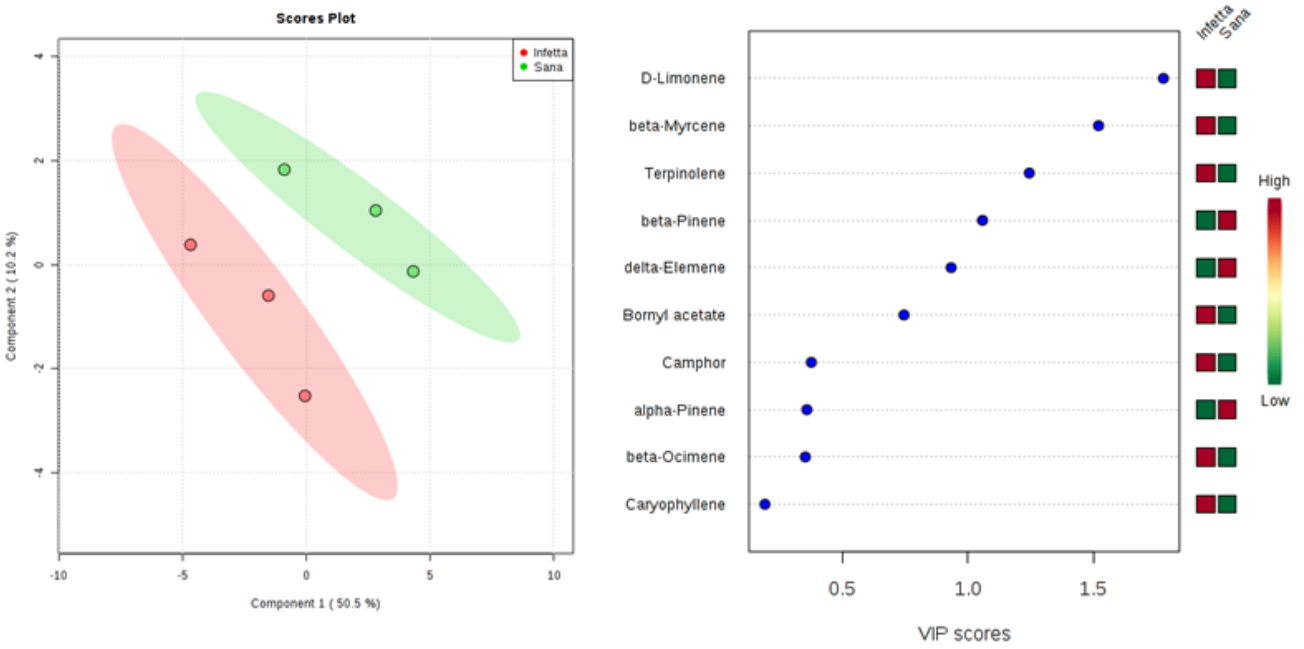


Fig. 4.12. *Partial Least Squares - Discriminant Analysis applied to the PROCE-1 sampling in Canolo. Parameters and symbols as in Fig. 4.3.*

In the PROCE-2 sampling in Canolo (Fig. 4.13), the components 1 and 2 explained 62.7% of the variability. In particular, the PC1 explained 51.5%, whereas the PC2 explained 11.2%. VIPS > 1.4 were found for α-pinene, germacrene D, D-limonene, β-ocimene and δ-cadinene, all of which were up-accumulated in the NI plants.

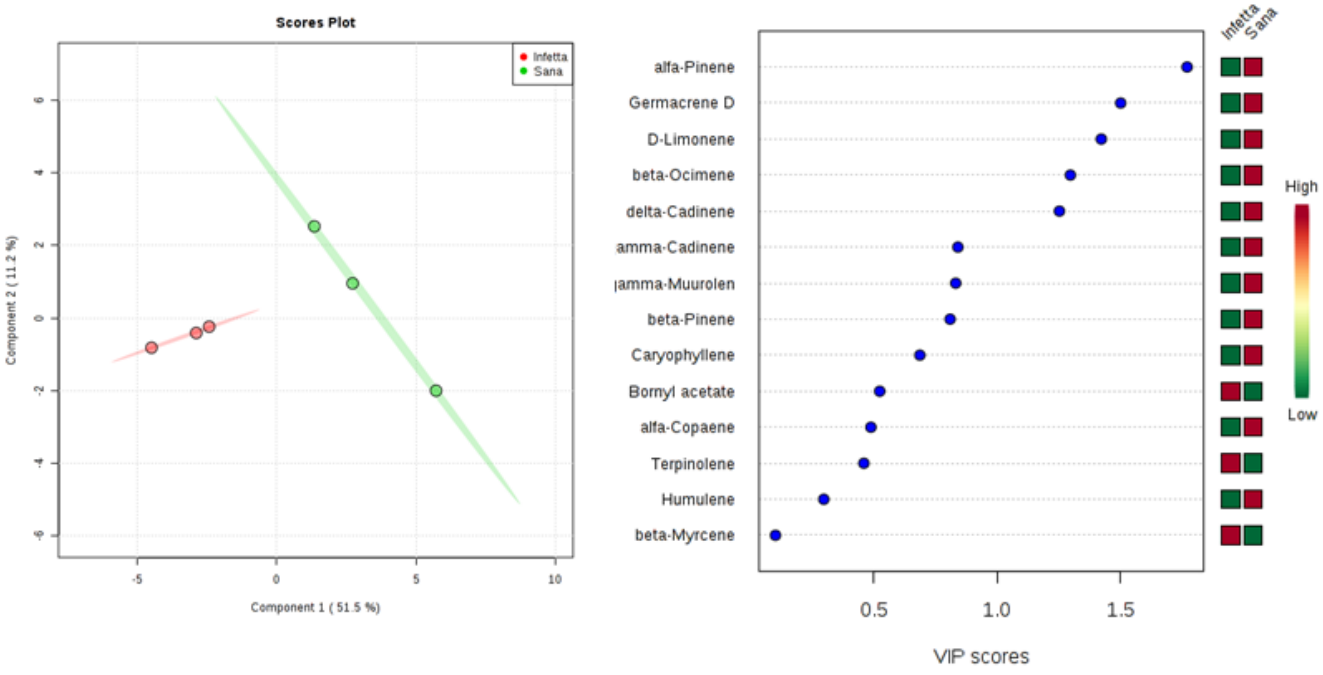


Fig. 2.13. *Partial Least Squares - Discriminant Analysis applied to the PROCE-2 sampling in Canolo. Parameters and symbols as in Fig. 4.3.*

In the OVI sampling at Canolo (Fig. 4.14), the components 1 and 2 explained 49.5% of the variability. In particular the PC1 explained 34.4%, whereas the PC2 explained 15.1%. VIPS > 1.4 were found for β- myrcene, camphor, β-terpinolene and α-cubebene. The first and the third metabolites were up-accumulated in the NI plants, whereas the second and the fourth in the PPM-infested plants.

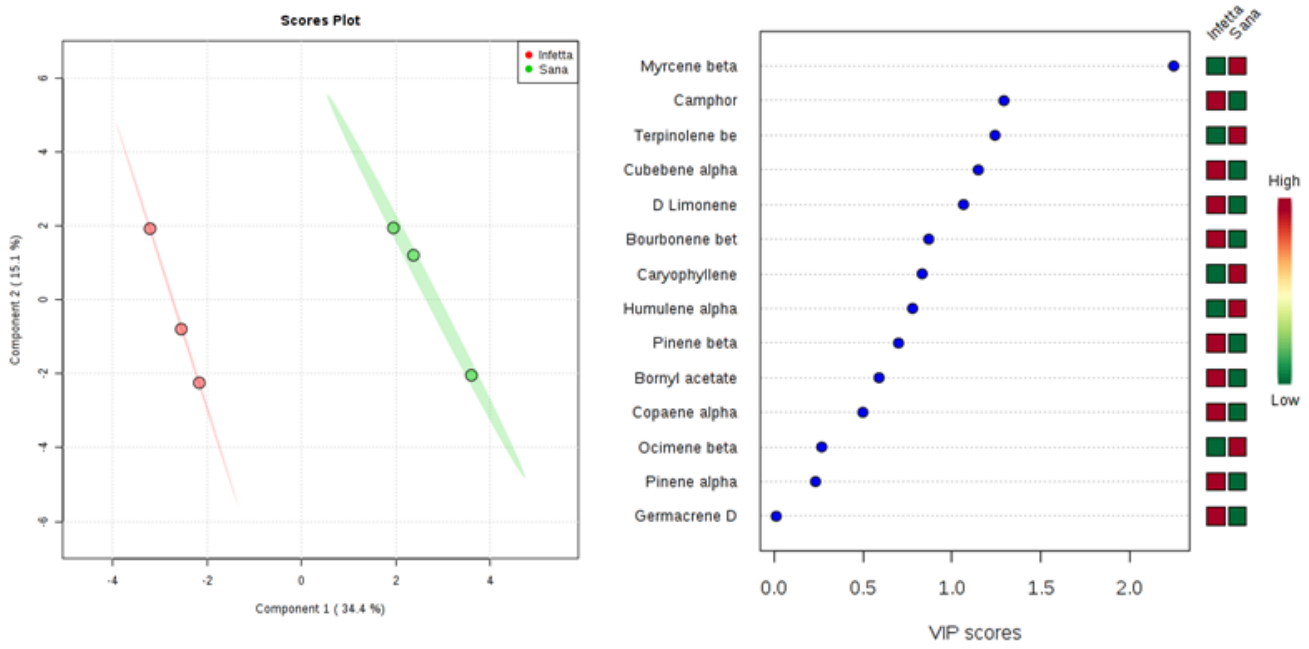


Fig. 4.14. *Partial Least Squares - Discriminant Analysis applied to the OVI sampling in Canolo. Parameters and symbols as in Fig. 4.3.*

The four heatmaps shown below (Fig. 4.15), one for each of the two experimental variants and for each of the two sampling sites, aim at presenting a synopsis of the temporal semi-quantitative changes observed in the emission of green leaf volatiles from Calabrian pine needles during the course of the sampling campaigns.

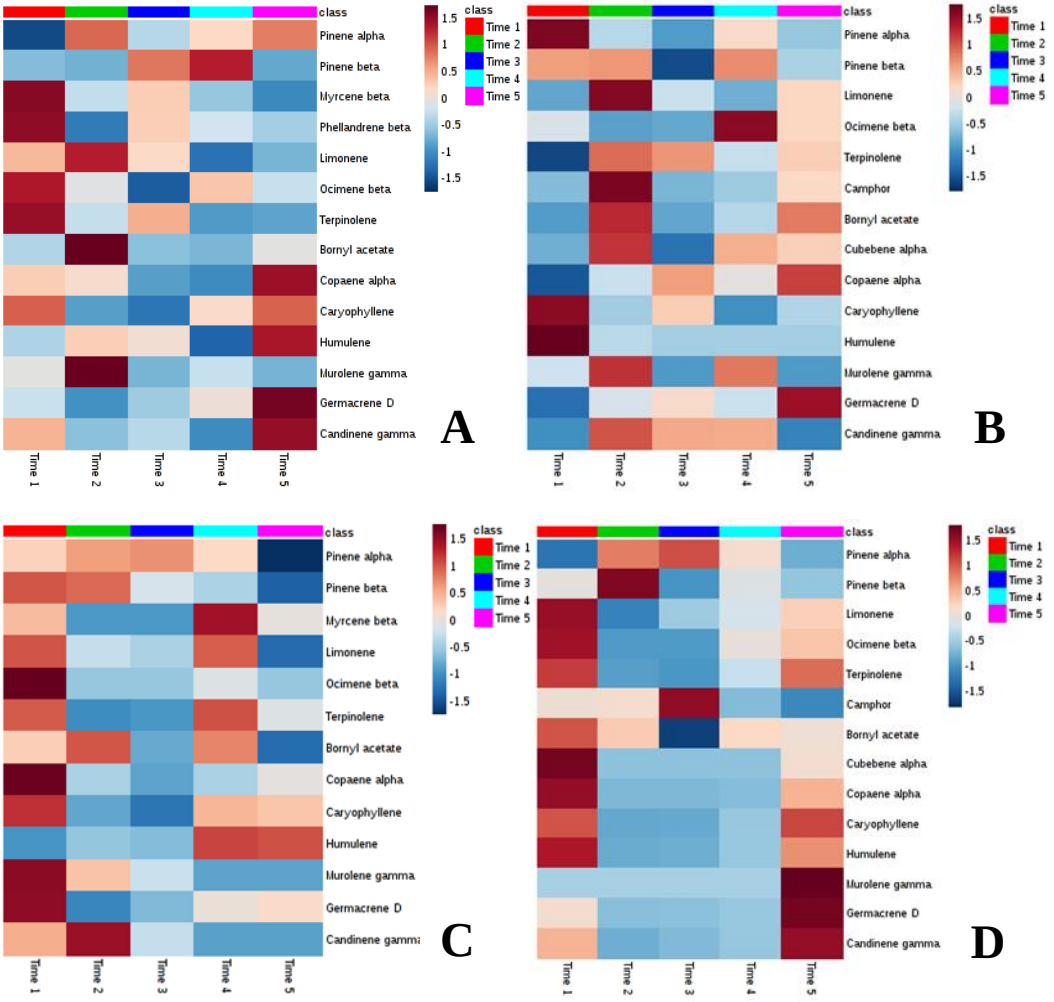


Fig. 4.15. *Heatmap synopsis of the changes in the emission of volatile terpenoids from Calabria pine needles (y axis) at five sampling points (x axis) during the trophic season of the pine processionary moth caterpillars. Brownish colors indicate increase, bluish colors indicate decrease. A, infested plants in Bova; B, noninfested plants in Bova; C, infested plants in Canolo; D, noninfested plants in Canolo.*

By considering only those of the aforementioned green leaf volatiles showing the highest VIPS, and by using a more restrictive criterion, the terpenoids shown to be present in all or almost all the headspaces analysed were the two monoterpenes bornyl acetate [(4,7,7-trimethyl-3-bicyclo[2.2.1]heptanyl) acetate] and β-ocimene (*cis/trans*-3,7-dimethyl-1,3,6-octatriene).

The data reported above suggest that, at both sampling sites, a differential emission of bornyl acetate took place from the needles I plants as compared to their respective NI counterparts, being higher in the former. Such difference was maximal during the period of more intense trophic activity of the PPM larvae, i.e., on DEFO-1 and DEFO-2 samplings, and tended to disappear thereafter.

4.4. DISCUSSION

The role of terpenes in insect-conifer interactions has been extensively studied since the 60s-70s of the previous century (Paine *et al.* 1997; Koricheva *et al.* 2004; Mumm *et al.* 2006). A topic attracting most attention has been the bark beetle – conifer interactions, due to the importance of terpenes in these relations, and the great economic and ecological impacts of these pests, especially in the American forests (Paine *et al.* 1997; Trapp and Croteau, 2001). Folivorous insects (insects feeding on conifer needles, such as Lepidoptera) also have important impacts on conifer forests (Jacquet *et al.* 2012).

Plant defenses and herbivorous attacks can raise (Sampedro *et al.* 2011) or decrease (Litvak and Monson, 1998) the concentration of terpenoids, and can induce changes in the composition and production of resins and in the emission of volatiles, including terpenes (Staudt and Lhoutellier, 2007). The emission of volatile terpenes by conifers also has important functions in indirect resistance, because these compounds act as airborne molecular messengers that deter herbivores, attract parasitoids of herbivores (Hilker *et al.* 2002) or warn other plants of attack, but may also be used as an olfactory cue by herbivores for their host selection (Peñuelas, 2004). this fascinating complexity makes the role of terpenes in the defense of plants against defoliators controversial, needing further study to understand the functioning of these interactions.

In the present work, green leaf volatiles emitted from Calabrian pine needles challenged by PPM infestation was carried out both qualitatively and semi-quantitatively. To the best of our knowledge, no previous study of the kind was carried out concerning such plant-host interaction. Indeed, in a study dating back to the Seventies of the previous Century, Arbez et al (1974) did measure terpenoids in the oleoresins extracted from Calabrian pine, but the analytical approach they adopted was traditional gas-chromatography, i.e., not interfaced with mass-spectrometry; furthermore, no differentiation among volatile and non-volatile terpenoids was attempted, and, most importantly, no plant-insect interaction was considered. On the other hand, and much more recently, the role of volatile terpenoids in the arm race among conifers and PPM was studied in detail by Peñuelas and Co-workers [Achotegui *et al.* 2015; Achotegui *et al.* 2013], but the plant species involved was Scots pine (*Pinus sylvestris* L.), instead of *P. nigra.*

The results reported here suggest that bornyl acetate and, to a lesser extent, β-ocimene foliar emissions differentiated PPM-infested plants from their respective non-infested controls, being higher in the former during the period of maximal PPM trophic activity. It is noteworthy that these same differences were observed at both sampling sites, namely Bova and Canolo, which are located at the opposite ends of the Aspromonte National Park, i.e., about 40 km apart from each other as the crow flies.

However, such differential terpenoids emission among infested and non-infested individuals was more pronounced in the Bova plants than in the Canolo ones. Speculatively speaking, this might have resulted from the higher degree of PPM-infestation observed in the Bova plants, respect to the Canolo ones, and/or from a different proximity of the PPM nests with respect to the branches chosen for samplings, being Canolo trees larger than Bova ones, thus bearing the PPM nests at higher heights respect to the sampled branches (see Materials and Methods for further details).

Bornyl acetate is a monoterpene known to be involved in plant defense, and the present study suggests in fact that its emission increased during the period of most intense trophic activity of PPM caterpillars. However, the previous studies available on such topic suggest that the role of such volatile terpenoid in plant-insect interactions might be complex. For example, Cates *et al.* (1987) carried out an agar diet study on Western spruce budworm populations, to determine the effects of varying concentrations of nitrogen, β-pinene, and bornyl acetate on larval growth and survival. Bornyl acetate reduced both growth and survival, suggesting that this compound may be functioning as a toxin or a feeding deterrent. β-pinene, instead, was associated with increased growth rate and may be functioning as a feeding stimulant. On the other hand, Ryan and Guerin (1982) found that (-)-bornyl acetate can act as a host-location cue for the carrot fly larva *Chamaepsila rosae* Fabricius, 1794, and Nishino and Manabe (1983) reported that. (+)-bornyl acetate is a mimic of the sex pheromone of the American cockroach (*Periplaneta americana* L.). A recent upsurge of interest for bornyl acetate and other components of essential oils from Lamiaceae, Lauraceae, and Valerianaceae as natural insecticides to combat insects feeding on storage products, has come from the studies of Rozman *et al.* (2007) and of Feng *et al.* (2020; 2019).

The reported results suggest that the presence of defoliator insects influences the emission of specific terpenes by the infested plants. According to the current knowledge, it remains to be ascertained if this might have defensive purposes, either as toxins/repellents towards the attacking insect/ attractants for its parasitoids (often referred to as tri-trophic interactions), or as an alarm signal to be communicated to the neighboring plants.

4.5. CONCLUSION

The present study, which was the first of its kind to be carried out on Calabrian pine, suggests once again how terpenes, despite being considered part of the so-called “secondary metabolism”, are profoundly engaged in all respects in the defense of the plant. Indeed, and despite a great variability among individual plants living in the open environment, it has been shown here that a specific monoterpene, namely bornyl acetate, is emitted in comparatively greater amounts by plants undergoing

pine processionary moth infestation, and especially so during the periods of most intense trophic activity of the insect. Calibrating the sampling of plant material over the biological cycle of the insect, as it was done in the present study, was another remarkable approach of the present research, to obtain a more realistic and reliable scenario of the plant-insect interaction dynamics. There is no doubt that a better understanding of plant defense mechanisms will be beneficial to agroforestry. More attention will be paid to the study of terpenoids that participate in plant defense responses and the isolation of new secondary metabolites that can be used to control pests, pathogens, and weeds.

5. CHAPTER 5: THE EXPRESSION OF TERPENE SYNTHASE GENES IN CALABRIAN PINE SAPLINGS INFESTED BY THE PINE PROCESSIONARY MOTH UNDER SEMI-NATURAL CONDITIONS

Abstract

The pine processionary moth, *Thaumetopoea pityocampa* (Denis & Schiffermüller, 1775), is one of the most important phytophagous of the Mediterranean forest, that in general proliferates on species belonging to *Pinus* spp. It is considered one of the most important defoliators of the pine trees, causing important economic and ecological losses. The terpenoids of needles play a central role in plant defence system, with the production of toxic substances for the host and as volatile signals in the context of plant-to-plant communication. In the present work we report about the temporal expression profiles of mono- and di- terpene synthase genes in the needles of *Pinus nigra* subsp. *laricio* (Calabrian black pine) facing the experimental infestation brought about by the pine processionary moth under semi-natural conditions. Transcript profiling of the Calabrian pine *DTPS* and *MTPS* genes revealed differential expression across the different sampling times during the infestation of pine processionary moth on pine trees, suggesting their potential roles in plant-insect interactions.

Keywords: monoterpene synthase; diterpene synthase; pine processionary moth; *Thaumetopoea pityocampa* (Denis & Schiffermüller, 1775); Calabrian black pine; *Pinus nigra* subsp. *laricio* (Poiret) Maire; defoliation.

5.1. INTRODUCTION

Because of global climate changes, the incidence of biotic and abiotic stresses on plant biology is expected to increase, and especially so on natural or semi-natural plant communities such as forest ecosystems, giving raise to decline in growth and deterioration in health, with negative consequences in terms of carbon sequestration, biodiversity, and ecosystem services. To respond to abiotic and biotic stimuli, conifers have evolved complex biochemical defence systems (Celedon and Bohlmann, 2019), including the production, either constitutive or inducible, of oleoresin and/or the emission of volatile terpenoids, which are directly or indirectly related to stress response. Indeed, these secondary metabolites are also known to mediate plant-insect interactions (Boncan *et al.* 2020), by acting as toxic substances against pests and pathogens and as biologically active agents to dissuade predation by insects (Phillips and Croteau, 1999). In contrast to broad-leaved trees, which emit mainly isoprene (C5), conifer species produce a large amount of terpene compounds (Kopaczyk *et al.* 2020; Andreani-Aksoyoglu and Keller, 1995) such as mono- (C10), sesqui- (C15) and di- (C20) terpenoids which are volatile, semi-volatile or non-volatile at ambient temperature, respectively. Both constitutive and induced terpenes are produced by a wide array of terpene synthase (TPSs), a large family of enzymes that catalyse the synthesis of hundreds terpenoids involved in general or specialized metabolism (Alicandri *et al.* 2020). In recent years, such topic has become the subject of intensive research, both because of the many potential applications of terpenoids in several technological fields (e.g., production of biofuels, fragrances, nutraceuticals and pharmaceuticals), and to exploit their desirable contribution in the biocontrol of herbivores and/or pathogens. Despite the efforts to fill the knowledge gap related to this last biological application field, systemic responses of conifer species to insect herbivory remain largely unexplored (Lòpez-Goldar *et al.* 2020).

Mediterranean pine forests are often under attack of a severe defoliating pest, the pine processionary moth (PPM) *Thaumetopoea pityocampa* (Dennis and Schiff., Lepidoptera: Thaumetopoeidae) that, in response to the climate warming, increased its spread on *Pinus* species during the last years, inflicting serious economic and ecological losses (Hòdar *et al.* 2003) as well as severe allergic reactions in humans and other mammals (Werno *et al.* 1990; Foti *et al.* 2020).

In the present work, we examined the transcriptional regulation of TPS genes in saplings of the non-model conifer *Pinus nigra* subsp. *laricio,* Calabrian black pine, facing the attack of PPM caterpillars under semi-natural conditions.

To achieve this aim, two ad hoc protective enclosures were provided by the Regional Observatory for Biodiversity in Santo Stefano in Aspromonte (RC) to host nine PPM-infested and as many healthy plants acting as the control. Three sampling times at larval stages 2-3 and 5, at 30-60 and 200 days after

hatching, respectively, were performed in order to monitor the expression level of mono- and di- terpene synthase genes.

5.2. MATERIAL AND METHODS

➢ *5.2.1. Study site and experimental design*

The present study was carried out at the Regional Observatory for Biodiversity in Santo Stefano in Aspromonte, locality of Cucullaro (Lat 38.172128°; Lon 15.816535°; 1.156,9 m a.s.l.), within the Aspromonte National Park, Southern Italy, (Fig. 5.1).

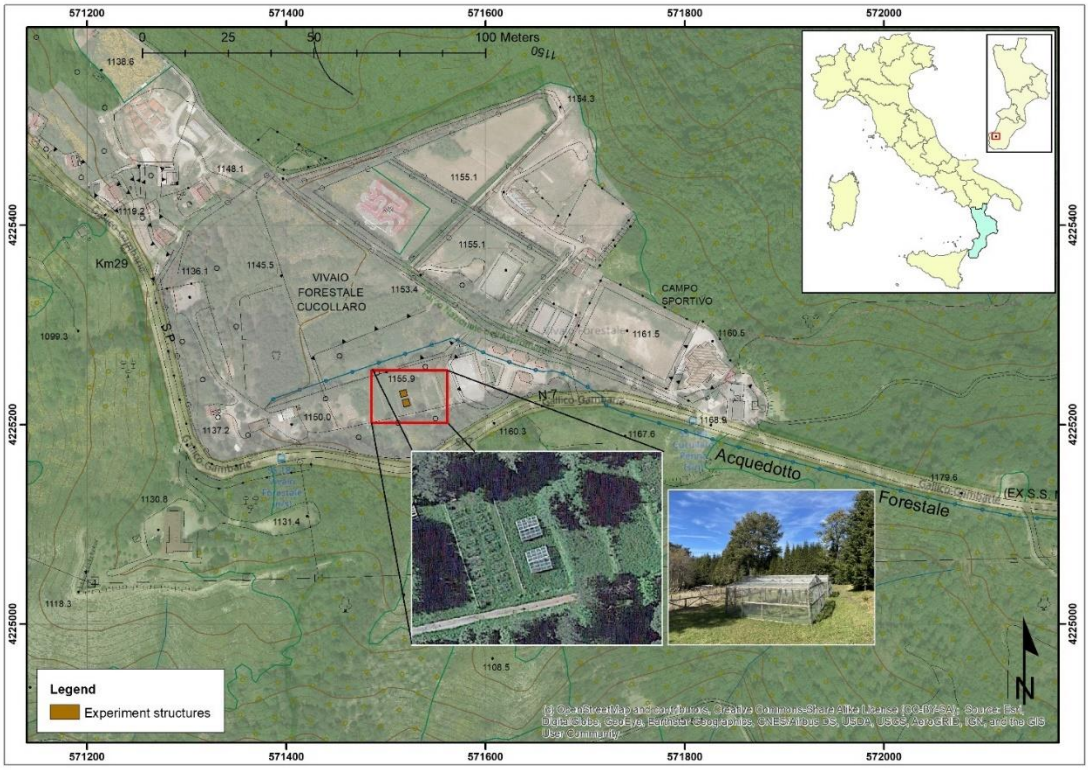


Fig. 5.1. *Map of the study site in Southern Italy, within the Aspromonte National Park in Calabria region (RC).*

Eighteen Calabrian pines saplings grown in pots were obtained from the Calabria Regional Forest nursery (Calabria Verde Agency, Catanzaro, Italy) and randomly assigned to two treatment groups at the experimental sites: 9 control pines (no PPM infestation) labelled as 1-2-3X, 1-2-3Y and 1-2-3Z, and 9 pines (assigned to future attack) labelled as 1-2-3A, 1-2-3B and 1-2-3C, were placed in two adjacent greenhouse-type enclosures, under otherwise undisturbed environmental conditions (Fig. 5.2), in which the plants were watered twice a week.

The experiment lasted from the end of October 2019 to the mid-May 2020 (Fig. 5.3).

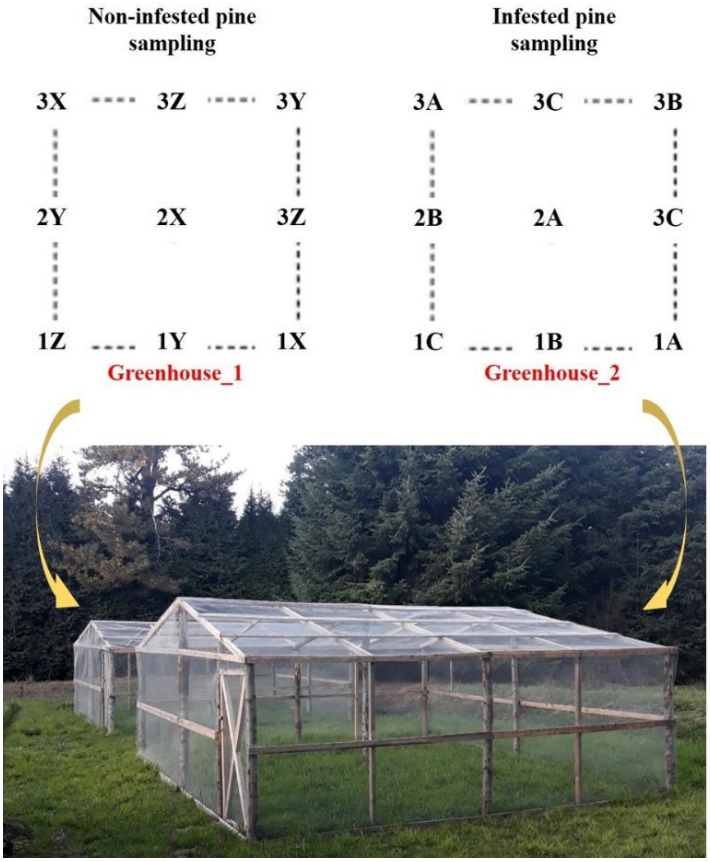


Fig. 5.2. *Experimental design of pine sampling in not infested (1) and infested (2) greenhouses.*

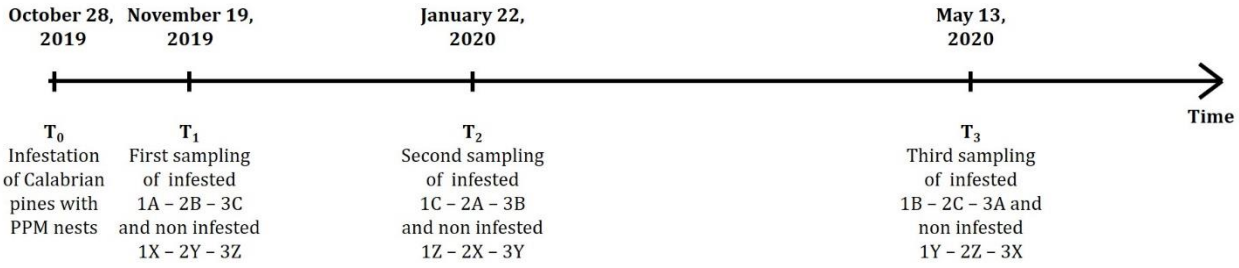


Fig. 5.3. *Timeline and sampling scheme of the experimental site.*

The study site is located in an area with pure stands of *Pinus nigra* subsp. *laricio*, normally showing a large presence of PPM colonies, from which, on October 28, 2019, PPM nests of uniform size and weight, previously checked for the effective presence and vitality of PPM *larvae*, were taken for the experimental infestation of the Calabrian pine saplings. One PPM nest was placed on each of the nine pine samplings selected for the PPM infestation and secured with the help of plastic ties (Fig 5.4A). At

the time of the experimental infestation, the processionary *larvae* were at their second stage and ranged from 10 to 15 mm in length (Fig. 5.4B).

Fig. 5.4. *Detail of a pine processionary nest placed on a Calabrian pine through plastic ties (A) and PPM larvae in the sampled nest (B).*

➢ *5.2.2. First sampling*

The first sampling of plant material was carried out from 9.00 am to 11.00 am on November 19th, 2019. Control plants 1X - 2Y - 3Z and infested plants 1A - 2B - 3C were sampled. For each infested plant, needles of attached branches were collected. Each of the nine infested plants showed signs of trophic activity (the plants were partially defoliated and the needles visibly damaged) suggesting that the infestation had been successful. Only the needles not visibly damaged were collected. Immediately after excision, the collected plant material was frozen in liquid nitrogen and then kept a -80 °C until analysis.

➢ *5.2.3. Second sampling*

The second sampling was carried out from 9.00 am to 11.00 am on January 22, 2020. Control plants 1Z - 2X - 3Y and infested plants 1C - 2A - 3B and were sampled according to the previously described procedure. All the infested plants showed severe signs of trophic activity except 1C and 3A plants,

indicating that the *larvae* had started their trophic activities on the original plants on which their nests had been placed, but that they moved to the neighbouring pine individuals during the 60 days between the first and second sampling for reasons to be ascertained. On the remaining infested plants, signs of activity were observed such as the presence of *larvae* on the first and secondary stems during their trophic activity (Fig. 5.5). To each of the infested plants, one or two (in case of severe defoliation) additional pine saplings were put aside, to provide enough needles for trophic activity of the *larvae* in order to complete the experiment.

Fig. 5.5. *Trophic procession of PPM larvae on infested Calabrian pine during the winter season (third larval stage).*

➢ *5.2.4. Third sampling*

The third sampling was carried out from 9.00 am to 11.00 am on May 13, 2020. Control plants 1Y - 2Z - 3X and infested plants 1B - 2C - 3A were sampled according to the previously described procedure. Because all the infested plants except 1C and 3A (see previously section) were completely defoliated through the first and the second stem, in this sampling the damaged needles were collected (Fig. 5.6).

Fig. 5.6. *Infested Calabrian pine tree during the third sampling, completely defoliated after the trophic activity of PPM larvae*

➢ *5.2.5. RNA isolation and cDNA synthesis*

Total RNA was extracted from 250 mg of needles according to Pavy *et al.* (2008). RNA concentration and integrity were checked by using a NanoDrop ND-1000 spectrophotometer (Labtech, East Sussex, UK). Only RNA samples with a 260/280 wavelength ratio between 1.9 and 2.1, and a 260/230 wavelength ratio greater than 2.0, were used for cDNA synthesis. First-strand cDNA was synthesized from 3 µg of total RNA of both treated and control samples using a Xpert cDNA Synthesis Kit (GRiSP Research Solution, Porto, Portugal) according to the manufacturer's instructions.

➢ *5.2.6. Gene expression analysis*

The quantitative real time (qRT-PCR) analysis to detect the expression patterns of the isolated *P. nigra* subsp. *laricio* di- and mono- terpene synthase genes in the control and infested needles collected during the three samplings was conducted as described in Chapter 2 for diterpene synthase genes and Chapter 3 for monoterpene synthase genes (see Sections 2.2.10 and 3.2.3) (Alicandri *et al.* 2021; Alicandri *et al.* 2022). These include samples replication, quantitative real time (qRT-PCR) conditions, the selection of the most appropriate and stable reference genes for normalization, the designing of primer pairs for both target and reference genes (see Tables 2.1 and 3.2), the evaluation of primers specificity and amplification efficiency, and the criteria used to calculate normalized relative values of gene expression and their standard deviation.

5.3. RESULTS AND DISCUSSION

➢ *5.3.1. Diterpene synthase genes expression analysis*

Among the diterpene synthase genes the *DTPS1* showed the high expression level in infested plants compared to control ones (Figure 5.7 and Table 5.1). In particular, *DTPS1* [likely involved in the production of palustradiene, levopimaradiene, abietadiene and neoabietadiene (Hall *et al.* 2013; Alicandri *et al.* 2021)] appear to have a strong increase in the second and third samplings during the most intense trophic activity of the caterpillars (larval stages 3 and 5). Indeed, infested plants labelled as 2B-2A-2C and 3C-3B-3A were found significantly different from all the three control plants samplings and from the first sampling of infested plants (1A-1C-1B) (Figure 5.7 and Table 5.1), suggesting a large involvement of *DTPS1* in the plant-insect interaction.

Contrary, the *DTPS2* [for which it was not possible to predict its possible role in DRAs biosynthesis (Hall *et al.* 2013; Alicandri *et al.* 2021)] showed a strong decrease in all the infested plant compared to control ones that were comparatively higher expressed in all the three-sampling time (Figure 5.7 and Table 5.1).

A similar trend was found for *DTPS4* [likely involved in the production of pimaradiene (Hall *et al.* 2013; Alicandri *et al.* 2021)] for which infested plants significantly decreased in the transcript abundance than control ones in all the three-sampling time (Figure 5.7 and Table 5.1).

Finally, *DTPS3* did not significantly differ between control and infested plants in the first and in the second sampling (Figure 5.7 and Table 5.1). A slight decrease was found in the infested plants of the third sampling (3C-3B-3A), when the pines were completely defoliated by the caterpillars, that significantly differ from the control ones of the 3rd sampling (larval stage 5) and from the control and infested plants of other 1st and 2nd samplings.

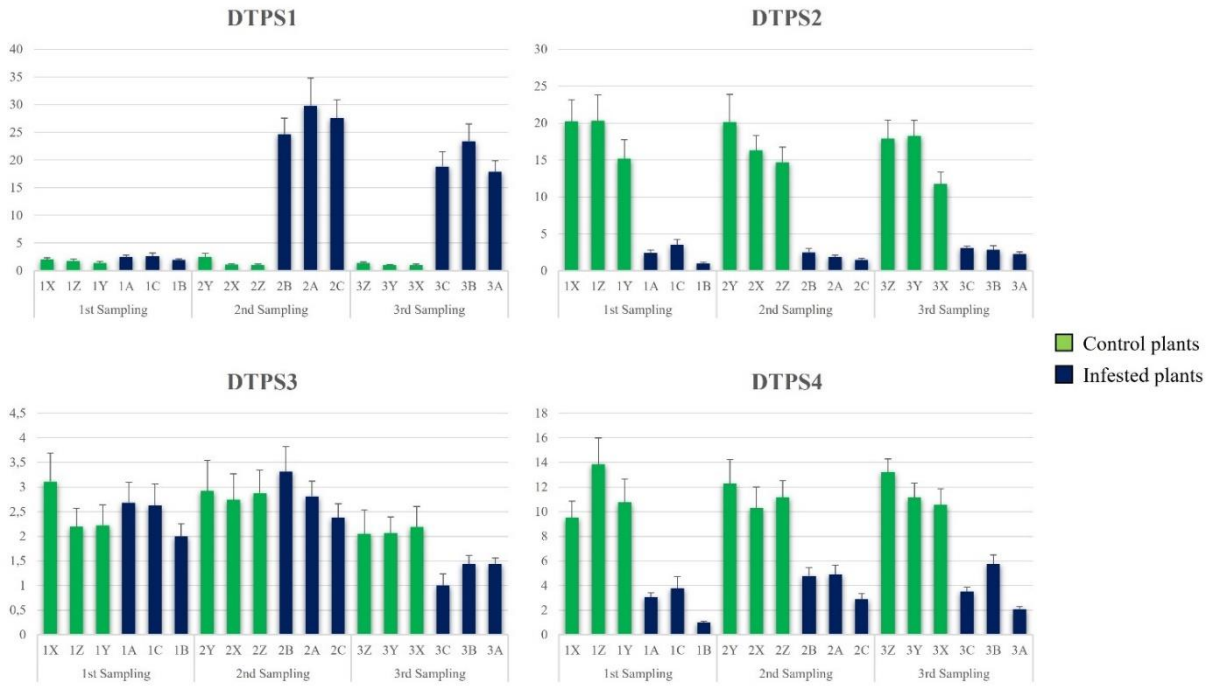


Fig. 5.7. *Relative expression levels of four diterpene synthase genes (DTPS1–4) in both control and infested Calabrian pine needles in the three-sampling time. The expression data of each gene were normalized using the geometric average of the two reference genes CYP and upLOC. Relative expression levels for each of the different DTPS genes were referred to a calibrator, set to the value 1, which was represented by the gene in the five tissues with the lowest expression.*

Tab. 5.1. *Statistical evaluation of the differences (one-way ANOVA followed by Tukey test) among the relative expression levels of DTPS genes in both control and infested Calabrian pine needles in the three-sampling time. Different letters denote significant differences according to the Tukey's test ($p < 0.01$).*

Samples		DTPS1		DTPS2		DTPS3		DTPS4	
1st sampling control plants	**1X**	2,012868	D	20,22762	A	3,113811	A	9,521089	B
	1Z	1,781069	D	20,31525	A	2,196123	ABC	13,85332	A
	1Y	1,413071	D	15,21077	AB	2,222572	ABC	10,74472	AB
1st sampling infested plants	**1A**	2,46631	D	2,398796	C	2,685203	AB	3,06443	CD
	1C	2,629023	D	3,521319	C	2,625294	AB	3,769407	CD
	1B	1,967223	D	1	C	1,994184	ABC	1	D
2nd sampling control plants	**2Y**	2,491727	D	20,14711	A	2,92041	AB	12,27893	AB
	2X	1,098467	D	16,32129	AB	2,744598	AB	10,28288	AB
	2Z	1,025002	D	14,68668	AB	2,880054	AB	11,16505	AB
2nd sampling infested plants	**2B**	24,67044	AB	2,483959	C	3,310389	A	4,784024	CD
	2A	29,80243	AB	1,850199	C	2,813757	AB	4,889335	CD
	2C	27,55694	AB	1,482044	C	2,375222	AB	2,87769	CD
3rd sampling control plants	**3Z**	1,412152	D	17,87826	A	2,049653	ABC	13,21413	A
	3Y	1,030081	D	18,26849	A	2,064116	ABC	11,14858	AB
	3X	1	D	11,72106	B	2,190633	ABC	10,53973	AB
3rd sampling infested plants	**3C**	18,83733	C	3,075091	C	1	C	3,505153	CD
	3B	23,40333	BC	2,83929	C	1,439471	BC	5,753326	C
	3A	17,83322	C	2,252961	C	1,433986	BC	2,084712	CD

➢ *5.3.2. Monoterpene synthase genes expression analysis*

Among the monoterpene synthase genes, *MBOS1* showed the highest expression level in infested plants compared to control ones (Figure 5.8 and Table 5.2). In particular, *MBOS1* [likely involved in the production of isoprene (Gray *et al.* 2011; Alicandri *et al.* 2022)] appear to have a strong increase in the 1st sampling time (larval stage 2), after 30 days from the nest infestation on pine trees, and a decreasing trend during the 2nd and 3rd sampling (larval stage 3 and 5 respectively) (Figure 5.8 and Table 5.2). A different trend was found by Chen *et al.* (2018) for *Pinus massoniana* under feeding pressure of beetle *Monochamus alternatus*. Indeed, this monoterpene synthase increase soon after beetle feeding and gradually returned to the normal level after 3–6 days (Chen *et al.* 2018). By contrast, in another conifer species, *Picea sitchensis* (Sitka spruce), the 2-methyl-3-buten-2-ol synthase genes induced by weevils increased within 1–2 days and remained at a high level for the next 2 weeks (Chen *et al.* 2018; Miller et al. 2005). The control plants for each of three samplings didn't show significant differences in *MBOS1* transcripts abundance.

Also in the case of *MTPS2* [likely involved in the production of α-terpineol (Hall *et al.* 2013b; Alicandri *et al.* 2022)] and of *MTPS4* [likely involved in the production of (−)-β-pinene, (−)-α-pinene and α-terpineol (Hall *et al.* 2013b; Phillips *et al.* 2003; Alicandri *et al.* 2022)] infested plants showed a high expression level in comparison to control ones, with an increasing trend from the first sampling to the last ones (Figure 5.8 and Table 5.2). Indeed, the highest transcripts abundance for both *MTPS2* and *MTPS4* has been reached in the third sampling by the plants 3C-3B-3A, during the most intense trophic activity of the caterpillars, when the pines had been defoliated by up to 90% (Figure 5.8 and Table 5.2).

A similar trend was found in *MTPS5* [likely involved in the production of (−)-α-pinene (Hall *et al.* 2013b; Phillips *et al.* 2003; Alicandri *et al.* 2022)] that showed the same expression level in all the control plants and in infested plants of the first sampling (1A-1C-1B) and was comparatively higher in the second and third samplings of infested plants. Indeed, the two last sampling of infested plants were found significant different to the first ones (Figure 5.8 and Table 5.2).

Contrary, transcripts abundance of *MTPS3* [likely involved in the production of (+)-3-carene (Hall *et al.* 2013b; Alicandri *et al.* 2022)] appeared higher in control plants compared to the infested ones (Figure 5.8 and Table 5.2). Indeed, control plants were comparatively high in the first (1X-1Z-1Y), second (2Y-2X-2Z) and third (3Z-3Y-3X) samplings, respect to very low amount detected for each of three sampling times of infested plants (Figure 5.8 and Table 5.2).

Finally, *MTPS6* and *MTPS7* [likely involved in the production of (-)-β-phellandrene, (-)-camphene and (+)-α-pinene for *MTPS6* and (+)-α-pinene for *MTPS7* (Hall *et al.* 2013b; Alicandri *et al.* 2022)] exhibited the same transcripts abundance trend with similar expression levels in the first and second

sampling for both control and infested plants, and highly expressed in the infested plants of the last sampling (3C-3B-3A) (Figure 5.8 and Table 5.2).

In summary, transcript profiling of the Calabrian pine *DTPS* and *MTPS* genes revealed differential expression across the different sampling times during the infestation of pine processionary moth on pine trees, suggesting their potential roles in plant-insect interactions.

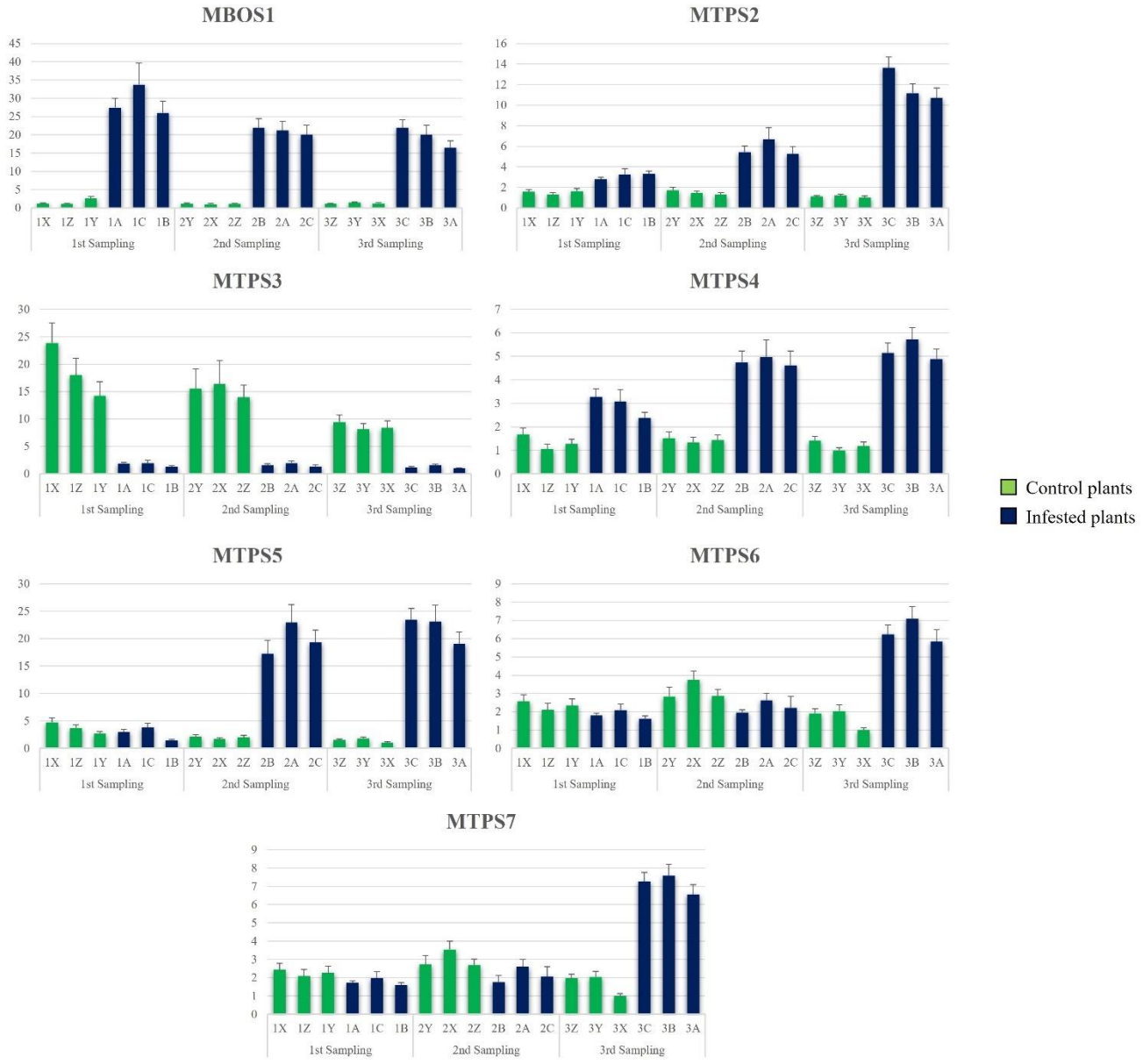


Fig. 5.8. *Relative expression levels of seven monoterpene synthase genes (MBOS1/MTPS2–7) in both control and infested Calabrian pine needles. The expression data of each gene were normalized using the geometric average of the two reference genes CYP and upLOC. Relative expression levels for each of the different MTPS genes were referred to a calibrator, set to the value 1, which was represented by the gene in the five tissues with the lowest expression.*

Tab. 5.2. *Statistical evaluation of the differences (one-way ANOVA followed by Tukey test) among the relative expression levels of MBOS and MTPS genes in both control and infested Calabrian pine needles in the three-sampling time. Different letters denote significant differences according to the Tukey's test ($p < 0.01$).*

Samples		**MBOS1**		**MTPS2**		**MTPS3**		**MTPS4**		**MTPS5**		**MTPS6**		**MTPS7**	
1st sampling control plants	**1X**	1,182356	E	1,558913	DEF	23,88354	A	1,675534	CD	4,727761	C	2,561933	BC	2,432793	BC
	1Z	1,085951	E	1,300517	EF	18,00153	AB	1,056421	D	3,715338	C	2,103172	CD	2,085784	CD
	1Y	2,578264	E	1,632741	DEF	14,22065	C	1,263473	CD	2,668977	C	2,335923	CD	2,259338	BCD
1st sampling infested plants	**1A**	27,42788	AB	2,798448	DE	1,888658	F	3,27195	B	2,963191	C	1,796652	CD	1,716278	CD
	1C	33,75244	A	3,253911	D	1,935301	F	3,064224	B	3,814333	C	2,090315	CD	1,977068	CD
	1B	26,02343	ABC	3,317384	D	1,307208	F	2,369347	BC	1,47095	C	1,628169	CD	1,591627	CD
2nd sampling control plants	**2Y**	1,158799	E	1,714979	DEF	15,50598	BC	1,513701	CD	2,087561	C	2,84077	BC	2,722768	BC
	2X	1	E	1,456095	EF	16,3661	B	1,33608	CD	1,672538	C	3,750005	B	3,548091	B
	2Z	1,101875	E	1,282985	EF	13,95585	BCDE	1,432471	CD	1,984807	C	2,866318	BC	2,695777	BC
2nd sampling infested plants	**2B**	21,94352	BCD	5,436982	C	1,549988	F	4,753184	A	17,2545	B	1,964108	CD	1,75804	CD
	2A	21,16063	BCD	6,642539	C	1,930528	F	4,972322	A	22,89934	A	2,622772	BC	2,599443	BC
	2C	20,0562	CD	5,269247	C	1,346413	F	4,611305	A	19,2883	AB	2,217974	CD	2,06465	CD
3rd sampling control plants	**3Z**	1,112455	E	1,109891	EF	9,455949	CDE	1,416384	CD	1,512515	C	1,915792	CD	1,974012	CD
	3Y	1,447546	E	1,195187	EF	8,139018	E	1	D	1,784669	C	2,042598	CD	2,041108	CD
	3X	1,262182	E	1	F	8,360855	DE	1,185894	D	1	C	1	D	1	D
3rd sampling infested plants	**3C**	21,89085	BCD	13,63559	A	1,098858	F	5,145062	A	23,435	A	6,238453	A	7,265589	A
	3B	20,0355	CD	11,17629	B	1,578446	F	5,721843	A	23,07892	A	7,081676	A	7,58441	A
	3A	16,45856	D	10,68128	B	1	F	4,886361	A	19,05527	AB	5,850331	A	6,557026	A

5.4. CONCLUSION

The present study, conducted on the non-model species *Pinus nigra* subsp. *laricio* aimed to fill a knowledge gap on the interactions between this Calabrian conifer and one of its most important phytophagous *Thaumetopoea pityocampa*.

Expression analyses of diterpene and monoterpene synthase genes showed that during PPM larval developmental (L2-L3 and L5), the relative abundances of the *DTPS* and *MTPS* were very different in the needle tissue. *DTPS1* and *MBOS1* appear to be the most largely activated genes under biotic stress, while *DTPS2*, *DTPS4* and *MTPS3* appeared to be the lowest expressed in response to the herbivory infestation. Other genes, such as *DTPS3* doesn't show significant differences between control and infested plants.

Further studies on the identification and production of diterpenes and monoterpenes could correlate the involvement of the *DTPS* and *MTPS* genes analyzed in this study with the putative corresponding metabolic products and elucidate about the function of these metabolites during the biotic stress.

The study of terpene synthase genes and terpenoids that participate in plant defense responses could contribute to developing bio-control strategies to control pests, pathogens, and weeds.

CONCLUSIONS AND FUTURE PERSPECTIVES

The increasing interest in new terpenoid products for their many potential uses makes dependence on natural resources alone not always sufficient or feasible. However, new possibilities exist nowadays for developing and improving the production of high value terpenoid compounds on an industrial scale, via metabolic engineering of the natural biosynthetic pathway in bacterial (*E. coli*) and yeast (*Saccharomyces cerevisiae*) systems, and in heterologous plant hosts. Therefore, novel and in-depth knowledge of the evolutionary diversification of members of conifer TPS family, their modular structure, and their putative functions appears to be important not only for a deeper understanding of their physiological and ecological roles, but also to foster metabolic engineering and synthetic biology tools for the production of high value terpenoid compounds.

In such context, an extensive characterization and phylogeny of all the known TPSs from different *Pinus* species is reported in the present thesis, which, to the best of our knowledge, represents for such genus the first effort to explore the evolutionary history of the large family of TPS genes involved in specialized metabolism.

A phylogeny-based approach allowed the isolation and sequencing of four di- and seven mono-terpene synthase genes belonging to the d3- and d1-TPS subclades, respectively, in five different tissues of Calabrian pine (young and mature needles, leader and interwhorl stems and roots). The analysis of the eleven deduced amino acid sequences allowed to predict their potential roles in the biosynthesis of diterpene resin acids and monoterpenoids in Calabrian pine.

In addition, for the first time, the metabolite profiles of di- and mono-terpenoids were also obtained from the same tissue types, to assign the isolated genes to their putative corresponding metabolite compounds.

Transcripts profiling of the Calabrian pine *DTPS* and *MTPS* genes revealed differential expression across the different tissues and were found to be consistent with the corresponding di- and mono-terpenoids profiles confirming a possible correlation.

Taken together, the above results seem to suggest a remarkable tissue-specificity, as well as species-specificity, of terpenoids composition in conifers, whose functional significance in terms of plant's biological performance, as well as in terms of possible exploitation for a variety of applications, awaits and deserves further studies.

In addition, to better understand the potential roles and employment of terpenoid in plant defence, and by considering the foreseeable increasing impact of biotic and abiotic stress factors as consequences of global changes, two experimental tests were conducted in natural and semi-natural conditions for the study of the Calabrian pine and *Thaumetopoea pityocampa* (pine processionary moth PPM)

interactions. In such contest, the determination of the quali-quantitative profiles of terpene synthase genes and terpenoid metabolites can help to obtain a more realistic and reliable scenario of the plant-insect interaction dynamics to better understanding of plant defence mechanisms that can occur in natural ecosystem. In prospect, this could contribute to improving and diversifying forest management options, as well as to developing bio-control strategies in a range of possible contexts, including forestry, agroforestry and urban greenery, so complying with the commitments of the United Nations Agenda 2030 and its sustainable development goals.

In conclusion, the DTPS and MTPS sequences identified in the present thesis provide a starting point to further investigate the complexity of the TPS gene family in Calabrian pine. For instance, most of the genes coding for enzymes involved in the synthesis of sesqui-terpenes remain to be discovered and characterized. 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.

ACKNOWLEDGEMENTS

This thesis represents the combined work of four different research groups that have supported and sustained me during these three years. First of all, the Reggio Calabria unit, my tutor *Prof. Agostino Sorgonà* and *Prof. Maurizio Badiani*, always ready for discussion and willing to give their valuable advice. My heartfelt “thanks” goes to you for believing in me when I didn't believe it either.

The Viterbo unit, and specifically the Plant Molecular Genetics Lab coordinated by *Prof. Mario Ciaffi* and supervised by *Dr. Annarita Paolacci*, has become my second home. In addition to the esteem for the extreme professionalism, a deep and sincere affection matured over the years binds me to them. My greatest “thanks” goes to you for having endured me with patience and encouraged me to always give my best.

The Reggio Calabria unit represented by *Prof. Carmelo Bonsignore* and his precious collaborators *Drs Francesco Manti* and *Elvira Castiglione* whom I met thanks to this experience. Professionalism and availability are their watchwords, together with a consistent level of passion for their territories. My "thanks" goes to you for always accompanying me with enthusiasm through the wonderful woods of Aspromonte Mountains, revealing me the beauties of this area.

The unit of Perugia, and specifically *Dr Bartolomeo Sebastiani* and *Dr Stefano Covino* which wisely guided me in the chemical analyses, directing me to a new and unknown sector with professionalism and competence. My “thanks” goes to you for always being ready to give me information and suggestions for the success of this project.

Finally, my "thanks" are addressed to my *family*, and in particular to *Ilaria*, my sister, for having lightened the load on my back when she saw me overflowing, and to *Tullio*, my partner, for sharing all the moments of this journey, giving me the awareness of never being alone.

Thanks also to my *families* scattered throughout Italy, between Sulmona, Viterbo, Novara and Reggio Calabria, because even if they are not mentioned, they know.

Thanks to you, *daddy*, my most avid supporter; the only one I would have liked to see smile at my achievement, and the only one who will not be able to do it. I dedicate this goal to *you*.

REFERENCES

1. Abbas, F., Ke, Y., Yu, R., Yue, Y., Amanullah, S., Jahangir, M. M. & Fan, Y. (2017). Volatile terpenoids: multiple functions, biosynthesis, modulation and manipulation by genetic engineering. *Planta*, 246(5), pp. 803–816, doi: 10.1007/s00425-017-2749-x
2. Achotegui Castells, A. (2015). The role of terpenes in the defensive responses of conifers against herbivores and pathogens. PhD Thesis. Universitat Autònoma de Barcelona.
3. Achotegui-Castells, A., Llusia, J., Hódar, J. A., & Peñuelas, J. (2013). Needle terpene concentrations and emissions of two coexisting subspecies of Scots pine attacked by the pine processionary moth (*Thaumetopoea pityocampa*). *Acta Physiologiae Plantarum*, 35(10), pp. 3047-3058. doi: 10.1007/s11738-013-1337-3
4. Adams, R. P. (2007). Identification of essential oil components by gas chromatography/mass spectrometry (Vol. 456). Carol Stream, IL: Allured publishing corporation.
5. Aharoni, A., Giri, A. P., Verstappen, F. W. A., Bertea, C. M., Sevenier, R., Sun, Z., Jongsma, M. A., Schwab, W. & Bouwmeester, H. J. (2004). Gain and loss of fruit flavor compounds produced by wild and cultivated strawberry species. *The Plant Cell*, 16(11), pp. 3110–3131, doi: 10.1105/tpc.104.023895
6. Akhtar, T. A., Matsuba, Y., Schauvinhold, I., Yu, G., Lees, H. A., Klein, S. E. & Pichersky, E. (2013). The tomato *cis*- prenyltransferase gene family. *The Plant Journal*, 73(4), pp. 640–652, doi: 10.1111/tpj.12063
7. Akiyama, K., Matsuzaki, K. & Hayashi, H. (2005). Plant sesquiterpenes induce hyphal branching in arbuscular mycorrhizal fungi. *Nature*, 435(7043), pp. 824–827, doi: 10.1038/nature03608
8. Akula, R., & Ravishankar, G. A. (2011). Influence of abiotic stress signals on secondary metabolites in plants. Plant signaling & behavior, 6(11), 1720-1731, doi: 10.4161/psb.6.11.17613
9. Alicandri, E., Paolacci, A. R., Osadolor, S., Sorgonà, A., Badiani, M. & Ciaffi, M. (2020). On the evolution and functional diversity of terpene synthases in the *Pinus* species: A Review. *Journal of Molecular Evolution*, 88(3), pp. 253–283, doi: 10.1007/s00239-020-09930-8
10. Alicandri, E.; Covino, S.; Sebastiani, B.; Paolacci, A.R.; Badiani, M.; Manti, F.; Bonsignore, C.P.; Sorgonà, A.; Ciaffi, M. Diterpene resin acids and olefins in Calabrian pine (*Pinus nigra* subsp. *laricio* (Poiret) Maire) oleoresin: GC-MS profiling of major diterpenoids in different

plant organs, molecular identification and expression analysis of diterpene synthase genes. *Plants*, 10(11), 2391. doi: 10.3390/plants10112391

11. Alicandri, E.; Covino, S.; Sebastiani, B.; Paolacci, A.R.; Badiani, M.; Sorgonà, A.; Ciaffi, M. Monoterpene synthase genes and monoterpene profiles in Calabrian pine [*Pinus nigra* subsp. *laricio* (Poiret) Maire] Submitted to *Plants* 2022-01-10.
12. Allenspach, M., & Steuer, C. (2021). α-Pinene: A never-ending story. *Phytochemistry*, *190*, 112857, doi: 10.1016/j.phytochem.2021.112857
13. Andersen, C. L., Jensen, J. L. & Ørntoft, T. F. (2004). Normalization of real-time quantitative reverse transcription-PCR data: a model-based variance estimation approach to identify genes suited for normalization, applied to bladder and colon cancer data sets. *Cancer Research*, 64(15), pp. 5245–5250, doi: 10.1158/0008-5472.CAN-04-0496
14. Andreani-Aksoyoglu, S. & Keller, J. (1995). Estimates of monoterpene and isoprene emissions from the forests in Switzerland. *Journal of Atmospheric Chemistry*, 20(1), 71-87.
15. Araniti, F., Lupini, A., Sunseri, F., & Abenavoli, M. R. (2017). Allelopatic potential of *Dittrichia viscosa* (L.) W. Greuter mediated by VOCs: A physiological and metabolomic approach. *PLoS One*, 12(1), e0170161. doi: 10.1371/journal.pone.0170161
16. Arbez, M., Bernard-Dagan, C., & Fillon, C. (1974). Variabilité intraspécifique des monoterpènes de *Pinus nigra* Arn.: Bilan des premiers résultats. *Annales des Sciences forestieres* 31(1), pp. 57-70). EDP Sciences.
17. Artola-Bordás, F., Arnedo-Pena, A., Romeu-García, M.A & Bellido-Blasco, J.B. (2008) Outbreak of dermatitis caused by pine processionary caterpillar (*Thaumetopoea pityocampa*) in schoolchildren. *Anales del sistema sanitario de Navarra*. 3, pp. 289-293. doi: 10.4321/s1137-66272008000500008.
18. Aubourg, S., Lecharny, A. & 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
19. Auerswald, L. & Gäde, G. (2002) Physiological and biochemical aspects of flight metabolism in cocoon-enclosed adults of the fruit beetle, *Pachnoda sinuate*. *Journal of Insect Physiology,* 48, pp. 239–248. doi: 10.1016/S0022-1910(01)00169-X
20. Avtzis, N. (1986) Development of *Thaumetopoea pityocampa* Schiff. (Lepidoptera:Thaumetopoeidae) in relation to food consumption. *Forest Ecology and Management* 15, pp. 65–68. doi: 10.1016/0378-1127(86)90090-3

21. Aydin T., Branco M., Güven Ö., Gonçalves H., Lima A., Karaca İ. & Butt T. (2018) Significant mortality of eggs and young larvae of two pine processionary moth species due to the entomopathogenic fungus *Metarhizium brunneum*. *Biocontrol Science and Technology*. 28: pp. 317–331. doi: 10.1080/09583157.2018.1447084

22. Baldwin, I. T. (2006). Volatile Signaling in Plant-Plant Interactions: 'Talking Trees' in the Genomics Era. *Science*, 311(5762), pp. 812–815, doi: 10.1126/science.1118446

23. Barbaro L. & Battisti A. (2011) Birds as predators of the pine processionary moth (Lepidoptera: Notodontidae). *Biological Control*. 56, pp. 107–114. doi: 10.1016/j.biocontrol.2010.10.009

24. Bartwal, A., Mall, R., Lohani, P., Guru, S.K., Arora, S. (2013). Role of secondary metabolites and brassinosteroids in plant defense against environmental stresses. *Journal Plant Growth Regulation*. 32, pp. 216-232. doi: 10.1007/s00344-012-9272-x

25. Battisti, A., Bernardi, M., and Ghiraldo, C. (2000), Predation by the hoopoe (*Upupa epops*) on pupae of *Thaumetopoea pityocampa* and the likely influence on other natural enemies. *BioControl*, 45, pp. 311–323. doi: 10.1023/A:1009992321465

26. Battisti A., Stastny M., Buffo, E. & Larsson, S. (2006) A rapid altitudinal range expansion in the pine processionary moth produced by the 2003 climatic anomaly. *Global Change Biology*, 12, pp. 662-671. https://doi.org/10.1111/j.1365-2486.2006.01124.x

27. Battisti, A., Stastny, M., Netherer, S., Robinet, C., Schopf, A., Roques, A., & Larsson, S. (2005). Expansion of geographic range in the pine processionary moth caused by increased winter temperatures. *Ecological applications*, 15(6), pp. 2084-2096. doi: 10.1890/04-1903

28. Behnke, K., Kleist, E., Uerlings, R., Wildt, J., Rennenberg, H. & Schnitzler, J.-P. (2009). RNAi-mediated suppression of isoprene biosynthesis in hybrid poplar impacts ozone tolerance. *Tree Physiology*, 29(5), pp. 725–736, doi: 10.1093/treephys/tpp009

29. Beran, F., Köllner, T. G., Gershenzon, J., & Tholl, D. (2019). Chemical convergence between plants and insects: biosynthetic origins and functions of common secondary metabolites. *New Phytologist*, 223(1), pp. 52-67. doi: 10.1111/nph.15718

30. Bernetti, G. (1995). Selvicoltura speciale. Unione tipografico-editrice Torinese.

31. Bick, J. A. & Lange, B. M. (2003). Metabolic cross talk between cytosolic and plastidial pathways of isoprenoid biosynthesis: unidirectional transport of intermediates across the chloroplast envelope membrane. *Archives of Biochemistry and Biophysics*, 415(2), pp. 146–154, doi: 10.1016/S0003-9861(03)00233-9

32. Birol, I., Raymond, A., Jackman, S. D., Pleasance, S., Coope, R., Taylor, G. A., Yuen, M. M. S., Keeling, C. I., Brand, D., Vandervalk, B. P., Kirk, H., Pandoh, P., Moore, R. A., Zhao, Y., Mungall, A. J., Jaquish, B., Yanchuk, A., Ritland, C., Boyle, B., Bousquet, J., Ritland, K., MacKay, J., Bohlmann, J. & Jones, S. J. M. (2013). Assembling the 20 Gb white spruce (*Picea glauca*) genome from whole-genome shotgun sequencing data. *Bioinformatics*, 29(12), pp. 1492–1497, doi: 10.1093/bioinformatics/btt178

33. Björkman, C., Kytö, M., Larsson, S. & Niemelä, P. (1998). Different responses of two carbon-based defenses in Scots pine needles to nitrogen fertilization. *Ecoscience*, 5(4) pp. 502-507. doi: 10.1080/11956860.1998.11682484

34. Blanch, J.S., Peñuelas, J., Sardans, J., Llusià, J. (2009). Drought, warming and soil fertilization effects on leaf volatile terpene concentrations in *Pinus halepensis* and *Quercus ilex*. *Acta Physiologiae Plantarum* 31, pp. 207-218. doi: 10.1007/s11738-008-0221-z

35. Bleeker, P. M., Mirabella, R., Diergaarde, P. J., VanDoorn, A., Tissier, A., Kant, M. R., Prins, M., de Vos, M., Haring, M. A. & Schuurink, R. C. (2012). Improved herbivore resistance in cultivated tomato with the sesquiterpene biosynthetic pathway from a wild relative. *Proceedings of the National Academy of Sciences*, 109(49), pp. 20124–20129, doi: 10.1073/pnas.1208756109

36. Bleeker, Petra M., Diergaarde, P. J., Ament, K., Schütz, S., Johne, B., Dijkink, J., Hiemstra, H., de Gelder, R., de Both, M. T. J., Sabelis, M. W., Haring, M. A. & Schuurink, R. C. (2011). Tomato-produced 7-epizingiberene and R-curcumene act as repellents to whiteflies. *Phytochemistry*, 72(1), pp. 68–73, doi: 10.1016/j.phytochem.2010.10.014

37. Bohlmann, J., Steele, C. L., & Croteau, R. (1997). Monoterpene synthases from grand fir (*Abies grandis*): cDNA isolation, characterization, and functional expression of myrcene synthase, (−)-(4S)-limonene synthase, and (−)-(1S, 5S)-pinene synthase. *Journal of Biological Chemistry*, *272*(35), 21784-21792, doi: 10.1074/jbc.272.35.21784

38. Bohlmann, J., Meyer-Gauen, G. & 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

39. Bohlmann, J. & Keeling, C. I. (2008). Terpenoid biomaterials. *Plant Journal*, 54(4), pp. 656–669, doi: 10.1111/j.1365-313X.2008.03449.x

40. Boncan, D. A. T., Tsang, S. S., Li, C., Lee, I. H., Lam, H. M., Chan, T. F. & Hui, J. H. (2020). Terpenes and terpenoids in plants: interactions with environment and insects. *International Journal of Molecular Sciences*, *21*(19), pp. 7382. doi: 10.3390/ijms21197382

41. Bonello, P., Capretti, P., Luchi, N., Martini, V., Michelozzi, M. (2008). Systemic effects of *Heterobasidion annosum* s.s. infection on severity of *Diplodia pinea* tip blight and terpenoid metabolism in Italian stone pine (*Pinus pinea*). *Tree Physiology* 28 pp. 1653-1660. doi: 10.1093/treephys/28.11.1653

42. Bonsignore, C.P. & Bellamy, C. (2007) Daily activity and flight behaviour of adults of *Capnodis tenebrionis* (Coleoptera: Buprestidae). *European Journal of Entomology*, 104, pp. 425–431.

43. Bonsignore, C.P., Manti, F., Castiglione, E. & Vacante, V. (2011). A study on the emergence sequence of pupal parasitoids of the pine processionary moth, *Thaumetopoea pityocampa*. *biocontrol science & technology*. 21, pp. 587–591. doi: 10.1080/09583157.2011.566322

44. Bonsignore, C.P. & Manti F. (2013) Influence of habitat and climate on the flight of male pine processionary moths. *Bulletin of insectology*, 66, pp. 27–34.

45. Bonsignore, C.P., Manti, F. & Castiglione, E. (2015). Interactions between pupae of the pine processionary moth (*Thaumetopoea pityocampa*) and parasitoids in a *Pinus* forest. *Bulletin of Entomological Research* 105, pp. 621–628. doi: 10.1017/S0007485315000541

46. Bonsignore, C. P., Manti, F., Castiglione, E. & Battisti, A. (2019). Pupal traits and adult emergence in the pine processionary moth *Thaumetopoea pityocampa* (Lepidoptera: Notodontidae) are affected by pupal density. *European Journal of Entomology*, 116, pp. 320-329. doi: 10.14411/eje.2019.035

47. Boone, C. K., Aukema, B. H., Bohlmann, J., Carroll, A. L., & Raffa, K. F. (2011). Efficacy of tree defense physiology varies with bark beetle population density: a basis for positive feedback in eruptive species. *Canadian Journal of Forest Research*, *41*(6), 1174-1188, doi: 10.1139/x11-041

48. Brown, J. W. S. & Simpson, C. G. (1998). SPLICE SITE SELECTION IN PLANT PRE-mRNA SPLICING. *Annual Review of Plant Physiology and Plant Molecular Biology*, 49(1), pp. 77–95, doi: 10.1146/annurev.arplant.49.1.77

49. Büchel, K., Malskies, S., Mayer, M., Fenning, T. M., Gershenzon, J., Hilker, M. & Meiners, T. (2011). How plants give early herbivore alert: Volatile terpenoids attract parasitoids to egg-infested elms. *Basic and Applied Ecology*, 12(5), pp. 403–412, doi: 10.1016/j.baae.2011.06.002

50. Buxton, R.D. (1990). The influence of host tree species on timing of pupation of *Thaumetopoea pityocampa* Schiff. (Lep., Thaumetopoeidae) and its exposure to parasitism by *Phryxe caudata*

Rond. (Dipt., Larvaevoridae). *Journal of Applied Entomology*, 109, pp. 302–310. doi: 10.1111/j.1439-0418.1990.tb00055.x

51. Byun-McKay, A., Godard, K.A., Toudefallah, M., Martin, D.M., Alfaro, R., King, J., Bohlmann, J. & Plant, A.L. (2006). Wound-induced terpene synthase gene expression in Sitka spruce that exibit resistance or susceptibility to attack by the white pine weevil. *Plant Physiology* 104, pp. 1009-1021. doi: 10.1104/pp.105.071803
52. Calvo, D. & Molina, J.M. (2005). Fecundity-body size relationship and other reproductive aspects of *Streblote panda* (Lepidoptera: Lasiocampidae). *Annals of the Entomological Society of America*, 98, pp. 191–196. doi: 10.1603/0013-8746(2005)098[0191:FSRAOR]2.0.CO;2
53. Camps, F., Canela, R., Coll, J., Guerrero, A. & Riba, M. (1981). Synthesis of the two isomers of the potential sex pheromone of *Thaumetopoea pityocampa* (Lepidoptera, Notodontidae) and related model compounds. *Chemistry Letters*, pp. 703-706. doi: 10.1246/cl.1981.703
54. Cao, R., Zhang, Y., Mann, F. M., Huang, C., Mukkamala, D., Hudock, M. P., Mead, M. E., Prisic, S., Wang, K., Lin, F.-Y., Chang, T.-K., Peters, R. J. & Oldfield, E. (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
55. Carus, S. (2004). Impact of defoliation by the Pine Processionary Moth (*Thaumetopoea pityocampa*) on radial, height and volume growth of Calabrian Pine (*Pinus brutia*) Trees in Turkey. *Phytoparasitica* 32, pp. 459-469.
56. Castagneyrol, B., Jactel, H., Brockerhoff, E.G., Perrette, N., Larter, M., Delzon, S. & Piou, D. (2016). Host range expansion is density dependent. *Oecologia*, 182, pp. 779–788. doi: 10.1007/s00442-016-3711-5
57. Cates, R. G., Henderson, C. B. & Redak, R. A. (1987). Responses of the western spruce budworm to varying levels of nitrogen and terpenes. *Oecologia*, 73(2), pp. 312-316.
58. Celedon, J. M. & Bohlmann, J. (2019). Oleoresin defenses in conifers: chemical diversity, terpene synthases and limitations of oleoresin defense under climate change. *New Phytologist*, 224(4), pp. 1444–1463, doi: 10.1111/nph.15984
59. Chang, Y. T., & Chu, F. H. (2011). Molecular cloning and characterization of monoterpene synthases from *Litsea cubeba* (Lour.) Persoon. *Tree Genetics & Genomes*, *7*(4), 835-844, doi: 10.1007/s11295-011-0377-3
60. Chen, F., Ro, D.-K., Petri, J., Gershenzon, J., Bohlmann, J., Pichersky, E. & Tholl, D. (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

61. Chen, F., Tholl, D., Bohlmann, J. & Pichersky, E. (2011). The family of terpene synthases in plants: a mid-size family of genes for specialized metabolism that is highly diversified throughout the kingdom: Terpene synthase family. *Plant Journal*, 66(1), pp. 212–229, doi: 10.1111/j.1365-313X.2011.04520.x

62. Chen, H., Yang, Z., Hu, Y., Tan, J., Jia, J., Xu, H. & Chen, X. (2016). Reference genes selection for quantitative gene expression studies in *Pinus massoniana* L. *Trees*, 30(3), pp. 685–696, doi: 10.1007/s00468-015-1311-3

63. Chen, R., He, X., Chen, J. *et al.* (2019). Traumatic resin duct development, terpenoid formation, and related synthase gene expression in *Pinus massoniana* under feeding pressure of *Monochamus alternatus*. *J Plant Growth Regul*, 38:897–908, doi: 10.1007/s00344-018-9900-1

64. Christianson, D. W. (2017). Structural and chemical biology of terpenoid cyclases. *Chemical Reviews*, 117(17), pp. 11570–11648, doi: 10.1021/acs.chemrev.7b00287

65. Ciaffi, M., Paolacci, A. R., Paolocci, M., Alicandri, E., Bigini, V., Badiani, M., & Muganu, M. (2019). Transcriptional regulation of stilbene synthases in grapevine germplasm differentially susceptible to downy mildew. *BMC plant biology*, *19*(1), 1-18, doi: 10.1186/s12870-019-2014-5

66. Cox, R. E., Yamamoto, S., Otto, A. & Simoneit, B. R. T. (2007). Oxygenated di- and tricyclic diterpenoids of southern hemisphere conifers. *Biochemical Systematics and Ecology*, 35(6), pp. 342–362, doi: 10.1016/j.bse.2006.09.013

67. Cristani, M., D'Arrigo, M., Mandalari, G., Castelli, F., Sarpietro, M. G., Micieli, D., ... & Trombetta, D. (2007). Interaction of four monoterpenes contained in essential oils with model membranes: implications for their antibacterial activity. *Journal of agricultural and food chemistry*, *55*(15), 6300-6308, doi: 10.1021/jf070094x

68. Croteau, R., Ketchum, R.E.B., Long, R.M., Kaspera, R. & Wildung, M.R. (2006). Taxol biosynthesis and molecular genetics. *Phytochemistry Reviews*, 5, pp. 75–97. doi: 10.1007/s11101-005-3748-2

69. Dahham, S. S., Tabana, Y. M., Iqbal, M. A., Ahamed, M. B., Ezzat, M. O., Majid, A. S., & Majid, A. M. (2015). The anticancer, antioxidant and antimicrobial properties of the sesquiterpene β-caryophyllene from the essential oil of *Aquilaria crassna*. *Molecules*, *20*(7), 11808-11829, doi: 10.3390/molecules200711808

70. D'amico, L.J., Davidowitz, G. & Nijhout, H.F. (2001). The developmental and physiological basis of body size evolution in an insect. *Proceedings of the Royal Society B*, 268, pp. 1589–1593. doi: 10.1098/rspb.2001.1698

71. Davidowitz, G., D'amico, L.J. & Nijhout, H.F. (2003). Critical weight in the development of insect body size. *Evolution & Development,* 5, pp. 188–197. doi: 10.1046/j.1525-142X.2003.03026.x

72. Démolin, G. (1969). Bioecologia della "Processionaria del pino" *Thaumetopoea pityocampa* Schiff. Incidencias de los factores climaticos. *Boletin del Servicio de Plagas Forestales* 12, pp. 9-24.

73. Demolin, G. (1971). Incidences de quelques facteurs agissant sur le comportement social des chenilles de *Thaumetopoea pityocampa* Schiff. (Lepidoptera) pendant la période des rocessions de nymphose. Répercussion sur l'efficacité des parasites. *Annales de Zoologie Ecologie Animale*, pp. 33–56.

74. Devkota, B. & Schmidt, G.H. (1990). Larval development of *Thaumetopoea pityocampa* (Den. and Schiff.) (Lep., Thaumetopoeidae) from Greece as influenced by different host plants under laboratory conditions. *Journal of Applied Entomology*, 109, 321 – 330. doi: 10.1111/j.1439-0418.1990.tb00059.x

75. Dong, L., Jongedijk, E., Bouwmeester, H. & Van Der Krol, A. (2016). Monoterpene biosynthesis potential of plant subcellular compartments. *New Phytologist*, 209(2), pp. 679–690, doi: 10.1111/nph.13629

76. Duan, J.J., Weber, D.C. & Dorn, S. (1998). Flight behaviour of pre- and post- diapause weevils in relation to ambient temperature. *Entomologia Experimentalis et Applicata*, 88, pp. 97–99. doi: 10.1023/A:1003269621622

77. Ducombs, G, Lamy, M, Mollard, S., Guillard J.M. & Maleville J., (1981). Contact dermatitis from processional pine caterpillar (*Thaumetopoea pityocampa* Den. and Schiff Lepidoptera). *Contact Dermatitis* 7, pp. 287-288. doi: 10.1111/j.1600-0536.1981.tb04081.x

78. Dudareva, N., Andersson, S., Orlova, I., Gatto, N., Reichelt, M., Rhodes, D., Boland, W. & Gershenzon, J. (2005). The non-mevalonate pathway supports both monoterpene and sesquiterpene formation in snapdragon flowers. *Plant Biology*, pp. 6. doi: 10.1073/pnas.0407360102

79. Dudareva, N., Flower, C., Cseke, L., Blanc, V.M. & Pichersky, E. (1996). Evolution of floral scent in Clarkia: novel patterns of S-linalool synthase gene expression in the *C. breweri* flower. *Plant Cell*, 8, pp. 1137-1148. doi: 10.1105/tpc.8.7.1137

80. Dudareva, N., Klempien, A., Muhlemann, J. K. & Kaplan, I. (2013). Biosynthesis, function and metabolic engineering of plant volatile organic compounds. *New Phytologist*, 198(1), pp. 16–32, doi: 10.1111/nph.12145

81. Dulaurent-Mercadal, A.M., Porté, A.J., Van Halder, I., Vetillard, F., Menassieu, P. & Jactel H. (2011). A case of habitat complementation in forest pests: Pine processionary moth pupae survive better in open areas. *Forest Ecology and Management*, 261, pp. 1069–1076. https://doi.org/10.1016/j.foreco.2010.12.029

82. Einhorn, J., Menassieu, P., Michelot, D. & Riom, J. (1983). The use of sex-traps baited with synthetic attractants against the pine processionary, *Thaumetopoea pityocampa* Schiff. (Lep., Notodontidae). First experiments in south-western France. *Agronomie*, 3, pp. 499-505.

83. Emanuelsson, O., Nielsen, H. & Heijne, G. V. (1999). ChloroP, a neural network-based method for predicting chloroplast transit peptides and their cleavage sites. *Protein Science*, 8(5), pp. 978–984, doi: 10.1110/ps.8.5.978

84. Enescu, C. M., de Rigo, D., Caudullo, G., Mauri, A., Houston Durrant, T., (2016). *Pinus nigra* in Europe: distribution, habitat, usage and threats. In: San-Miguel-Ayanz, J., de Rigo, D., Caudullo, G., Houston Durrant, T., Mauri, A. (Eds.), European Atlas of Forest Tree Species. Publ. Off. EU, Luxembourg, pp. e015138+.

85. Er, M.K., Tunaz, H. & Gockce, A. (2007). Pathogenicity of entomopathogenic fungi to *Thaumetopoea pityocampa* (Schiff.) (Lepidoptera: Thaumatopoeidae) larvae in laboratory conditions. *Journal of Pest Science*, 80, pp. 235–239. doi: 10.1007/s10340-007-0177-6

86. Falara, V., Akhtar, T. A., Nguyen, T. T. H., Spyropoulou, E. A., Bleeker, P. M., Schauvinhold, I., Matsuba, Y., Bonini, M. E., Schilmiller, A. L., Last, R. L., Schuurink, R. C. & Pichersky, E. (2011). The tomato terpene synthase gene family. *Plant Physiology*, 157(2), pp. 770–789, doi: 10.1104/pp.111.179648

87. Fantinou, A.A., Perdikis, D. & Stamogiannis, N. (2008). Effect of larval crowding on the life history traits of *Sesamia nonagrioides* (Lepidoptera: Noctuidae). *European Journal of Entomology*, 105, pp. 625–630.

88. Ferro, D.N., Alyokhin, A.V. & Tobin, D.B. (1999). Reproductive status and flight activity of the overwintered Colorado potato beetle. *Entomologia Experimentalis et Applicata*, 91, pp. 443–448.

89. Feng, Y. X., Wang, Y., Geng, Z. F., Zhang, D., Almaz, B. & Du, S. S. (2020). Contact toxicity and repellent efficacy of Valerianaceae spp. to three stored-product insects and synergistic interactions between two major compounds camphene and bornyl acetate. *Ecotoxicology and environmental safety*, 190, pp. 110106. doi: 10.1016/j.ecoenv.2019.110106

90. Feng, Y. X., Wang, Y., Chen, Z. Y., Guo, S. S., You, C. X., & Du, S. S. (2019). Efficacy of bornyl acetate and camphene from *Valeriana officinalis* essential oil against two storage insects. *Environmental Science and Pollution Research*, 26(16), pp. 16157-16165. doi: 10.1007/s11356-019-05035-y

91. Flügge, U. -I. & Gao, W. (2005). Transport of Isoprenoid Intermediates Across Chloroplast Envelope Membranes. *Plant Biology*, 7(1), pp. 91–97, doi: 10.1055/s-2004-830446

92. Foti, V., Araniti, F., Manti, F., Alicandri, E., Giuffrè, A. M., Bonsignore, C. P., Castiglione E., Covino S., Paolacci A.R., Ciaffi, M. & Badiani, M. (2020). Profiling volatile terpenoids from Calabrian pine stands infested by the pine processionary moth. *Plants*, 9(10) pp. 1362. doi: 10.3390/plants9101362

93. Fox, C.W., Mclennan, A.L., & Mousseau, T.A. (1995). Male body size affects female lifetime reproductive success. *Animal Behaviour*, 50, pp. 281–284.

94. Gao, Y., Honzatko, R. B. & Peters, R. J. (2012). Terpenoid synthase structures: a so far incomplete view of complex catalysis. *Natural Product Reports*, 29(10), p. 1153, doi: 10.1039/c2np20059g

95. Gennadios, H. A., Gonzalez, V., Di Costanzo, L., Li, A., Yu, F., Miller, D. J., Allemann, R. K. & Christianson, D. W. (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

96. Geron, C., Rasmussen, R., Arnts, R. R. & Guenther, A. (2000). A review and synthesis of monoterpene speciation from forests in the United States. *Atmospheric Environment*, 34(11), pp. 1761-1781. doi. 10.1016/S1352-2310(99)00364-7

97. Gershenzon, J. & Dudareva, N. (2007). The function of terpene natural products in the natural world. *Nature Chemical Biology*, 3(7) pp. 408-414. doi: 10.1038/nchembio.2007.5

98. Gleizes, M., Paulym, G., Bernard-Dagan, C., Jacques, R. (1980). Effects of light on terpene hydrocarbon synthesis in *Pinus pinaster*. *Physiologia Plantarum*, 50, pp. 16-20. doi: 10.1111/j.1399-3054.1980.tb02676.x

99. Gols, R. (2014). Direct and indirect chemical defenses against insects in a multitrophic framework: Plant chemical defenses against insects. *Plant, Cell & Environment*, 37(8), pp. 1741–1752, doi: 10.1111/pce.12318.

100. Gonzalez-Cano, J.M. (1981). Predacion de "procesionaria del pino" por vertebrados en la zona de Mora de Rubielos (Teruel). *Boletin de la Estacion Central de Ecologia*, 10, pp. 53-77.

101. Gray, D. W., Breneman, S. R., Topper, L. A. & Sharkey, T. D. (2011). Biochemical characterization and homology modeling of methylbutenol synthase and implications for understanding hemiterpene synthase evolution in plants. *Journal of Biological Chemistry*, 286(23), 20582-20590. doi:10.1074/jbc.M111.237438

102. Guerrero, A., Camps, F., Coll, J., Riba, M., Einhorn, J., Descoins, C. & Lallemand, J.Y. (1981). Identification of a potential sex pheromone of the processionary moth *Thaumetopoea pityocampa*. *Tetrahedron Letters*, 22, pp. 2013-2016. doi: 10.1016/S0040-4039(01)92892-8

103. Gutensohn, M., Orlova, I., Nguyen, T. T. H., Davidovich-Rikanati, R., Ferruzzi, M. G., Sitrit, Y., Lewinsohn, E., Pichersky, E. & Dudareva, N. (2013). Cytosolic monoterpene biosynthesis is supported by plastid-generated geranyl diphosphate substrate in transgenic tomato fruits. *The Plant Journal*, 75(3), pp. 351–363, doi: 10.1111/tpj.12212

104. Hamberger, B., Ohnishi, T., Séguin, A. & Bohlmann, J. (2011). Evolution of diterpene metabolism: Sitka spruce CYP720B4 catalyzes multiple oxidations in resin acid biosynthesis of conifer defense against insects. *Plant Physiology*, 157(4), pp. 1677–1695, doi: 10.1104/pp.111.185843

105. Hamberger, B., Hall, D., Yuen, M., Oddy, C., Hamberger, B., Keeling, C. I., ... & Bohlmann, J. (2009). Targeted isolation, sequence assembly and characterization of two white spruce (*Picea glauca*) BAC clones for terpenoid synthase and cytochrome P450 genes involved in conifer defence reveal insights into a conifer genome. *BMC plant biology*, *9*(1), 1-13, doi: 10.1186/1471-2229-9-106

106. Hall, D. E., Robert, J. A., Keeling, C. I., Domanski, D., Quesada, A. L., Jancsik, S., Kuzyk, M. A., Hamberger, B., Borchers, C. H. & Bohlmann, J. (2011). An integrated genomic, proteomic and biochemical analysis of (+)-3-carene biosynthesis in Sitka spruce (*Picea sitchensis*)

genotypes that are resistant or susceptible to white pine weevil: (+)-3-carene biosynthesis in Sitka spruce. *Plant Journal*, 65(6), pp. 936–948, doi: 10.1111/j.1365-313X.2010.04478.x

107. Hall, D. E., Yuen, M. M. S., Jancsik, S., Quesada, A. L., Dullat, H. K., Li, M., Henderson, H., Arango-Velez, A., Liao, N. Y., Docking, R. T., Chan, S. K., Cooke, J. E., Breuil, C., Jones, S. J., Keeling, C. I. & Bohlmann, J. (2013b). 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*, 13(1), p. 80, doi: 10.1186/1471-2229-13-80

108. Hall, D. E., Zerbe, P., Jancsik, S., Quesada, A. L., Dullat, H., Madilao, L. L., Yuen, M. & Bohlmann, J. (2013a). 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

109. Harley, P., Fridd-Stroud, V., Greenberg, J., Guenther, A., & Vasconcellos, P. (1998). Emission of 2-methyl-3-buten-2-ol by pines: A potentially large natural source of reactive carbon to the atmosphere. *Journal of Geophysical Research: Atmospheres*, *103*(D19), 25479-25486, doi: 10.1029/98JD00820

110. Hayashi, K., Horie, K., Hiwatashi, Y., Kawaide, H., Yamaguchi, S., Hanada, A., Nakashima, T., Nakajima, M., Mander, L. N., Yamane, H., Hasebe, M. & Nozaki, H. (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

111. Hayashi, K., Kawaide, H., Notomi, M., Sakigi, Y., Matsuo, A. & Nozaki, H. (2006). Identification and functional analysis of bifunctional *ent*-kaurene synthase from the moss *Physcomitrella patens*. *FEBS Letters*, 580(26), pp. 6175–6181, doi: 10.1016/j.febslet.2006.10.018

112. Hebsgaard, S. (1996). Splice site prediction in Arabidopsis thaliana pre-mRNA by combining local and global sequence information. *Nucleic Acids Research*, 24(17), pp. 3439–3452, doi: 10.1093/nar/24.17.3439

113. Heil, M. (2014). Herbivore-induced plant volatiles: targets, perception and unanswered questions. *New Phytologist*, 204(2), pp. 297–306, doi: 10.1111/nph.12977

114. Hemmerlin, A. (2013). Post-translational events and modifications regulating plant enzymes involved in isoprenoid precursor biosynthesis. *Plant Science*, 203–204, pp. 41–54, doi: 10.1016/j.plantsci.2012.12.008

115. Herde, M., Gärtner, K., Köllner, T. G., Fode, B., Boland, W., Gershenzon, J., Gatz, C. & Tholl, D. (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*, 20(4), pp. 1152–1168, doi: 10.1105/tpc.106.049478

116. Heyworth, C.J., Iason, G.R., Temperton, V., Jarvis, P.G. & Duncan, A.J. (1998). The effect of elevated CO_2 concentration and nutrient supply on carbon-based plant secondary metabolites in *Pinus sylvestris*. L. *Oecologia* 115 pp. 344-350. doi: 10.1007/s004420050526

117. Hilker, M. & Meiners, T. (2002). Induction of plant responses to oviposition and feeding by herbivorous arthropods: a comparison. In Proceedings of the 11th International Symposium on Insect-Plant Relationships (pp. 181-192). Springer, Dordrecht. doi: 10.1007/978-94-017-2776-1_21

118. Hillwig, M. L., Xu, M., Toyomasu, T., Tiernan, M. S., Wei, G., Cui, G., Huang, L. & Peters, R. J. (2011). Domain loss has independently occurred multiple times in plant terpene synthase evolution: Complex evolutionary origins for terpene synthases. *The Plant Journal*, 68(6), pp. 1051–1060, doi: 10.1111/j.1365-313X.2011.04756.x

119. Ho, C.-L., Yang, S.-S., Chang, T.-M. & Su, Y.-C. (2012). Composition, antioxidant, antimicrobial and anti-wood-decay fungal activities of the twig essential oil of *Taiwania cryptomerioides* from Taiwan. *Natural Product Communications*, 7(2), p. 1934578X1200700, doi: 10.1177/1934578X1200700239

120. Hódar, J. A., Castro, J. & Zamora, R. (2003). Pine processionary caterpillar *Thaumetopoea pityocampa* as a new threat for relict Mediterranean Scots pine forests under climatic warming. Biological conservation, 110(1), 123-129. doi: 10.1016/S0006-3207(02)00183-0

121. Hodar, J.A. & Zamora, R. (2004). Herbivory and climatic warming: a Mediterranean outbreaking caterpillar attacks a relict, boreal pine species. *Biodiversity & Conservation*, 13, pp. 493–500. doi: 10.1023/B:BIOC.0000009495.95589.a7

122. Horton, P., Park, K.-J., Obayashi, T., Fujita, N., Harada, H., Adams-Collier, C. J. & Nakai, K. (2007). WoLF PSORT: protein localization predictor. *Nucleic Acids Research*, 35(Web Server), pp. W585–W587, doi: 10.1093/nar/gkm259

123. Hu, X.-G., Liu, H., Jin, Y., Sun, Y.-Q., Li, Y., Zhao, W., El-Kassaby, Y. A., Wang, X.-R. & Mao, J.-F. (2016). De Novo Transcriptome Assembly and Characterization for the Widespread and Stress-Tolerant Conifer *Platycladus orientalis*. *PLOS ONE*, 11(2), p. e0148985, doi: 10.1371/journal.pone.0148985

124. Huang, M., Abel, C., Sohrabi, R., Petri, J., Haupt, I., Cosimano, J., Gershenzon, J. & Tholl, D. (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

125. Huang, M., Sanchez-Moreiras, A. M., Abel, C., Sohrabi, R., Lee, S., Gershenzon, J. & Tholl, D. (2012). The major volatile organic compound emitted from *Arabidopsis thaliana* flowers, the sesquiterpene (*E*)-β-caryophyllene, is a defense against a bacterial pathogen. *New Phytologist*, 193(4), pp. 997–1008, doi: 10.1111/j.1469-8137.2011.04001.x

126. Hyatt, D. C., Youn, B., Zhao, Y., Santhamma, B., Coates, R. M., Croteau, R. B. & Kang, C. (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

127. Hyatt, David C. & 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

128. Irmisch, S., Jiang, Y., Chen, F., Gershenzon, J. & 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), p. 270, doi: 10.1186/s12870-014-0270-y

129. Isajev, V., Fady, B., Semerci, H. & Andonovski, V. (2004). EUFORGEN Technical Guidelines for genetic conservation and use for European black pine (*Pinus nigra*). IPGRI Rome, Italy

130. Jacquet, J. S., Orazio, C. & Jactel, H. (2012). Defoliation by processionary moth significantly reduces tree growth: a quantitative review. *Annals of Forest Science*, 69(8), pp. 857-866. doi: 10.1007/s13595-012-0209-0

131. Jactel, H., Menassieu, P., Vétillard, F., Barthélémy, B., Piou, D., Frérot, B., Rousselet, J., Goussard, F., Branco, M. & Battisti, A. (2006). Population monitoring of the pine processionary moth (Lepidoptera: Thaumetopoeidae) with pheromone baited traps. *Forest Ecology and Management*, 235, pp. 96–106. https://doi.org/10.1016/j.foreco.2006.08.002

132. Jia, Q. Computational Identification of Terpene Synthase Genes and Their Evolutionary Analysis, p. 221.

133. Jia, Q., Köllner, T. G., Gershenzon, J. & Chen, F. (2018). MTPSLs: New Terpene Synthases in Nonseed Plants. *Trends in Plant Science*, 23(2), pp. 121–128, doi: 10.1016/j.tplants.2017.09.014

134. Jüttner, F. (1988). Changes of monoterpene concentration in needles of pollution-injured *Picea abies* exhibiting montane yellowing. *Physiologia Plantarum*, 72, pp. 48-56. doi: 10.1111/j.1399-3054.1988.tb06621.x

135. Keeling, C. I., Dullat, H. K., Yuen, M., Ralph, S. G., Jancsik, S. & Bohlmann, J. (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

136. Keeling, C. I., Weisshaar, S., Ralph, S. G., Jancsik, S., Hamberger, B., Dullat, H. K. & Bohlmann, J. (2011). Transcriptome mining, functional characterization, and phylogeny of a large terpene synthase gene family in spruce (*Picea* spp.). *BMC Plant Biology*, 11(1), p. 43, doi: 10.1186/1471-2229-11-43

137. Kerdelhué, C., Magnoux, E., Lieutier, F., Roques, A. & Rousselet J. (2006). Comparative population genetic study of two oligophagous insects associated with the same hosts. *Heredity*, 97, pp. 38–45 doi: 10.1038/sj.hdy.6800836

138. Kirby, J., Nishimoto, M., Park, J. G., Withers, S. T., Nowroozi, F., Behrendt, D., Rutledge, E. J. G., Fortman, J. L., Johnson, H. E. & Anderson, J. V. (2010). Cloning of casbene and neocembrene synthases from Euphorbiaceae plants and expression in Saccharomyces cerevisiae. *Phytochemistry*, 71(13), pp. 1466–1473, doi: 10.1016/j.phytochem.2010.06.001

139. Köksal, M., Hu, H., Coates, R. M., Peters, R. J. & Christianson, D. W. (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

140. Köksal, M., Jin, Y., Coates, R. M., Croteau, R. & Christianson, D. W. (2011). Taxadiene synthase structure and evolution of modular architecture in terpene biosynthesis. *Nature*, 469(7328), pp. 116–120, doi: 10.1038/nature09628

141. Köksal, M., Potter, K., Peters, R. J. & Christianson, D. W. (2014). Å-resolution structure of *ent*-copalyl diphosphate synthase and exploration of general acid function by site-directed mutagenesis. *Biochimica et Biophysica Acta (BBA) - General Subjects*, 1840(1), pp. 184–190, doi: 10.1016/j.bbagen.2013.09.004

142. Komenda, M. & Koppmann, R. (2002). Monoterpene emissions from Scots pine (*Pinus sylvestris*): field studies of emission rate variabilities. *Journal of Geophysical Research Atmospheres*, 107. https://doi.org/10.1029/2001JD000691

143. Kopaczyk, J. M., Warguła, J. & Jelonek, T. (2020). The variability of terpenes in conifers under developmental and environmental stimuli. *Environmental and Experimental Botany*, pp. 104197. doi: 10.1016/j.envexpbot.2020.104197

144. Koricheva, J., Nykänen, H. & Gianoli, E. (2004). Meta-analysis of trade-offs among plant antiherbivore defenses: are plants jacks-of-all-trades, masters of all?. *The American Naturalist*, 163(4), E64-E75. doi: 10.1086/382601

145. Kshatriya, K., Whitehill, J. G. A., Madilao, L., Henderson, H., Kermode, A., Kolotelo, D. & Bohlmann, J. (2018). Histology of resin vesicles and oleoresin terpene composition of conifer seeds. *Canadian Journal of Forest Research*, 48(9), pp. 1073–1084, doi: 10.1139/cjfr-2018-0164

146. Külheim, C., Padovan, A., Hefer, C., Krause, S. T., Köllner, T. G., Myburg, A. A., Degenhardt, J. & Foley, W. J. (2015). The Eucalyptus terpene synthase gene family. *BMC Genomics*, 16(1), p. 450, doi: 10.1186/s12864-015-1598-x

147. Kumar, S., Stecher, G., Li, M., Knyaz, C. & Tamura, K. (2018). MEGA X: molecular evolutionary genetics analysis across computing platforms. *Molecular Biology and Evolution*, 35(6), pp. 1547–1549, doi: 10.1093/molbev/msy096

148. Kupcinskiene, E., Stikliene, A. & Judzentiene, A. (2008). The essential oil qualitative and quantitative composition in the needles of *Pinus sylvestris* L. growing along industrial transect. *Environmental Pollution* 155 pp. 481-491. doi: 10.1016/j.envpol.2008.02.001

149. Lah, L., Haridas, S., Bohlmann, J. & Breuil, C. (2013). The cytochromes P450 of *Grosmannia clavigera*: Genome organization, phylogeny, and expression in response to pine host chemicals. *Fungal Genetics and Biology*, 50, pp. 72–81, doi: 10.1016/j.fgb.2012.10.002

150. Lamy, M. (1990). Contact dermatitis (Erucism) produced by processionary caterpillars (Genus *Thaumetopoea*). *Journal of Applied Entomology*, 110, pp. 425-437. doi: 10.1111/j.1439-0418.1990.tb00142.x

151. Laothawornkitkul, J., Paul, N. D., Vickers, C. E., Possell, M., Taylor, J. E., Mullineaux, P. M. & Hewitt, C. N. (2008). Isoprene emissions influence herbivore feeding decisions. *Plant, Cell & Environment*, 31(10), pp. 1410–1415, doi: 10.1111/j.1365-3040.2008.01849.x

152. Ledesma, L. (1971), Notas relativas a la distribucion y predacion de puestas de procesionaria del pino (*Thaumetopoea pityocampa* Schiff.) sobre pies en edades de monte bravo y latizal de pino negral (*Pinus laricio* Poir.). *Boletin del Servicio de Plagas Forestales*, 14,71-80.

153. Lehnert, A. S., Perreca, E., Gershenzon, J., Pohnert, G., & Trumbore, S. E. (2020). Simultaneous real-time measurement of isoprene and 2-methyl-3-buten-2-ol emissions from trees using SIFT-MS. *Frontiers in plant science*, *11*, 1867, doi: 10.3389/fpls.2020.578204

154. Lerdau, M., & Gray, D. (2003). Ecology and evolution of light-dependent and light-independent phytogenic volatile organic carbon. *New Phytologist*, *157*(2), 199-211, doi: 10.1046/j.1469-8137.2003.00673.x

155. Lesburg, C. A., Zhai, G., Cane, D. E., & Christianson, D. W. (1997). Crystal structure of pentalenene synthase: mechanistic insights on terpenoid cyclization reactions in biology. *Science*, *277*(5333), 1820-1824, doi: 10.1126/science.277.5333.1820

156. Li, G., Kollner, T. G., Yin, Y., Jiang, Y., Chen, H., Xu, Y., Gershenzon, J., Pichersky, E. & Chen, F. (2012). Nonseed plant *Selaginella moellendorffii* has both seed plant and microbial types of terpene synthases. *Proceedings of the National Academy of Sciences*, 109(36), pp. 14711–14715, doi: 10.1073/pnas.1204300109

157. Li, W. *Molecular evolution*. Sinauer associates incorporated, 1997.

158. Litvak, M. E. & Monson, R. K. (1998). Patterns of induced and constitutive monoterpene production in conifer needles in relation to insect herbivory. *Oecologia*, 114(4), pp. 531-540. doi: 10.1007/s004420050477

159. Loreto, F., Dicke, M., Schnitzler, J.-P. & Turlings, T. C. J. (2014). Plant volatiles and the environment: Plant volatiles and the environment. *Plant, Cell & Environment*, 37(8), pp. 1905–1908, doi: 10.1111/pce.12369

160. Loreto, F., Fischbach, R. J., Schnitzler, J.-P., Ciccioli, P., Brancaleoni, E., Calfapietra, C. & Seufert, G. (2001). Monoterpene emission and monoterpene synthase activities in the Mediterranean evergreen oak *Quercus ilex* L. grown at elevated CO_2 concentrations: monoterpene emission at E levated co $_2$ by holm oak. *Global Change Biology*, 7(6), pp. 709–717, doi: 10.1046/j.1354-1013.2001.00442.x

161. Loreto, F., Förster, A., Dürr, M., Csiky, O. & Seufert, G. (1998). On the monoterpene emission under heat stress and on the increased thermotolerance of leaves of *Quercus ilex* L. fumigated with selected monoterpenes. *Plant, Cell & Environment* 21: 101-107. doi: 10.1046/j.1365-3040.1998.00268.x

162. Loreto, F. & Schnitzler, J.-P. (2010). Abiotic stresses and induced BVOCs. *Trends in Plant Science*, 15(3), pp. 154–166, doi: 10.1016/j.tplants.2009.12.006

163. López-Goldar, X., Lundborg, L., Borg-Karlson, A. K., Zas, R. & Sampedro, L. (2020). Resin acids as inducible chemical defenses of pine seedlings against chewing insects. *Plos One*, 15(5), pp. e0232692, doi: 10.1371/journal.pone.0232692

164. Ma, L., Lee, Y., Tsao, N., Wang, S., Zerbe, P. & Chu, F. (2019). Biochemical characterization of diterpene synthases of *Taiwania cryptomerioides* expands the known functional space of specialized diterpene metabolism in gymnosperms. *Plant Journal*, 100(6), pp. 1254–1272, doi: 10.1111/tpj.14513

165. Ma, L.-T., Wang, C.-H., Hon, C.-Y., Lee, Y.-R. & Chu, F.-H. (2021). Discovery and characterization of diterpene synthases in *Chamaecyparis formosensis* Matsum. which participated in an unprecedented diterpenoid biosynthesis route in conifer. *Plant Science*, 304, pp. 110790, doi: 10.1016/j.plantsci.2020.110790

166. Mafu, S., Hillwig, M. L. & Peters, R. J. (2011). A novel labda-7,13E-dien-15-ol-producing bifunctional diterpene synthase from *Selaginella moellendorffii*. *ChemBioChem*, 12(13), pp. 1984–1987, doi: 10.1002/cbic.201100336

167. Mafu, S., Karunanithi, P. S., Palazzo, T. A., Harrod, B. L., Rodriguez, S. M., Mollhoff, I. N., O'Brien, T. E., Tong, S., Fiehn, O., Tantillo, D. J., Bohlmann, J. & Zerbe, P. (2017). Biosynthesis of the microtubule-destabilizing diterpene pseudolaric acid B from golden larch involves an unusual diterpene synthase. *Proceedings of the National Academy of Sciences*, 114(5), pp. 974–979, doi: 10.1073/pnas.1612901114

168. Mageroy, M.H., Christiansen, E., Långström, B., Borg-karlson, A., Solheim, H., Björklund, N., Zhao, T., Schmidt, A., Fossdal, C.G. & Krokene, P. (2019). Priming of inducible defenses protects Norway spruce against tree-killing bark beetles. *Plant, Cell & Environment*, 43, pp. 420-430. doi: 10.1111/pce.13661

169. Markalas, S. (1989). Influence of soil moisture on the mortality, fecundity and diapause of the pine 363 processionary moth (*Thaumetopoea pityocampa* Schiff.). *Journal of Applied Entomology*, 107, pp. 211-215. doi: 10.1111/j.1439-0418.1989.tb00250.x

170. Martin, D. M., Aubourg, S., Schouwey, M. B., Daviet, L., Schalk, M., Toub, O., Lund, S. T. & Bohlmann, J. (2010). Functional annotation, genome organization and phylogeny of the grapevine (*Vitis vinifera*) terpene synthase gene family based on genome assembly, FL cDNA cloning, and enzyme assays. *BMC Plant Biology*, 10(1), p. 226, doi: 10.1186/1471-2229-10-226

171. Martin, D. M., Fäldt, J. & Bohlmann, J. (2004). Functional characterization of nine Norway spruce *TPS* genes and evolution of gymnosperm terpene synthases of the *TPS-d* subfamily. *Plant Physiology*, 135(4), pp. 1908–1927, doi: 10.1104/pp.104.042028

172. Masutti, L. & Battisti, A., (1990). *Thaumetopoea pityocampa* (Den. & Schiff.) in Italy - Bionomics and perspectives of integrated control. *Journal of Applied Entomology*, 110, pp. 229-234. doi: 10.1111/j.1439-0418.1990.tb00117.x

173. Mau, C. J. & West, C. A. (1994). Cloning of casbene synthase cDNA: evidence for conserved structural features among terpenoid cyclases in plants. *Proceedings of the National Academy of Sciences*, 91(18), pp. 8497–8501, doi: 10.1073/pnas.91.18.8497

174. May, B., Lange, B. M. & Wüst, M. (2013). Biosynthesis of sesquiterpenes in grape berry exocarp of *Vitis vinifera* L.: Evidence for a transport of farnesyl diphosphate precursors from plastids to the cytosol. *Phytochemistry*, 95, pp. 135–144, doi: 10.1016/j.phytochem.2013.07.021

175. McAndrew, R. P., Peralta-Yahya, P. P., DeGiovanni, A., Pereira, J. H., Hadi, M. Z., Keasling, J. D. & Adams, P. D. (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

176. Mewalal, R., Rai, D. K., Kainer, D., Chen, F., Külheim, C., Peter, G. F. & Tuskan, G. A. (2017). Plant-derived terpenes: a feedstock for specialty biofuels. *Trends in biotechnology*, 35(3) pp. 227-240. doi: 10.1016/j.tibtech.2016.08.003

177. Miller, B., Madilao, L.L., Ralph, S., Bohlmann, J. (2005) Insect-induced conifer defense. White pine weevil and methyl jasmonate induce traumatic resinosis, de Novo formed volatile emissions, and accumulation of terpenoid synthase and putative octadecanoid pathway transcripts in Sitka spruce. *Plant Physiol* 137:369–382, doi: 10.1104/pp.103.022723

178. Mo, J., Xu, J., Jin, W., Yang, L., Yin, T. & Shi, J. (2019). Identification of reference genes for quantitative gene expression studies in *Pinus massoniana* and its introgression hybrid. *Forests*, 10(9), pp. 787, doi: 10.3390/f10090787.

179. Monson, R. K., Jones, R. T., Rosenstiel, T. N. & Schnitzler, J.-P. (2013). Why only some plants emit isoprene: phylogeny of isoprene. *Plant, Cell & Environment*, 36(3), pp. 503–516, doi: 10.1111/pce.12015

180. Morrone, D., Chambers, J., Lowry, L., Kim, G., Anterola, A., Bender, K. & Peters, R. J. (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

181. Muirhead, J.R., Leung, B., Van Overdijk, C., Kelly, D.W., Nandakumar, K., Marchant, K.R. & Macisaac, H.J. (2006). Modelling local and long-distance dispersal of invasive emerald ash borer *Agrilus planipennis* (Coleoptera) in North America. *Diversity and Distributions*, 12, 71–79. doi: 10.1111/j.1366-9516.2006.00218.x

182. Mukrimin, M., Kovalchuk, A., Ghimire, R. P., Kivimäenpää, M., Sun, H., Holopainen, J. K., & Asiegbu, F. O. (2019). Evaluation of potential genetic and chemical markers for Scots pine tolerance against *Heterobasidion annosum* infection. *Planta*, *250*(6), 1881-1895, doi: 10.1007/s00425-019-03270-8

183. Mumm, R., & Hilker, M. (2006). Direct and indirect chemical defense of pine against folivorous insects. *Trends in plant science*, 11(7), 351-358. doi: 10.1016/j.tplants.2006.05.007

184. Mumm, R., Posthumus, M. A. & Dicke, M. (2008). Significance of terpenoids in induced indirect plant defense against herbivorous arthropods. *Plant, Cell & Environment*, 31(4), pp. 575–585, doi: 10.1111/j.1365-3040.2008.01783.x

185. Nagegowda, D. A., Gutensohn, M., Wilkerson, C. G. & Dudareva, N. (2008). Two nearly identical terpene synthases catalyze the formation of nerolidol and linalool in snapdragon flowers. *The Plant Journal*, 55(2), pp. 224–239, doi: 10.1111/j.1365-313X.2008.03496.x

186. Neale, D. B., Wegrzyn, J. L., Stevens, K. A., Zimin, A. V., Puiu, D., Crepeau, M. W., Cardeno, C., Koriabine, M., Holtz-Morris, A. E., Liechty, J. D., Martínez-García, P. J., Vasquez-Gross, H. A., Lin, B. Y., Zieve, J. J., Dougherty, W. M., Fuentes-Soriano, S., Wu, L.-S., Gilbert, D., Marçais, G., Roberts, M., Holt, C., Yandell, M., Davis, J. M., Smith, K. E., Dean, J. F., Lorenz, W., Whetten, R. W., Sederoff, R., Wheeler, N., McGuire, P. E., Main, D., Loopstra, C. A., Mockaitis, K., deJong, P. J., Yorke, J. A., Salzberg, S. L. & Langley, C. H. (2014). Decoding the massive genome of loblolly pine using haploid DNA and novel assembly strategies. *Genome Biology*, 15(3), p. R59, doi: 10.1186/gb-2014-15-3-r59

187. Nelson, D. R. (2011). Progress in tracing the evolutionary paths of cytochrome P450. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics*, 1814(1), pp. 14–18, doi: 10.1016/j.bbapap.2010.08.008

188. Nicolaci, A., Travaglini, D., Menguzzato, G., Nocentini, S., Veltri, A. & Iovino, F. (2015). Ecological and anthropogenic drivers of Calabrian pine (*Pinus nigra* J.F. Arn. ssp. *laricio* (Poiret) Maire) distribution in the Sila mountain range. *iForest - Biogeosciences and Forestry*, 8(4), pp. 497–508, doi: 10.3832/ifor1041-007

189. Nicole, M.C., Zeneli, G., Lavallée, R., Rioux, D., Bauce, É., Morency, M.J., Fenning, T.M. & Séguin, A. (2006). White pine weevil (*Pissodes strobi*) biological performance in un-affected

by the jasminic acid or wound-induced defense response in Norway spruce (*Picea abies*). *Tree Physiol*, 26, pp. 1377-1389. doi: 10.1093/treephys/26.11.1377

190. Niinemets, Ü. & Monson, R. K. eds. (2013). *Biology, Controls and Models of Tree Volatile Organic Compound Emissions*. Dordrecht: Springer Netherlands.

191. Nishino, C. & Manabe, S. (1983). Olfactory receptor systems for sex pheromone mimics in the American cockroach, *Periplaneta americana* L. *Experientia*, 39(12), 1340-1342. doi: 10.1007/BF01990092

192. Nystedt, B., Street, N. R., Wetterbom, A., Zuccolo, A., Lin, Y.-C., Scofield, D. G., Vezzi, F., Delhomme, N., Giacomello, S., Alexeyenko, A., Vicedomini, R., Sahlin, K., Sherwood, E., Elfstrand, M., Gramzow, L., Holmberg, K., Hällman, J., Keech, O., Klasson, L., Koriabine, M., Kucukoglu, M., Käller, M., Luthman, J., Lysholm, F., Niittylä, T., Olson, Å., Rilakovic, N., Ritland, C., Rosselló, J. A., Sena, J., Svensson, T., Talavera-López, C., Theißen, G., Tuominen, H., Vanneste, K., Wu, Z.-Q., Zhang, B., Zerbe, P., Arvestad, L., Bhalerao, R., Bohlmann, J., Bousquet, J., Garcia Gil, R., Hvidsten, T. R., de Jong, P., MacKay, J., Morgante, M., Ritland, K., Sundberg, B., Lee Thompson, S., Van de Peer, Y., Andersson, B., Nilsson, O., Ingvarsson, P. K., Lundeberg, J. & Jansson, S. (2013). The Norway spruce genome sequence and conifer genome evolution. *Nature*, 497(7451), pp. 579–584, doi: 10.1038/nature12211

193. Orlova, I., Nagegowda, D. A., Kish, C. M., Gutensohn, M., Maeda, H., Varbanova, M., Fridman, E., Yamaguchi, S., Hanada, A., Kamiya, Y., Krichevsky, A., Citovsky, V., Pichersky, E. & Dudareva, N. (2010). 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

194. Ormeño, E., Baldy, V., Ballini, C. & Fernandez, C. (2008). Production and diversity of volatile terpens from plants on calcareous and siliceous soil: effect of soil nutrients. *Journal of Chemical Ecology*, 34, pp. 1219-1229. doi: 10.1007/s10886-008-9515-2

195. Ovaskainen, O. (2004). Habitat-specific movement parameters estimated using mark–recapture data and a diffusion model. *Ecology*, 85, pp. 242–257. doi: 10.1890/02-0706

196. Ovaskainen, O. & Cornell, S.J. (2003). Biased movement at boundary and conditional occupancy times for diffusion processes. *Journal of Applied Probability*, 40, pp. 557–580. doi: 10.1239/jap/1059060888

197. Paine, T. D., Raffa, K. F. & Harrington, T. C. (1997). Interactions among scolytid bark beetles, their associated fungi, and live host conifers. *Annual Review of Entomology*, 42(1), pp. 179-206. doi: 10.1146/annurev.ento.42.1.179

198. Paolacci, A. R., Catarcione, G., Ederli, L., Zadra, C., Pasqualini, S., Badiani, M., Musetti, R., Santi, S. & Ciaffi, M. (2017). Jasmonate-mediated defense responses, unlike salicylate-mediated responses, are involved in the recovery of grapevine from bois noir disease. *BMC Plant Biology*, 17(1), p. 118, doi: 10.1186/s12870-017-1069-4

199. Parlak, S., Özçankaya, İ.M., Batur, M., Akkaş, M.E., Boza, Z. & Toprak, Ö. (2019) Determining the edge effect of pine processionary moth (*Thaumetopoea pityocampa*) in its horizontal distribution in the stand. *Journal of Forestry Research*, 30, pp. 347–352. doi: 10.1007/s11676-018-0634-5

200. Pavy, N., Boyle, B., Nelson, C., Paule, C., Giguère, I., Caron, S., Parsons, L. S., Dallaire, N., Bedon, F., Bérubé, H., Cooke, J. & Mackay, J. (2008). Identification of conserved core xylem gene sets: conifer cDNA microarray development, transcript profiling and computational analyses. *New Phytologist*, 180(4), pp. 766–786, doi: 10.1111/j.1469-8137.2008.02615.x

201. Pazouki, L. & Niinemets, Ü. (2016). Multi-substrate terpene synthases: their occurrence and physiological significance. *Frontiers in Plant Science*, 7, doi: 10.3389/fpls.2016.01019

202. Peñuelas, J. & Llusià, J. (2004). Plant VOC emissions: making use of the unavoidable. Trends in ecology & evolution, 19(8), pp. 402-404. doi: 10.1016/j.tree.2004.06.002

203. Perez-Mendoza, J., Hagstrum, D.W., Dover, B.A., Hopkins, T.L. & Baker, J.E. (1999). Flight response, body weight, and lipid content of *Rhyzoperta dominica* (F) (Coleoptera: Bostrichidae) as influenced by strain, season and phenotype. *Journal of Stored Products Resource*, 35, pp. 183–196. doi: 10.1016/S0022-474X(98)00044-7

204. Peters, R. J. (2010). Two rings in them all: The labdane-related diterpenoids. *Natural Product Reports*, 27(11), pp. 1521, doi: 10.1039/c0np00019a

205. Peters, R. J., Carter, O. A., Zhang, Y., Matthews, B. W. & Croteau, R. B. (2003). Bifunctional abietadiene synthase: mutual structural dependence of the active sites for protonation-initiated and ionization-initiated cyclizations. *Biochemistry*, 42(9), pp. 2700–2707, doi: 10.1021/bi020492n

206. Phillips, M.A., Wildung, M.R., Williams, D.C., Hyatt, D.C. & Croteau, R. (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, pp. 267-276. doi: 10.1016/S0003-9861(02)00746-4

207. Phillips, M. A. & Croteau, R. B. (1999). Resin-based defenses in conifers. *Trends in plant science*, 4(5), pp. 184-190. doi: 10.1016/S1360-1385(99)01401-6

208. Pierik, R., Ballaré, C. L. & Dicke, M. (2014). Ecology of plant volatiles: taking a plant community perspective: Ecology of plant volatiles. *Plant, Cell & Environment*, 37(8), pp. 1845–1853, doi: 10.1111/pce.12330

209. Pokorska, O., Dewulf, J., Amelynck, C., Schoon, N., Šimpraga, M., Steppe, K., & Van Langenhove, H. (2012). Isoprene and terpenoid emissions from *Abies alba*: identification and emission rates under ambient conditions. *Atmospheric environment*, *59*, 501-508, doi: 10.1016/j.atmosenv.2012.04.061

210. Pollastrini, M., Luchi, N., Michelozzi, M., Gerosa, G., Marzuoli, R., Bussotti, F. & Capretti, P. (2015). Early physiological responses of *Pinus pinea* L. seedlings infected by *Heterobasidion* sp. pl. in an ozone-enriched atmospheric environment. *Tree Physiology*, 35(3) pp. 331-340. doi: 10.1093/treephys/tpv008

211. Porres-Martínez, M., González-Burgos, E., Carretero, M. E., & Gómez-Serranillos, M. P. (2016). In vitro neuroprotective potential of the monoterpenes α-pinene and 1, 8-cineole against H2O2-induced oxidative stress in PC12 cells. *Zeitschrift für Naturforschung C*, *71*(7-8), 191-199, doi: 10.1515/znc-2014-4135

212. Prisic, S., Xu, M., Wilderman, P. R. & Peters, R. J. (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

213. Quero, C. Malo, E. A., Fabrias, G., Camps, F. Lucas, P., Renou, M. &, Guerrero, A. (1997). Reinvestigation of female sex pheromone of processionary moth (*Thaumetopoea pityocampa*): no evidence for minor components. *Journal of Chemical Ecology*, 23, 713-726. doi: 10.1023/B:JOEC.0000006406.30444.69

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

215. Raffa, K. F. (2014). Terpenes tell different tales at different scales: glimpses into the chemical ecology of conifer-bark beetle-microbial interactions. *Journal of Chemical Ecology*, *40*(1), 1-20. doi: 10.1007/s10886-013-0368-y

216. Rasmann, S., Köllner, T. G., Degenhardt, J., Hiltpold, I., Toepfer, S., Kuhlmann, U., Gershenzon, J. & Turlings, T. C. J. (2005). Recruitment of entomopathogenic nematodes by insect-damaged maize roots. *Nature*, 434(7034), pp. 732–737, doi: 10.1038/nature03451

217. Rasulov, B., Talts, E., Kännaste, A. & Niinemets, Ü. (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

218. Ro, D.-K. & 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

219. Robinet, C., Laparie, M., & Rousselet, J. (2015). Looking beyond the large-scale effects of global change: local phenologies can result in critical heterogeneity in the pine processionary moth. *Frontiers in Physiology*, 6, 334, pp. 5. doi: 10.3389/fphys.2015.00334

220. Roques, A. (Ed.). (2015). *Processionary moths and climate change: an update* (Vol. 427). Springer Netherlands.

221. Rosenkranz, M., Chen, Y., Zhu, P. & Vlot, A. C. (2021). Volatile terpenes–mediators of plant-to-plant communication. *The Plant Journal*. doi: 10.1111/tpj.15453

222. Rozman, V., Kalinovic, I. & Korunic, Z. (2007). Toxicity of naturally occurring compounds of Lamiaceae and Lauraceae to three stored-product insects. *Journal of stored products research*, 43(4), pp. 349-355. doi: 10.1016/j.jspr.2006.09.001

223. Ryan, M. F. & Guerin, P. M. (1982). Behavioural responses of the carrot fly larva, *Psila rosae*, to carrot root volatiles. *Physiological Entomology*, 7(3), pp. 315-324. doi: 10.1111/j.1365-3032.1982.tb00304.x

224. Saitou, N. & Nei, M. (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution*, 4(4), pp. 406-425 doi: 10.1093/oxfordjournals.molbev.a040454

225. Samalens, J.C. & Rossi, J.P. (2011). Does landscape composition alter the spatiotemporal distribution of the pine processionary moth in a pine plantation forest? *Population Ecology*, 53, pp. 287–296. doi: 10.1007/s10144-010-0227-4

226. Sallas, L., Luomala, E. M., Utriainen, J., Kainulainen, P. & Holopainen, J. K. (2003). Contrasting effects of elevated carbon dioxide concentration and temperature on Rubisco activity, chlorophyll fluorescence, needle ultrastructure and secondary metabolites in conifer seedlings. *Tree Physiology*, 23(2) pp. 97-108. doi: 10.1093/treephys/23.2.97

227. Sallaud, C., Rontein, D., Onillon, S., Jabès, F., Duffé, P., Giacalone, C., Thoraval, S., Escoffier, C., Herbette, G., Leonhardt, N., Causse, M. & Tissier, A. (2009). A novel pathway for sesquiterpene biosynthesis from *Z,Z* -farnesyl pyrophosphate in the wild tomato *Solanum habrochaites*. *The Plant Cell*, 21(1), pp. 301–317, doi: 10.1105/tpc.107.057885

228. Sampedro, L., Moreira, X. & Zas, R. (2011). Costs of constitutive and herbivore-induced chemical defenses in pine trees emerge only under low nutrient availability. *Journal of Ecology*, 99(3), pp. 818-827. doi: 10.1111/j.1365-2745.2011.01814.x

229. Schilmiller, A. L., Schauvinhold, I., Larson, M., Xu, R., Charbonneau, A. L., Schmidt, A., Wilkerson, C., Last, R. L. & Pichersky, E. (2009). Monoterpenes in the glandular trichomes of tomato are synthesized from a neryl diphosphate precursor rather than geranyl diphosphate. *Proceedings of the National Academy of Sciences*, 106(26), pp. 10865–10870, doi: 10.1073/pnas.0904113106

230. Schnitzler, W. H. & Krauss, S. (2010). Quality and health promoting compounds of tomato fruit (*Lycopersicon esculentum* Mill.) under salinity. *Acta Horticulturae*, (856), pp. 21–30, doi: 10.17660/ActaHortic.2010.856.2

231. Schultz, C.B. & Crone, E.E. (2001). Edge-mediated dispersal behaviour in a prairie butterfly. *Ecology*, 82, pp. 1879–1892. doi: 10.1890/0012-9658(2001)082[1879:EMDBIA]2.0.CO;2

232. Shalev, T. J., Yuen, M. M. S., Gesell, A., Yuen, A., Russell, J. H. & Bohlmann, J. (2018). An annotated transcriptome of highly inbred *Thuja plicata* (Cupressaceae) and its utility for gene discovery of terpenoid biosynthesis and conifer defense. *Tree Genetics & Genomes*, 14(3), p. 35, doi: 10.1007/s11295-018-1248-y

233. Sharkey, T. D. & Yeh, S. (2001). Isoprene e mission from plants. *Annual Review of Plant Physiology and Plant Molecular Biology*, 52(1), pp. 407–436, doi: 10.1146/annurev.arplant.52.1.407

234. Sharkey, T. D., Gray, D. W., Pell, H. K., Breneman, S. R., & Topper, L. (2013). Isoprene synthase genes form a monophyletic clade of acyclic terpene synthases in the Tps-b terpene synthase family. *Evolution: International Journal of Organic Evolution*, *67*(4), 1026-1040, doi: 10.1111/evo.12013

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

236. Small, I., Peeters, N., Legeai, F. & Lurin, C. (2004). Predotar: A tool for rapidly screening proteomes forN-terminal targeting sequences. *Proteomics*, 4(6), pp. 1581–1590, doi: 10.1002/pmic.200300776

237. Smart, R. C., & Hodgson, E. (Eds.). (2018). Molecular and biochemical toxicology. John Wiley & Sons.

238. Sobral, M. V., Xavier, A. L., Lima, T. C., & de Sousa, D. P. (2014). Antitumor activity of monoterpenes found in essential oils. *The Scientific World Journal*, *2014*, doi: 10.1155/2014/953451

239. Staudt, M. & Lhoutellier, L. (2007). Volatile organic compound emission from holm oak infested by gypsy moth larvae: evidence for distinct responses in damaged and undamaged leaves. *Tree physiology*, 27(10), 1433-1440. doi: 10.1093/treephys/27.10.1433

240. Steele, C. L., Katoh, S., Bohlmann, J. & Croteau, R. (1998). Regulation of oleoresinosis in grand fir (*Abies grandis*) differential transcriptional control of monoterpene, sesquiterpene, and diterpene synthase genes in response to wounding. *Plant Physiology*, 116(4) pp. 1497-1504. doi: 10.1104/pp.116.4.1497

241. Stofer Vogel, B., Wildung, M. R., Vogel, G. & Croteau, R. (1996). Abietadiene synthase from grand fir (*Abies grandis*) cDNA isolation, characterization, and bacterial expression of a bifunctional diterpene cyclase involved in resin acid biosynthesis. *Journal of Biological Chemistry*, 271(38), pp. 23262–23268, doi: 10.1074/jbc.271.38.23262

242. Supuka, J., Berta, F. & Chladná, A. (1997). The influence of the urban environment on the composition of terpenes in the needles of Black pine (*Pinus nigra* Arnold). *Tres*, 11, pp. 176-182. doi: 10.1007/PL00009666

243. Surmacz, L. & Swiezewska, E. (2011). Polyisoprenoids – Secondary metabolites or physiologically important superlipids? *Biochemical and Biophysical Research Communications*, 407(4), pp. 627–632, doi: 10.1016/j.bbrc.2011.03.059

244. Takahashi, S. & Koyama, T. (2006). Structure and function ofcis-prenyl chain elongating enzymes. *The Chemical Record*, 6(4), pp. 194–205, doi: 10.1002/tcr.20083

245. Tarasco, E. (1995). Morfologia larvale e biologia di *Coelichneumon rudis* (Boyer de Fonscolombe) (Hymenoptera: Ichneumonidae), endoparassitoide delle crisalidi della *Thaumetopoea pityocampa* (Denis et Schiffermüller) (Lepidoptera: Thaumetopoeidae). *Entomologica, Bari*, 29, 5-51.

246. Tarshis, L. C., Proteau, P. J., Kellogg, B. A., Sacchettini, J. C., & Poulter, C. D. (1996). Regulation of product chain length by isoprenyl diphosphate synthases. *Proceedings of the National Academy of Sciences*, *93*(26), 15018-15023, doi: 10.1073/pnas.93.26.15018

247. Thiel, D., Nagy, L., Beierkuhnlein, C., Huber, G., Jentsch, A., Konnert, M. & Kreyling, J. (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

248. Thimmappa, R., Geisler, K., Louveau, T., O'Maille, P. & Osbourn, A. (2014). Triterpene biosynthesis in plants. *Annual Review of Plant Biology*, 65(1), pp. 225–257, doi: 10.1146/annurev-arplant-050312-120229

249. Tholl, D., Lee, S. (2011). Terpene specialized metabolism in *Arabidopsis thaliana*. The *Arabidopsis* Book/American Society of Plant Biologists, 9.

250. Tholl, D. (2015). biosynthesis and biological functions of terpenoids in plants. In: Schrader, J. & Bohlmann, J. (eds.) *Biotechnology of Isoprenoids*. Cham: Springer International Publishing, pp. 63–106.

251. Thompson, J. (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

252. Tiberi R. & Niccoli A. (1984) Osservazioni pluriennali sull'impiego di trappole con il feromone sessuale di *Thaumetopoea pityocampa* (Den. et Schiff.) (Lepidoptera, Thaumetopoeidae). *Redia*, 67, 129-144.

253. Tingey, D.T., Manning, M., Grothaus, L.C. & Burns, W.F. (1979). The influence of light and temperature on isoprene emission rates from live oak. *Physiologia Plantarum*, 47, pp. 112-118. doi: 10.1111/j.1399-3054.1979.tb03200.x

254. Torres-Muros, L., Hódar, J.A. & Zamora, R. (2017). Effect of habitat type and soil moisture on pupal stage of a Mediterranean forest pest (*Thaumetopoea pityocampa*). *Agricoltural and Forest Entomology*, 19, pp. 130–138. doi: 10.1111/afe.12188

255. Townsend, B. J., Poole, A., Blake, C. J. & Llewellyn, D. J. (2005). Antisense suppression of a (+)- δ -cadinene synthase gene in cotton prevents the induction of this defense response gene during bacterial blight infection but not its constitutive expression. *Plant Physiology*, 138(1), pp. 516–528, doi: 10.1104/pp.104.056010

256. Trapp, S. C. & Croteau, R. B. (2001). Genomic organization of plant terpene synthases and molecular evolutionary implications. *Genetics*, 158(2), pp. 811–832, doi: 10.1093/genetics/158.2.811

257. Turner, G.W., Parrish, A.N., Zager, J.J., Fischedick, J.T., Lange, B.M. (2019). Assessment of flux through oleoresin biosynthesis in epithelial cells of loblolly pine resin ducts. *Journal of Experimental Botany*, 70(1), 217–230. https://doi.org/10.1093/jxb/ery338

258. Turtola, S. (2005). The effects of drought stress and enhanced UV-B radiation on the growth and secondary chemistry of boreal conifer and willow seedlings. PhD Diss Biol 82: 1457-2486.

259. Turtola, S., Manninen, A.M., Rikala, R. & Kainulainen, P. (2003). Drought stress alters the concentartion of wood terpenoids in Scots pine and Norway spruce seedlings. *Journal of Chemical Ecology*, 29, pp. 1981-1995. doi: 10.1023/A:1025674116183

260. Ueda, H., Kikuta, Y. & Matsuda, K. (2012). Plant communication: mediated by individual or blended VOCs?. *Plant Signaling & Behavior*, 7(2) pp. 222-226. doi: 10.4161/psb.18765

261. Vaughan, M. M., Wang, Q., Webster, F. X., Kiemle, D., Hong, Y. J., Tantillo, D. J., Coates, R. M., Wray, A. T., Askew, W., O'Donnell, C., Tokuhisa, J. G. & Tholl, D. (2013). Formation of the unusual semivolatile diterpene rhizathalene by the *Arabidopsis* class I terpene synthase TPS08 in the root stele is involved in defense against belowground herbivory. *The Plant Cell*, 25(3), pp. 1108–1125, doi: 10.1105/tpc.112.100057

262. Velikova, V., Ghirardo, A., Vanzo, E., Merl, J., Hauck, S. M. & Schnitzler, J.-P. (2014). Genetic manipulation of isoprene emissions in poplar plants remodels the chloroplast proteome. *Journal of Proteome Research*, 13(4), pp. 2005–2018, doi: 10.1021/pr401124z

263. Viiri, H., Annila, E., Kitunen, V. & Niemelä, P. (2001). Induced responses in stilbenes and terpens in fertilized Norway spruce after inoculation with blue-stain fungus, *Ceratocystis polonica*. *Trees Structure and Function*, 15, pp. 112-122. doi: 10.1007/s004680000082

264. Vranová, E., Coman, D. & 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

265. Warren, R. L., Keeling, C. I., Yuen, M. M. S., Raymond, A., Taylor, G. A., Vandervalk, B. P., Mohamadi, H., Paulino, D., Chiu, R., Jackman, S. D., Robertson, G., Yang, C., Boyle, B., Hoffmann, M., Weigel, D., Nelson, D. R., Ritland, C., Isabel, N., Jaquish, B., Yanchuk, A., Bousquet, J., Jones, S. J. M., MacKay, J., Birol, I. & Bohlmann, J. (2015). Improved white spruce (*Picea glauca*) genome assemblies and annotation of large gene families of conifer terpenoid and phenolic defense metabolism. *The Plant Journal*, 83(2), pp. 189–212, doi: 10.1111/tpj.12886

266. Wang, Y., Lim, L., DiGuistini, S., Robertson, G., Bohlmann, J. & Breuil, C. (2013). A specialized ABC efflux transporter GcABC-G1 confers monoterpene resistance to *Grosmannia clavigera*, a bark beetle-associated fungal pathogen of pine trees. *New Phytologist*, 197(3), pp. 886–898, doi: 10.1111/nph.12063

267. Way, M.J., Paiva, M.R. & Cammell, M.E. (1999). Natural biological control of the pine processionary moth *Thaumetopoea pityocampa* (Den. & Schiff.) by the Argentine ant *Linepithema humile* (Mayr) in Portugal. *Agricultural and Forest Entomology*, 1, pp. 27-31. doi: 10.1046/j.1461-9563.1999.00005.x

268. Werno, J. & Lamy, M. (1994). Daily cycles for emission of urticating hairs from the pine processionary caterpillar (*Thaumetopoea pityocampa* S.) and the brown tail moth (*Euproctis chrysorrhoea* L) (Lepidoptera) in laboratory conditions. *Aerobiologia*, 10, pp. 147-151. doi: 10.1007/BF02459229

269. Werno, J. & Lamy, M. (1990). Atmospheric pollution of animal origin: the urticating hairs of the processionary caterpillar (*Thaumetopoea pityocampa* Schiff.) (Insects, Lepidoptera). Comptes rendus de l'Academie des sciences. *Serie III, Sciences de la vie*, 310(8), 325-331.

270. Whittington, D. A., Wise, M. L., Urbansky, M., Coates, R. M., Croteau, R. B. & Christianson, D. W. (2002). Bornyl diphosphate synthase: structure and strategy for carbocation manipulation by a terpenoid cyclase. *Proceedings of the National Academy of Sciences of the United*, 99, pp. 15375–15380. doi: 10.1073/pnas.232591099

271. Wildung, M. R. & 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

272. Wise, M. L., Savage, T. J., Katahira, E., & Croteau, R. (1998). Monoterpene Synthases from Common Sage (*Salvia officinalis*)*: cDNA isolation, characterization, and functional expression of (+)-sabinene synthase, 1, 8-cineole synthase, and (+)-bornyl diphosphate synthase. *Journal of Biological Chemistry*, *273*(24), 14891-14899, doi: 10.1074/jbc.273.24.14891

273. Wright, G. A. & 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*, pp. 11. doi: 10.1111/j.1365-2435.2009.01627.x

274. Xu, M., Galhano, R., Wiemann, P., Bueno, E., Tiernan, M., Wu, W., Chung, I., Gershenzon, J., Tudzynski, B., Sesma, A. & Peters, R. J. (2012). Genetic evidence for natural product-mediated plant–plant allelopathy in rice (*Oryza sativa*). *New Phytologist*, 193(3), pp. 570–575, doi: 10.1111/j.1469-8137.2011.04005.x

275. Xu, M., Ross Wilderman, P., Morrone, D., Xu, J., Roy, A., Margis-Pinheiro, M., Upadhyaya, N. M., Coates, R. M. & Peters, R. J. (2007). Functional characterization of the rice kaurene synthase-like gene family. *Phytochemistry*, 68(3), pp. 312–326, doi: 10.1016/j.phytochem.2006.10.016

276. Yokouchi, Y. & Ambe, Y. (1984) Factors affecting the emission of monoterpenes from red pine (*Pinus densiflora*). *Plant Physiology*, 75, pp. 1009-1012. doi: 10.1104/pp.75.4.1009

277. Zhao, T., Krokene, P., Hu, J., Christiansen, E., Björklund, N., Långström, B., Solheim, H. & Borg-Karlson, A.K. (2011). Induced terpene accumulation in Norway spruce inhibits bark beetle colonization in a dose-dependent manner. *Plos one*, 6. doi: 10.1371/journal.pone.0026649

278. Zera, J.A., Sall, J. & Otto, K. (1999). Biochemical aspects of flight and flight lessness in Gryllus: Flight fuels, enzyme activities and electrophoretic profiles of flight muscles from flight-capable and flightless morphs. *Journal of Insect Physiology*, 45, pp. 275–285. doi: 10.1016/S0022-1910(98)00123-1

279. Zerbe, P. & Bohlmann, J. (2015). Enzymes for synthetic biology of ambroxide-related diterpenoid fragrance compounds. In: Schrader, J. & Bohlmann, J. (eds.) *Biotechnology of Isoprenoids*. Cham: Springer International Publishing, pp. 427–447.

280. Zerbe, P., Hamberger, B., Yuen, M. M. S., Chiang, A., Sandhu, H. K., Madilao, L. L., Nguyen, A., Hamberger, B., Bach, S. S. & Bohlmann, J. (2013). Gene discovery of modular diterpene metabolism in non-model systems. *Plant Physiology*, 162(2), pp. 1073–1091, doi: 10.1104/pp.113.218347

281. Zerbe, P. & Bohlmann, J. (2014). Bioproducts, biofuels, and perfumes: conifer terpene synthases and their potential for metabolic engineering. In *Phytochemicals–biosynthesis, function and application* (pp. 85-107). Springer, Cham. doi: 10.1007/978-3-319-04045-5_5

282. Zerbe, P., Chiang, A., Yuen, M., Hamberger, B., Hamberger, B., Draper, J. A., Britton, R. & Bohlmann, J. (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

283. Zhang, Q.H., & Paiva, M.R., (1998). Female calling behaviour and male response to the sex pheromone in *Thaumetopoea pityocampa* (Den. & Schiff.) (Lep. Thaumetopoeidae). *Journal of Applied Entomology*, 122, pp. 353–360. doi: 10.1111/j.1439-0418.1998.tb01512.x

284. Zhou, K., Gao, Y., Hoy, J. A., Mann, F. M., Honzatko, R. B. & Peters, R. J. (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

285. Zi, J., Matsuba, Y., Hong, Y. J., Jackson, A. J., Tantillo, D. J., Pichersky, E. & Peters, R. J. (2014). Biosynthesis of lycosantalonol, a *cis* -prenyl derived diterpenoid. *Journal of the American Chemical Society*, 136(49), pp. 16951–16953, doi: 10.1021/ja508477e

286. Zulak, K. G. & Bohlmann, J. (2010). Terpenoid biosynthesis and specialized vascular cells of conifer defense. *Journal of Integrative Plant Biology*, 52(1), pp. 86–97, doi: 10.1111/j.1744-7909.2010.00910.x

LIST OF TABLES AND FIGURES

contribute to the activity of class-I and class-II TPSs, are indicated. The N-terminal plastid transit peptide (TP) and "RR(X8)W" motif are also shown.

Fig. 1.4. Proposed model for the structural evolution of diterpene synthases (DTPSs) (Cao *et al.* 2010; Gao *et al.* 2012) and hypothesis on evolution of sesquiterpene synthases (STPSs) and monoterpene synthases (MTPSs) according to two potential routes based on the analysis of multisubstrate enzymes (Pazouki and Niinemets 2016).

Table 1.1. Size of typical plant TPS family and subfamilies in selected plant species.

Table 1.2. Function and taxonomic distribution of plant terpene synthases (TPSs) subfamilies.

Table 1.3. Full-length cDNA sequences retrieved from the NCBI database upon which the phylogenetic analysis of terpene synthases in gymnosperms was carried out (Fig. 1.5). The ent-kaurene synthase from the moss *Physcomitrella patens* was included as an outgroup.

Fig. 1.5. Phylogenetic tree of terpene synthases (TPSs) in gymnosperms: Cupressaceae (red diamonds), Pinaceae (black), Taxaceae (green), Ginkgoaceae (brown) and Cycadaceae (heavenly). The *Physcomitrella patens ent*-kaurene synthase (PtTPS-entKS; violet diamond) was used to root the tree. Branches indicated with dots represent bootstrap support more than 80% (100 repetitions). Modifications in the typical γβα-domain architecture of TPS and the presence of functional active sites (a yellow cross indicate loss of function) are illustrated corresponding to the different subfamilies of the TPS plant family and to the different group within the TPS-d3 subfamily. For acronyms denoting plants species. Abbreviations: DTPSs, diterpene synthases; MTPSs, monoterpene synthases; STPSs, sesquiterpene synthases.

Table 1.4. Full-length cDNA sequences of functionally characterized terpene synthases (TPSs) employed for the BLAST search in the NCBI database of the putative TPSs of *Pinus* spp.

Table 1.5. Full-length cDNA sequences of putative terpene synthases retrieved from the NCBI database upon which the phylogenetic analysis of terpene synthases in *Pinus* spp. was carried out (Fig. 1.7). The *ent*-kaurene synthase from the moss *Physcomitrella patens* was included as an outgroup (continue on the next page).

Fig. 1.6. Phylogenetic tree for the deduced amino acid sequences of the 2-methyl-3-buten-2-ol synthases (MBOSs) from *Pinus* species identified in NCBI database (Table 1.5). *Physcomitrella patens* ent-kaurene synthase (Pt TPS-entKS) was used to root the tree.

Fig. 1.7. Phylogenetic tree of the deduced amino acid sequences of terpene synthases (TPSs) in *Pinus* spp. The *Physcomitrella patens ent*-kaurene synthase (Pt TPS-entKS) was used to root the tree. Branches indicated with dots represent bootstrap support more than 80% (100 repetitions). Colour rectangles denote TPS subfamilies and groups within the TPS-d subfamily. For acronyms denoting plants species, see Table 1.3. Abbreviations: CPS, ent-copalyl diphosphate synthase; DTPS, diterpene

synthase; KS, *ent*-kaurene synthase; MBOS, 2-methyl-3-buten-2-ol synthase; MTPS, monoterpene synthase; STPS, sesquiterpene synthase.

Fig. 1.8. Alignment of deduced amino acid sequences of full-length 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) isolated from *Pinus nigra* subsp. *laricio* needles. The black-shaded residues are highly conserved ones; the gray-shaded residues are identical in at least three of the six sequences shown. The brace indicates the putative N-terminal transit peptide region. The "RRX8W" and the "DDxxD" motifs are indicated with red open rectangles. Abbreviations: MBOS, 2-methyl-3-buten-2-ol synthase; MTPS, monoterpene synthase.

Fig. 1.9. Strategy adopted for the genomic amplification of a putative *Pinus nigra* subsp. *laricio* gene belonging to the phylogenetic TPS-d1 Group 3. A) Schematic representation of the full-length cDNA of a representative member of the phylogenetic TPS-d1 Group 3 (PcMTPS4 from *Pinus contorta*, in the present case, see Fig. 1.7 and Table 1.5) in which the positions of the forward (F2) and the reverse (R3) primers used in the amplification of genomic DNA are shown. B) Intron (yellow)/exon (blue) structure of the amplified *Pinus nigra* subsp. *laricio* genomic sequence. The positions of the primers used to amplify the genomic fragment are also shown.

Fig. 1.10. Phylogenetic tree of the deduced amino acid sequences obtained by combining monoterpene synthases (MTPSs) and the five selected 2-methyl-3-buten-2-ol synthases (MBOSs) identified in different *Pinus* species (Fig. 1.7 and Table 1.5) and the six sequences isolated from *Pinus nigra* subsp. *laricio* (outlined in red). *Physcomitrella patens ent*-kaurene synthase (Pt TPS-entKS) was used to root the tree. Branches indicated with dots represent bootstrap support more than 80% (100 repetitions). The seven phylogenetic groups identified in the pine members of TPS-d1 clade are highlighted with square brackets. For acronyms denoting plants species, see Table 1.5.

Figure 2.1. Chemical structures of the most represented diterpenoids in Pinus spp. [R = CH3 olefins constituents; R = CH2OH alcoholic constituents; R = CHO aldehydic constituents; R = COOH diterpene resin acid (DRA) constituents].

Table 2.1. List of primer pairs of Calabrian pine DTPS and reference (CYP and upLOC) genes used in qRT-PCR analyses.

Fig. 2.2. A representative example of the quantitative relationships among acidic (diterpene resin acids, DRAs) and neutral (olefins) components of the diterpenes extracted from *Pinus nigra* subsp. *laricio* (Calabrian pine) tissues, visualized by overlapping GC-MS ion chromatograms at selected m/z, i.e., 374/359 for DRA and 272/257 for olefins (magnified inset on the bottom left side of the item).

Fig. 2.3. A representative GC-MS profile of the diterpene resin acids extracted from the leader stem of Calabrian pine. The single ion monitoring at m/z 374 (red line, non-dehydrogenated species) was

overlapped with the single ion monitoring at m/z 372 (blue line, dehydrogenated species). (1), pimaric acid; (2), sandaracopimaric acid; (3), isopimaric acid; (4), palustric acid; (5), levopimaric acid; (6), dehydroabietic acid; (7), abietic acid; (8), neoabietic acid; (9), non-identified dehydroisomer.

Fig. 2.4. Mass spectra of the nine diterpene resin acids identified in *Pinus nigra* subsp. *laricio*. Pimaric acid (A), Sandaracopimaric acid (B), Isopimaric acid (C), Palustric acid (D), Levopimaric acid (E), Dehydroabietic acid (F), Abietic acid (G), Neoabietic acid (H) and Dehydroisomer acid (I).

Fig. 2.5. A representative GC-MS profile of the diterpene olefins and aldehydes in the extract obtained from leader stem tissue of Calabrian pine. The selected m/z 272 and 257 for diterpene olefins (red number above peaks) were overlapped with the selected m/z 286 for aldehydes (black number above peaks). (1) sandaracopimaradiene; (2) levopimaradiene; (3) palustradiene; (4) abietadiene; (5) neoabietadiene; (6) sandaracopimaradienal; (7) palustradienal; (8) isopimara-dienal; (9) abietadienal; (10) neoabietadienal.

Fig. 2.6. Total diterpene resin acids (DRAs, panel at the top) and levels of individual DRAs in different tissues of 3-year-old *Pinus nigra* subsp. *laricio* (Calabrian pine) saplings. Error bars indicate the standard deviation of the mean. The statistical significance of the differences was evaluated by one-way ANOVA, followed by the Tukey's test. Different letters denote statistical significance of the difference at $p < 0.01$. YN, young needle; MN, mature needle; LS, leader stem (LS); IS, interwhorl stem; R, root; DW, dry weight.

Table 2.2. Full length cDNA sequences identified in the National Center for Biotechnology Information (NCBI) database coding for putative diterpene synthases (DTPS) in the *Pinus* species. ORF, open reading frame; bp, base pair.

Fig. 2.7. Schematic representation of the full length cDNAs of four representative diterpene synthase (DTPS) members of the phylogenetic d3 clade of genes, two from *Pinus banksiana* (Pb) and two from *Pinus contorta* (Pc) and the positions of their Forward and Reverse primers used in the present study for the isolation of the partial transcripts coding for the orthologous genes (F1c/R1c) in Calabrian pine. The position of the specific primers used for 5' and 3' RACE (Rapid Amplification of cDNA Ends) extensions of partial transcripts of Calabrian pine DTPS (groups 1-4) are indicated in blue. UTR, untranslated region.

Table 2.3. Forward and Reverse primers used for the isolation of cDNAs and genomic diterpene synthase sequences in *Pinus nigra* subsp. *laricio*. RACE, Rapid Amplification of cDNA Ends.

Fig. 2.8. Alignment of deduced amino acid sequences of the four putative diterpene synthases from Calabrian pine (Pnl DTPS 1-4) isolated in the present study. Amino acid residues with black background indicate highly conserved regions, while amino acid residues which are identical in more

than 50% of the proteins are in grey background. The DTPS class II (DxDD) and class I (DDxxD, NSE/DTE) signature motifs are indicated.

Fig. 2.9. Phylogenetic tree of the deduced amino acid sequences of 13 diterpene synthases (DTPSs) identified in different *Pinus* species and the four DTPSs from Calabrian pine isolated in the present study (red squares). The *ent*-kaurene synthase from *Physcomitrella patens* (Pt TPS-entKS, BAF61135) was used to root the tree. Branches marked with dots represent bootstrap support more than 80% (1000 repetitions). The four phylogenetic groups identified in the pine members of the d3 clade of terpene synthases are indicated by square brackets.

Fig. 2.10. Alignment of deduced amino acid sequences of diterpene synthase (DTPS) belonging to the phylogenetic group 1. Amino acid residues with black background indicate highly conserved regions, while amino acid residues which are identical in more than 50% of the proteins are in grey background. The DTPS class II (DxDD) and class I (DDxxD, NSE/DTE) signature motifs are indicated. Pb, *Pinus banksiana*; Pc, *Pinus contorta*; Pt, *Pinus taeda*; Pnl, *Pinus nigra* subsp. *laricio*.

Table 2.4. Amino acid sequence identity matrix of DTPS candidate genes from *P. laricio* (in red) with previously characterized DTPSs from P. *taeda* (Pt), *P. contorta* (Pc) and *P. banksiana* (Pb).

Fig. 2.11. Alignment of deduced amino acid sequences of diterpene synthase (DTPS) belonging to the phylogenetic group 2. Amino acid residues with black background indicate highly conserved regions, while amino acid residues which are identical in more than 50% of the proteins are in grey background. The DTPS class II (DxDD) and class I (DDxxD, NSE/DTE) signature motifs are indicated.

Fig. 2.12. Alignment of deduced amino acid sequences of diterpene synthase (DTPS) belonging to the phylogenetic group 3. Amino acid residues with black background indicate highly conserved regions, while amino acid residues which are identical in more than 50% of the proteins are in grey background. The DTPS class II (DxDD) and class I (DDxxD, NSE/DTE) signature motifs are indicated.

Fig. 2.13. Alignment of deduced amino acid sequences of diterpene synthase (DTPS) belonging to the phylogenetic group 4. Amino acid residues with black background indicate highly conserved regions, while amino acid residues which are identical in more than 50% of the proteins are in grey background. The DTPS class II (DxDD) and class I (DDxxD, NSE/DTE) signature motifs are indicated.

Fig. 2.14. Schematic representation of the exon/intron structures of the four diterpene synthase (DTPS) genes isolated from Calabrian pine (Pnl) in the present study. For both exons (blue boxes) and introns (black lines) the lengths in bp are indicated. Introns were numbered (Roman numerals) starting from the 5' end of each genomic sequences.

Fig. 2.15. Genomic organization of plant DTPS genes. Black vertical bars represent introns (indicated by Roman numerals) and are separated by colored boxes with specified lengths in aa (amino acids) representing exons. The number above the intron Roman numeral represents the intron phase number

and demonstrates conservation throughout the plant DTPS genes. Introns are classified into three phase types according to Li (1997). Schematic, intron numbers, and exon coloring scheme are based upon Trapp and Croteau (2001). Genomic DNA sequences compared are as follows: AgAS, *A. grandis* abietadiene synthase (NCBI accession no. AF326516); GbLS, *G. biloba* levopimaradiene synthase (AY574248); TbTXS, *Taxus brevifolia* taxadiene synthase (AF326519); PgKS, *P. glauca ent*-kaurene synthase (GU059905); AtCPS, *A. thaliana* copalyl diphosphate synthase (AT4G02780); Pnl DTPS 1-4 isolated in this study.

Fig. 2.16. Relative expression levels of four diterpene synthase genes (DTPS1-4) in five different tissues of Calabrian pine. The expression data of each gene were normalized using the geometric average of the two reference genes CYP and upLOC. Relative expression levels of the different DTPS genes were referred to a calibrator, set to the value 1, which was represented by the gene in the five tissues with the lowest expression (DTPS3 in YN). YN, young needles; MN, mature needles; LS, bark and xylem combined from the leader stem; IS, bark and xylem combined from the interwhorl stems; R, roots. Different letters denote significant differences according to the Tukey's test ($p < 0.01$).

Table 3.1. Forward and Reverse primers used for the isolation of cDNAs and genomic sequences coding for MBOS (Group 1) and MTPSs (Groups 2-7) in *Pinus nigra* subsp. *laricio.* RACE, Rapid Amplification of cDNA Ends.

Table 3.2. List of primer pairs of Calabrian pine *MBOS*, *MTPS* and reference (*CYP* and *upLOC*) genes used in qRT-PCR analyses

Fig. 3.1. A representative GC-MS profile of the monoterpenes extracted from the young needles of Calabrian pine. Single ion monitoring at m/z 136, 121 and 93. (1) tricyclene, (2) α-thujene, (3) α-pinene, (4) α-fenchene, (5) camphene, (6) sabinene, (7) β-pinene, (8) myrcene, (9) δ-3-carene, (10) β-phellandrene, (11) limonene, (12) terpinolene, (13) α-terpineol and (14) bornyl acetate.

Fig. 3.2. Chemical structures of the monoterpenes found in different tissues of Calabrian pine.

Fig. 3.3. Total monoterpene contents (top left panel A) and levels of individual monoterpenes (panels B-F) in different tissues of three-years old Calabrian pine saplings. Error bars indicate the standard deviation of the mean. The statistical significance of the differences was evaluated by one-way ANOVA, followed by the Tukey's test. Different letters denote statistical significance of the difference at $p < 0.01$.

Table 3.3. Full-length cDNA sequences of putative MBOSs and MTPSs in *Pinus* spp. retrieved from the NCBI database

Fig. 3.4. Alignment of deduced amino acid sequences of the seven putative hemi- and mono-terpene synthases from Calabrian pine (Pnl MBOS1/Pnl MTPS 2-7) isolated in the present study. Amino acid residues with black background indicate highly conserved regions, while amino acid residues which are

identical in more than 50% of the proteins are in grey background. The horizontal square bracket indicates the putative N-terminal transit peptide region. The "RRX8W" and class-I "DDxxD" signature motifs are indicated with red open rectangles.

Fig. 3.5. Phylogenetic tree of the deduced amino acid sequences of *MTPS* and *MBOS* genes identified in different *Pinus* species (Table S1) and those from the seven Calabrian pine *MBOS* and *MTPS* genes isolated in the present study (yellow triangles). The ent-kaurene synthase from Physcomitrella patens (Pt TPS-entKS, BAF61135) was used to root the tree. Branches marked with dots represent bootstrap support more than 80% (1000 repetitions). The seven phylogenetic groups identified in the pine members of the TPS- d1- clade are indicated by square brackets.

Fig. 3.6. Alignment of deduced amino acid sequences of MTPSs belonging to the phylogenetic group 3. Amino acid residues with blue background indicate highly conserved regions, while amino acid residues which are identical in more than 50% of the proteins are in pink background. Pb, *Pinus banksiana*; Pc, *Pinus contorta*; Pnl, *Pinus nigra* subsp. *laricio* (Calabrian pine).

Fig. 3.7. Alignment of deduced amino acid sequences of MTPSs belonging to the phylogenetic group 5. Amino acid residues with blue background indicate highly conserved regions, while amino acid residues which are identical in more than 50% of the proteins are in pink background. Pb, *Pinus banksiana*; Pc, *Pinus contorta*; Pt, *Pinus taeda*; Pnl, *Pinus nigra* subsp. *laricio* (Calabrian pine).

Table 3.4. Comparison of the introns for the hemi-terpene synthase (*MBOS*) and of the monoterpene synthases (*MTPS*) genes isolated in the present study: position (the letters are referred to the last coded amino acid of the exon), size (in parentheses), total number and phase. No asterisk indicates no interruption between codons; single asterisk indicates intron inserted between the first and the second nucleotide; double asterisk indicates intron inserted between the second and third nucleotide; hyphen indicates intron not present. The introns are numbered according to Trapp and Croteau (2001).

Fig. 3.8. Genomic organization of class I, class II and III terpene synthase genes *sensu* Trapp and Croteau (2001). Black vertical slashes represent introns (indicated by Roman numerals) and are separated among each other by colored boxes with indicated lengths in amino acids, representing exons. The numbers above the introns of the first row from the top represent the intron phase type classification according to Li (1997) and indicate conservation throughout the plants' TPS genes. Schematization, intron numbers, and exon coloring scheme are based upon Trapp and Croteau (2001). Genomic DNA sequences compared are as follows: AgAS, *Abies grandis* abietadiene synthase (NCBI accession no. AF326516), AgLS, *Abies grandis* limonene synthase (AF326518); AgLS, *Abies grandis* pinene synthase (AF326517); AtLS, *Arabidopsis thaliana* putative limonene synthase (Z97341); Pnl MBOS1/Pnl MTPS2-7 denote the *MBOS* and *MTPS* genes isolated from Calabrian pine in the present study.

Fig. 3.9. Heat maps of the relative expression levels of the hemi-terpene synthase (*MBOS*) and of the monoterpene synthases (*MTPS*) genes isolated from five different tissues of Calabrian pine. The expression data of each gene were normalized using the geometric average of the two reference genes *CYP* and *upLOC*. Relative expression levels were calculated by setting a value of 1 for the lowest value among the seven genes in the five tissues considered (*MTPS3* in YN). For each gene, the differences in the relative expression levels were shown in colour according to the scale and statistical evaluation of the differences among the seven genes in the five tissues is reported in Table 3.5. IS: bark and xylem combined from the interwhorl stems; LS: bark and xylem combined from the leader stem; MN: mature needles; R: roots; YN: young needles.

Table. 3.5 Statistical evaluation of the differences (one-way ANOVA followed by Tukey test) among the relative expression levels of *MBOS* and *MTPS* genes in five different tissues of Calabrian pine. The expression data of each gene were normalized using the geometric average of the two reference genes *CYP* and *upLOC*. Relative expression levels of the different *MBOS* and *MTPS* genes were referred to a calibrator, set to the value 1, which was represented by the gene in the five tissues with the lowest expression (*MTPS3* in YN). Different letters denote significant differences according to the Tukey's test ($p < 0.01$). YN: young needles; MN: mature needles; LS: bark and xylem combined from the leader stem; IS: bark and xylem combined from the interwhorl stems; R: roots.

Table 4.1. Volatile organic compounds identified by GC/MS analysis in the head space of Calabrian pine needles. RT, retention time; KI, retention index.

Table 4.4. Volatile organic compounds from the needles of not infested Calabrian pines detected in at least two replicates out of three at each sampling time in the two sampling sites of Canolo and Bova.

Table 4.3. Volatile organic compounds from the needles of pine processionary moth-infested Calabrian pines detected in at least two replicates out of three at each sampling time in the two sampling sites of Canolo and Bova

Fig. 4.1a. A typical headspace-GC/MS chromatogram obtained from Calabrian pine needles collected at Bova. Upper panel, plant infested by the pine processionary moth; lower panel, not infested (control) plant.

Fig. 4.1b. A typical headspace-GC/MS chromatogram obtained from Calabrian pine needles collected at Canolo Nuova. Upper panel, plant infested by the pine processionary moth; lower panel, not infested (control) plant.

Fig. 4.2. Principal Component Analysis applied to volatiles emission data obtained from pine processionary moth-infested- (red circles) or not infested (green circles) pine needles collected in Bova (left) or in Canolo (right) during six samplings along the insect biological cycle. (Continues).

Fig. 4.2. Continued from the previous page.

Fig. 5.5. Trophic procession of PPM larvae on infested Calabrian pine during the winter season (third larval stage).

Fig. 5.6. Infested Calabrian pine tree during the third sampling, completely defoliated after the trophic activity of PPM *larvae*.

Fig. 5.7. Relative expression levels of four diterpene synthase genes (*DTPS1–4*) in both control and infested Calabrian pine needles in the three-sampling time. The expression data of each gene were normalized using the geometric average of the two reference genes *CYP* and *upLOC*. Relative expression levels for each of the different *DTPS* genes were referred to a calibrator, set to the value 1, which was represented by the gene in the five tissues with the lowest expression.

Table 5.1. Statistical evaluation of the differences (one-way ANOVA followed by Tukey test) among the relative expression levels of *DTPS* genes in both control and infested Calabrian pine needles in the three-sampling time. Different letters denote significant differences according to the Tukey's test ($p < 0.01$).

Fig. 5.8. Relative expression levels of seven monoterpene synthase genes (*MBOS1/MTPS2–7*) in both control and infested Calabrian pine needles. The expression data of each gene were normalized using the geometric average of the two reference genes *CYP* and *upLOC*. Relative expression levels for each of the different *MTPS* genes were referred to a calibrator, set to the value 1, which was represented by the gene in the five tissues with the lowest expression.

Table 5.2. Statistical evaluation of the differences (one-way ANOVA followed by Tukey test) among the relative expression levels of *MBOS* and *MTPS* genes in both control and infested Calabrian pine needles in the three-sampling time. Different letters denote significant differences according to the Tukey's test ($p < 0.01$).