On the Evolution and Functional Diversity of Terpene Synthases in the *Pinus* Species: A Review

Enrica Alicandri[1] (e.alicandri@gmail.com), Anna Rita Paolacci[2] (arpaolacci@unitus.it), Samson Osadolor[2] (soooct24@gmail.com), Agostino Sorgonà[1] (asorgona@unirc.it), Maurizio Badiani[1] (mbadiani@unirc.it), Mario Ciaffi[2] (ciaffi@unitus.it)

1Dipartimento di Agraria, Università Mediterranea di Reggio Calabria, Loc. Feo di Vito, I-89129 Reggio Calabria, Italy

2Dipartimento per la Innovazione nei Sistemi Biologici, Agroalimentari e Forestali, Università della Tuscia, Via S. Camillo De Lellis, s.n.c, I-01100 Viterbo, Italy

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 medium 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 diterpenoid specialized metabolites, which are found in volatile emissions and oleoresin secretions. Such substances are involved in the defence against pathogens and herbivores and can help to protect against abiotic stress. Oleoresin terpenoids can be also profitably used in a number of different fields, from traditional and modern medicine to fine chemicals, fragrances, and flavours, and, in the last years, in biorefinery 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, can be seen as 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 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 defence, Gymnosperms, Gene family evolution

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 (C_5) isoprenoid units, isopentenyl diphosphate (IPP) and its isomer and dimethylallyl diphosphate (DMAPP), can be categorized as hemiterpenoids (C_5), monoterpenoids (C_{10}), sesquiterpenoids (C_{15}), diterpenoids (C_{20}), triterpenoid (C_{30}), tetraterpenoid (C_{40}) or polyterpenoids (C_{5n}), based on the number of C_5 units they contain (Tholl and Lee 2011). While a number of plant 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 medium-size to large family with approximately 30–100 functional members in the genomes of nearly all the plant species sequenced so far (Chen et al. 2011; Warren et al. 2015; Kumar et al. 2018; Karunanithi and Zerbe 2019). 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, conifers 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 diterpenoid specialized metabolites, which are found in volatile emissions and oleoresin secretions. These specialized metabolites are involved in the defence against pathogens and herbivores and can help to protect against abiotic stress (Zulak and Bohlman 2010; Hall et al. 2011; Tholl et al. 2015; Celedon and Bohlmann 2019). Oleoresin terpenoids are also important for the production of flavours and fragrances, therapeutics, solvents, coatings and resins, and more recently have been taken into consideration as potential precursors of biofuels (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 MTPSs gene sequences for the first time 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.

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, semi-volatile or non- volatile at ambient temperature, respectively, are involved in plant defence 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 against biotic stress as well, being part of the constitutive and/or inducible defence 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 many plant species (Mumm et al. 2008). In cotton, for instance, gossypol and its related sesquiterpene aldehydes play a role in the inducible and constitutive defence 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 oleo- resin or emitted as volatile compounds, play an important role in the physical and chemical defence responses against pathogens and herbivores (Zulak and Bohlman 2010; Cele- don and Bohlmann 2019). Oleoresin, whose main components are mono- and diterpenes (including diterpene resin acids, DRAs), 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; Celedon and Bohlmann 2019). The importance of terpenoids in the defence 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; Pierik et al. 2014). Such indirect defence strategy is used by plants to protect their photosynthetic tissues from pathogens and herbivores, as well as to limit insect oviposition. For instance, eggs deposition by the elm leaf beetle (*Xan- thogaleruca luteola*) on the leaves of *Ulmus minor* causes 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 terpe- noids can act as interspecific, intraspecific, and intraplant signals to promote defence 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 the plant-to-plant communication mediated by volatile compounds. Not every organisms 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 from flowers and fruits can be involved in mutualistic interactions with plant pollinators and seed dispersal agents (Abbas et al. 2017). For instance, many studies demonstrated the role of volatile terpenoids as constituents of floral scent in communication between plants and pollinators (Baldwin et al. 2006; 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). Monoterpenes and sesquiterpenes, as the major components of floral volatiles, are particularly suited as long-distance chemical messengers, because of their low-molecular-weight, high vapour pressure at ordinary temperatures, and lipophilic nature, which facilitate their interactions with membrane systems (Tholl 2015; Abbas et al. 2017).

Although terpenes have been mostly studied in the above- ground tissues, similar functions in direct and indirect defence responses have been also identified in the below-ground environment. For instance, *Arabidopsis* roots produce semi-volatile diterpene hydrocarbons, known as rhizathalenes, able to limit root damage by acting as local antifeedant towards herbivores (Vaughan et al. 2013). Similarly, the triterpene saponins known as avenacins are powerful phytoalexins exuded by oat roots (Thimmappa et al. 2014). Indirect defence brought about by volatile terpenes is also seen below ground; for instance, maize roots attacked by the western corn rootworm (*Dia- brotica 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 func- tions, are also able to promote the beneficial root infection brought about by mycorrhiza (Akiyama 2005).

Biosynthesis of Terpenoids in Plants

Isopentenyl diphosphate (IPP) and dimethylallyl diphosphate (DMAPP) are the C_5 momomeric precursors of all terpenes. They derive from two distinct metabolic routes, namely the mevalonate (MVA) pathway, and the methyl- erythritol 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 C_5 building blocks, yielding intermediates such as geranyl diphosphate (GPP, C_{10}), farnesyl diphosphate (FPP, C_{15}), geranylgeranyl diphosphate (GGPP, C_{20}), and squalene (C_{30}), respectively (Tholl 2015; Pazouki and Niinemets 2016). These intermediates are finally converted into terpenes by the action of TPSs, which are named according to the length of their respective reaction products: hemiterpene- or isoprene- (C_5), monoterpene- (C_{10}), sesquiterpene- (C_{15}), and diterpene- (C_{20}) synthases (Chen et al. 2011; Tholl 2015; Pazouki and Niinemets 2016).

The First Step: The Terpenoid Basic Units, Namely IPP and DMAPP, are Synthetized in Two Compartmentally Separated Metabolic Pathways

Both the IPP and DMAPP C_5 precursors, which can be inter- converted into each other, can derive from both the MVA and the MEP pathways; the former pathway is localized into the cytosol, although it also operates in the endoplasmic reticulum and peroxisomes, whereas the latter operates inside 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, 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-di- phosphate. 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, 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 reductoisomerase (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 stoichi- ometry of about 5:1 (Tholl 2015; Abbas et al. 2017).

The biosynthesis of terpenoids containing more than five carbon atoms (Fig. 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, left), whereas in the MEP pathway, whose final products are both IPP and DMAPP (Fig. 1, right), IDI is thought to ensure an optimal IPP/DMAPP ratio for the assemblage of the C_5 units leading to terpenoid pre- cursors and/or to fuel export from plastids to cytosol. It is worth nothing (Fig. 2, right) that DMAPP can be also used as a substrate for hemiterpene (C_5) biosynthesis, by the activity of isoprene synthase (IPS in Fig. 2). A single and two distinct IDI genes were identified in the genomes of *Catharanthus roseus* and *Arabidopsis thaliana*, respectively, and found to be transcribed as splice variants. In both species, the 'long' protein variant is transported into chloroplasts and/or mitochondria, while the 'short' protein variant, missing a targeting signal, is localized into peroxisomes (Guirimand et al. 2012), underlining the involve- ment of different subcellular compartments in isoprenoid biosynthesis in plants.

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 path- ways, IPP and DMAPP are used as building blocks for the assembling of the prenyldiphosphates (see below), from which all the terpenes derive. A large group of prenyltransferases (Fig. 2, see below), whose general name is also isoprenyl diphosphate synthases, are the enzymes in charge for producing prenyldiphosphates, whose bio- synthesis 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 C_{10} prenyl diphosphate, to which one or more further IPP units can be added, again by head-to-tail condensation reactions, to produce short- chain (C_{15}–C_{25}), medium-chain (C_{30}–C_{35}), and long-chain (C_{40}–C_n) prenyl diphosphates.

Since the double bonds of the prenyl diphosphate to be formed can be either in *cis*- or in *trans* configuration, distinct families of *cis*- or *trans*-prenyltransferases, respectively, have to come into play (Kharel 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 C_{70}–C_{120} dehydro- dolichol diphosphates (Takahashi and Koyama 2006; Surmacz and Swiezewska 2011) (Fig. 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. 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. 2).

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, allocated to either the primary or the secondary metabolism (Fig. 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, respectively, are acted upon by specific TPSs, to yield specialized metabolites such as monoterpenes, sesquiterpenes, and diterpenes, respectively (Fig. 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. 2). A similar condensation reaction of two molecules of GGPP produces phytoene, fueling the downstream synthesis of carotenoids (Fig. 2). Geranylgeranyl diphosphate is also involved in the biosynthesis of *ent*-kaurene, from which all the plant gibberellins derive. Such conversion is carried out in sequence by two structurally related TPSs, first *ent*-copalyl diphosphate (*ent*-CPP) synthase (CPS), which transforms 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. 2).

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. 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 sup- ported 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 $C_{10–20}$ 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 C_{10}–C_{40} 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-eucalyptol and 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 ^{13}C 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 interchange 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 path- ways (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 activ- ity 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 channeled 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) and the IP kinases (IPKs), which catalyze the hydrolysis and phosphorylation of IPP and IP, respectively (Fig. 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 cytosol and seems to control the pro- duction of both MVA- and MEP-derived terpenoids (Henry et al. 2015). Indeed, *Arabidopsis* T-DNA insertion knockout lines 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. More attempts to understand the formation of IP/DMAP in plants recognized the function of Nudix hydrolases (Nudxs), a large family of two-domain hydrolases/peptidases widely detected in bacteria, plants and animals (Henry et al. 2018). Studies on the two cytosolic Nudxs found in the genome of *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).

Plant Terpene Synthases

As described above, TPSs transform acyclic *cis*- or *trans*- prenyl diphosphate intermediates bearing 5 to 20 carbon atoms into hemiterpenes (C_5), such as isoprene, monoterpenoids (C_{10}), sesquiterpenoids (C_{15}), or diterpenoids (C_{20}) (Fig. 2). These terpenoid compounds can be further modified to produce biologically active final products of greater structural diversity by means of secondary enzy- matic 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 present in 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 con- sequence of minor structural changes (Tholl 2015; Pazouki and Niinemets 2016; Karunanithi and Zerbe 2019). 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).

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. 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; Karunanithi and Zerbe 2019).

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 oper- ated 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 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. 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. 3). They include the diterpene synthases (CPS/KS) found in the mosses *Physcomitrella patens* and *Jungermannia subulata* and in the lycophyte *Selaginella moellendorffii*, which catalyse 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 possess 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 which lost the activity of the α- or the β-domain, respectively (Fig. 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. 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. 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. 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. 3).

Most MTPSs and DTPSs, unlike the cytosolic STPSs, have obvious N-terminal plastid transit peptides (Fig. 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. 3).

A new class of plant TPSs designated as microbial TPS-like proteins (MTPSLs) 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 much more strongly related, from the evolutionary point of view, to microbial TPSs rather than to the typical plant ones (Jia et al. 2018). According to their different origin (see below), MTPSLs and typical plant TPSs differ from each others 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 3 classes depending on the presence of 12–14 introns, 9 introns, or 6 introns (Trapp and Croteau 2001). In the MTPSLs genes of *S. moellendorffii*, by contrast, the number of introns can vary from none, or just one, in the majority of 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. 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).

Origin and Evolution of Plant Terpene Synthases

Figure 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 TPSs in plants.

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. Later on, Hayashi et al. (2010) and Keeling et al. (2010) proposed the class-I/II DTPS from *P. patens* to be the common ancestor. Notably, two diterpene synthases has been found in the bacterium *Bradyrhizobium japonicum*, namely *ent*-copalyldiphosphate 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, 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. 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. 4, central part). Following 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 (Hillwiget al. 2011; Köksal et al. 2011a, 2011b; Rajabiet al. 2013) (Fig. 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 pep- tide in TPSs (Pazouki and Niinemets 2016) (Fig. 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 (C_{20}) or (*E*,*E*)-FDP for the production of (*E*)-nerolidol (C_{15}) (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 to use C_{20} substrates. This hypothesis is supported by the finding of multisubstrate (*E*)-nerolidol (C_{15})/(*E*,*E*)-geranyllinalool (C_{20}) synthases in *Vitis vinifera*, namely VvPNLNGl1 and VvCSENerGl, which are able to use both C_{15} and C_{20} 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. 4, bottom part). Indeed, mixed-substrate TPS have been found in a number of species, among which: *A. majus*, which has two (*E*)- nerolidol (C_{15})/linalool (C_{10}) synthases, i.e. AmNES/LIS-1 and AmNES/LIS-2 (Nagegowda et al. 2008); *A. thaliana*, which has two (*E*,*E*)-α-farnesene (C_{15})/(*E*)-β-ocimene (C_{10}) synthases, i.e. AtTPS02 and AtTPS03 (Huang et al. 2010); and *Fragaria ananassa*, which has two (*E*)-nerolidol (C_{15})/ linalool (C_{10}) 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. 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. 4).

In gymnosperms, on the other hand, there is evidence for a different evolutionary scenario for the formation of STPSs. Here, a tri-domain (γβα) cytosol-active TPS might have arisen after the loss of the transit peptide from a class-I DTPS, followed by the loss of the γ-domain and leading to the formation of a bi-domain (βα) cytosol-localized TPS. Evidence for such hypothesis came from the analysis of three *A. grandis* C_{10}/C_{15} 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. This suggests that during the evolution of gymnosperm STPSs, the transit pep- tide could have been lost first, followed by the loss of the γ-domain (Fig. 4, bottom part).

New insights concerning the evolution of plant TPSs came from the identification in cryptogamae of the

afore- mentioned 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-domain (βα) typical plant TPSs from the three-domain (γβα) proteins (Fig. 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 relatedness to bacterial 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 harbour 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 monoterpe- nes 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.

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 sizes range from 20 to 170 members, counting both full-length genes and pseudogenes, the only notable exception being the moss *P. patens*, whose genome contain a single functional TPS gene (Table 1). In particular, full- length genes coding for typical plant TPSs range from 14 in *S. moellendorffii* to 113, 106 and 69 in *Eucalyptus gran- dis*, *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). 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). 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), i.e. comparable to the number of potentially active TPS genes found in the sequenced genomes of angiosperms (Table 1). The above data suggest that the evolution of typical plant TPS families in higher plants involved several gene dupli- cation events, followed by sub- and neo-functionalizations. 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 intercalated. 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 subfamily (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 an highly compact gene cluster encompassing a stretch of 690 kb on chromosome 18 (Mar- tin 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 sub- family in sequenced plant genomes is reported in Table 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, 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, 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 both 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 2).

Analysis of several flowering plants whose genomes have 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 the TPS family in each species (Chen et al. 2011) (Table 1). Phylogenetic analyses suggest that the growth of the TPS-a gene subfamily took place after the separation of the mono- cot 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 2). Moreover, the positions of *Arabidopsis*, *P. trichocarpa*, *Eucalyptus* spp. and *V. vinifera* genes on the branches of the TPS-a1 clade suggest that many of them originated from 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 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). 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 "$RR(X)_8$ W" 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 Origin and evolution of plant terpene synthases section 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* (see "Structure of Plant Terpene Synthases" and "Origin and Evolution of Plant Terpene Synthases" sections), class-II CPSs involved in primary metabolism (gibberellin biosynthesis) in both angio- sperms 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 monofunctional CPSs (Chen et al. 2011). This subfamily also contains class- II DTPSs

involved in the specialized metabolism of angio- sperms (Table 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 defence (Peters 2006). The TPS-e/f subfamily contains mainly class-I KSs from gymnosperms and angiosperms involved in the primary metabolism (gibberellin biosynthesis) (Table 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 indicate 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). 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 11, 10 and 9 TPS-e/f genes, respectively, whereas the other eight species have 1–5 members (Table 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 defence 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 2).

Additional lines of evidence support evolutionary diversification within the TPS-e/f subfamily. For instance, func- tionally 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 *S. 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 *Solanum 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).

Selaginella 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, 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 ("Origin and Evolution of Plant Terpene Synthases" section and Fig. 4), gymnosperm class-I/II DTPSs are likely to have evolved from a CPS/KS ancestor probably prior to the

620 divergence from angiosperms, since neither lineage appears to contain a CPS/KS similar to that of *P. patens*
621 (Keeling et al. 2010). By the same manner, the putative class-I/II TPSs in the newly established TPS-h
622 subfamily is likely to have evolved from a prototypic TPS similar to the CPS/KS of *P. patens*, and may be
623 implicated in the specialized terpenoid metabolism of *S. moellendorffii.*
624
625 **Terpene Synthase Genes in Gymnosperms**

626 As mentioned before, conifer trees produce complex mixtures of terpenoids, prevalently in the form of
627 oleoresin, which can play multiple roles in the physical and chemical defence against insects and pathogens
628 (Keeling and Bohlmann 2006; Zulak and Bohlmann 2010; Celedon and Bohlmann 2019). The diversity of
629 conifer terpenoids con- sists mainly of monoterpenes, sesquiterpenes and diterpenes, which, besides the primary
630 catalysis brought about by a large family of TPSs, arise from the action of other enzymes, among which cytochrome
631 P450 monoxygenases and transferases. This, by introducing additional functionalities to the TPS products, could
632 further increase the structural complexity of conifer terpenoids (Ro et al. 2005; Keeling and Bohlmann 2006;
633 Nelson 2011).
634 In the preceding sections, it has been highlighted that, while the evolutionary diversification of angiosperms
635 TPSs is thought to have arisen from DTPSs involved in gibberellin biosynthesis, the diversity amongst
636 gymnosperm TPSs appears to have evolved from class-I/II DTPS ancestors sharing structural and functional
637 similarities with the CPS/KS from *P. patens* (see "Origin and Evolution of Plant Terpene Synthases" section).
638 According to the different scenario in the evolution of TPSs among angiosperms and gymno- sperms, the
639 many MTPSs, STPSs and DTPSs of special- ized metabolism in conifers form the gymnosperm-specific TPS-
640 d subfamily (Chen et al. 2011; Warren et al. 2015; Sha- lev et al. 2018; Celedon and Bohlmann 2019). The
641 functional variety of gymnosperm TPSs seems to have evolved via repeated gene duplication events and further
642 sub- and neo-functionalization, contributing to the expansion of the TPS-d multigene family (Chen et al. 2011;
643 Warren et al. 2015; Shalev et al. 2018; Celedon and Bohlmann 2019), which represents the key player behind
644 the chemical complexity of specialized terpenes in conifers. In contrast, gymnosperm DTPSs involved in primary
645 metabolism are members of the TPS-c and TPS-e/f subfamilies, which include also angiosperm orthologous
646 genes, suggesting that DTPSs involved in gibberellin biosynthesis are conserved across the two phyla (Keeling et
647 al. 2010; Chen et al. 2011).
648 Until about a decade ago, our knowledge concerning the number, structural and functional complexity and
649 phylogeny of gymnosperm TPSs was based on targeted cDNA cloning and characterization carried out in only two
650 conifer species, namely *A. grandis* and *P. abies*, together with a few TPSs in other gymnosperms (Keeling and
651 Bohlmann 2006; Chen et al. 2011). In *A. grandis*, 11 distinct TPS genes were functionally characterized
652 (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 TPSs of Cupressaceae received far less consideration, in the present review 20 putative full-length TPS sequences from several Cupressaceae species, such as *T. 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 S1) to construct a maximum likelihood phylogeny (Fig. 5). 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) (see Experimental Procedures in the Supplementary Material).

Confirming previous phylogenetic analyses (Chen et al. 2011; Shalev et al. 2018), the 82 gymnosperm TPSs con- sidered here were divided into three major clades (Fig. 5), corresponding to 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. 5), whereas the TPS-e/f sub- family, 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. 5).

The vast majority of the TPSs considered here fell into the TPS-d subfamily (Fig. 5), which, as already said, are reputed to be gymnosperm-specific and contains all the TPSs involved in specialized metabolism. On the whole, Fig. 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 mofunctional 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 diver- sification and evolution of specific TPS functions in the two gymnosperm families (Fig. 5).

As previously discussed (see "Origin and Evolution of Plant Terpene Synthases" section), 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. 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. 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. 5). The above three TPSs from *P. contorta* are closely related to three *T. plicata* sequences (Tp STS3, Tp STS5 and TPS25) (Fig. 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 them 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 into the clade of the putative DTPSs (Fig. 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. 5). In general, the phylogenetic analysis presented here suggests that events of mono-functionalization, i.e. a form of subfunctionalization 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 deter- mined whether the class-I/II DTPSs have been lost in the Cupressaceae. Apart from DTPSs, mono-functionalization ultimately also led to

719 the ancestral tri-domain (γβα) STPSs, and to the large family of bi-domain gymnosperm MTPSs and STPSs
720 constituting the TPS-d1 and TPS-d2 groups (Fig. 5). Within these two groups, it is evident that MTPSs and
721 STPSs of the Cupressaceae clustered apart from those of the Pinaceae (Fig. 5), indicating that, as stated
722 before, the diversification and the evolution of these two TPS functions occurred independently in the two
723 gymnosperm families. While the TPS-d2 group included only the bi-domain (βα) STPSs, TPS-d1 contained all
724 the known gymnosperm MTPSs, in addition to Pinaceae TPSs that use DMAPP as a substrate to produce
725 hemiterpenes, and two bi-domain STPSs isolated from *P. taeda* and *P. abies* which produce the acyclic
726 sesquiterpene *E*,*E*-α-farnesene (Phillips et al. 2003; Martin et al. 2004; Fig. 5).
727
728 **Identification and Phylogeny of TPS Gene Sequences in *Pinus* Species**

729 Despite of its economic and ecological relevance, as well as of the importance of oleoresin terpenes in
730 defence and as bioproducts, a comprehensive analysis of the functional diversity and evolution of TPSs in
731 the *Pinus* genus still awaits to be carried out, to the best of our knowledge. There- fore, an extensive in silico
732 search was performed here, to identify all the putative full-length TPSs for primary and specialized
733 metabolisms in different *Pinus* species, and to describe their general characteristics, functional properties and
734 phylogenetic relationships.
735 The identification of full-length cDNAs coding for putative pine TPSs was based on the BLAST search of
736 the NCBI database, by using selected and functionally characterized TPSs from different conifer species
737 (Table S2, see also Experimental Procedures in the Supplementary Material). Database search identified a
738 total of 93 full-length cDNA sequences coding for putative TPSs from 28 different *Pinus* species (Table 3).
739 BLAST searches using as queries the CPS and KS from *P. glauca* (Table S2), assumed to represent DTPSs
740 involved in primary metabolism (gibberellin biosynthesis, see above), detected orthologous full-length cDNA
741 sequences only in *Pinus tabuliformis* (Pta CS1 and Pta KS1 in Table 3). It is worth noting that gymnosperm
742 CPS and KS gene sequences have been previously isolated and characterized only in *P. glauca* and *P. sitchensis*
743 (Keeling et al. 2010).
744 Five STPSs were identified for *Pinus* species in the NCBI database, of which four from *P. sylvestris* (Ps
745 STPS1-4 in Table 3) and one from *P. taeda* (Pt STPS1 in Table 3). As described before, this latter STPS
746 produces the acyclic sesquiterpene (*E*,*E*)-α-farnesene (Phillips et al. 2003) and clusters into the TPS-d1
747 group (Fig. 5). Heterologous expression in *Escherichia coli* allowed to find out that the first three STPSs
748 from *P. sylvestris* produce longifolene and α-longipinene (Ps STPS1), 1(10),5-germacradiene-4-ol (Ps
749 STPS2), and (*E*)-β-caryophyllene and α-humulene (Ps STPS3), as their main products (Köpke et al. 2008).
750 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 iden- tity (70–80%) was noticed with other conifer STPS, such as longifolene synthase from [*P. engelmannii* × *P. glauca*] and a α-humulene synthase from *P. glauca* (Keeling et al. 2011). This indicates the presence of putative hortologous 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 *Pinus banksiana*, respectively, and one each in *P. taeda* and *Pinus densiflora* (Table 3). 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; Mar- tin 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 chromatogra- phy–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

784 synthase (Pg KS) of gibberellin metabolism (Keeling et al. 2010), indicating their involvement in the
785 specialized metabolism, rather than in the primary one. Showing 99% amino acid sequence identity to each
786 other, Pc DTPS mISO1 and Pb DTPS mISO1, as well as Pc DTPS mPIM1 and Pb DTPS mPIM1, are
787 probably two pairs of orthologous genes from *P. contorta* and *P. banksiana*, respectively. The other class-
788 I DTPSs (namely Pc MDTPS1, Pc MDTPS2, PcmdiTPS3, and Pb MDTPS1), though very similar among
789 each others (97% to 98% protein sequence identity), exhibited a low identity (71–75%) with the other identified
790 pine DTPSs. Functional characterization of four of the eight class-I DTPSs, identified the putative
791 orthologous pair Pc DTPS mPIM1 and Pb DTPS mPIM1 as single product pimaradiene synthases, whereas
792 the orthologous pair of Pc DTPS mISO1 and Pb DTPS mISO1 were found to produce isopimaradiene as
793 main product, whit small amounts of sandaracopimaradiene (Hall et al. 2013a). Interestingly, class-I DTPSs
794 of specialized DRA metabolism have not been previously identified in Pinaceae. The only other known
795 examples of class-I DTPSs of specialized metabolism in gymnosperms are the two TPSs identified in the
796 Cupressaceae species *T. plicata* (Tp TPS10) and *T. cryptomerioides* (Tcr TPS4), whose functions remain to
797 be determined, and taxadiene synthases in *Taxus* ssp. (Wildung and Croteau 1996), which specifically
798 transforms GGPP into the macrocyclic taxadiene backbone without the need for a bicyclic diphosphate
799 intermediate. The aforementioned class-I DTPS characterized by Hall et al. (2013a) are the first examples
800 in gymnosperms of enzymes able to synthetize mainly pimaradiene; as such, they add themselves to the
801 already known ISO and LAS conifers DTPS participating in the DRA specialized metabolism.
802 BLAST searches using as queries the 7 selected MTPSs from *P. contorta*, *P. abies* and *P. glauca* (Table S2)
803 detected 74 putative full-length cDNAs coding for MTPSs from 26 different *Pinus* species (Table 3). However,
804 only 32 of them could be classified as true MTPSs. The deduced amino acid sequences of the remaining 42
805 cDNA sequences, from 18 different *Pinus* species, were predicted to synthetize 2-methyl-3-buten-2-ol
806 (MBO), a C_5 alcohol produced and emitted by several pine species (Lerdau and Gray 2003). MBO is related
807 to isoprene by a structural and biosynthetic point of view and both derive from DMAPP (Gray et al. 2011).
808 The gene for MBO synthase (MBOS) was first isolated from *Pinus sabiniana* and found to encode for a
809 bifunctional enzyme able to produce both MBO and isoprene in a 90:1 ratio (Gray et al. 2011). Conifer MBOSs
810 appear to have evolved independently from their homologous proteins in angiosperms, i.e. IPSs (Sharkey et al.
811 2013). Indeed, phylogenetic analysis showed that MBOSs fall into the TPS-d1 group, together with the
812 gymnosperm MTPSs, and are most closely related to linalool synthases from *P. abies* (Martin et al. 2004) and
813 *P. sitchensis* (Keeling et al. 2011) (Fig. 5). The 42 full-length MBOS sequences identified here showed a high
814 level of homology among each other (93–99% amino acid sequence identity) as shown in the phylogenetic tree
815 reported in Fig. S1.
816 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 2 *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. 6.

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. 6), consistent with their assignment to distinct TPS subfamilies, namely TPS-d for the formers, and TPS-c and TPS-e/f for the latters (Fig. 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).

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. 6), thus confirming the previous findings of Gray et al. (2011). The five iden- tified STPSs in *Pinus* species were found to distribute in all of the three TPS-d clades (Fig. 6). 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 bifunctional DTPSs identified in the *Pinus* species (Fig. 6).

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. 6). 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. 6 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 diversi- fication. 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. 6), *P. abies*, *P. sitchensis*, *Abies balsamea* and *A. grandis* (Fig. 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 5 selected pine MBOSs clustered into 7 distinct groups (Fig. 6). 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. 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. 6) 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. 5) showed that the MBOSs from several Pinaceae species are closely related to linalool synthases from *P. abies* and *P. sitchensis*.

Group 2 (Fig. 6) included only two proteins from *P. con- torta* (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 has only a 62% sequence identity with Pt MTPS2, a *P. taeda* protein that also pro- duces α-terpineol (Phillips et al. 2003) but was assigned to the Group 4 (Fig. 6). Indeed, Fig. 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. 6) 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. 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. 6) contained four MTPSs from *P. con- torta* 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. 6), 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 and Croteau 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 *Pinus 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. 6) included 10 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. 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. 6) 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. 6) 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 *Pinus kesiya*, although reported to be a α-pinene synthase in the NCBI database, was not function- ally 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. 5). In the phylogenetic tree of Fig. 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. 6 with Fig. 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.

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. 6).

Pinus nigra subsp. *laricio* (Poiret) is one of the six subspecies of *P. 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 on 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. 6) 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. 7). The Group 4 partial transcript of 1132 bp in length encoded an incomplete protein of 376 aa (Fig. 7).

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 (see

Experimental Procedures in the Supplementary Material). 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 10 exons (with the 1st and the 10th incomplete) and 9 introns (Fig. S2), being consistent with the previously characterized genomic sequences of conifer MTPSs (Trapp and Croteau 2001; Hamberger et al. 2009; Hall et al. 2011). In this context, it is noteworthy the high conservation of the exon size detected between genes coding for MTPS in *A. grandis* (Trapp and Croteau 2001), *P. glauca* (Hamberger et al. 2009), *P. sitchensis* (Hall et al. 2011) and the *P. laricio* genomic sequence isolated in 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 amino acid 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. S3), 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. 7). 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.

Conclusions and Prospects

Conifers developed a variety of physical and chemical defences against pathogens and herbivores, among which one of the most significant is the production of oleoresin, a complex blend of volatile mono- and sesquiterpenes, along with non-volatile diterpene resin acids. In fact, the complex defence 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, pharmaceuticals, 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 (*Sac- charomyces 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 coni- fer 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.

Funding The present work was carried out in the framework of the "ALForLab" Project (PON03PE_00024_1), co-funded by the National Operational Programme for Research and Competitiveness (PON R&C) 2007–2013, through the European Regional Development Fund (ERDF) and National Resource (Revolving Fund-Cohesion Action Plan (CAP) MIUR).

Data Availability The cDNA and genomic sequences of the monoter- pene synthases (MTPSs) from *Pinus nigra* subsp. *laricio* will appear in the GenBank database with the following accession num- bers: MN088807 (Pnl_MBOS_1.1), MN088808 (Pnl_MTPS_1.2), MN088809 (Pnl_MTPS_1.5), MN088810 (Pnl_MTPS_1.7), MN088811 (Pnl_MTPS_1.4) and MN088812 (Pnl_MTPS_1.3).

Compliance with Ethical Standards

Conflict of interest The authors declare that they have no conflict of interest.

References

Abbas F, Ke Y, Yu R, Yue Y, Amanullah S, Jahangir MM, Fan Y (2017) Volatile terpenoids: multiple functions, biosynthesis, modulation and manipulation by genetic engineering. Planta 246:803–816. https://doi.org/10.1007/s00425-017-2749-x

Aharoni A, Giri AP, Verstappen FWA, Bertea CM, Sevenier R (2004) Gain and loss of fruit flavor compounds produced by wild and cultivated strawberry species. Plant Cell 16:3110–3131. https://doi.org/10.1105/tpc.104.023895.native

Akhtar TA, Matsuba Y, Schauvinhold I, Yu G, Lees HA, Klein SE, Pichersky E (2013) The tomato *cis*-prenyltransferase gene family. Plant J 73:640–652. https://doi.org/10.1111/tpj.12063

Akiyama K, Matsuzaki K, Hayashi H (2005) Plant sesquiterpenes induce hyphal branching in arbuscular mycorrhizal fungi. Nature 435:824. https://doi.org/10.1038/nature03608

Aubourg S, Lecharny A, Bohlmann J (2002) Genomic analysis of the terpenoid synthase (AtTPS) gene family of *Arabidopsis thali- ana*. Mol Genet Genomics 267:730–745. https://doi.org/10.1007/ s00438-002-0709-y

Baier M, Hartung W (1988) Movement of abscisic acid across the plasmalemma and the tonoplast of guard cells of *Valerianella locusta*. Plant Biol 101:332–337. https://doi. org/10.1111/j.1438-8677.1988.tb00053.x

Baldwin IT, Halitschke R, Paschold A, Von Dahl CC, Preston CA (2006) Volatile signaling in plant–plant

interactions: “talking trees” in the genomics era. Science 311:812–815. https://doi.org/10.1126/science.1118446

Behnke K, Wildt R, Kleist E, Uerlings R (2009) RNAi-mediated suppression of isoprene biosynthesis in hybrid poplar impacts ozone tolerance. Tree Physiol 29:725–736. https://doi.org/10.1093/treephys/tpp009

Bick JA, Lange BM (2003) Metabolic cross talk between cytosolic and plastidial pathways of isoprenoid biosynthesis: unidirectional transport of intermediates across the chloroplast envelope membrane. Arch Biochem Biophys 415:146–154. https://doi.org/10.1016/S0003-9861(03)00233-9

Birol I, Raymond A, Jackman SD, Pleasance S, Coope R, Taylor GA et al (2013) Genome analysis assembling the 20 Gb white spruce (*Picea glauca*) genome from whole-genome shotgun sequencing data. Bioinformatics 29:1492–1497. https://doi.org/10.1093/bioinformatics/btt178

Bleeker PM, Diergaarde PJ, Ament K, Schütz S, Johne B, Dijkink J et al (2011) Tomato-produced 7-epizingiberene and R-curcumene act as repellents to whiteflies. Phytochemistry 72:68–73. https://doi.org/10.1016/j.phytochem.2010.10.014

Bohlmann J, Keeling CI (2008) Terpenoid biomaterials. Plant J 54:656–669. https://doi.org/10.1111/j.1365-313X.2008.03449.x

Bohlmann J, Steele CL, 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. J Biol Chem 272:21784–21792. https://doi.org/10.1074/jbc.272.35.21784

Bohlmann J, Meyer-Gauen G, Croteau R (1998) Plant terpenoid synthases: molecular biology and phylogenetic analysis. Proc Natl Acad Sci USA 95:4126–4133. https://doi.org/10.1073/pnas.95.8.4126

Bohlmann J, Phillips M, Ramachandiran V, Katoh S, Croteau R (1999) cDNA cloning, characterization, and functional expression of four new monoterpene synthase members of the Tpsd gene family from grand fir (*Abies grandis*). Arch Biochem Biophys 368:232–243. https://doi.org/10.1006/abbi.1999.1332

Büchel K, Malskies S, Mayer M, Fenning TM, Gershenzon J, Hilker M, Meiners T (2011) How plants give early herbivore alert: volatile terpenoids attract parasitoids to egg-infested elms. Basic Appl Ecol 12:403–412. https://doi.org/10.1016/j.baae.2011.06.002

Cao R, Zhang Y, Mann FM, Huang C, Mukkamala D, Hudock MP et al (2010) Diterpene cyclases and the nature of the isoprene fold. Proteins 78:2417–2432. https://doi.org/10.1002/prot.22751

Celedon JM, Bohlmann J (2019) Oleoresin defenses in conifers: chemical diversity, terpene synthases, limitations of oleoresin defense under climate change. N Phytol 224:1444–1463. https://doi.org/10.1111/nph.15984

Chen F, Pichersky E, Ro D, Petri J, Gershenzon J, Tholl D (2004) characterization of a root-specific *Arabidopsis* terpene synthase responsible for the formation of the volatile monoterpene 1, 8-cineole. Plant Physiol 135:1956–1966. https://doi.org/10.1104/pp.104.044388.1956

Chen F, Tholl D, Bohlmann J, Pichersky E (2011) The family of terpene synthases in plants: a mid-size family of genes for spe- cialized metabolism that is highly diversified throughout the kingdom. Plant J 66:212–229. https://doi.org/10.1111/j.1365- 313X.2011.04520.x

Cherian S, Ryu SB, Cornish K (2019) Natural rubber biosynthesis in plants, the rubber transferase complex, and metabolic engineer- ing progress and prospects. Plant Biotechnol J 1–21: https://doi. org/10.1111/pbi.13181

Christianson DW (2017) Structural and chemical biology of terpenoid cyclases. Chem Rev 117:11570–11648. https://doi.org/10.1021/ acs.chemrev.7b00287

Cordoba E, Salmi M, Leo P (2009) Unravelling the regulatory mechanisms that modulate the MEP pathway in higher plants. J Exp Bot 60:2933–2943. https://doi.org/10.1093/jxb/erp190

Dellas N, Thomas ST, Manning G, Noel JP (2013) Discovery of a metabolic alternative to the classical mevalonate pathway. Elife 2:e00672. https://doi.org/10.7554/eLife.00672

Dong L, Jongedijk E, Bouwmeester H, Krol A (2016) Monoterpene biosynthesis potential of plant subcellular compartments. N Phytol 209:679–690. https://doi.org/10.1111/nph.13629

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

Dudareva N, Andersson S, Orlova I, Gatto N, Reichelt M, Rhodes D et al (2005) The nonmevalonate pathway supports both monoterpene and sesquiterpene formation in snapdragon flowers. Proc Natl Acad Sci USA 102:933–938. https://doi.org/10.1073/ pnas.0407360102

Dudareva N, Klempien A, Muhlemann K, Kaplan I (2013) Tansley review biosynthesis, function and metabolic engineering of plant volatile organic compounds. N Phytol 198:16–32. https://doi. org/10.1111/nph.12145

Falara V, Akhtar TA, Nguyen TTH, Spyropoulou EA, Bleeker PM, Schauvinhold I et al (2011) The tomato terpene synthase gene family. Plant Physiol 157:770–789. https://doi.org/10.1104/ pp.111.179648

Flügge U, Gao W (2005) Transport of isoprenoid intermediates across chloroplast envelope membranes. Plant Biol 7:91–97. https://doi. org/10.1055/s-2004-830446

Gao Y, Honzatko RB, Peters RJ (2012) Terpenoid synthase structures: a so far incomplete view of complex catalysis. Nat Prod Rep 29:1153–1175. https://doi.org/10.1039/C2NP20059G

Gennadios HA, Gonzalez V, Costanzo L, Li A, Yu F, Allemann RK et al (2009) Crystal structure of (+)-δ-cadinene synthase from *Gossypium arboreum* and evolutionary divergence of metal binding motifs for catalysis. Biochemistry 48:6175–6183. https://doi. org/10.1021/bi900483b

Ghassemian M, Lutes J, Tepperman JM, Chang HS, Zhu T, Wang X, Quail PH, Lange BM (2006) Integrative analysis of transcript and metabolite profiling data sets to evaluate the regulation of bio- chemical

pathways during photomorphogenesis. Arch Biochem Biophys 448:45–59. https://doi.org/10.1016/j.abb.2005.11.020

Gols R (2014) Direct and indirect chemical defences against insects in a multitrophic framework. Plant Cell Environ 37:1741–1752. https://doi.org/10.1111/pce.12318

Gray DW, Breneman SR, Topper LA, Sharkey TD (2011) Biochemical characterization and homology modeling of methylbutenol synthase and implications for understanding hemiterpene synthase evolution in plants. J Biol Chem 286:20582–20590. https://doi. org/10.1074/jbc.M111.237438

Guirimand G, Guihur A, Phillips MA, Oudin A, Glevarec G, Melin C, Papon N, Clastre M, St-Pierre B, Rodríguez-Concepción M, Burlat V, Courdavault V (2012) A single gene encodes isopentenyl diphosphate isomerase isoforms targeted to plastids, mitochondria and peroxisomes in *Catharanthus roseus*. Plant Mol Biol 79:443–459. https://doi.org/10.1007/s11103-012-9923-0

Gutensohn M, Orlova I, Nguyen TTH, Davidovich-Rikanati R, Ferruzzi MG, Sitrit Y et al (2013) Cytosolic monoterpene biosynthesis is supported by plastid-generated geranyl diphosphate substrate in transgenic tomato fruits. Plant J 75:351–363. https://doi. org/10.1111/tpj.12212

Hall DE, Robert JA, Keeling CI, Domanski D, Quesada AL, Jancsik S et al (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. Plant J 65:936–948. https://doi.org/10.1111/j.1365-313X.2010.04478.x

Hall DE, Oa DSW, Zerbe P, Jancsik S, Quesada AL, Dullat H, Madilao LL (2013a) Evolution of conifer diterpene synthases: diterpene resin acid biosynthesis in lodgepole pine and jack pine involves monofunctional and bifunctional diterpene synthases. Plant Physiol 161:600–616. https://doi.org/10.1104/pp.112.208546

Hall DE, Yuen MM, Jancsik S, Quesada AL, Dullat HK, Li M et al (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 Biol 13:80. https://doi. org/10.1186/1471-2229-13-80

Hamberger B, Hall D, Yuen M, Oddy C, Hamberger B, Keeling CI et al (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 Biol 9:106. https://doi.org/10.1186/1471-2229-9-106

Hampel D, Mosandl A, Wust M (2005) Biosynthesis of mono- and sesquiterpenes in carrot roots and leaves (*Daucus carota* L.): metabolic cross talk of cytosolic mevalonate and plastidial methylerythritol phosphate pathways. Phytochemistry 66:305–311. https://doi.org/10.1016/j.phytochem.2004.12.010

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 Lett 580:6175–6181. https://doi.org/10.1016/j.febsl et.2006.10.018

Hayashi K, Horie K, Hiwatashi Y, Kawaide H, Yamaguchi S (2010) Endogenous diterpenes derived from *ent*-kaurene, a common gibberellin precursor, regulate protonema differentiation of the moss *Physcomitrella patens*. Plant Physiol 153:1085–1097. https
://doi.org/10.1104/pp.110.157909

Heil M (2014) Herbivore-induced plant volatiles: targets, perception and unanswered questions. N Phytol 204:297–306. https://doi. org/10.1111/nph.12977

Hemmerlin A (2013) Post-translational events and modifications regulating plant enzymes involved in isoprenoid precursor bio- synthesis. Plant Sci 203:41–54. https://doi.org/10.1016/j.plant sci.2012.12.008

Hemmerlin A, Hoeffler JF, Meyer O, Tritsch D, Kagan IA, Grosdemange-Billiard C, Rohmer M, Bach TJ (2003) Cross-talk between the cytosolic mevalonate and the plastidial methylerythritol phosphate pathways in tobacco bright yellow-2 cells. J Biol Chem 278:26666–26676. https://doi.org/10.1074/jbc. M302526200

Henry LK, Gutensohn M, Thomas ST, Noel JP, Dudareva N (2015) Orthologs of the archaeal isopentenyl phosphate kinase regu- late terpenoid production in plants. Proc Natl Acad Sci USA 112:10050–10055. https://doi.org/10.1073/pnas.1504798112

Henry LK, Thomas ST, Widhalm JR, Lynch JH, Davis TC, Kessler SA, Bohlmann J, Noel JP, Dudareva N (2018) Contribution of isopentenyl phosphate to plant terpenoid metabolism. Nat Plants 4:721–729. https://doi.org/10.1038/s41477-018-0220-z

Herde M, Fode B, Boland W, Gershenzon J, Ga K, Gatz C, Tholl D (2008) Identification and regulation of TPS04/GES, an *Arabidopsis* geranyl linalool synthase catalyzing the first step in the formation of the insect-induced volatile C_{16}-homoterpene TMTT. Plant Cell 20:1152–1168. https://doi.org/10.1105/tpc.106.049478

Hillwig ML, Xu M, Toyomasu T, Tiernan MS, Wei G, Cui G et al (2011) Domain loss has independently occurred multiple times in plant terpene synthase evolution. Plant J 68:1051–1060. https://doi.org/10.1111/j.1365-313X.2011.04756.x

Hu XG, Liu H, Jin YQ, Sun YQ, Li Y, Zhao W et al (2016) De novo transcriptome assembly and characterization for the widespread and stress-tolerant conifer *Platycladus orientalis*. PLoS ONE 11:1–19. https://doi.org/10.1371/journal.pone.0148985

Huang M, Abel C, Sohrabi R, Petri J, Haupt I, Cosimano J et al (2010) Variation of herbivore-induced volatile terpenes among *Arabidopsis* ecotypes depends on allelic differences and subcellular targeting of two terpene. Plant Physiol 153:1293–1310. https:// doi.org/10.1104/pp.110.154864

Huang M, Sanchez-Moreiras AM, Abel C, Sohrabi R, Lee S, Gershenzon J, Tholl D (2012) The major volatile

organic compound emit- ted from *Arabidopsis thaliana* flowers, the sesquiterpene (E)-β- caryophyllene, is a defense against a bacterial pathogen. N Phytol 193:997–1008. https://doi.org/10.1111/j.1469-8137.2011.04001.x

Huber DP, Philippe RN, Godard KA, Sturrock RN, Bohlmann J (2005) Characterization of four terpene synthase cDNAs from methyl jasmonate-induced Douglas-fir, *Pseudotsuga menziesii*. Phytochemistry 66:1427–1439. https://doi.org/10.1016/j.phyto chem.2005.04.030

Hyatt DC, Croteau R (2005) Mutational analysis of a monoterpene synthase reaction: altered catalysis through directed mutagenesis of (−)-pinene synthase from *Abies grandis*. Arch Biochem Biophys 439:222–233. https://doi.org/10.1016/j. abb.2005.05.017

Hyatt DC, Youn B, Zhao Y, Santhamma B, Coates RM, Croteau RB (2007) Structure of limonene synthase, a simple model for terpenoid cyclase catalysis. Proc Natl Acad Sci USA 104:5360–5365. https://doi.org/10.1073/pnas.0700915104

Irmisch S, Chen F, Gershenzon J, Jiang Y, Köllner TG (2014) Terpene synthases and their contribution to herbivore-induced volatile emission in western balsam poplar (*Populus trichocarpa*). BMC Plant Biol 14:1–16. https://doi.org/10.1186/s12870-014-0270-y

Jia Q, Köllner TG, Gershenzon J, Chen F (2018) MTPSLs: new terpene synthases in non seed plants. Trends Plant Sci 23:121–128. https://doi.org/10.1016/j.tplants.2017.09.014

Kai G, Xu H, Zhou C, Liao P, Xiao J, Luo X, You L, Zhang L (2011) Metabolic engineering Tanshinone biosynthetic pathway in *Salvia miltiorrhiza* hairy root cultures. Metab Eng 13:319–327. https://doi.org/10.1016/j.ymben.2011.02.003

Kampranis SC, Ioannidis D, Purvis A, Mahrez W, Ninga E, Katerelos NA, Goodenough PW et al (2007) Rational conversion of substrate and product specificity in a *Salvia* monoterpene synthase: structural insights into the evolution of terpene synthase function. Plant Cell 19:1994–2005. https://doi.org/10.1105/ tpc.106.047779

Karunanithi PS, Zerbe P (2019) Terpene synthases as metabolic gate- keepers in the evolution of plant terpenoid chemical diversity. Front Plant Sci 10:1166. https://doi.org/10.3389/fpls.2019.01166

Kasahara H, Hanada A, Kuzuyama T, Takagi M, Kamiya Y, Yamaguchi S (2002) Contribution of the mevalonate and methylerythritol phosphate pathways to the biosynthesis of gibberellins in *Arabidopsis*. J Biol Chem 277:45188–45194. https://doi.org/10.1074/ jbc.M208659200

Katoh S, Hyatt D, Croteau R (2004) Altering product outcome in *Abies grandis* (−)limonene synthase and (−)-limonene/(−)-α-pinene synthase by domain swapping and directed mutagenesis. Arch Biochem Biophys 425:65–76. https://doi.org/10.1016/j. abb.2004.02.015

Keeling CI, Bohlmann J (2006) Genes, enzymes and chemicals of terpenoid diversity in the constitutive and induced defence of conifers against insects and pathogens. N Phytol 170:657–675. https://doi.org/10.1111/j.1469-8137.2006.01716.x

Keeling CI, Dullat HK, Yuen M, Ralph SG, Jancsik S (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 Physiol 152:1197–1208. https://doi.org/10.1104/pp.109.151456

Keeling CI, Weisshaar S, Ralph SG, Jancsik S, Hamberger B, Dullat HK (2011) Transcriptome mining, functional characterization, and phylogeny of a large terpene synthase gene family in spruce (*Picea* spp.). BMC Plant Biol 11:43. https://doi. org/10.1186/1471-2229-11-43

Kharel Y, Koyama T (2003) Molecular analysis of *cis*-prenyl chain elongating enzymes. Nat Prod Rep 20:111–118. https://doi. org/10.1039/B108934J

Kim YJ, Lee OR, Oh JY, Jang MG, Yang DC (2014) Functional analysis of 3-hydroxy-3-methylglutaryl coenzyme A reductase encoding genes in triterpene saponin-producing ginseng. Plant Physiol 165:373–387. https://doi.org/10.1104/pp.113.222596

Kirby J, Nishimoto M, Park JG, Withers ST, Nowroozi F, Behrendt D et al (2010) Cloning of casbene and neocembrene synthases from Euphorbiaceae plants and expression in *Saccharomyces cerevi- siae*. Phytochemistry 71:1466–1473. https://doi.org/10.1016/j. phytochem.2010.06.001

Köksal M, Hu H, Coates RM, Peters RJ, Christianson DW (2011a) Structure and mechanism of the diterpene cyclase *ent*-copalyl diphosphate synthase. Nat Chem Biol 7:431. https://doi. org/10.1038/NChemBio.578

Köksal M, Jin Y, Coates RM, Croteau R, Christianson DW (2011b) Taxadiene synthase structure and evolution of modular architecture in terpene biosynthesis. Nature 469:116. https://doi. org/10.1038/nature09628

Köksal M, Potter K, Peters RJ, Christianson DW (2014) 1.55 Å resolution structure of *ent*-copalyl diphosphate synthase and exploration of general acid function by site-directed mutagenesis. Biochim Biophys Acta 1840:184–190. https://doi.org/10.1016/j. bbagen.2013.09.004

Köpke D, Schröder R, Fischer HM, Gershenzon J, Hilker M, Schmidt A (2008) Does egg deposition by herbivorous pine sawflies affect transcription of sesquiterpene synthases in pine? Planta 228:427–438. https://doi.org/10.1007/s00425-008-0747-8

Külheim C, Padovan A, Hefer C, Krause ST, Köllner TG, Myburg AA et al (2015) The *Eucalyptus* terpene synthase gene family. BMC Genomics 16: https://doi.org/10.1186/s12864-015-1598-x

Kumar S, Hahn FM, Baidoo E, Kahlon TS, Wood DF, McMahan CM, Cornish K, Keasling JD, Daniell H, Whalen MC (2012) Remodeling the isoprenoid pathway in tobacco by expressing the cytoplasmic mevalonate pathway in chloroplasts. Metab Eng 14:19–28. https://doi.org/10.1016/j.ymben.2011.11.005

Kumar S, Stecher G, Li M, Knyaz C, Tamura K (2018) MEGA X: molecular evolutionary genetics analysis

across computing plat- forms. Mol Biol Evol 35:1547–1549. https://doi.org/10.1093/ molbev/msy096

Laby RJ, Kincaid MS, Kim D, Gibson SI (2000) The *Arabidopsis* sugar-insensitive mutants sis4 and sis5 are defective in abscisic acid synthesis and response. Plant J 23:587–596. https://doi. org/10.1046/j.1365-313x.2000.00833.x

Lange I, Poirier BC, Herron BK, Lange BM (2015) Comprehensive assessment of transcriptional regulation facilitates metabolic engineering of isoprenoid accumulation in *Arabidopsis*. Plant Physiol 169:1595–1606. https://doi.org/10.1104/pp.15.00573

Laothawornkitkul J, Paul ND, Vickers CE, Possell M, Taylor JE, Mullineaux PM et al (2008) Isoprene emissions influence herbivore feeding decisions. Plant Cell Environ 31:1410–1415. https://doi. org/10.1111/j.1365-3040.2008.01849.x

Laule O, Furholz A, Chang HS, Zhu T, Wang X, Heifetz PB, Gruissem W, Lange M (2003) Crosstalk between cytosolic and plastidial pathways of isoprenoid biosynthesis in *Arabidopsis thaliana*. Proc Natl Acad Sci USA 100:6866–6871. https://doi. org/10.1073/pnas.1031755100

Lerdau M, Gray D (2003) Ecology and evolution of light-dependent and light-independent phytogenic volatile organic carbon. N Phytol 157:199–211. https://doi.org/10.1046/j.1469-8137.2003.00673.x

Lesburg CA, Zhai G, Cane DE, Christianson DW (1997) Crystal structure of pentalenene synthase: mechanistic insights on terpenoid cyclization reactions in biology. Science 277:1820–1824. https://doi.org/10.1126/science.277.5333.1820

Li G, Köllner TG, Yin Y, Jiang Y, Xu Y, Gershenzon J et al (2012) Nonseed plant *Selaginella moellendorffii* has both seed plant and microbial types of terpene synthases. Proc Natl Acad Sci USA 109:14711–14715. https://doi.org/10.1073/pnas.1219604110

Loreto F, Fischbach RJ, Schnitzler JP, Ciccioli P, Brancaleoni E, Calfapietra C, Seufert G (2001) Monoterpene emission and monoterpene synthase activities in the Mediterranean ever- green oak *Quercus ilex* L. grown at elevated CO_2 concentrations. Glob Change Biol 7:709–717. https://doi.org/10.1046/j.1354-1013.2001.00442.x

Loreto F, Dicke M, Schnitzler JP, Turlings TC (2014) Plant volatiles and the environment. Plant Cell Environ 37:1905–1908. https:// doi.org/10.1111/pce.12369

Ma D, Li G, Zhu Y, Xie DY (2017) Overexpression and suppression of *Artemisia annua* 4-hydroxy-3-methylbut-2-enyl diphosphate reductase 1 gene (AaHDR1) differentially regulate artemisinin and terpenoid biosynthesis. Front Plant Sci 8:77. https://doi. org/10.3389/fpls.2017.00077

Mafu S, Hillwig ML, Peters RJ (2011) A novel labda-7, 13-E-dien- 15-ol-producing bifunctional diterpene synthase from *Selaginella moellendorffii*. ChemBioChem 12:1984–1987. https://doi.org/10.1002/cbic.201100336

Martin DM, Fäldt J, Bohlmann J (2004) Functional characterization of nine Norway spruce TPS genes and evolution of gymno- sperm terpene synthases of the TPS-d subfamily. Plant Physiol 135:1908–1927. https://doi.org/10.1104/pp.104.042028.1908 Martin DM, Aubourg S, Schouwey MB, Daviet L, Schalk M, Toub O et al (2010) Functional annotation, genome organization and phylogeny of the grapevine (*Vitis vinifera*) terpene synthase gene family based on genome assembly, FLcDNA cloning, and enzyme assays. BMC Plant Biol 10:226. https://doi. org/10.1186/1471-2229-10-226

Mau JD, West CA (1994) Cloning of casbene synthase cDNA: evidence for conserved structural features among terpenoid cyclases in plants. Proc Natl Acad Sci USA 91:8497–8501. https://doi.org/10.1073/pnas.91.18.8497

May B, Lange BM, Wüst M (2013) Biosynthesis of sesquiterpenes in grape berry exocarp of *Vitis vinifera* L.: evidence for a trans- port of farnesyl diphosphate precursors from plastids to the cytosol. Phytochemistry 95:135–144. https://doi.org/10.1016/j. phytochem.2013.07.021

Mcandrew RP, Peralta-Yahya PP, Degiovanni A, Pereira JH, Hadi MZ, Keasling JD, Adams PD (2011) Structure of a three- domain sesquiterpene synthase: a prospective target for advanced biofuels production. Structure 19:1876–1884. https://doi.org/10.1016/j.str.2011.09.013

Meier S, Tzfadia O, Vallabhaneni R, Gehring C, Wurtzel ET (2011) A transcriptional analysis of carotenoid, chlorophyll and plastidial isoprenoid biosynthesis genes during development and osmotic stress responses in *Arabidopsis thaliana*. BMC Syst Biol. https://doi.org/10.1186/1752-0509-5-77

Mendoza-Poudereux I, Kutzner E, Huber C, Segura J, Eisenreich W, Arrillaga I (2015) Metabolic cross-talk between pathways of terpenoid backbone biosynthesis in spike lavender. Plant Physiol Biochem 95:113–120. https://doi.org/10.1016/j.plaph y.2015.07.029

Mongelard G, Seemann M, Boisson AM, Rohmer M, Bligny R, Rivasseau C (2011) Measurement of carbon flux through the MEP pathway for isoprenoid synthesis by ^{31}P-NMR spectroscopy after specific inhibition of 2-C-methyl-D-erythritol 2,4-cyclodiphosphate reductase. Effect of light and temperature. Plant Cell Environ 34:1241–1247. https://doi.org/10.111 1/j.1365-3040.2011.02322.x

Monson RK, Jones RT, Rosenstiel TN, Schnitzler JP (2013) Why only some plants emit isoprene. Plant Cell Environ 36:503–516. https://doi.org/10.1111/pce.12015

Morrone D, Chambers J, Lowry L, Kim G, Anterola A, Bender K, Peters RJ (2009) Gibberellin biosynthesis in bacteria: separate *ent*-copalyl diphosphate and *ent*-kaurene synthases in *Bradyrhizobium japonicum*. FEBS Lett 583:475–480. https:// doi.org/10.1016/j.febslet.2008.12.052

Mumm R, Posthumus MA, Dicke M (2008) Significance of terpenoids in induced indirect plant defence against herbivorous arthropods. Plant Cell Environ 31:575–585. https://doi.org/10.1111/j.1365-3040.2008.01783.x

Nagegowda DA, Gutensohn M, Wilkerson CG, Dudareva N, Lafayette W, Consortium MP, Lansing E (2008) Two nearly identical terpene synthases catalyze the formation of nerolidol and linalool in snapdragon flowers. Plant J 55:224–239. https://doi. org/10.1111/j.1365-313X.2008.03496.x

Neale DB, Wegrzyn JL, Stevens KA, Zimin AV, Puiu D, Crepeau MW et al (2014) Decoding the massive genome of loblolly pine using haploid DNA and novel assembly strategies. Genome Biol 15:R59. https://doi.org/10.1186/gb-2014-15-3-r59

Nelson DR (2011) Progress in tracing the evolutionary paths of cytochrome P450. Biochim Biophys Acta Proteins Proteomics 1814:14–18. https://doi.org/10.1016/j.bbapap.2010.08.008

Nicolaci A, Travaglini D, Menguzzato G, Nocentini S, Veltri A, Iovino F (2014) Ecological and anthropogenic drivers of Calabrian pine (*Pinus nigra* JF Arn. ssp. *laricio* (Poiret) Maire) distribution in the Sila Mountain Range. iForest 8:497. https:// doi.org/10.3832/ifor1041-007

Nystedt B, Street NR, Wetterbom A, Zuccolo A, Lin YC, Scofield DG et al (2013) The Norway spruce genome sequence and conifer genome evolution. Nature 497:579–584. https://doi. org/10.1038/nature12211

Opitz S, Nes WD, Gershenzon J (2014) Both methylerythritol phosphate and mevalonate pathways contribute to biosynthesis of each of the major isoprenoid classes in young cotton seedlings. Phytochemistry 98:110–119. https://doi.org/10.1016/j.phyto chem.2013.11.010

Orlova I, Varbanova M, Fridman E, Yamaguchi S, Hanada A, Kamiya Y, Krichevsky A (2009) The small subunit of snap- dragon geranyl diphosphate synthase modifies the chain length specificity of tobacco geranylgeranyl diphosphate synthase in planta. Plant Cell 21:4002–4017. https://doi.org/10.1105/ tpc.109.071282

Pazouki L, Niinemets Ü (2016) Multi-substrate terpene synthases: their occurrence and physiological significance. Front Plant Sci 7:1019. https://doi.org/10.3389/fpls.2016.01019

Peters RJ (2006) Uncovering the complex metabolic network underlying diterpenoid phytoalexin biosynthesis in rice and other cereal crop plants. Phytochemistry 67:2307–2317. https://doi. org/10.1016/j.phytochem.2006.08.009

Peters RJ, Flory JE, Jetter R, Ravn MM, Lee HJ, Coates RM, Croteau RB (2000) Abietadiene synthase from grand fir (*Abies grandis*): characterization and mechanism of action of the "pseudomature" recombinant enzyme. Biochemistry 39:15592–15602. https://doi. org/10.1021/bi001997l

Peters RJ, Carter OA, Zhang Y, Matthews BW, Croteau RB (2003) Bifunctional abietadiene synthase: mutual structural dependence of the active sites for protonation-initiated and ionization- initiated cyclizations. Biochemistry 42:2700–2707. https://doi. org/10.1021/bi020492n

Phillips MA, Wildung MR, Williams DC, Hyatt DC, Croteau R (2003) cDNA isolation, functional

1342 expression, and characterization of (+)-α-pinene synthase and (−)-α-pinene synthase from loblolly pine (*Pinus*
1343 *taeda*): stereocontrol in pinene biosynthesis. Arch Biochem Biophys 411:267–276. https://doi.org/10.1016/S0003-
1344 9861(02)00746-4
1345 Pierik R, Ballaré CL, Dicke M (2014) Ecology of plant volatiles: taking a plant community perspective. Plant
1346 Cell Environ 37:1845– 1853. https://doi.org/10.1111/pce.12330
1347 Polge C, Thomas M (2007) SNF1/AMPK/SnRK1 kinases, global regulators at the heart of energy control?
1348 Trends Plant Sci 12:20–28. https://doi.org/10.1016/j.tplants.2006.11.005
1349 Prisic S, Xu M, Wilderman PR, Peters RJ (2004) Rice contains two disparate *ent*-copalyl diphosphate synthases
1350 with distinct metabolic functions. Plant Physiol 136:4228–4236. https://doi.org/10.1104/ pp.104.050567.4228
1351 Rajabi MH, Pazouki L, Niinemets Ü (2013) The biochemistry and molecular biology of volatile messengers
1352 in trees. Biology, controls and models of tree volatile organic compound emissions. Springer, Dordrecht, pp
1353 47–93
1354 Rasmann S, Köllner TG, Degenhardt J, Hiltpold I, Toepfer S, Kuh- lmann U et al (2005) Recruitment
1355 of entomopathogenic nematodes by insect-damaged maize roots. Nature 434:732.
1356 https://doi.org/10.1038/nature03451DO
1357 Rasulov B, Talts E, Kännaste A, Niinemets Ü (2015) Bisphosphonate inhibitors reveal a large elasticity of
1358 plastidic isoprenoid syn- thesis pathway in isoprene-emitting hybrid aspen. Plant Physiol 168:532–548.
1359 https://doi.org/10.1104/pp.15.00470
1360 Ro DK, Bohlmann J (2006) Diterpene resin acid biosynthesis in loblolly pine (*Pinus taeda*): functional
1361 characterization of abietadiene/levopimaradiene synthase (PtTPS-LAS) cDNA and sub- cellular targeting of
1362 PtTPS-LAS and abietadienol/abietadienal oxidase (PtAO, CYP720B1). Phytochemistry 67:1572–1578.
1363 https://doi.org/10.1016/j.phytochem.2006.01.011
1364 Ro DK, Arimura GI, Lau SY, Piers E, Bohlmann J (2005) Loblolly pine abietadienol/abietadienal oxidase
1365 PtAO (CYP720B1) is a multifunctional, multisubstrate cytochrome P450 monooxygenase. Proc Natl Acad
1366 Sci USA 102:8060–8065. https://doi. org/10.1073/pnas.0500825102
1367 Rodríguez-Concepción M (2006) Early steps in isoprenoid biosynthesis: multilevel regulation of the supply
1368 of common precursors in plant cells. Phytochem Rev 5:1–15. https://doi.org/10.1007/ s11101-005-3130-4
1369 Rodríguez-Concepción M, Boronat A (2015) Breaking new ground in the regulation of the early steps of
1370 plant isoprenoid biosynthesis. Curr Opin Plant Biol 25:17–22. https://doi.org/10.1016/j. pbi.2015.04.001
1371 Rodríguez-Concepción M, Fores O, Martinez-Garcia JF, Gonzalez V, Phillips MA, Ferrer A, Boronat A
1372 (2004) Distinct light-mediated pathways regulate the biosynthesis and exchange of isoprenoid precursors
1373 during *Arabidopsis* seedling development. Plant Cell 16:144–156. https://doi.org/10.1105/tpc.016204
1374 Sallaud C, Rontein D, Onillon S, Jabès F, Duffé P, Giacalone C et al (2009) A novel pathway for

sesquiterpene biosynthesis from Z, Z-farnesyl pyrophosphate in the wild tomato *Solanum habrochaites*. Plant Cell 21:301–317. https://doi.org/10.1105/ tpc.107.057885

Schilmiller AL, Schauvinhold I, Larson M, Xu R, Charbonneau AL, Schmidt A et al (2009) Monoterpenes in the glandular trichomes of tomato are synthesized from a neryl diphosphate precursor rather than geranyl diphosphate. Proc Natl Acad Sci USA 106:10865–10870. https://doi.org/10.1073/pnas.0904113106

Schnitzler JP, Louis S, Behnke K, Loivamäki M (2010) Poplar volatiles—biosynthesis, regulation and (eco) physiology of isoprene and stress-induced isoprenoids. Plant Biol 12:302–316. https:// doi.org/10.1111/j.1438-8677.2009.00284.x

Shalev TJ, Yuen MM, Gesell A, Yuen A, Russell JH, Bohlmann J (2018) An annotated transcriptome of highly inbred *Thuja pli- cata* (Cupressaceae) and its utility for gene discovery of terpenoid biosynthesis and conifer defense. Tree Genet Genomes 14:35. https://doi.org/10.1007/s11295-018-1248-y

Sharkey TD, Yeh S (2001) Isoprene emission from plants. Annu Rev Plant Biol 52:407–436. https://doi.org/10.1146/annurev.arpla nt.52.1.407

Sharkey TD, Gray DW, Pell HK, Breneman SR, Topper L (2013) Isoprene synthase genes form a monophyletic clade of acyclic terpene synthases in the tps-b terpene synthase family. Evolution 67:1026–1040. https://doi.org/10.1111/evo.12013

Shi M, Luo X, Ju G, Yu X, Hao X, Huang Q, Xiao J, Cui L, Kai G (2014) Increased accumulation of the cardio-cerebrovascular disease treatment drug Tanshinone in *Salvia miltiorrhiza* hairy roots by the enzymes 3-hydroxy-3-methylglutaryl CoA reductase and 1-deoxy-D-xylulose 5-phosphate reductoisomerase. Funct Integr Genomics 14:603–615. https://doi.org/10.1007/s1014 2-014-0385-0

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

Surmacz L, Swiezewska E (2011) Biochemical and biophysical research communications polyisoprenoids—secondary metabolites or physiologically important superlipids? Biochem Biophys Res Commun 407:627–632. https://doi.org/10.1016/j. bbrc.2011.03.059

Takahashi S, Koyama T (2006) Structure and function of *cis*-prenyl chain elongating enzymes. Chem Rec 6:194–205. https://doi. org/10.1002/tcr.20083

Tarshis LC, Proteau PJ, Kellogg BA, Sacchettini JC, Poulter CD (1996) Regulation of product chain length by isoprenyl diphosphate synthases. Proc Natl Acad Sci USA 93:15018–15023. https://doi. org/10.1073/pnas.93.26.15018

Thimmappa R, Geisler K, Louveau T, Maille PO, Osbourn A (2014) Triterpene biosynthesis in plants. Annu Rev Plant Biol 65:225–257. https://doi.org/10.1146/annurev-arplant-050312-120229

Tholl D (2015) Biosynthesis and biological functions of terpenoids in plants. Biotechnology of isoprenoids.

Springer, Cham, pp 63–106

Tholl D, Lee S (2011) Terpene specialized metabolism in *Arabidopsis thaliana*. Arabidopsis Book 9:e0143. https://doi.org/10.1043/ tab.0143

Townsend BJ, Poole A, Blake CJ, Llewellyn DJ (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 Physiol 138:516–528. https://doi.org/10.1104/pp.104.056010

Trapp SC, Croteau RB (2001) Genomic organization of plant terpene synthases and molecular evolutionary implications. Genetics 158:811–832

Vannice JC, Skaff DA, Keightley A, Addo JK, Wyckoff GJ, Miziorko HM (2014) Identification in *Haloferax volcanii* of phosphomevalonate decarboxylase and isopentenyl phosphate kinase as catalysts of the terminal enzyme reactions in an archaeal alternate mevalonate pathway. J Bacteriol 196:1055–1063. https://doi.org/10.1128/JB.01230-13

Vaughan MM, Wang Q, Webster FX, Kiemle D, Hong YJ, Tantillo DJ et al (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. Plant Cell 25:1108–1125. https://doi.org/10.1105/tpc.112.100057

Velikova V, Ghirardo A, Vanzo E, Merl J, Hauck SM, Schnitzler JP (2014) Genetic manipulation of isoprene emissions in poplar plants remodels the chloroplast proteome. J Proteome Res 13:2005–2018

Vranova E, Coman D, Gruissem W (2013) Network analysis of the MVA and MEP pathways for isoprenoid synthesis. Annu Rev Plant Biol 64:665–700. https://doi.org/10.1146/annurev-arpla nt-050312-120116

Warren RL, Keeling CI, Yuen MM, Raymond A, Taylor GA, Vandervalk BP et al (2015) Improved white spruce (*Picea glauca*) genome assemblies and annotation of large gene families of conifer terpenoid and phenolic defense metabolism. Plant J 83:189–212. https://doi.org/10.1111/tpj.12886

Whittington DA, Wise ML, Urbansky M, Coates RM, Croteau RB, Christianson DW (2002) Bornyl diphosphate synthase: structure and strategy for carbocation manipulation by a terpenoid cyclase. Proc Natl Acad Sci USA 99:15375–15380. https://doi. org/10.1073/pnas.232591099

Wildung MR, Croteau R (1996) A cDNA clone for taxadiene synthase, the diterpene cyclase that catalyzes the committed step of taxol biosynthesis. J Biol Chem 27:9201–9204. https://doi. org/10.1074/jbc.271.16.9201

Wright GA, Schiestl FP (2009) The evolution of floral scent: the influence of olfactory learning by insect pollinators on the honest signalling of floral rewards. Funct Ecol 23:841–851. https://doi. org/10.1111/j.1365-2435.2009.01627.x

Xu M, Wilderman PR, Morrone D, Xu J, Roy A, Margis-Pinheiro M et al (2007) Functional

characterization of the rice kaurene synthase-like gene family. Phytochemistry 68:312–326. https://doi.org/10.1016/j.phytochem.2006.10.016

Xu M, Galhano R, Wiemann P, Bueno E, Tiernan M, Wu W, Peters RJ et al (2012) Genetic evidence for natural product-mediated plant–plant allelopathy in rice (*Oryza sativa*). N Phytol 193:570–575. https://doi.org/10.1111/j.1469-8137.2011.04005.x

Zerbe P, Chiang A, Yuen M, Hamberger B, Draper JA, Bohlmann J et al (2012) Bifunctional *cis*-abienol synthase from *Abies balsamea* discovered by transcriptome sequencing and its implications for diterpenoid fragrance production. J Biol Chem 287:12121–12131. https://doi.org/10.1074/jbc.M111.317669

Zerbe P, Hamberger B, Yuen MM, Chiang A, Sandhu HK, Madilao LL et al (2013) Gene discovery of modular diterpene metabolism in nonmodel systems. Plant Physiol 162:1073–1091. https://doi.org/10.1104/pp.113.218347

Zhou K, Gao Y, Hoy JA, Mann FM, Honzatko RB, Peters RJ (2012) Insights into diterpene cyclization from structure of bifunctional abietadiene synthase from *Abies grandis*. J Biol Chem 287:6840–6850. https://doi.org/10.1074/jbc.M111.337592

Zi J, Matsuba Y, Hong YJ, Jackson AJ, Tantillo DJ, Pichersky E, Peters RJ (2014) Biosynthesis of lycosantalonol, a *cis*-prenyl derived diterpenoid. J Am Chem Soc 136:16951–16953. https://doi.org/10.1021/ja508477e

Zulak KG, Bohlmann J (2010) Terpenoid biosynthesis and specialized vascular cells of conifer defense. J Integr Plant Biol 52:86–97. https://doi.org/10.1111/j.1744-7909.2010.00910.x

Table 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 retrived 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.

Table 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

Table 3 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. 6)

Species	Function	Abbreviation	Accession mRNA sequence	ORF (bp)	Accession protein sequence	Amino acids

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

Table 3 (continued)

Species	Function	Abbreviation	Accession mRNA sequence	ORF (bp)	Accession protein sequence	Amino acids
	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*	Terpene 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 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

1509

Table 3 (continued)

Species	Function	Abbreviation	Accession mRNA sequence	ORF (bp)	Accession protein sequence	Amino acids
Pinus tabuliformis	*ent*-Copalyl diphosphate synthase	Pta CPS1	KJ158966	2391	AHW42450	796
	entkaurene 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

The *ent*-kaurene synthase from the moss *Physcomitrella patens* was included as outgroup (continue on the next page)

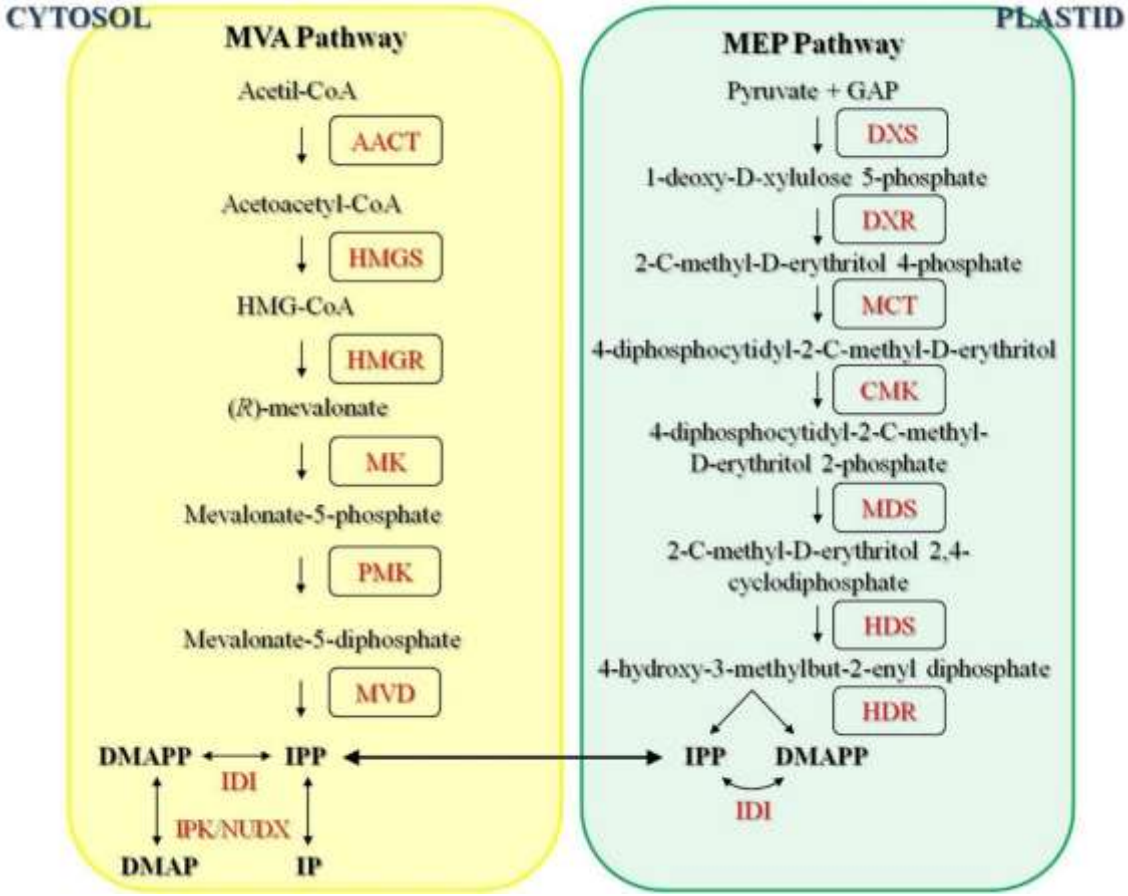

Fig. 1 The two biochemical pathways leading to isopentenyl diphos- phate (IPP), the basic unit for terpenoid biosynthesis in plants. Acro- nyms are in red for enzymes and in black for metabolites. *AACT* acetoacetyl-CoA thiolase, *CMK* 4-diphosphocytidyl-2-C-methyl- D-erythritol kinase, *DMAP* dimethylallyl monophosphate, *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, *IP* isopentenyl monophosphate, *IPK* isopentenyl kinase,*MCT* 4-diphosphocytidyl-2-C-methyl-D-erythritol synthase, *MDS* 2-C-methyl-D-erythritol 2,4-cyclodiphosphate synthase, *MEP* methy- lerythritol phosphate, *MK* mevalonate kinase, *MVA* mevalonic acid, *MVD* mevalonate diphosphate decarboxylase, *NUDX* Nudix hydro- lase, *PMK* phosphomevalonate kinase. The double arrow denotes the cross talk between cytosol and plastids based on the exchange of IPP and possibly of $C_{10–20}$ prenyl diphosphate intermediates (GPP, FPP, and GGPP). The mutual conversion of IPP and DMAPP with their respective monophosphate products IP and DMAP by IP kinase (IPK) and Nudix hydrolase (NUDX) enzymes can also affect the pathway flux in terpenoid metabolism. Adapted from Tholl (2015) and Abbas et al. (2017)

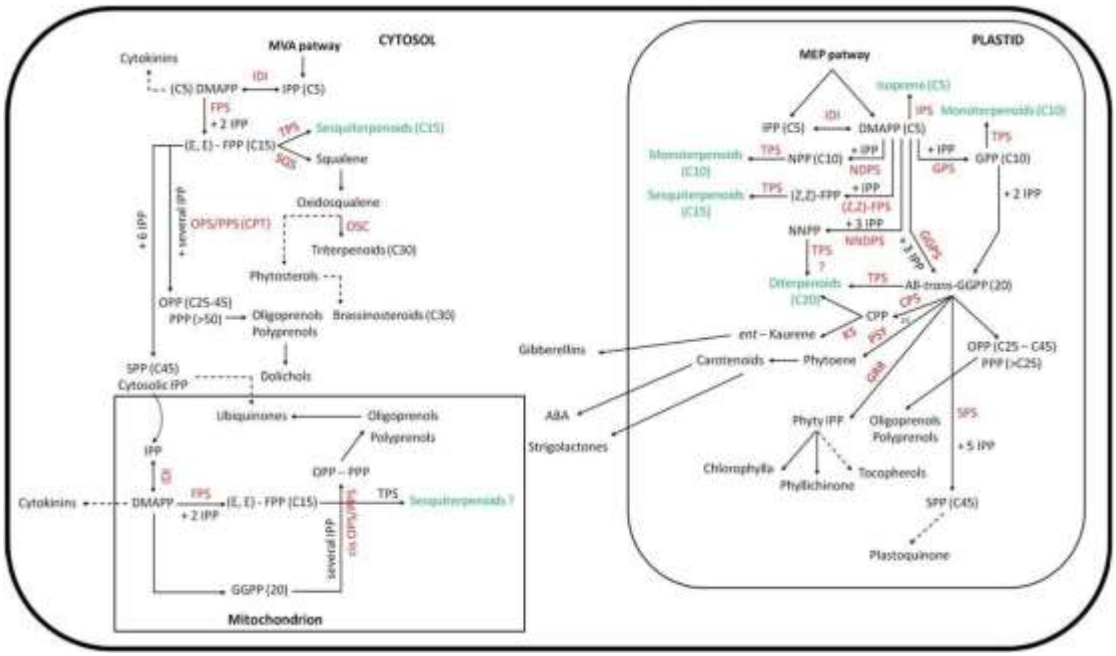

Fig. 2 Terpenoid biosynthetic pathways and their subcellular localization. Enzymes are marked in red and specialized terpenoids are marked in green; all other intermediates and terpenoid end products are in black. Solid and dashed arrows indicate single and multiple enzymatic steps, respectively. ABA abscisic acid, CPS ent-copalyl diphosphate synthase, CPT cis-prenyltransferase, 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)

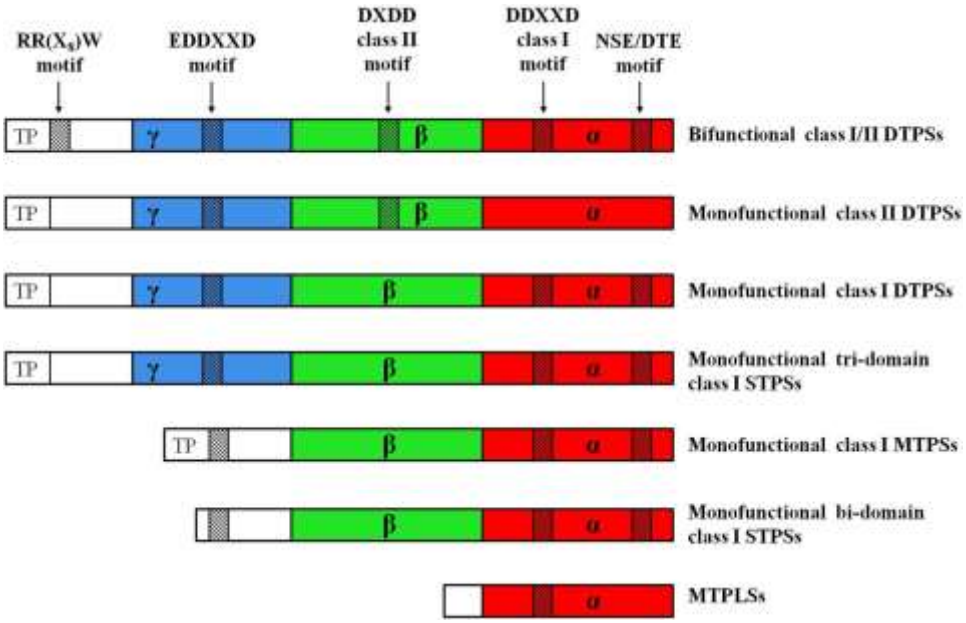

Fig. 3 Structural features of plant monoterpene- (MTPSs), diterpene- (DTPSs), sesquiterpene (STPSs) synthases, and of microbial terpene synthase-like proteins (MTPSLs) based on the combination of the γ, β, and α domains and the presence of distinctive aminoacidic motifs. The two highly conserved aspartate-rich 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(X_8)W" motif are also shown

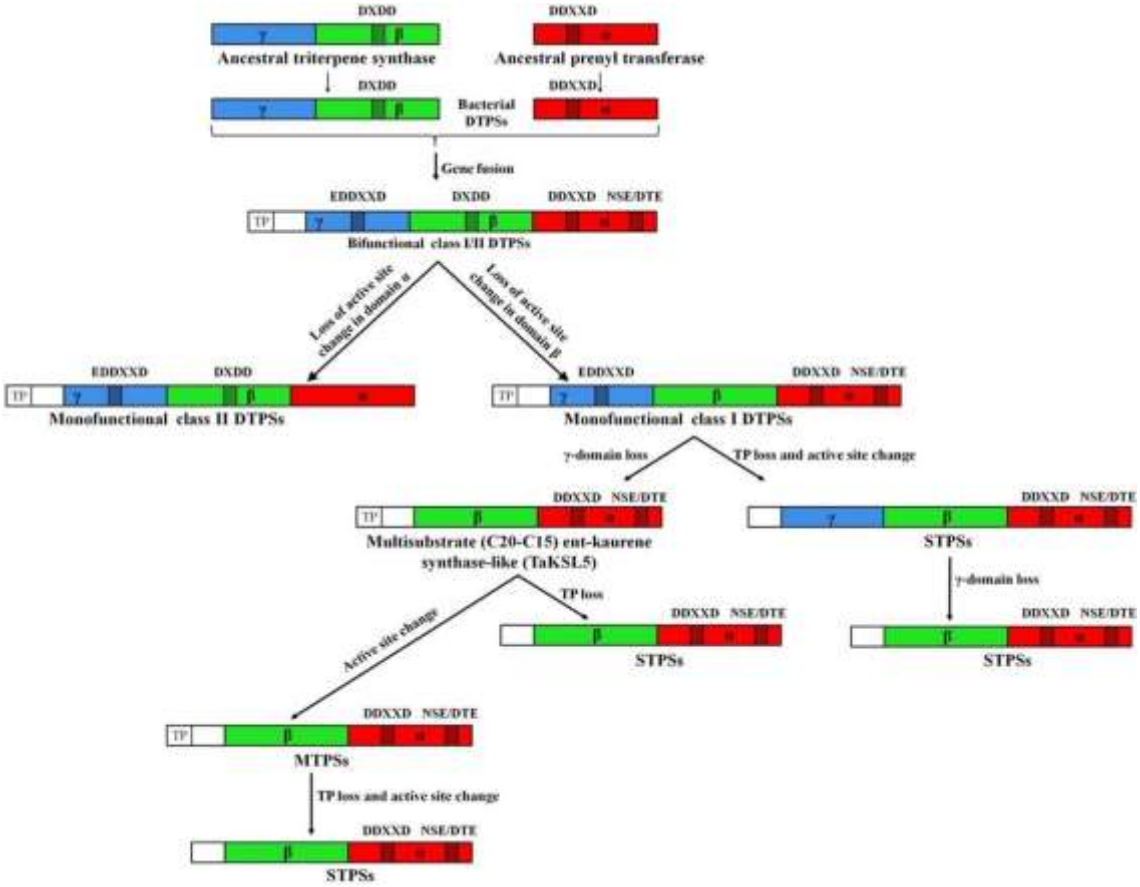

Fig. 4 Proposed model for the structural evolution of diterpene synthases (DTPSs) (Cao et al. 2010; Gao et al. 2012) and hypothesis on evolution of sesquiterpene synthases (STPSs) and monoterpene synthases (MTPSs) according to two potential routes based on the analysis of multisubstrate enzymes (Pazouki and Niinemets 2016)

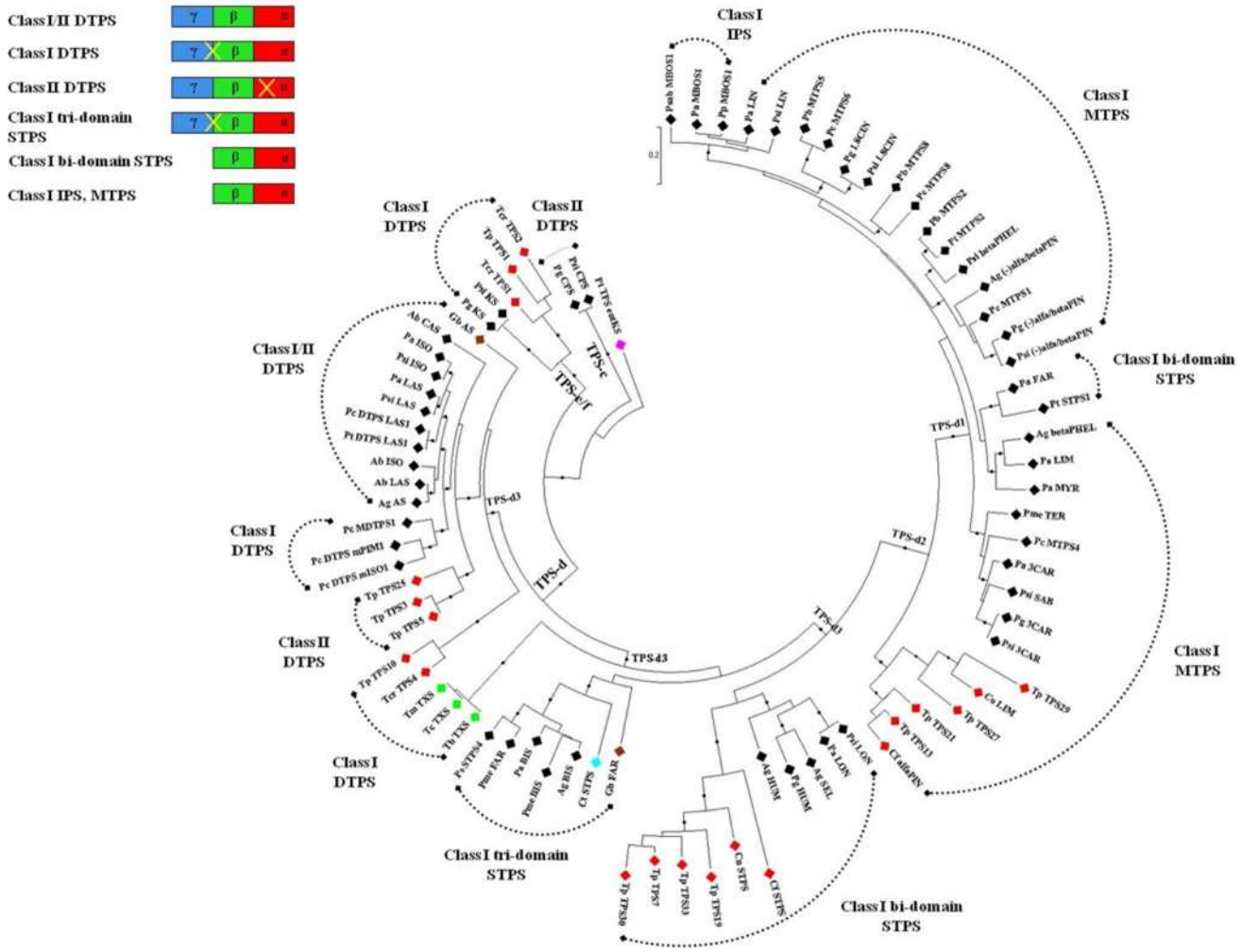

Fig. 5 Phylogenetic tree of terpene synthases (TPSs) in gymno- sperms: Cupressaceae (red diamonds), Pinaceae (black), Taxaceae (green), Ginkgoaceae (brown) and Cycadaceae (heavenly). The *Physcomitrella patens ent*-kaurene synthase (PtTPS-entKS; vio- let diamond) was used to root the tree. Branches indicated 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 groups within the TPS-d3 sub- family. For acronyms denoting plants species, see Table S1. *DTPSs* diterpene synthases, *MTPSs* monoterpene synthases, *STPSs* sesquiterpene synthases

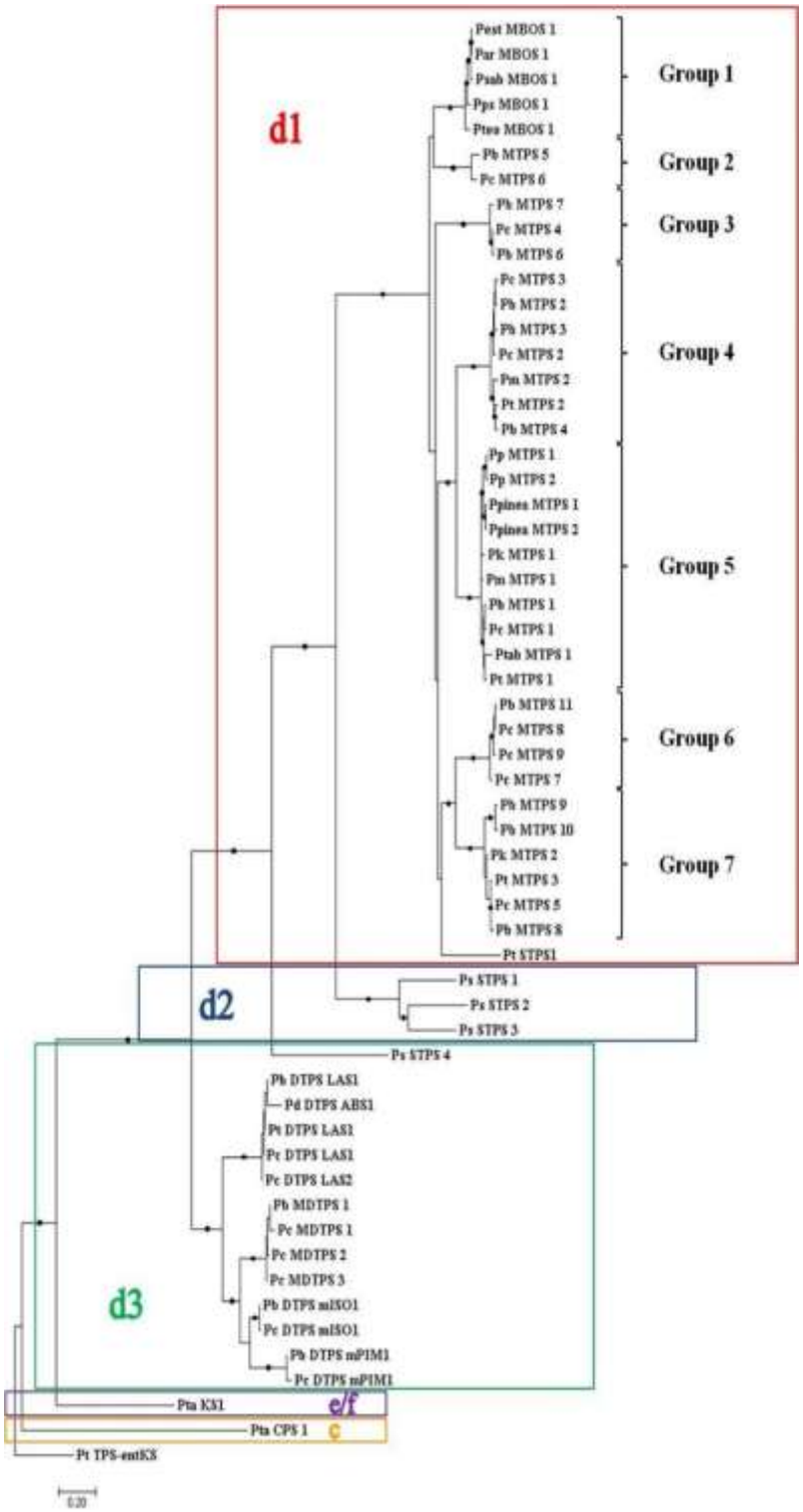

Fig. 6 Phylogenetic tree of the deduced amino acid sequences of terpene synthases (TPSs) in *Pinus* spp. The *Physcomitrella patens ent*-kaurene synthase (Pt TPS-entKS) was used to root the tree. Branches indicated with dots represent bootstrap support more than 80% (100 repetitions). Colour rectangles denote TPS subfamilies and groups within the TPS-d subfamily. For acronyms denoting plants species, see Table 3. *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

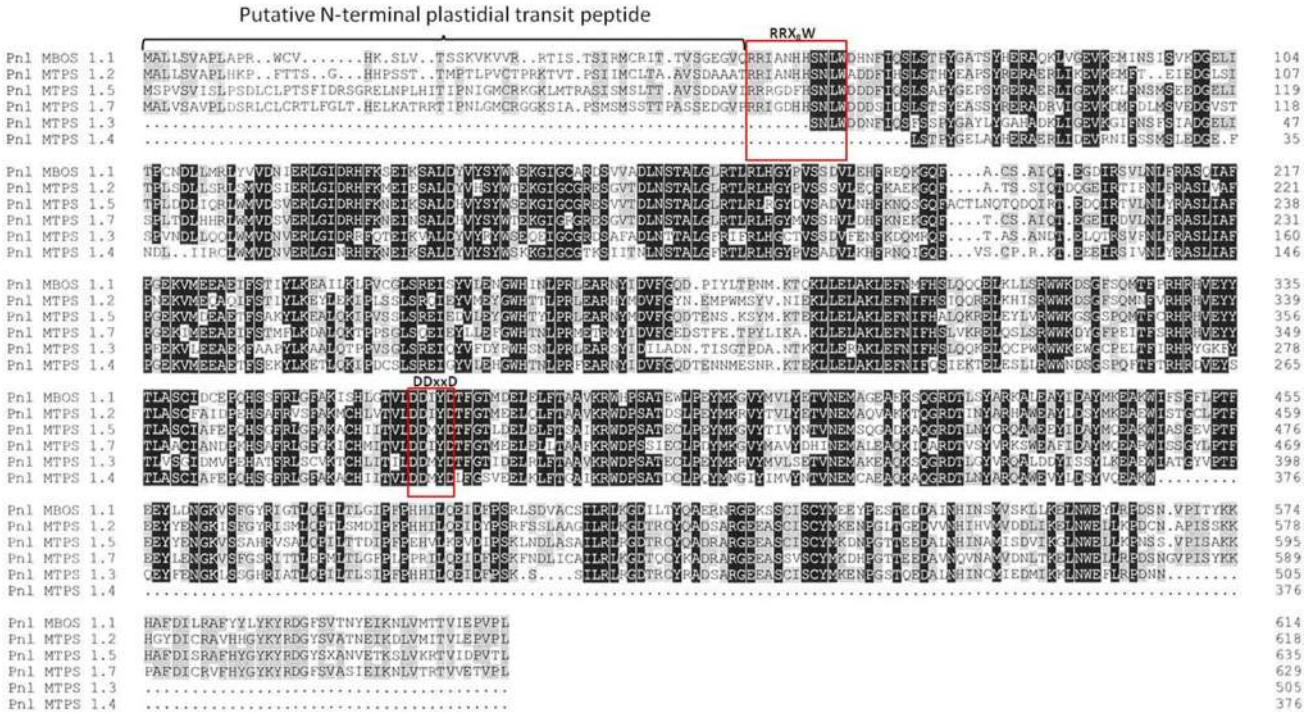

Fig. 7 Alignment of deduced amino acid sequences of full-length cDNAs (Pnl MBOS-1.1, Pnl MTPS-1.2, -1.5 and -1.7) and partial genomic and cDNA sequences (Pnl MTPS-1.3 and-1.4) isolated from *Pinus nigra* subsp. *laricio* needles. The black-shaded residues are highly conserved ones; the grey-shaded residues 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. *MBOS* 2-methyl-3-buten-2-ol synthase, *MTPS* monoterpene synthase

Manuscript title: On the evolution and functional diversity of terpene synthases in the *Pinus* species: a review.

Journal name: Journal of Molecular Evolution

Author names: Enrica Alicandri[1] (e.alicandri@gmail.com), Anna Rita Paolacci[2] (arpaolacci@unitus.it), Samson Osadolor[2] (soooct24@gmail.com), Agostino Sorgonà[1] (asorgona@unirc.it), Maurizio Badiani[1] (mbadiani@unirc.it), Mario Ciaffi[2*] (ciaffi@unitus.it)

Affiliation:
[1]Dipartimento di Agraria, Università Mediterranea di Reggio Calabria, Loc. Feo di Vito, I-89129 Reggio Calabria, Italy
[2]Dipartimento per la Innovazione nei Sistemi Biologici, Agroalimentari e Forestali, Università della Tuscia, Via S. Camillo De Lellis, s.n.c, I-01100 Viterbo, Italy

* Corresponding author:
Prof. Mario Ciaffi
Dipartimento per la Innovazione nei Sistemi Biologici, Agroalimentari e Forestali, Università della Tuscia, Via S. Camillo De Lellis, s.n.c, I-01100 Viterbo, Italy.
Email address: ciaffi@unitus.it, Tel: +39 0761 357 424, Fax: +39 0761 357 242.

The following Supplementary Material is available for the aforementioned manuscript in the present additional file 6:

Experimental procedures

Experimental procedures

1) Identification of terpene synthase (TPS) gene sequences belonging to Pinus species

The putative full-length cDNA sequences for mono-, sesqui- and diterpene synthases (MTPSs, STPSs and DTPSs, respectively), and for *ent*-copalyl diphosphate- and *ent*-kaurene-synthases (CPS and KS) in *Pinus* species were identified by a BLAST search in the National Center for biotechnology Information (NCBI) database, using selected and functionally characterized TPSs from different conifer species (Table S2). The search was restricted to the TPS sequences in NCBI database that correspond to the taxid 3337 (*Pinus*). For each putative identified gene, the corresponding mRNA and protein sequences were retrieved (Table 3).

2) Phylogenetic analysis

Multiple sequence alignment of the TPS proteins was performed by Multiple Sequence Comparison by Log-Expectation (MUSCLE) algorithm (gap open, -2.9; gap extended, 0; hydrophobicity multiplier, 1.5; clustering method, upgmb) (Edgar, 2004), implemented in MEGAX (Kumar et al. 2018). *Physcomitrella patens ent*-kaurene synthase (Pt TPS-entKS, BAF61135) was also included in the analysis as outgroup. A phylogenetic tree was generated with the Maximum Likelihood method using MEGAX 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 modeled with a gamma distribution (shape parameter = 1). Reliability of the tree obtained was tested using bootstrapping with 100 replicates.

Four different phylogenetic trees were computed using amino acid sequences of: 1) 20 full-length TPS sequences from several Cupressaceae species and a representative set of 62 functionally characterized TPSs from species of the Pinaceae, Taxaceae, Ginkgoaceae and Cycadaceae families (Table S1 and Fig. 5); 2) 50 sequences of pine TPSs for specialized metabolism (32 MTPSs, 5 STPSs and 13 DTPSs), five selected pine 2-methyl-3-buten-2-ol synthases (MBOSs), and the two *P. tabuliformis* class-I (KS) and class-II (CPS) of gibberellin biosynthesis (Table 3 and Fig. 6); 3) 42 full-lenght MBOSs identified in different *Pinus* species (Table 3 and Fig. S1); 4) 37 MTPSs and MBOSs identified in different *Pinus* species (Table 3) and the six isolated from *P. laricio* (Fig. S3).

3) Plant material

The sampling of needles from *Pinus nigra* subsp. *laricio* was carried out on 10/10/2017 within the Bonis basin, near Acri (CS) in Calabria. This is an area located in the so-called “Sila Greca Cosentina”

with an extension of 139 hectares. The needle samples were collected from five individuals. For each plant, three needle samples were collected (3-5 g for each sample) on branches located between 2.5 and 4 m of height. The needle samples were immediately frozen in liquid nitrogen and stored at -80 °C until use for DNA and RNA isolation.

4) DNA extraction

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

5) RNA extraction and cDNA preparation

Total RNA was extracted from needles, following the method described by Paolacci et al. (2017). The RNA samples were treated with RNase-free DNase I (Promega, Madison, WI, USA), according to the manufacturer's protocol. Following digestion, nucleotides were removed from RNA using a G50 Sepharose buffer exchange column (Amersham, Pittsburgh, PA, USA). The RNA concentration and integrity were checked, using a NanoDrop ND-1000 spectrophotometer (Labtech, East Sussex, UK). Only RNA samples with a 260/280 ratio (an index of protein contamination) between 1.9 and 2.1, and a 260/230 ratio (an index of reagent contamination) greater than 2.0, were used for cDNA synthesis. The quality of RNA samples was also assessed by electrophoresis on 1% formaldehyde agarose gels. First-strand cDNA was synthesized from 3 μg of total RNA using Expand Reverse Transcriptase (Roche Diagnostics, Milano, Italy), according to the manufacturer's protocol, and the resulting cDNA was used for RT-PCR analyses.

6) Isolation of partial and full length cDNAs coding for MTPS in Pinus nigra subsp. laricio

Deduced amino acid and nucleotide sequences of pine MTPSs belonging to each of the seven identified phylogenetic groups in the TPS-d1 clade (Fig. 6) were aligned by using the MUSCLE algorithm in order to identify highly conserved regions among members of each group. The nucleotide sequences in the identified conserved regions for each group were then used to design specific primers for the isolation by RT-PCR of partial transcripts of orthologous genes in *P. laricio* (see below). Figure S4 schematically outlines the full-length cDNAs for six representative members of the seven phylogenetic groups in which the TPS-d1 clade can be subdivided, and the positions of

their specific forward and reverse primers; the complete list of the same primers is reported in Table S4.

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

By using such strategy, we were able to isolate and sequence partial MTPS transcripts of putative *P. laricio* orthologous genes belonging to five out of seven phylogenetic groups. Moreover, four partial *P. laricio* transcripts of Groups 1, 2, 5 and 7, were used as templates for 5' and 3' RACE (Rapid Amplification of cDNA Ends) extensions using the 5'/3' RACE kit from ROCHE following manufacturer's instructions. The sequences of RACE primers are reported in Table S4 and their positions indicated in Fig. S4.

7) Isolation of genomic sequence coding for P. laricio MTPS of phylogenetic Group 3

Putative orthologous genes for the phylogenetic TPS-d1 Group 3 were not found in the transcriptome (i.e. cDNA) of needles of *P. laricio*, despite extensive efforts to amplify by PCR cDNA fragments of these genes, suggesting that they were not expressed in the needles. To assess the presence of Group 3 genes within the *P. laricio* genome, we used the primers designed in conserved regions of pine members of the phylogenetic group 3 (Fig. S4 and Table S4) and the genomic DNA extracted from *P. laricio* needles as a template. The PCR reactions and conditions were the same as described in section 6.

8) Cloning and sequencing of cDNA, RACE and genomic amplification products

Samples (5-10 μL) of the amplification products of RACE, partial cDNA and genomic DNA were separated on 1.2 % (w/v) agarose gels and visualized under UV radiation after staining with ethidium bromide (0.001%) and analyzed using the UVITEC Essential V6 Gel Imaging and Documentation System (Cleaver Scientific, Rugby, United Kingdom). PCR products of expected size were excised from the gel, purified using the High Pure Purification kit (ROCHE) according to manufacturer's instructions, and cloned into the pGEM-T easy plasmid vector (Promega, Madison, WI, USA) following the manufacturer's instructions. Two independent PCR amplifications were performed for

each cDNA, genomic and RACE amplicon, their products were cloned and for each reaction multiple clones were sequenced. Plasmid DNA for sequencing reaction was prepared from 3 mL overnight cultures using a plasmid mini-prep kit (QIAGEN). Sequencing was performed by a private company (MWG, Biotech AG, Germany). Recombinant positive plasmids were sequenced on both strands by the ABI PRISM 377 capillary sequencer (PE Applied Biosystem) using an ABI Prism Dye Terminator sequencing kit (PE Applied Biosystem) and either vector or sequence specific primers. All sequences were analyzed by DNAMAN Sequence Analysis Software (Version 3, Lynnon Biosoft) and their homologies were scored using the BLASTX program (Altschul et al. 1997) through the NCBI GeneBank database. The software developed by Hesbsgaard et al. (1996) was used for the prediction of intron splice sites within the genomic sequence coding for *P. laricio* MTPS of phylogenetic Group 3 (Pnl_MTPS_1.3). The predicted protein sequences were analyzed by searching for conserved motifs in CDD (Conserved Domain Database in the NCBI) and SMART (Simple Modular Architecture Research Tool, EMBL, Universitat Heidelberg) databases; their subcellular locations were predicted by TargetP 1.1, ChloroP 1.1 and Predotar.

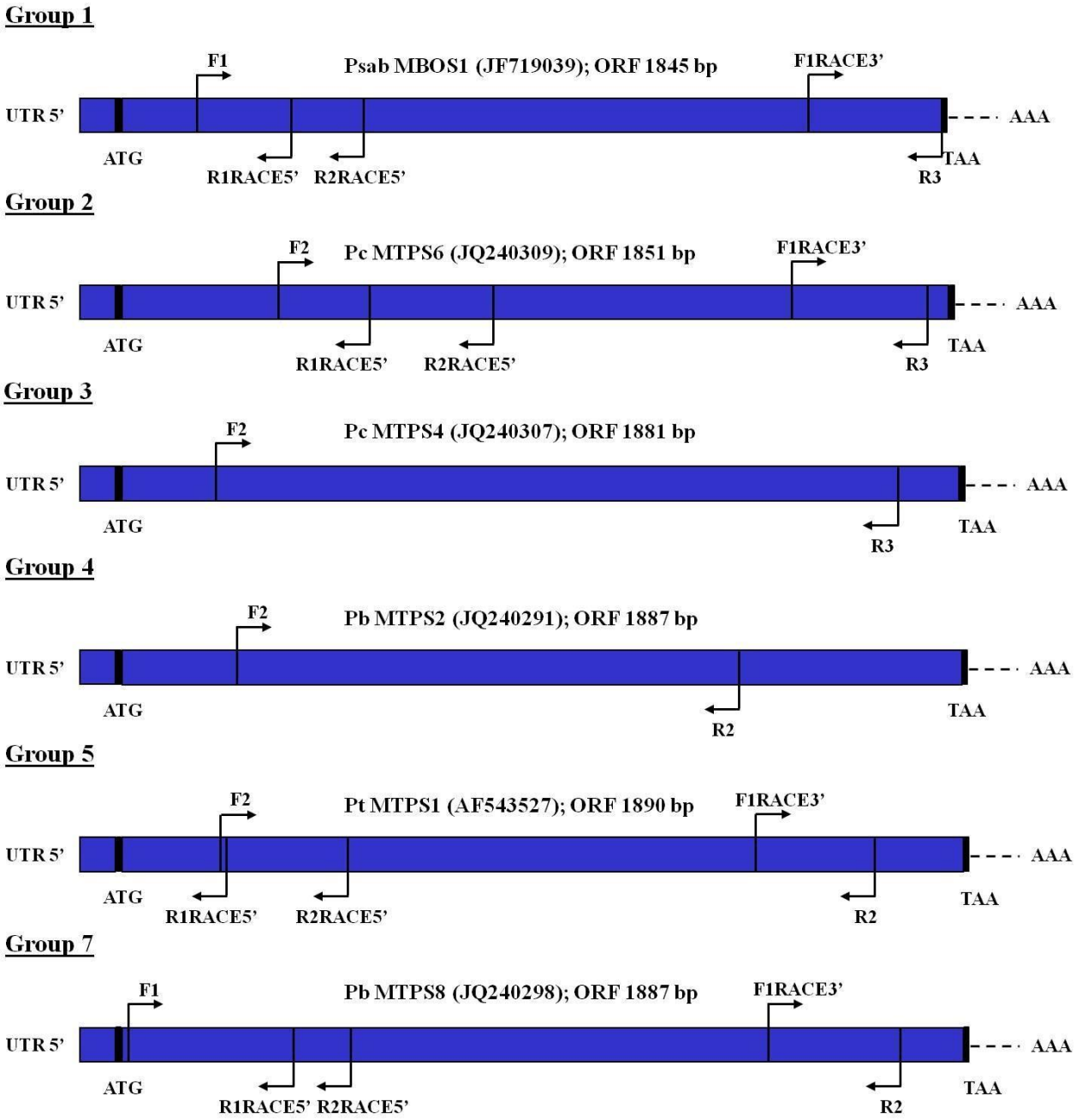

Figure S4. Schematic representation of the full-length cDNAs for six representative members of the seven phylogenetic groups of the TPS-d1 clade, in which are indicated the positions of the forward and reverse primers used for the isolation of the partial transcripts coding for orthologous genes in *P. nigra* subsp. *laricio*. The position of specific primers used for 5' and 3' RACE (Rapid Amplification of cDNA Ends) extensions of partial transcripts of *P. laricio* of Groups 1, 2, 5 and 7 are also reported.

Table S4. Forward and reverse primers used for the isolation of partial cDNAs coding for MBOS and MTPS in *P. nigra* subs. *laricio* and specific primers used for 5' and 3' RACE (Rapid Amplification of cDNA Ends) extensions of partial transcripts of *P. laricio* of Groups 1, 2, 5 and 7.

Phylogenetic groups	**Partial transcripts**	
	Forward primers	**Reverse primers**
1	Gr1-F1: 5'-CATCATTCCAACCTCTGGGA-3'	Gr1-R3: 5'-AGGCACAGGCTCAATGACG-3'
2	Gr2-F2: 5'-CCTTTCCATGGTCGATAGCA-3'	Gr2-R3: 5'-ATTGGTGGCGACGCTGTA-3'
3	Gr2-F2: 5'-TTCTAACCTGTTGGGACGACAA-3'	Gr3-R3: 5'-GCATTGTTGTCCGGTCTAAGA-3'
4	Gr4-F2: 5'-TTCTGTCAACGCCTTATGGG-3'	Gr4-R2: 5'-CGATCCACTTTGCTTCTTGC-3'
5	Gr5-F1: 5'-AACTTGCAAAGTTGGAGTTCAAC-3'	Gr5-R2: 5'-TTGATATGATTGAGAGCATCT-3'
7	Gr7-F1: 5'-GGTTTCTGCTGTCCCGTTGG-3'	Gr7-R2: 5'-CCGTTGCTGTCGGGTCTAAGTAAC-3'
	RACE	
	Race 5'	**Race 3'**
1	Gr1-R2: 5'-ATCTGAAGACACCGGGTATTCC-3' Gr1-R1: 5'-TCGATTCCCAAACGTTCAA-3'	Gr1-F1: 5'-TACCAGGCTGAGAGGAACCG-3'
2	Gr2-R2: 5'-GACGTCCATGTAATGCCTTGC-3' Gr2-R1: 5'-AACGCTTGAAGACACCGGG-3'	Gr2-F1: 5'-TTCAGTAGCTTGGCGGCTG-3'
5	Gr5-R2: 5'-GAGGGAAGCCCGATATAAATT-3' Gr5-R1: 5'-ATCATCGTCCCACAGGTTGGAAT-3'	Gr5-F1: 5'-ACTATGAGAACGGGAAAGTTAG-3'
7	Gr7-R2: 5'-CCACTTTCTCTCCCACGTCC-3' Gr7-R1: 5'-CGGTGATGGAGGTCAGTGA-3'	Gr7-F1: 5'-TCTCGCATAACCACGCTCG-3'

References

Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 25:3389-3402.

ChloroP 1.1[http://cbs.dtu.dk/services/ChloroP/]

Conserved domains and protein classification (CDD) [http://www.ncbi.nlm.nih.gov/Structure/cdd/cdd.shtml]

Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res 32:1792-1797.

Hebsgaard SM, Korning PG, Tolstrup N, Engelbrecht J, Rouze P, Brunak S (1996) Splice site prediction in *Arabidopsis thaliana* pre-mRNA by combining local and global sequence information. Nucleic Acids Res. 24:3439-3452. https://doi.org/10.1093/nar/24.17.3439

Kumar S, Stecher G, Li M, Knyaz C, Tamura K (2018) MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. Mol biol evoln 35:1547-1549. https://doi.org/10.1093/molbev/msy096

National Center for Biotechnology Information (NCBI) GeneBank database [http://www.ncbi.nlm.nih.gov/]

Paolacci AR, Catarcione G, Ederli L, Zadra C, Pasqualini S, Badiani M, Musetti R, Santi S, Ciaffi, M. (2017). Jasmonate-mediated defence responses, unlike salicylate-mediated responses, are involved in the recovery of grapevine from bois noir disease. BMC plant biol 17:118. https://doi.org/10.1186/s12870-017-1069-4

Predotar [https://urgi.versailles.inra.fr/predotar/predotar.html]

Simple Modular Architecture Research Tool (SMART) [http://smart.embl-heidelberg.de/]

TargetP 1.1 [http://www.cbs.dtu.dk/services/TargetP/]

Table S1. Full-length cDNA sequences retrieved from the NCBI database upon which the phylogenetic analysis of terpene synthases in gymnosperms was carried out (Fig. 5). The *ent*-kaurene synthase from the moss *Physcomitrella patens* was included as outgroup.

Species	Function	Abbreviation	Accession mRNA sequence	ORF (bp)	Accession protein sequence	Amino acids
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

References

Shalev TJ, Yuen MM, Gesell A, Yuen A, Russell JH, 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 14:35.

Table S2. 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

REFERENCES

Hall DE, Oa DSW, Zerbe P, Jancsik S, Quesada AL, Dullat H, Madilao LL (2013a) Evolution of conifer diterpene synthases: diterpene resin acid biosynthesis in lodgepole pine and jack pine involves monofunctional and bifunctional diterpene synthases. Plant physiology 161:600-616.

Hall DE, Yuen MM, Jancsik S, Quesada AL, Dullat HK, Li M, et al (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 13:80.

Köpke D, Schröder R, Fischer HM, Gershenzon J, Hilker M, Schmidt A (2008) Does egg deposition by herbivorous pine sawflies affect transcription of sesquiterpene synthases in pine? Planta 228:427-438.

Phillips MA, Wildung MR, Williams DC, Hyatt DC, 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. Arch Biochem Biophys 411:267-276.

Martin DM, 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 135:1908-1927.

Keeling CI, Dullat HK, Yuen M, Ralph SG, Jancsik S (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 152:1197-1208.

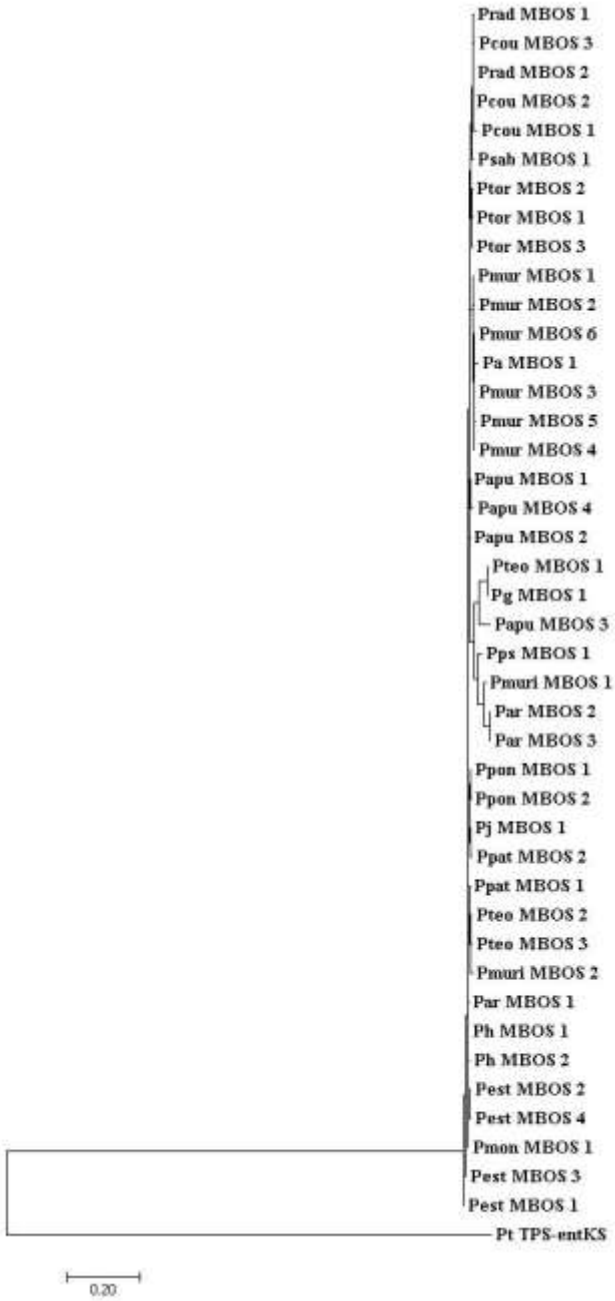

Fig. S1 Phylogenetic tree for the deduced amino acid sequences of the 2-methyl-3-buten-2-ol synthases (MBOSs) from Pinus species identified in NCBI database (Table 3). Physcomitrella patens ent-kaurene synthase (Pt TPS-entKS) was used to root the tree

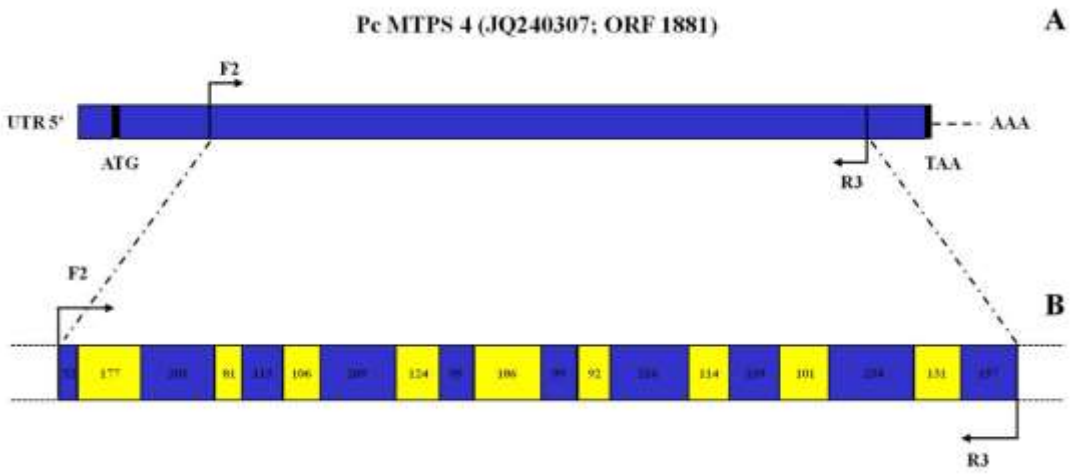

Fig. S2 Strategy adopted for the genomic amplification of a putative Pinus nigra subsp. laricio gene belonging to the phylogenetic TPS-d1 Group 3. a Schematic representation of the full-length cDNA of a representative member of the phylogenetic TPS-d1 Group 3 (PcMTPS4 from Pinus contorta, in the present case, see Fig. 6; Table 3) 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

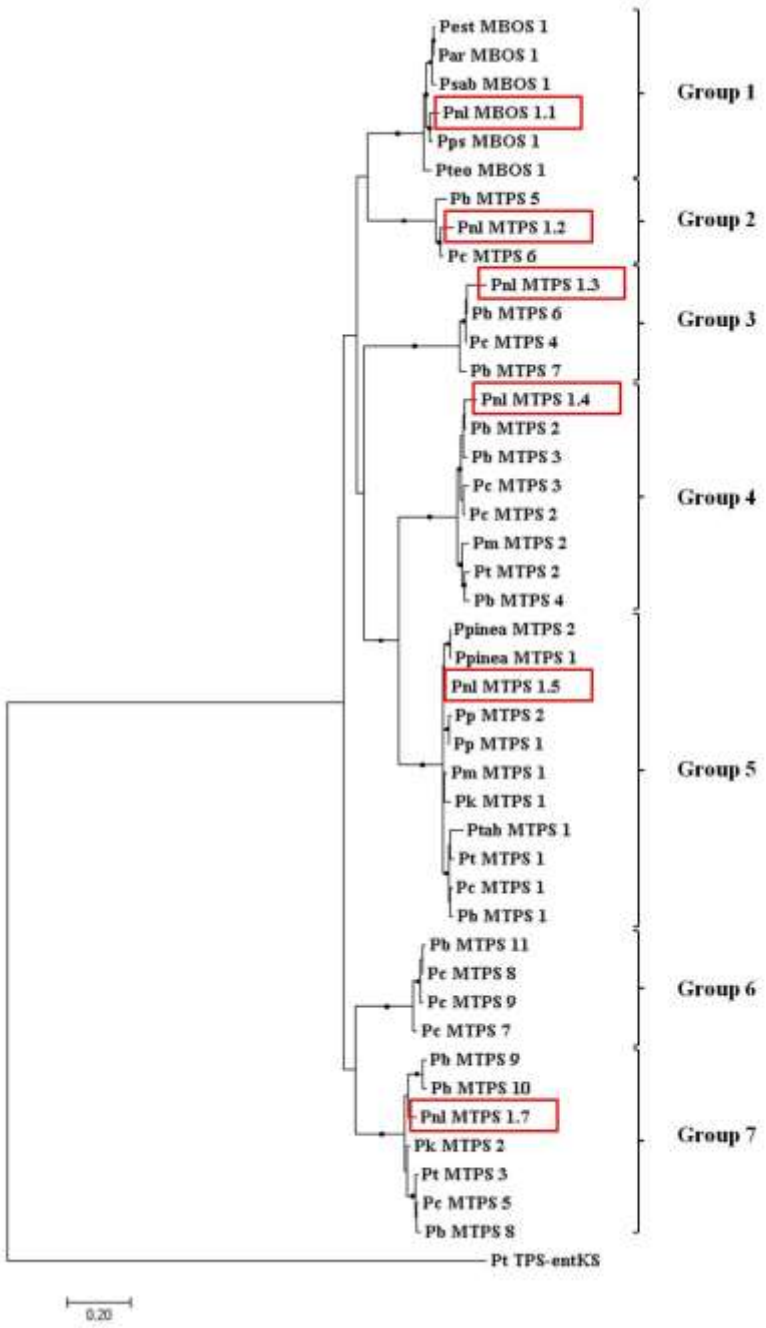

Fig. S3 Phylogenetic tree of the deduced amino acid sequences obtained by combining monoterpene synthases (MTPSs) and the five selected 2-methyl-3-buten-2-ol synthases (MBOSs) identified in different Pinus species (Fig. 6; Table 3) and the six sequences isolated from Pinus nigra subsp. laricio (outlined 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