Evolution, biosynthesis and regulation of diterpene resin acids in Norway spruce (*Picea abies*) specialized defense metabolism

Dissertation

To Fulfill the Requirements for the Degree of

"doctor rerum naturalium" (Dr. rer. nat.)

Submitted to the Council of the Faculty of Biological Sciences

of Friedrich Schiller University Jena

by Andrew John O'Donnell, M.Sc.

born on the 27th of August, 1986 in Michigan, USA

Date of Public Defense: 21.10.2021

Reviewers

Prof. Jonathan Gershenzon, Max Planck Institute for Chemical Ecology, Department of Biochemistry (Germany)

Prof. Dr. Günter Theißen, Friedrich-Schiller-University Jena, Matthias Schleiden Institute Department of Genetics (Germany)

Prof. Reuben J. Peters, Iowa State University, Roy J. Carver Department of Biochemistry, Biophysics and Molecular Biology (USA)

Table of Contents

List of Figures

List of Tables

Chapter I Introduction

"In three words I can sum up everything I've learned about life: it goes on."
-Robert Frost

I.I General

This thesis attempts to use classical processes and patterns of evolution as an explanation for (as opposed to a story about) chemical diversity in living land plant species. The featured plant species is the conifer *Picea abies* (Norway spruce), partially because this species is amenable to genetic manipulation. Norway spruce additionally has a published genome, is of economic importance, and produces abietane- and pimarane-type terpenoid compounds in high abundance that are unique to the Pinaceae family of conifers (Figure 1.1). Norway spruce is a large coniferous tree with a wide growth range across northern, central, and eastern Europe. This species is cultivated for ornamental purposes and provides a source of softwood lumber. Unfortunately, bark beetle outbreaks have devastated many hectares of Norway spruce forests in recent years (Biedermann et al., 2019). Thus, investigations into the evolutionary, molecular, and regulatory bases of Norway spruce terpenoids are crucial.

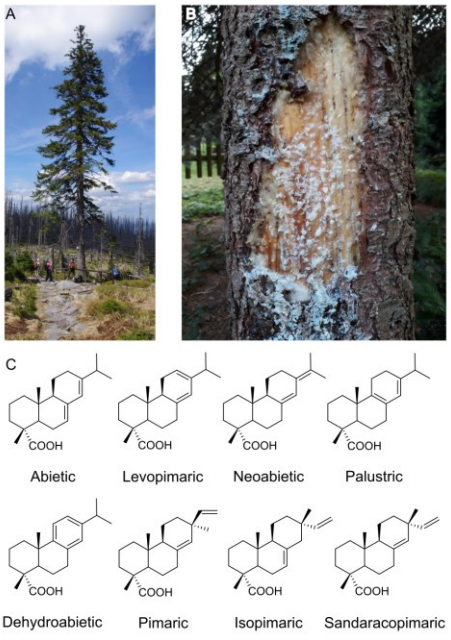


Figure 1.1: *Picea abies* (Norway spruce) produces a defensive mixture of diterpenoids constitutively and in response to tissue damage. A) Adult Norway spruce tree in a devastated population in the Bohemian Forest in Central Europe. B) Damaged Norway spruce bark exuding defensive resin. C) Eight known diterpene resin acids that form a component of Norway spruce resin and resin in other conifers of the family Pinaceae.

As will be presently reviewed, terpenoids come in volatile and non-volatile varieties, act as hormones in plant primary growth, and defend plants against enemies. Some classes of terpenoids that are restricted

to single taxonomic levels have evolved via novel biosynthetic steps between enzymes encoded by two distinct gene families, as will also be discussed. However, it remains unclear how enzymes encoded by separate multi-gene families evolve the ability to catalyze novel biochemical routes and to what extent specific routes could be predicted if life were to evolve again.

This introduction begins with an overview of terpene biosynthesis, provides examples of some of the many diverse lineage-specific specialized diterpenoid products in land plants, and finally details the evolutionary concepts that make this area of metabolism an exciting area of research for biochemists and evolutionary biologists alike. Finally, the specific questions and hypotheses, scope, aims and methods of this work will be addressed.

I.II Biosynthetic origins of diterpenoids in land plants

For the present purposes, three major stages of terpene biosynthesis will be considered, categorized by the families of enzymes responsible for carrying out each stage. The first stage consists of sequential condensation reactions of prenyl diphosphates carried out by *trans*-isoprenyl diphosphate synthases (IDSs) to form linear products of varying size (C_{10}, C_{15} and C_{20}) and double bond geometry (Figure 1.2). In the second stage, terpene synthases (TPSs) convert the phosphorylated products to cyclic and acyclic olefins with monoterpene (C_{10}), sesquiterpene (C_{15}) and diterpene (C_{20}) carbon skeletons. Alternatively, some plants emit the C_5 volatile isoprene (Monson et al., 2013), with use of isoprene synthases (ISs) that yield isoprene by dephosphorylation of IPP (Köksal et al., 2010; Miller et al., 2001). The third stage considered here is oxidation of diterpene products from the first two stages to biologically active diterpenoids via cytochrome P450 monooxygenases (P450s).

I.II.I Isoprenyl diphosphate biosynthesis

Isoprenyl diphosphate biosynthesis can be considered the first step in terpenoid metabolism as it provides substrates that are committed to diverging pathways involving specific downstream TPS reactions. Isopentenyl diphosphate (IPP) and dimethylallyl diphosphate (DMAPP) serve as 5-carbon universal precursors to terpenoids, and are derived from the 2-*C*-methyl-D-erythritol 4-phosphate (MEP) pathway in plastids (Vranová et al., 2012). The mevalonate (MEV) pathway in the cytosol supplies IPP to cytosolic enzymes. In both pathways, interconversion between IPP and DMAPP by isopentenyl diphosphate isomerases (IDIs) then provides isoprenyl diphosphates as substrates for further metabolism as reviewed by Vranová et al., and intracellular IPP and

DMAPP concentrations are proposed to be tightly regulated via IDI (Krause et al., 2020). Since the majority of plant TPSs involved in downstream biosynthetic steps in diterpene biosynthesis (the focus of the present work) encode N-terminal plastid-targeting peptides, the IPP utilized as substrate as discussed here is presumed to be derived from the MEP pathway unless otherwise noted. However, exchange of IPP from plastids to the cytosol has been reported (Hemmerlin et al., 2003; Laule et al., 2003), and such exchanges might be more widespread than previously thought (Yu & Utsumi, 2009).

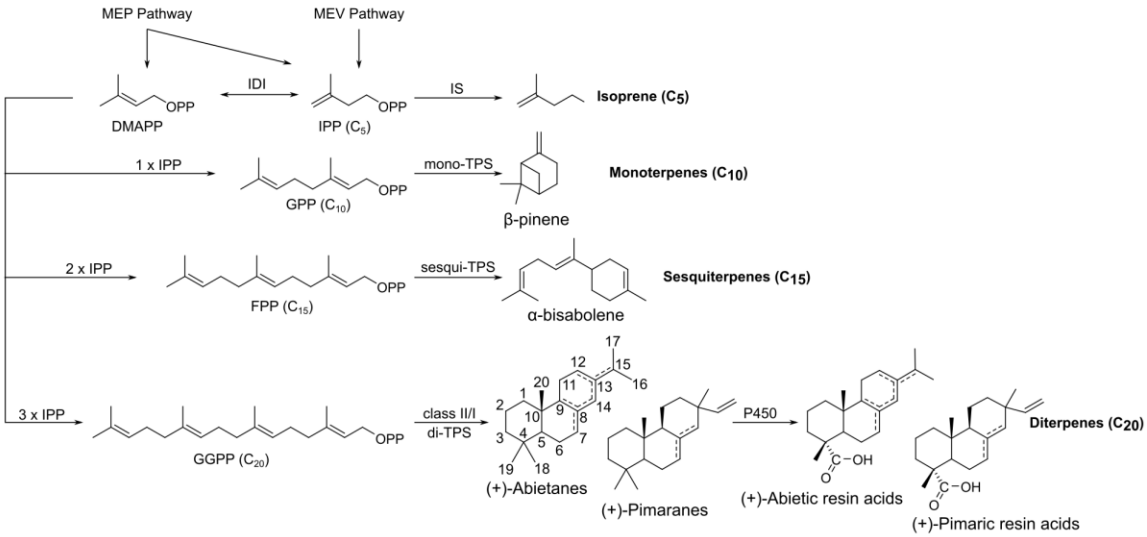


Figure 1.2: Terpenoids of various sizes are produced by sequential condensation reactions of isoprenyl diphosphates and are further converted by isoprene synthases, terpene synthases and cytochrome P450. di-TPS, diterpene synthase; DMAPP, dimethylallyl diphosphate; FPP, farnesyl diphosphate; GGPP, geranylgeranyl diphosphate; GPP, geranyl diphosphate; IPP, isopentenyl diphosphate; IS, isoprene synthase; MEP, 2-*C*-methyl-D-erythritol 4-phosphate; MEV, mevalonate; mono-TPS, monoterpene synthase; P450, cytochrome P450; sesqui-TPS, sesquiterpene synthase.

Plastid-localized plant IDSs bind DMAPP and condense it with one or more molecules of IPP to yield isoprenyl diphosphates that vary in length by five carbons (Nagel et al., 2019). PaIDS2 from Norway spruce, for example, converts DMAPP and one molecule of IPP to geranyl diphosphate (GPP; C_{10}) (Schmidt & Gershenzon, 2007b). PaIDS1 from the same species condenses DMAPP with one molecule of IPP to yield GPP and also sequentially adds two further IPP units to the growing allylic chain until the reaction terminates with the release of geranylgeranyl diphosphate (GGPP), a 20-carbon isoprenyl diphosphate (Schmidt et al., 2010). However, among Norway spruce IDSs, product specificity can be high, as is the case with

PaIDS5 and PaIDS6 which produce solely GGPP without release of any GPP or FPP intermediates *in vitro* (Schmidt & Gershenzon, 2007a). As an example of FPP biosynthesis, which occurs in the cytosol, Schmidt and Gershenzon (2007a) reported that heterologously expressed PaIDS4 catalyzes the condensation of DMAPP with one or two units of IPP to produce GPP and relatively larger quantities of FPP.

Condensation proceeds via generation of a carbocation intermediate following Mg^{2+}-dependent cleavage of the diphosphate group of IPP, followed by enzyme-mediated nucleophilic attack by deprotonated DMAPP (Liang et al., 2002). To achieve this, Mg^{2+} ions are coordinated by two aspartate-rich DDxxD motifs, with one motif coordinating DMAPP and the other stabilizing the negatively charged diphosphate group of IPP. Allylic chain length is determined by a variety of structural features, including bulky amino acids downstream of the DDxxD motif associated with IPP, resulting in steric hindrance of further condensation reactions (Liang et al., 2002; Nagel et al., 2019).

GPP (C_{10}), FPP (C_{15}), and GGPP (C_{20}) together comprise the major substrates for the terpene synthases, which can be classified as mono-, sesqui-, and di-TPSs, respectively. Isoprenyl diphosphates are also substrates for enzymes outside of the TPS families that will be discussed. For example, FPP is converted to squalene (C_{30}) by squalene synthases in steroid biosynthesis (Hou et al., 2018; Unland et al., 2018; B. Zhang et al., 2018).

I.II.II Overview of plant terpene synthase substrates and product outcomes

As plant TPSs have already received widespread attention in the literature and reaction products lead to an unprecedented range of plant terpenoid products, only a subset of TPS reactions will be mentioned here. This dissertation focuses primarily on diterpenes, derived from di-TPS catalysis with GGPP. The reader is therefore provided primarily with insight into the reaction products of plant di-TPSs. However, some attention will be paid to mono- and sesqui-TPSs because these enzymes share structural and mechanistic features in common with many of the di-TPSs.

Plant TPSs typically catalyze class I reactions that ionize the diphosphate group of isoprenyl diphosphates, yielding a carbocation intermediate that can deprotonate and rearrange to myriad terpene skeletal structures (Degenhardt et al., 2009; Karunanithi & Zerbe, 2019; Peters, 2010; Zhou & Pichersky, 2020). Class I TPSs possess a conserved C-terminal DDxxD motif that binds Mg^{2+} involved in the coordination and ionization of the diphosphate group, as for the IDSs

discussed previously. In some cases, carbocation intermediates that result from diphosphate lysis can be quenched with water, resulting in hydroxylated terpenes. Plant TPSs are often regarded as a "mid-size" family of plant enzymes (Chen et al., 2011), usually consisting of tens or scores of genomic copies. It is widely argued that actions by the multiple divergent families of plant TPSs are primarily responsible for the diversity of terpenoid carbon skeletons that ultimately lead to the >80,000 (and presumably bioactive) known terpenoid end products (Degenhardt et al., 2009; Zhou & Pichersky, 2020).

Plant mono-TPSs synthesize relatively small and volatile C_{10} cyclic and acyclic terpene hydrocarbons from GPP via class I diphosphate ionization (Degenhardt et al., 2009). For instance, PaTPS-Pin from Norway spruce converts GPP to β-pinene (Martin et al., 2004), a bicyclic volatile monoterpene with a characteristic pine-like odor. Acyclic monoterpenes can also be produced when carbocation intermediates are quenched with water, as described above for IDSs. When water quenching occurs, as is the case with PaTPS-Lin from Norway spruce (Martin et al., 2004), hydroxylated terpene products are produced. PaTPS-Lin thusly converts GPP to linalool, a monoterpene alcohol, as demonstrated by Martin and colleagues (2004). Although two mono-TPSs from Norway spruce were used here as examples, mono-TPSs are widely distributed among angiosperms and gymnosperms and in many cases appear to converge on the same products. For example, mono-TPSs catalyzing the formation of β-pinene have also been reported in *Citrus* spp. and *Artemisia annua*, which are angiosperms, the gymnosperm *Abies grandis*, and other plant species (Bohlmann et al., 1997; Lu et al., 2002; Lücker et al., 2002; Shimada et al., 2004).

Sesquiterpenes are similarly synthesized by class I reactions facilitated by sesqui-TPSs that utilize FPP as substrate (Degenhardt et al., 2009). The larger C_{15} isoprenyl diphosphate, as well as a higher degree of unsaturation in FPP relative to GPP, is generally implicated as a cause of the higher skeletal diversity of sesquiterpenes than monoterpenes. As the potential reaction products of mono- and sesqui-TPSs are too vast to review in detail here, attention will now be turned to the di-TPS class of plant terpene synthases.

Plant di-TPSs are split here into two further categories: di-TPSs that uniquely catalyze class II cyclization reactions with GGPP, yielding C_{20} compounds that retain the diphosphate group, and those that perform the typical class I reactions using the products of class II di-TPSs as substrates.

Class II reactions are mediated by an N-terminal DxDD motif that usually results in the generation of bicyclic diterpenes (Peters, 2010). Although additional class II reaction products have been identified, the most commonly formed by plant di-TPSs are (+)-copalyl diphosphate (also known as normal-CPP) and two stereoisomers thereof, *syn*-CPP and *ent*-CPP. Accordingly, most final diterpenoid products are built from one of these three stereoisomers of CPP as thoroughly outlined by Peters (2010).

The diphosphate groups of the various CPP stereoisomers are subsequently cleaved in the generation of diterpene hydrocarbon and alcohol terpenes by class I di-TPS reactions. Often, carbocation intermediates generated by class I reactions result in a second cyclization step to yield tricyclic products. The most common of these products are the kaurene-related terpenes, which are produced ubiquitously among plants by class I di-TPSs from *ent*-CPP and serve as intermediates in gibberellin biosynthesis (Zi et al., 2014). Whereas gibberellins are known hormones that regulate plant cell elongation and division (Hedden & Sponsel, 2015), many class I di-TPSs are responsible for producing diterpene olefins in response to environmental stressors that simulate pathogen or insect attack, as will be discussed further in a following section. Some class I di-TPSs have been identified that react with just one stereoisomer of CPP, while others react with multiple stereoisomers, indicating a role for broad substrate specificity in the evolutionary history of this enzyme class.

Interestingly, di-TPSs from nonseed plants and conifers from the family Pinaceae have been characterized that possess both class II and C-terminal class I aspartate-rich motifs. In the model moss *Physcomitrella patens*, for instance, a class II/I bifunctional di-TPS converts GGPP all the way to *ent*-kaurene via *ent*-CPP (Hayashi et al., 2006). Another example can be found in the nonseed plant *Selaginella moellendorffii*, which utilizes a bifunctional di-TPS in the conversion of GGPP to miltiradiene (Sugai et al., 2011). In the conifer *P. abies*, a bifunctional di-TPS (PaTPS-LAS) converts GGPP instead to 13-hydroxy-8,14-abietene via normal-CPP, which is a thermally unstable tertiary alcohol that can rearrange non-enzymatically to four abietane-type diterpene olefins (Keeling et al., 2011a). A comprehensive mechanistic study of a similar enzyme from the conifer *Abies grandis* concluded that these reactions occur via class II cyclization of GGPP, followed by diffusion of the bicyclic normal-CPP product to the class I active site where the diphosphate group is cleaved and conversion to the tricyclic product occurs (Peters et al., 2001).

Diversity among di-TPSs drives evolutionary novelty as these enzymes provide a wide range of substrates for downstream enzymes that further expand range of plant diterpenoid metabolites (Karunanithi & Zerbe, 2019). Together with the diverse diterpene olefins produced by di-TPSs, the actions of P450s are further responsible for the astounding expansion of chemical diversity among diterpenoids and many other metabolite classes across plant lineages.

I.II.III Diversity in cytochrome P450 reaction types

P450s comprise a superfamily of plant enzymes ranging from nearly 250 encoding genes in *Arabidopsis thaliana* to over 300 genes in *Populus trichocarpa*, *Glycine max* and *Oryza sativa* (Nelson & Werck-Reichhart, 2011). This plant super-family is widely known for NADPH-dependent oxidation of substrates yielding various oxygenated functional groups such as alcohols, aldehydes, carboxylic acids, ketones and lactones. Substrate specificities, sites and degrees of oxidation, and products vary widely among even closely related P450s and therefore often cannot be predicted by phylogenetic analyses or sequence identities alone.

Examples of broad plant P450 reaction types are abundant in brassinolide, gibberellin, and resin acid biosynthesis, some of which will be discussed in further detail in Chapter IV. To provide a few initial examples here, the CYP90 and CYP724 subfamilies of P450s in *A. thaliana* hydroxylates a carbon atom in the side chain of campesterol, a C_{30} steroid precursor to brassinolide (Ohnishi et al., 2006b). CYP85A1 from *Solanum lycopersicum* generates a ketone intermediate at a ring carbon of a different brassinolide intermediate. Alternatively, the closely related CYP85A3 generates the same ketone as CYP85A1 but further inserts an oxygen into the substrate ring, resulting in a lactone (Nomura et al., 2005). In gibberellin metabolism (Figure 1.3), CYP701A1 carboxylates a methyl group of *ent*-kaurene (C_{20}) to produce *ent*-kaurenoic acid, which is a precursor to gibberellic acid (Helliwell et al., 1999). CYP720Bs in conifers oxidize normally configured diterpene olefins at a similar methyl group to produce the resin acids shown in Figure 1.1.

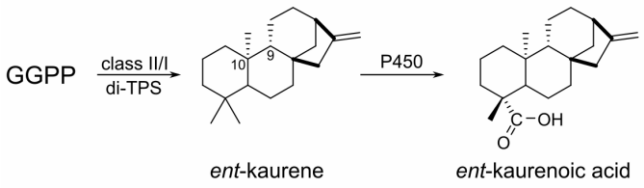


Figure 1.3: Plants ubiquitously use diterpene synthases to produce *ent*-kaurene from geranylgeranyl diphosphate (GGPP, C_{20}). Cytochrome P450s are responsible for oxidation of *ent*-kaurene to *ent*-kaurenoic acid, which is a precursor to gibberellins in plant primary metabolism. di-TPS, diterpene synthase; P450, cytochrome P450. Note the alternative configurations at carbons 9 and 10 (numbering based on (+)-abietane structure in Figure 1.2).

These reactions result in only a few of the hundreds of P450 products verified to date by *in vitro* assays. However, owing to the vast lineage-specific diversification of plant P450s, the characterized enzymes represent only a fraction of the thousands of different P450s present in the plant kingdom. An intriguing notion that will receive further attention later is that novel biosynthetic pathways that combine di-TPS and P450 reaction steps must have repeatedly evolved in order to complete the various diterpenoid pathways seen in modern-day plants. It should additionally be mentioned that such biochemical associations have been made possible by widespread primary metabolic repertoires across plants as found in gibberellin and brassinosteroid metabolic routes (Nelson & Werck-Reichhart, 2011; Zi et al., 2014). To illustrate these points, the divergent metabolic fates of several diterpenoids in land plants (determined by the concerted actions of di-TPS and P450 subfamilies) will be reviewed next. The pathways will also be framed in known or suspected ecological contexts when possible.

I.III Lineage-specific diversity and ecological roles of diterpenoids

Terpene research in plants from all major land plant lineages has revealed that diverse structures often occur at limited taxonomic levels and are even frequently limited to single species. It should also be noted that microbial-type terpene synthases (i.e. those terpene synthases that share closer phylogenetic relationships to enzymes of microbial origin than to the prototypical plant terpene synthases of focus here) have been found to have wide distribution in nonseed plants (Jia et al., 2016) and contribute to diterpenoid diversity along with P450 enzymes. However, a broad overview of pathways formed by the prototypical plant di-TPS and P450 enzymes will remain the focus here in order to set the stage for the evolutionary patterns and processes investigated in this work.

I.III.I Momilactones

OsKSL4 from *Oryza sativa* (rice) and HpDTC3 from the moss *Hypnum plumiformae* are specialized di-TPSs that convert GGPP and *syn*-CPP (respectively) to the labdane diterpene *syn*-pimaradiene (Okada et al., 2016; Otomo et al., 2004). At least in rice, a P450 member of the CYP99A subfamily is responsible for the first oxidative step of momilactone biosynthesis by converting a methyl group of *syn*-pimaradiene to a carboxyl function (Wang et al., 2011), while further oxidative steps by yet-unidentified P450s are implicated in the complete biosynthetic pathway (Schmelz et al., 2014). Concerning biological activity, rice seedlings were shown to dramatically stunt the hypocotyl and root development of nearby germinating *Lepidum sativum* and *Lactuca sativa* (garden cress and lettuce, respectively) (Kato-Noguchi et al., 2002). Momilactone B, which is exuded by *O. sativa* roots, was demonstrated to be responsible for this allelopathic activity when seeds were germinated on filter paper containing increasing concentrations of the compound. Interestingly, in the moss *Hypnum plumiformae*, the only other land plant reported to produce momilactones, these diterpene lactones appear to serve a similar allelopathic function against a broad range of seed and nonseed plants (Nozaki et al., 2007).

I.III.II Tanshinones

Members of the flowering plant family Lamiaceae are widely reported to produce miltiradiene (sometimes detected as dehydroabietadiene after non-enzymatic deprotonation), which is a normally configured tricyclic diterpene olefin synthesized from normal-CPP by the kaurene-synthase-like (KSL) family of di-TPSs (Božić et al., 2015; Brückner et al., 2014; Gao et al., 2009; Heskes et al., 2018; Jin et al., 2017; Zerbe et al., 2014). In *Sal*via *miltiorrhiza*, following rearrangement of miltiradiene to dehydroabietadiene, SmCYP76AH1 hydroxylates an aromatic carbon of dehydroabietadiene to yield a phenolic diterpene as a precursor to the tanshinones (Guo et al., 2013; Zi & Peters, 2013). A subsequent series of hydroxylations and both non-enzymatic and enzyme-mediated dehydration reactions by SmCYP76AH3 and SmCYP76AK1 from *S. miltiorrhiza* roots produces at least two precursors to distinct bioactive tanshinone products (Guo et al., 2016). Although an ecological benefit for tanshinone biosynthesis has not yet been conclusively identified, some evidence supports an effect of abiotic factors and links to geological distribution on *in planta* tanshinone abundance (Zhang et al., 2019; Zhang et al., 2018).

I.III.III Diterpene diversity in gymnosperms

Members of the gymnosperms provide further examples of how actions of specialized di-TPSs and P450s lead to diverse diterpenoid product sets that provide ecological advantages. Congeners of the conifer family Pinaceae often possess bifunctional di-TPS enzymes from the TPS-d subfamily that lead to normally configured diterpene products that are intermediates to diterpene resin acids. For instance, conversion of GGPP to hydroxy-8,14-abietene by PaTPS-LAS as mentioned above or to isopimaradiene by PaTPS-ISO in *P. abies* provides diterpene olefins as precursors to resin acids (Keeling et al., 2011a; Martin et al., 2004). Characterization of P450s in other *Picea* (spruce) members and *Pinus* spp. show that oxidized diterpenoids are produced by subsequent stereo-specific oxidation of a ring methyl group corresponding to C-18 in Figure 1.2 to a carboxylic acid via three successive oxidative steps catalyzed by a single enzyme (Geisler et al., 2016; Hamberger et al., 2011; Semiz & Sen, 2015).

Resin acids accumulate in constitutive resin ducts in needles and bark, and in induced bark resin ducts in species that can form traumatic tissue following damage or spraying with methyl jasmonate (MeJA) (Zulak & Bohlmann, 2010). Among conifers, the phylogenetic distribution of resin duct formation (both constitutive and inducible) appears to be associated with increased aggressiveness of invading bark beetles. For instance, conifers from the families Podocarpaceae and Taxaceae appear to lack structures dedicated to resin acid storage and also lack aggressive bark beetle attackers (Franceschi et al., 2005). While volatile mono- and sesquiterpenes are known as anti-feedants and toxic agents to adult beetles that feed on multiple Pinaceae species (*Ips typographus*, *Dendroctonus ponderosae*, and *Scolytus ventralis*) (Chiu et al., 2017; Faccoli et al., 2005; Raffa et al., 1985), a defensive role of the non-volatile diterpene resin acids against bark beetles seems to be less clear. However, abietic acid and isopimaric acid have been shown to strongly inhibit the germination rates of symbiotic fungi associated with the bark beetle *Ips pini*, which feeds on the North American pine species *Pinus resinosa* (Kopper et al., 2005). Diterpene resin acids are further thought to act as physical barriers to herbivorous insects by trapping, flushing out, and/or suffocating invaders (Franceschi et al., 2005; Keeling & Bohlmann, 2006).

Other examples of specialized diterpenoids from gymnosperms include lactones from the tree *Ginkgo biloba* (family Ginkgoaceae) and *Podocarpus* spp. from the family Podocarpaceae. While ecological roles for the lactones from *G. biloba*, known as ginkgolides, remain unknown, the

formation of these unique diterpenes is catalyzed by the actions of specialized di-TPSs and P450s. In *G. biloba*, a bifunctional di-TPS from the TPS-d family (GbTPS-LPS) converts GGPP to the diterpene olefin levopimaradiene (Schepmann et al., 2001). Biosynthesis appears to occur in the roots and involves P450s as implied by experiments with P450 inhibitors (Chen et al., 2015). Conifers from the family Podocarpaceae uniquely produce podocarpic acid, which is structurally similar to but distinct from the resin acids of the Pinaceae (Cox et al., 2007). Given the structural similarity and site of oxidation of podocarpic acid, this compound might be synthesized by TPS-d enzymes and P450s related to the 720B subfamily mentioned above in Pinaceae. However, the ecological roles and biosynthesis of this compound are yet to be determined.

I.IV Regulation of biosynthesis

Relatively little is known about how the formation of labdane-related diterpenoids is regulated in plants, but two conclusions can be drawn. First, those diterpenes presumed or demonstrated to be defensive, such as momilactones from *O. sativa* and the moss *H. plumaeforme*, are highly inducible by chemical stressors such as $CuCl_2$, chitosan, or MeJA (Okada et al., 2016; Wilderman et al., 2004). In studies of both species, the responsible di-TPSs were induced around 24 hours before accumulation of momilactones. In contrast, diterpene resin acids in MeJA-treated *P. abies* and *P. sitchensis* are induced after about two weeks following a 10 day post-induction period for associated di-TPSs and P450s (Hamberger et al., 2011; Martin et al., 2002), probably due to the requirement for complex anatomical changes associated with formation of traumatic resin ducts in this species (Franceschi et al., 2005). Secondly, despite successful RNA-interference of terminal pathway genes such as P450s in *P. sitchensis* (Hamberger et al., 2011), only minimal effects on diterpene abundance were seen, suggesting possible upstream control over pathway flux.

Indeed, it appears that gibberellic acid abundance in *A. thaliana* and *Nicotiana tabacum*, and tanshinones in *S. miltiorrhiza* can be controlled by upregulating IDSs upstream of both TPS and P450 enzymes (Shi et al., 2016; Tata et al., 2016). However, even when a known GGPP synthase such as PaIDS1 was overexpressed in *P. abies*, diterpene resin acid accumulation was unaffected, leading instead to an overabundance of geranylgeranyl fatty acid esters (Nagel et al., 2014).

Although the regulatory mechanisms of diterpene metabolism and its responses to environmental factors would be expected to vary between species and between growth and defense metabolism, the results of these experiments are surprising. A better understanding of the genetic underpinnings of diterpene accumulation is clearly needed. Nevertheless, the repeated evolution of unique pathways involving specialized di-TPS and P450 enzymes requires complementary enzymatic coupling between these two gene families; that is, novel di-TPS products must have existed within the same timeframe as recruitment of compatible P450s in order for novel beneficial pathways to evolve. This raises several evolutionary paradoxes that will be addressed in the next section.

I.V Evolutionary concepts of this thesis

In this thesis, evolution will be regarded on four conceptual levels. The first level is that of the number of genes involved in the evolutionary process and the second level concerns gene maintenance following duplication. A third level refers to the assembly of genes and their encoded proteins into biochemical pathways, while a fourth level covers the interaction between the sites of an individual biosynthetic protein, referred to as epistasis. Before these evolutionary levels are reviewed, however, definitions of orthology and parology are given that are used throughout this work.

I.V.I Orthology and parology

Genes that exist as multiple copies in nature fall primarily within two relational categories: orthology and parology (Figure 1.4). Orthologous genes (or the encoded proteins) are related sequences that arise via species descent. An example of orthologous genes would be the same copy of a hemoglobin subunit in separate tetrapod species, inherited by each species through the common ancestor of all tetrapods (Storz, 2016b). Alternatively, gene copy number can be amplified via duplication events, which results in paralogs. Three major mutagenic processes that give rise to paralogs are 1) unequal crossing over during meiosis, often resulting in tandem gene duplicates arranged on the same chromosome 2) reverse-transcription of mRNA copies that are re-inserted into the genome, and 3) whole-genome duplication (Zhang, 2003).

It is reasonable to argue that orthologs between species are often fixed under purifying selection for a role that is analogous between the species in question. All tetrapods would maintain their orthologous copies and functions of hemoglobin subunits, for example, due to the need for all

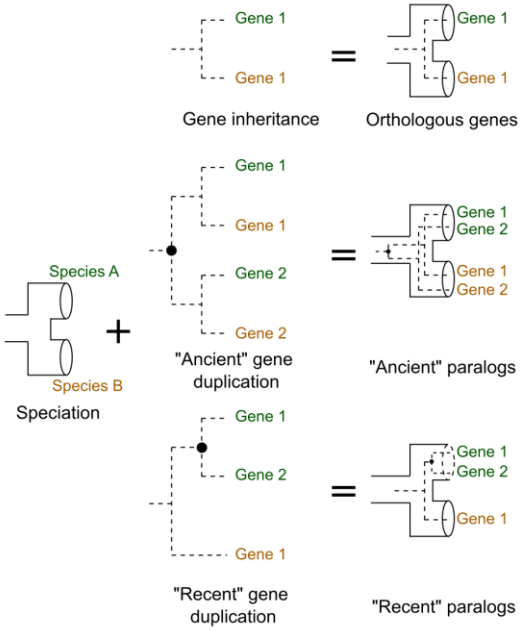


Figure 1.4: Multiple gene copies arise in nature primarily via speciation and gene duplication. Phylogenetic relationships between two hypothetical species (Species A and Species B in green and orange, respectively) are represented by a hollow three-dimensional dendrogram. Relationships between hypothetical genes from each species (Gene 1 and Gene 2 of the same color scheme) are represented by dotted lines. Hypothetical gene duplication events are represented by black dots. Lineage-specific clades of paralogous genes typically arise via recent gene duplications, whereas multi-gene families shared by diverse sets of species indicate ancient gene duplication in their common ancestor(s).

tetrapods to transport blood oxygen. Thus, phylogenetic relationships between orthologs often closely resemble or are completely congruent with the ancestry of the species in question, whereas paralogous genes are usually identified when one or more species is represented by multiple sequences. If paralogous genes existed in the ancestor of two or more species, the gene tree might consist of multiple subfamilies where the branching pattern within each subfamily resembles the species phylogeny. Clades of multiple sequences represented by only a single species further indicates species-specific (and often relatively recent) gene duplications, assuming the tree was constructed with sequences from a diverse set of organisms (see Figure 1.4 for these distinctions). Several models have been proposed for the fates of paralogous genes following duplication as has been thoroughly reviewed previously (Fani & Fondi, 2009; Innan & Kondrashov, 2010). The next section will highlight the evolutionary questions raised by some of these models.

I.V.II Gene fate models

Neo- and subfunctionalization (Figure 1.5) are two early models explaining the diversifying fates of paralogs after gene duplication. Neofunctionalization was first formally proposed by Ohno (Ohno, 1970) as a model for how a paralogous gene copy is free to acquire new functional properties while the other copy maintains an ancestral role. However, a current problem in evolutionary biology is how paralogous gene copies can be maintained intact as open reading frames in the genome long enough for novel beneficial mutations to become fixed by natural selection. This problem is two-fold, firstly because neofunctionalization predicts that the duplication itself and presence of redundant gene copies are neutral with respect to selection.

Duplicates would therefore need to drift to a high enough frequency in the population to make gain-of-function mutations more likely (Innan & Kondrashov, 2010). Secondly, paralogs are removed from populations at a higher rate than beneficial mutations are introduced (Bergthorsson et al., 2007; Kimura, 1983), meaning redundant genes are not protected from mutations that lead to pseudogenes or complete gene loss.

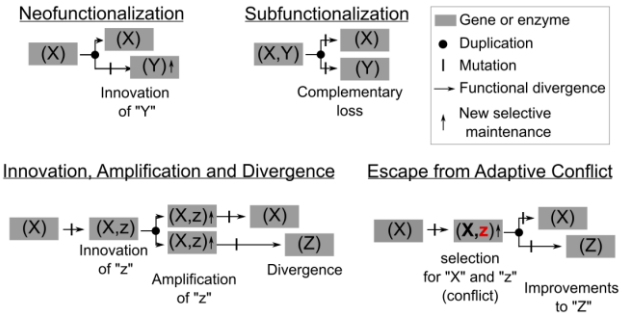


Figure 1.5: Multiple models of gene maintenance provide alternatives to earlier models. Showing fates of paralogs following gene duplication under four different models. Inset provides a key for symbols used. Capital letters represent hypothetical primary function, defined here as the function for which the allele has been fixed in a population. Lower-case letters represent a novel secondary trait. Red letter indicates that improvement to a secondary function would be deleterious to the gene's primary role.

Subfunctionalization was introduced as an alternative to neofunctionalization and predicts that ancestral functions are partitioned between paralogs via complementary loss-of-function mutations such that both descendant copies are required to maintain the ancestral phenotype (Lynch & Force, 2000). This resolved the problem of high-frequency null mutations that would otherwise lead to pseudogenes, as both copies can then be maintained by purifying selection for the ancestral function. Subfunctionalization, however, stops short of explaining how adaptive phenotypes could evolve *de novo*, as all of the functions in descendant genes were once performed by a single ancestor. Additionally, subfunctionalization is by definition a neutral process, meaning fixation of paralogs would need to occur via neutral forces. Nevertheless, the portion of the subfunctionalization model that specifies phenotypic partitioning of ancestral properties has been widely substantiated and is a fundamental concept when reviewing the importance of enzyme promiscuity (also known as "moonlighting") in evolutionary processes (Aharoni et al., 2005; Espinosa-Cantú et al., 2015; Fani & Fondi, 2009; Hughes, 1994; Tawfik, 2010).

The problems mentioned above regarding fates of paralogs are termed "Ohno's dilemma" and gave rise to the innovation, amplification and divergence (IAD) model of gene maintenance following duplication in addition to alternative models (Figure 1.5) (Bergthorsson et al., 2007; Innan & Kondrashov, 2010). IAD posits that fortuitous mutations conferring novel secondary activities of a gene or its encoded peptide could accumulate prior to duplication and be

maintained by selection for that gene's primary role. In the model, the innovation step refers to the first appearance of a new trait encoded by a gene prior to duplication (ex. a novel secondary enzyme product) no matter the magnitude of the novel phenotype. Amplification refers to an increase in magnitude of the new trait following gene duplication and it is in this stage that a gene dosage effect may confer a selective advantage for the secondary phenotype. Finally, descendant gene copies diverge in function under positive selection for the novel trait, reminiscent of subfunctionalization but with adaptive improvements. The IAD model addresses Ohno's dilemma because a gene can be maintained under purifying selection for a primary role long enough for gain-of-function mutations to accumulate, leading to a secondary role that is maintained by selection immediately following gene duplication. Improvements upon one copy to specialize or become more efficient in producing the novel trait would then follow.

Until now, the models presented focus primarily on neutral evolution or the role of selection post-duplication. However, given the high frequency of moonlighting proteins in nature and the paradoxes introduced by neo- and subfunctionalization as stand-alone models, the question arose whether selective pressures on individual promiscuous gene products prior to duplication would provide a primary role for gene duplication (Hughes, 1994; Piatigorsky, 1991). These ideas were later refined to detail how adaptively conflicted proteins could escape evolutionary dilemmas via gene duplication (Des Marais & Rausher, 2008; Hittinger & Carroll, 2007).

Hittinger & Carroll originally used the well-characterized yeast galactose (GAL) induction system in *Saccharomyces cerevisiae*, in which GAL3 activates transcription of its paralog possessing galactokinase activity, GAL1, in the presence of galactose. Interestingly, GAL3 and GAL1 serve regulatory and galactokinase functions, respectively, despite being closely related paralogs (Bhat & Murthy, 2001; Hittinger & Carroll, 2007; Johnston, 1987). The ancestral enzyme was presumed to be a bifunctional regulator and galactokinase, as is the case for the GAL3/GAL1-like bifunctional GAL protein from the yeast *Kluyveromyces lactis* (Meyer et al., 1991). In the study by Hittinger & Carroll, it was ultimately found that GAL3 and GAL1 promoters in *S. cerevisiae* resulted in poor fitness when driving transcription of the reciprocal gene product. However, the promoter for the bifunctional *K. lactis* GAL had differential fitness effects when driving *S. cerevisiae* GAL3 and GAL1; the *K. lactis* GAL promoter was fully functional in driving *S. cerevisiae* GAL3 expression but at a fitness cost due to poor GAL1 expression.

The authors concluded that promoters for ancestral bifunctional GAL proteins were adaptively conflicted in that they were incapable of simultaneously improving both regulatory and galactokinase functions. Molecular dissection of the various GAL promoters showed that improvements to transcriptional promotion of GAL1 came at a fitness cost to promotion of GAL3 (Hittinger & Carroll, 2007). Darwinian fitness of diverging paralogs could then explain fixation of mutant gene copies that improved upon a previous secondary function while simultaneously forgoing the putative ancestral primary function. This role for gene duplication was appropriately named "Escape from Adaptive Conflict" (EAC) and is largely compatible with the idea that proteins cannot be improved for multiple functions universally under selective conditions, and provides a means to circumventing evolutionary tradeoffs at the molecular level (Dalla & Dobler, 2016; Des Marais & Rausher, 2008).

I.V.III Assembly of novel biochemical pathways

The previous section dealt with the fates of gene duplicates and the differences between models that implicate selection or neutral evolution during different stages of evolving paralogs. Attention is now turned to the level of biochemical pathway assembly in the evolution of chemical diversity. The first attempt at explaining the origin of novel biochemical pathways was by Horowitz who suggested that the end products of modern pathways evolved first, and gene duplications that provided simpler precursors followed (Horowitz, 1945). This model assumes that simpler precursors of conversion products in early organisms once existed abundantly in the environment (Figure 1.6). As substrate provided by the environment became limiting, selection would favor paralogs that could synthesize substrates themselves, providing the first simple biochemical pathways. This came to be known as the retrograde hypothesis and predicts that genes belonging to a biochemical pathway are paralogs of each other (Fani & Fondi, 2009). However, as modern pathways are often completed by enzymes of different classes that facilitate widely dissimilar biochemical reactions (as is the case for di-TPSs and P450s in diterpenoid biosynthesis, for example) further models have been proposed.

The cumulative hypothesis (Granick, 1965), provided as an alternative to the retrograde hypothesis, posits that evolutionary pathways evolved in the forward-direction; early enzymes could utilize simple substrates provided by the environment and diverging gene duplicates could catalyze the formation of increasingly complex products (Figure 1.6). Granick proposed that this accounted for the evolution of more complex chlorophylls in true bacteria from structurally

simpler chlorophyll a, which is still used by plants and cyanobacteria. However, the cumulative hypothesis predicts that each new intermediate would provide a selective advantage. Although Granick was able to rationalize this in the chlorophyll biosynthetic pathway by early use of simpler chlorophylls, which evolved increased complexity later, it is insufficient to explain the assembly of biochemical pathways via intermediates that alone do not provide fitness benefits.

Another proposal, the patchwork hypothesis (Jensen, 1976; Yčas, 1974) of biochemical pathway evolution suggests that novel pathway enzymes evolved from ancestors with larger biochemical capabilities and were recruited into divergent pathways following selection for more specific biochemical routes (Figure 1.6). The patchwork hypothesis is compatible with gene fate models that implicate promiscuous ancestral enzymes, as well as with either 1) biochemical pathways consisting of sequential steps performed by paralogous enzymes, or 2) enzymes encoded by multiple gene families that perform sequential reaction steps.

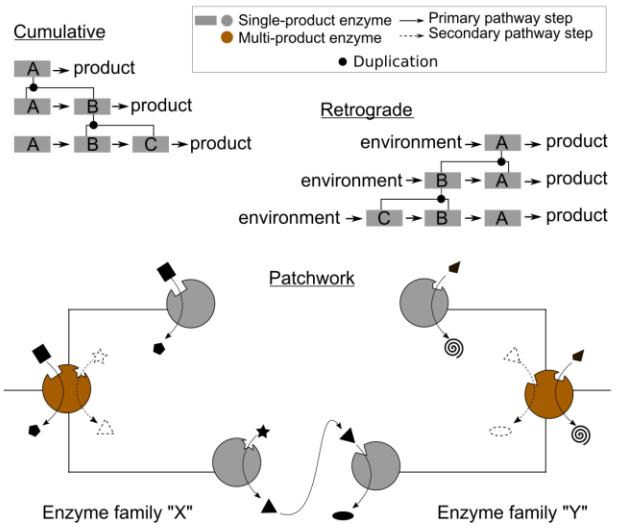


Figure 1.6: Alternative hypotheses of biochemical pathway assembly emphasize the roles of ancestral enzymes. Showing the cumulative and retrograde hypotheses under neofunctionalization conditions, and the patchwork hypothesis under moonlighting conditions. Legend provides symbol key. Solid and dotted polygons represent primary and secondary substrates and products (respectively) of enzymes. Enzyme families "X" and "Y" are encoded by separate gene families, such as diterepene synthases and cytochrome P450, respectively.

I.V.IV Epistatic interactions within biosynthetic genes and enzymes

Thus far, evolutionary outcomes have been discussed on three levels: the number of involved genes and their protein products, the fates of paralogs, and biochemical pathway assembly. One further layer that has guided modern evolutionary thought is that of protein primary sequence and tertiary structure and how each is capable of determining which phenotypic fates are available to evolving proteins. For decades, biologists have debated the relative contributions of chance and natural selection in shaping the phenotypic outcomes we see today, as is evident from reviews on historical contingency (Blount et al., 2008, 2018; Storz, 2016a). Contingency (or dependence) on the past

can be exemplified by chance events limiting the total number of mutational paths that a protein can traverse during evolution. This can arise via restrictive mutations at one locus that forbid certain mutations at another locus. Contingency can also arise via permissive mutations at one locus that are necessary to accommodate beneficial mutations at another locus.

Epistatically interacting residues, defined here as those sites of a protein that exert differential effects on alternative substitutions at another site, have been strongly implicated in determining the evolution of vertebrate hormone receptors and fungal heat shock proteins (Bridgham et al., 2009; Ortlund et al., 2007; Starr et al., 2018). Importantly, these studies found that permissive and restrictive mutations had little or no effect on function when introduced to ancestral proteins, consistent with the notion that contingency arises from processes invisible to selection. However, these functionally silent mutations had profound consequences on the effects of subsequent function-altering mutations. After beneficial mutations arise that contribute to organismal fitness, natural selection is certain to be strong regardless of mutational shaping via epistasis. However, there is a much lower rate of amino acid substitution in nature than would be expected if all proteins evolved neutrally or under selection, strongly suggesting that epistasis is the primary force in determining which phenotypic outcomes are ultimately available (Breen et al., 2012).

I.VI Contents of this work

The evolutionary concepts just outlined raise several questions about how diterpenoid pathways assembled from di-TPS and P450 enzymes arose in nature and generate such enormous chemical diversity. This section discusses specific questions addressed in this thesis, the associated hypotheses, and outlines the general aims and formats for the following chapters. Although a detailed methods section is included, the concepts of orthology searching and statistical estimation of ancestral sequences will be reviewed to provide the reader with sufficient background for interpreting results in the following chapters.

I.VI.I Questions and hypotheses

One question addressed in this thesis is to what extent specific phenotypic outcomes of diterpenoid metabolism depends on the past. One hypothesis is that biosynthesis of specific diterpene scaffolds is mutationally facile, and that diverse structures seen today can evolve rapidly in response to natural selection with little dependence on differing genetic backgrounds. Indeed, as will be discussed in more detail in Chapter II, di-TPSs are catalytically plastic such

that a few mutations can have dramatic functional consequences. An alternative viewpoint is that few mutations controlling product outcome may indicate a high degree of constraint, given that those rare function-altering mutations would be vastly outnumbered by silent or null mutations. These two scenarios are evaluated in Chapter II in the context of the repeated evolution of normally configured diterpene scaffolds across the land plant tree of life. Attention will also be drawn to the evolutionary consequences of chance events in shaping chemical phenotypes within these two compound classes when novel biochemical routes depend on recruitment of both class II and class I di-TPSs.

As biosynthetic routes towards specialized diterpenoids in land plants are often composed by di-TPSs and P450s that arose via lineage-specific gene duplications, land plants must have repeatedly formed novel pathways by combining these two gene families, leading to the diverse phenotypic outcomes we see today. However, it is unclear how the stepwise nature of evolution would result in compatible biosynthetic capabilities in unrelated gene families. Studies of diterpene metabolic genes in several angiosperm genomes indicate that TPSs and P450s involved in the same biosynthetic routes are in some cases clustered proximally within the genome (Boutanaev et al., 2015; Shimura et al., 2007; Swaminathan et al., 2009). Although it remains unknown to what extent and by which mechanisms selection would favor such clustering in general (Zi et al., 2014), it is apparent that novel biosynthetic relationships are followed by layered genetic changes from codon substitutions up to genomic organization.

For novel diterpenoid routes to evolve, the recruited di-TPS and P450 lineages must both acquire compatible derived phenotypes in order for selection to maintain paralogs from each family. The next question that arises is how new biochemical steps between P450s and di-TPSs evolve in cases where di-TPS products of modern pathways provide no fitness advantage on their own. Under the patchwork hypotheses, a novel side product formed by di-TPSs prior to duplication is predicted to be usable by P450s from a different metabolic pathway via fortuitous side activities from both gene families. This hypothesis would predict that broad product and substrate specificities of di-TPSs and P450s, respectively, predate the duplication events that gave rise to the gene families participating in modern-day resin acid biosynthesis in Pinaceae. Alternatively, secondary functional capabilities may be ancient in one gene lineage followed by change-of-function mutations in the other, leading to either cumulative or retrograde pathway assembly. To test these hypotheses, estimation and experimental resurrection of a ca. 170 million-year-old di-

TPS, ancestral to the Pinaceae lineage, is reported along with characterization of a bifunctional di-TPS from *S. moellendorffii* in Chapter III. This work is followed by phylogenetic and comparative biochemical studies of P450s in Chapter IV.

Chapter V focuses on diterpene metabolism in the modern-day conifer Norway spruce and explores how diterpene metabolism is regulated. A complete picture of metabolic control over the various terpenoid classes in Norway spruce resin is lacking, yet a better understanding of the pathway control points could aid in identifying factors associated with resistance to *I. typographus* invasion in natural populations. To this end, potential metabolic control points corresponding to the three stages of terpenoid biosynthesis outlined previously are examined with use of RNA-interference and gene overexpression.

I.VI.II Methods

Two tools employed in this work that are not typically part of informatics and biochemical characterization pipelines are OrthoFinder (Emms & Kelly, 2015) and the use of ancestral sequence reconstruction (Thornton, 2004).

OrthoFinder is a bioinformatics tool used to compare predicted genomes or proteomes across any number of chosen species, placing all sequences from the supplied genomes (or proteomes) into groups of descendants of single gene ancestors. To illustrate this concept, suppose that only two paralogous genes of the TPS family existed in the common ancestor of all land plants. Suppose then that these two genes were inherited by all species of interest, yet had differing patterns of duplication in each species. OrthoFinder uses a search algorithm of choice (e.g. BLAST) reciprocally on all modern sequences from the provided genomes to detect which genes descended from each paralog. Further bioinformatics steps such as phylogenetic analysis can then be performed on each group independently. As the TPS genes and encoded proteins of interest in this study form a multi-gene family, although not all subfamilies are of interest, OrthoFinder mitigates user bias in determining which sequences belong to a subfamily. Working with smaller subfamilies limits both the number of gaps introduced in multiple sequence alignments and saves computational resources in downstream analyses. The reduced sequence noise can also allow the inference of gene relationships that are otherwise obscured by large datasets.

Experimental resurrection of the ancestral di-TPS in Chapter III relied on maximum likelihood (ML) sequence estimation. ML is a widely used statistical method in phylogenetics for finding the tree topology most likely to give rise to the set of sequences observed; in other words, ML in phylogenetics finds the phylogenetic tree which results in the highest probability for the observed modern-day sequences such that, conceptually:

$$\boldsymbol{L = p(d|h)}$$

where ***L*** is the maximum likelihood calculation of each possible tree, and ***p*** is the probability of the data (***d***) under a given hypothesis (***h***). Note that in ML phylogenetics of proteins, ***d*** represents modern-day sequence data and ***h*** is substituted for a given tree structure (with associated branch lengths) and an amino acid substitution model. Many hypothetical trees are generated in ML phylogenetics, and likelihood is calculated for each. The tree chosen results in the highest probability for the observed set of sequences. The same principle is applied to protein sequences at internal tree nodes to infer ancestral amino acids states at a branch point of interest. In this case,

$$\boldsymbol{L_\theta = p(d|c,t,m)}$$

where $\boldsymbol{L_\theta}$ is the maximum calculated likelihood value for an ancestral character state (***c***) at a certain internal tree node. In ancestral sequence resurrection, each possible amino acid state of ***c*** is tested with a given a tree topology and branch lengths (***t***) and an optimal amino acid substitution model (***m***). The ancestral character state chosen results in the highest probability of generating the observed modern-day sequence data. Confidence is expressed as posterior probability and applies to each site independently of one another. For instance, any reconstructed ancestral peptide would have a posterior probability of 1.0 for the starting methionine, whereas a highly variable site where all 20 possible amino acids are equally likely might have a posterior probability of just 0.05. Fortunately, ancestral sequence reconstruction is largely robust to phylogenetic uncertainty and to low posterior probabilities at highly variable sites, as these sites are often silent with respect to function (Eick et al., 2017; Gaucher et al., 2003; Jermann et al., 1995). The estimated sequence can then be synthesized, heterologously expressed, and biochemically studied just as any modern-day protein.

I.VI.III Format, aims and scope

This thesis consists of four experimental chapters beginning with three chapters that investigate the evolution of diterpene resin acids in Norway spruce using comparative sequence analysis, ancestral sequence resurrection, enzyme biochemistry and homology modeling. The aim of these chapters is to dissect the genetic and biochemical layers underlying evolution of unique diterpenoid structures that may serve as a framework for future biochemical and mutational investigations into the basis of catalysis by present day enzymes. The first of these chapters employs a single phylogenetic method and reviews the last ca. two decades of di-TPS research in an attempt to offer new insight into the direction of evolution in di-TPS genes and their encoded enzymes. In the chapters thereafter, the hypotheses tested rely on laboratory experiments that offer insight into the genetic history of normally configured abietane and pimarane scaffold biosynthesis in conifers and expand the known biochemical function of P450s in Norway spruce. The last chapter of the thesis focuses on plant regulation of terpenoid formation in Norway spruce and aims to help identify the limiting steps of mono-, sesqui- and diterpene biosynthesis. The results should improve our understanding of the genetics underlying defensive compound accumulation such that better ecological predictions can be made for natural populations of Norway spruce under bark beetle attack.

Chapter II Land plant diterpenoid diversity is frequently constrained by chance events

II.I Background

The hypothesis that species trait diversity is contingent on chance events in lineage history is a classical alternative to adaptive positive selection for specific phenotypic outcomes (Blount et al., 2018). In fact, if forces that operate based on chance alone, such as neutral evolution and genetic drift, are as prevalent as argued by Jay Gould over three decades ago, then a large portion of the specific phenotypic fates we see in organisms today are likely not the result of selection (Erwin, 2006). Recent improvements to molecular phylogenetics tools and the increased availability of sequencing information has allowed modern molecular and biochemical approaches to be employed in testing the predictions of contingency on protein evolution.

This is not to say that the role of positive adaptive selection diminishes when contingency is high. By utilizing modern molecular tools, one study in evolving *Escherichia coli* populations demonstrated that adaptive metabolism of citrate as a carbon source can occur in glucose-limited media (Blount et al., 2008). However, this only occurred after an exceptionally large number of generations and was restricted to one of the late experimental populations. It was therefore concluded that adaptation to citrate metabolism was contingent on a lineage-specific set of permissive (and probably mostly neutral) mutations unique to that population. Several later studies demonstrated that interaction between protein residues, referred to as intramolecular epistasis, can constrain the direction of evolution by only allowing certain substitutions at one site given a particular amino acid status at another (Bridgham et al., 2009; Gong et al., 2013; Shah et al., 2015; Starr et al., 2018). Thus, it appears that the historical order in which mutations occur influences the probability of certain phenotypic outcomes because function-changing substitutions can often only exert their effects in the presence of mutations that were previously fixed. Epistasis between protein residues is therefore highly consequential for the fate of trait evolution, and in the search for function-changing amino acids in mutational studies, due to the growing support that mutational effects are contingent on genetic history.

There is empirical support that in cases of convergent evolution, defined broadly here as the repeated and independent acquisition of identical or highly similar traits, mutations that confer the same novel phenotype are often localized to identical or similar protein sites (Dalla & Dobler, 2016; Projecto-Garcia et al., 2013; J. Zhang, 2006; Zhen et al., 2012). The predictability of mutational effects can therefore be partly gauged, as the mutational experiment has already occurred naturally in the repeated evolution of protein phenotypes. However, if the constraining mutations discussed above are lineage-specific, they will lessen the probability that convergence will occur (Storz, 2016a), even in the presence of selection, and hamper attempts to dissect the mutational basis of enzyme catalysis (Hochberg & Thornton, 2017).

Plant diterpene metabolism offers an unprecedented opportunity to investigate the molecular basis of trait diversity and the roles of contingency, epistasis and constraint. As will be discussed in more detail, a relatively small number of plant di-TPSs have repeatedly converged on the biosynthesis of two classes of diterpene intermediate compounds: (+)-abietane and pimarane-type diterpene intermediates. Coupled with a much larger set of P450s, these convergently evolved intermediates were recruited into myriad divergent pathways, serving a broad range of ecological roles (Bathe & Tissier, 2019; Karunanithi & Zerbe, 2019; Nelson & Werck-Reichhart, 2011). This chapter highlights selected studies from the last two decades of di-TPS research, which in lieu of the land plant phylogeny, demonstrates a pattern of repeated and predictable trait evolution (that is, chemical convergence). It is also argued here that the repeated evolution of specific core diterpene structures was wrought by episodes of contingency on gene loss, duplication, and epistatically interacting protein residues, rendering the mutational processes underlying specific evolutionary outcomes difficult to predict.

II.II Mechanisms of labdane diterpenene biosynthesis

Labdane-related diterpenoids, which exist as a superfamily of thousands of diverse metabolites, are so named due to a common decalin core derived by the actions of class II di-TPSs (Peters, 2010). Class II di-TPSs possess a conserved N-terminal aspartate-rich DxDD motif and initiate cyclization reactions by protonating GGPP. Enzyme-mediated protonation of GGPP (**1**) initiates a cascade of isomerizing double bond rearrangements that ultimately results in bicyclic labdenyl diphosphates, most commonly as *syn*-, *ent*-, or "normal" configurations of CPP (Peters, 2010; Figure 2.1). Stereo-chemical control is facilitated by binding GGPP in various conformations.

Class II di-TPS are often referred to as copalyl diphosphate synthases (CPSs) or CPS-like enzymes and are encoded by the TPS-c subfamily of di-TPS. Class I di-TPSs, which are encoded by multiple di-TPS subfamilies, catalyze a second cyclization step mediated by a C-terminal DDxxD motif by ionizing the diphosphate group of the labdenyl intermediate, which results in its removal (Peters, 2010; Figure 2.1A). It is widely accepted that all major land plant lineages possess class II and class I di-TPSs capable of producing *ent*-kaurene via *ent*-CPP (**15**), as *ent*-kaurene (**17**) is utilized as a substrates by P450s in the biosynthesis of ubiquitous gibberellin plant hormones (Chen et al., 2011; Karunanithi & Zerbe, 2019; Zi et al., 2014).

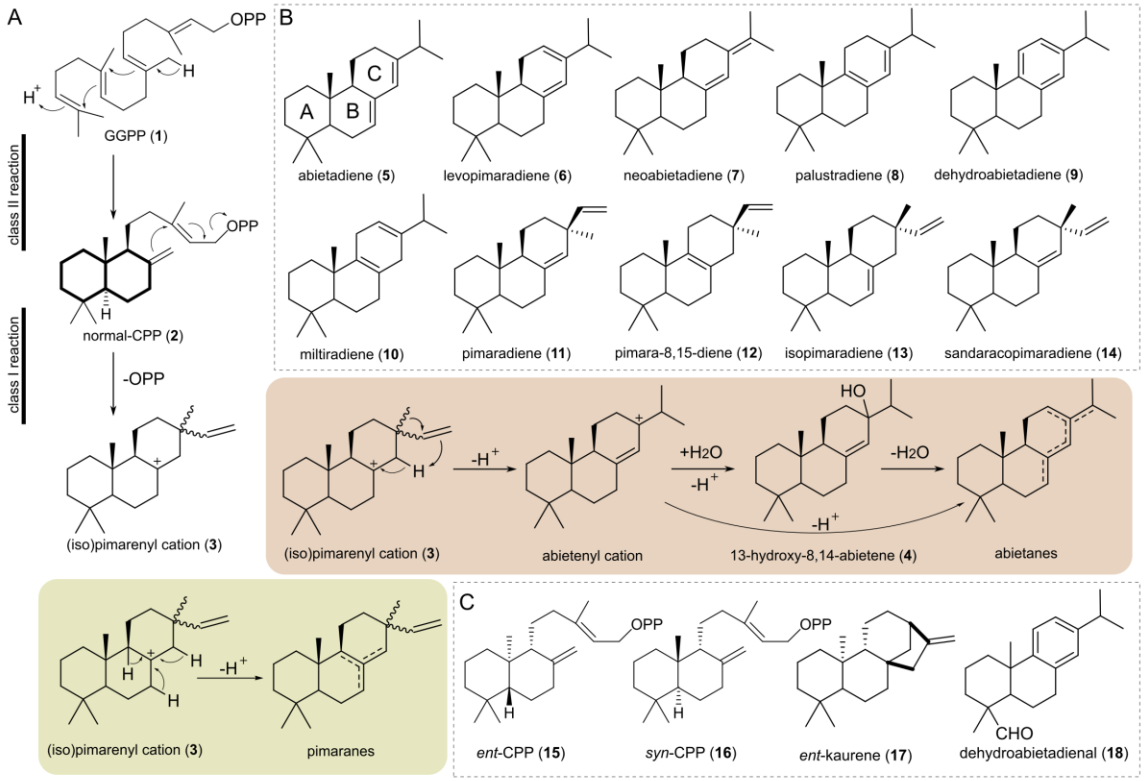


Figure 2.1: Biosynthesis of (+)-abietane and pimarane-type olefins is carried out by class II and class I diterpene synthases via carbocation intermediates. A) Biosynthetic pathway and simplified mechanism of protonation-initiated cycloisomerization of geranylgeranyl diphosphate (GGPP) and phosphate cleavage of (+)-copalyl diphosphate (normal-CPP), resulting in (iso)pimarenyl cation intermediates. Decalin core is shown in bold in normal-CPP . (+)-Abietane biosynthesis (pale orange) follows double bond protonation of (3, methyl and vinyl groups of either configuration) and a 1,2-methyl shift, and the abietenyl cation is quenched by direct deprotonation or water capture. (+)-Pimaranes (yellow-green) result from immediate deprotonation of (3) from three possible hydrogen positions. B) Normally-configured (+)-abietane and pimarane-type diterpene products of focus here. C) Other diterpene structures discussed here.

Normally configured (+)-abietane and (+)-pimarane olefins and alcohols, which will hereafter be referred to simply as (+)-abietanes and pimaranes, represent a class within the broader category of labdane diterpenes that are derived by the specific actions of specialized class I diterpene cyclases

(Peters, 2010; Figure 2.1A). The second cyclization in these reactions proceeds via epimers of reactive pimarenyl carbocation intermediates (**3**), which differ stereochemically in their C-ring methyl and vinyl groups and are usually quenched by deprotonation. However, the carbocation intermediates can also be quenched by water, leading to hydroxylated diterpenes (Keeling et al., 2011a; Peters, 2010). The specific class I cyclase determines both the intermediate produced (either pimarenyl or isopimarenyl carbocations) and which proton is lost in stabilizing the cationic carbon. Peters (2010) provides a detailed mechanistic explanation for how these transformations occur to yield multiple diterpene products that vary subtly in their arrangement of double bonds and specific stereochemistry (Figure 2.1A). Thus, a diverse set of product-specific class I diterpene cyclases, or a single nonspecific cyclase, could in principle produce a variety of abietanes and pimaranes from a single substrate. Plants possess both, as will be discussed in the next section.

II.III Recurring evolution of (+)-abietanes and pimaranes across the plant tree of life

This section is concerned with the mono- and bifunctional di-TPS enzymes whose actions across the land plant tree of life indicate the widespread, repeated evolution of similar diterpene scaffolds. Where appropriate, the olefins and diterpene alcohols of interest are indicated in Figure 2.1B & C.

II.III.I (+)-Abietane and pimarane synthases in angiosperms

Monocot species widely express *syn*- and *ent*-abietane and/or pimarane-type di-TPSs from the kaurene synthase-like (KSL) family, as has been well reviewed (Murphy & Zerbe, 2020; Schmelz et al., 2014; Zhou et al., 2012). KS and KSL enzymes belong to the distinct TPS-e/f subfamily of di-TPSs ubiquitous across the land plant phylogeny. TaKSL2 from *Triticum aestivum* (wheat), a monocot from the family Poaceae, converts *ent*-CPP to *ent*-pimara-8(14),15-diene but also catalyzes the formation of abietadiene (**5**) from normal-CPP (**2**) (Zhou et al., 2012). TaKSL1 and TaKSL4 convert normal-CPP to isopimaradiene (**13**) and pimara-8,15-diene (**12**). TaKSL1 and TaKSL4 appear to have arisen independently within wheat via gene duplication, as they are paralogs of each other (Figure 2.2). These two enzymes are in turn orthologous to (i.e. appeared

via species descent from) OsKSL4 from *Oryza sativa*, which instead converts *syn*-CPP (**16**) to *syn*-pimaradiene (Otomo et al., 2004). Zhou et al. (2012) argue that substrate promiscuity in monocot KSL enzymes likely contributes strongly to the apparently facile nature of functional divergence in this gene family and that these relatively recent gene duplications were the early mutations that gave way to neofunctionalizing mutations. Indeed, *Zea mays* (maize) is another monocot species that has recently-derived paralogs of KSL enzymes capable of producing pimarane-type diterpenes. Duplication in this lineage gave rise to ZmKSL4, which is capable of cyclizing *ent*- and *syn*-CPP to *ent*- and *syn*-pimarane-type diterpenes, and also converts normal-CPP to pimaradiene (**11**) (Mafu et al., 2018). This duplication event in maize also yielded ZmKSL2, which is not yet characterized. However, both ZmTPS2 and ZmTPS4 are paralogous to the *Z. mays ent*-kaurene synthase (ZmTPS1), strongly implying *Zea*-specific acquisition of pimaradiene biosynthesis (Fu et al., 2016).

Independent acquisition of (+)-abietanes and pimaranes among dicots is also apparent, supported by widespread occurrence of miltiradiene synthases across several genera following duplication of KSLs independently within the Lamiaceae (Brückner et al., 2014; Cui et al., 2015; Gao et al., 2009; Jin et al., 2017; Pateraki et al., 2014; Pelot et al., 2017; Zerbe et al., 2014). Substrate specificity also appears to vary among these enzymes, which arose in the Lamiaceae via gene duplications of KSL enzymes. MvELS, a KSL from *Marrubium vulgare*, for example, catalyzes the formation of miltiradiene (**10**) from normal-CPP, but also forms different products with other labdane-related diphosphate intermediates (Zerbe et al., 2014). This KSL lineage from Lamiaceae is orthologous to other angiosperm KS and KSL enzymes, including those that produce non-abietane/pimarane-type diterpenoids, mono- and sesquiterpenoids (Brückner et al., 2014; Sallaud et al., 2009, 2012; Schilmiller et al., 2009). For brevity, only the relevant KSL enzymes from *M. vulgare* are represented in Figure 2.2 and only the orthologous relationship to monocot KSL is shown. However, it should be noted that the KSL enzymes appear to be quite diverse in the Lamiaceae; in addition to miltiradiene synthases isolated from several Lamiaceae species, IrTPS2 from *Isodon rubescens* converts normal-CPP to nezukol, a pimarane-type diterpene alcohol (Pelot et al., 2017). Another KSL from *I. rubescens* (IrKSL6) belongs to the same subclade as IrTPS2 and MvELS and converts normal-CPP to isopimaradiene, which is produced specifically by some di-TPSs from conifer species, as will be discussed in the following sections.

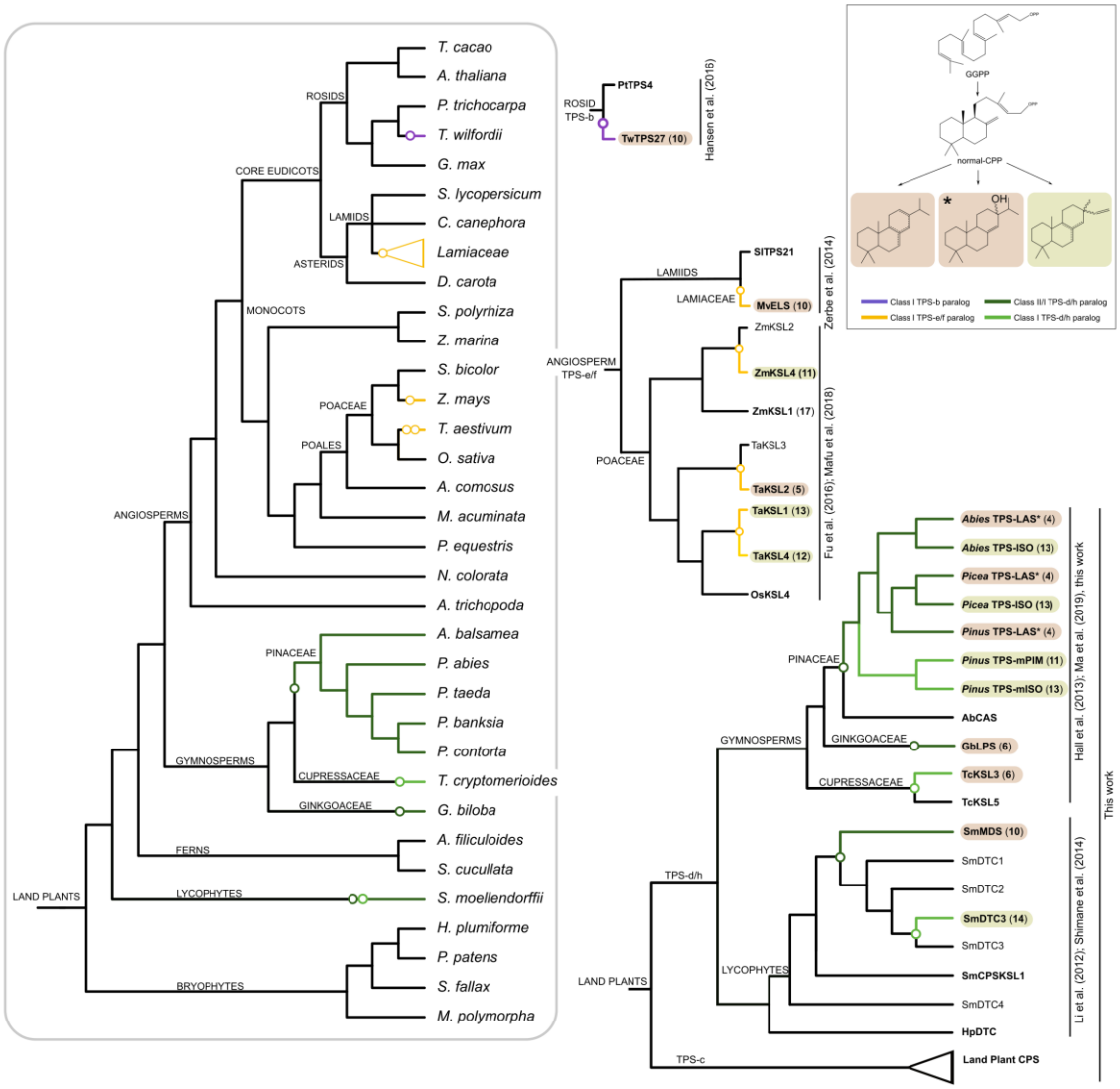


Figure 2.2: Similar diterpene olefins repeatedly evolved across land plants via recruitment of alternative gene families. Left) Land plant phylogeny of representative plants with available genome-predicted proteomes used in this work, as well as plants with reported (+)-abietane and pimarane-type diterpene synthases. Branch colors denote plant lineages with genes available for relevant biosynthetic steps (see inset for gene family names). Circles depict the earliest known gene duplication events inferred to have led to a paralog with at least some (+)-abietane or pimarane synthase activity. Full species binomials (from top to bottom) are *Theobroma cacao, Arabidopsis thaliana, Populus trichocarpa, Tripterigium wilfordii, Glycine max, Solanum lycopersicum, Coffea canephora, Daucus carota, Spirodela polyrhiza, Zostera marina, Sorghum bicolor, Zea mays, Triticum aestivum, Ananas comosus, Musa acuminata, Phalaenopsis equestris, Nymphaea colorata, Amborella trichopoda, Abies balsamea, Picea abies, Pinus taeda, Pinus banksia, Pinus contorta, Taiwania cryptomerioides, Ginkgo biloba, Azolla filiculoides, Selaginella moellendorffii, Hypnum plumiforme, Physcomitrella patens, Sphagnum fallax, Marchantia polymorpha*. Right) Reported relationships of diterpene synthase genes with descendants that encode either (+)-abietane-type (pale orange) or (+)-pimarane-type (yellow-green) diterpene synthases. Encoded enzymes that are characterized are in bold. Note that no paralog of GbLPS has been identified in *G. biloba*. Inset) Names of relevant gene families and product class. Asterisks denote release of unstable tertiary alcohol from the active site before further rearrangement.

Tripterygium wilfordii (Celastraceae) is a member of the Rosid clade of dicots and therefore only distantly related to both monocots and the Lamiaceae, part of the Asterid clade (Figure 2.2). *T. wilfordii* surprisingly uses TwTPS27 from the divergent TPS-b subfamily of plant di-TPSs to catalyze the conversion of normal-CPP to miltiradiene (Hansen et al., 2017). Other dicot species express terpene synthases that are represented in the TPS-b lineage. For instance, Pt*TPS4* from the tree *Populus trichocarpa* is orthologous to *Tw*TPS27 (Hansen et al., 2017; Figure 2.2) and is active with both GPP and FPP, converting them respectively to mono- and sesquiterpene hydrocarbons and alcohols (Danner et al., 2011). Prior to the study by Hansen and coworkers, however, this subfamily was viewed as a clade of monoterpene synthases, and TPS-b members are therefore not routinely assayed with cyclic prenyl diphosphate intermediates; it remains to be seen if more TPS-b members are capable of performing class I cyclization reactions with labdenyl diphosphates. Intriguingly, the oxidized abietane-type diterpenoid dehydroabietadienal (**18**) has been reported in *Arabidopsis thaliana* leaves, albeit with unconfirmed stereochemistry (Chaturvedi et al., 2012). As no di-TPS to date has been implicated in normal- or *syn*-CPP biosynthesis in *A. thaliana*, and due to the ability of miltiradiene to rearrange spontaneously to dehydroabietadiene (**9**) (Zi & Peters, 2013), it is feasible that *ent*-miltiradiene is an intermediate to DA, with its biosynthesis carried out by TPS-b members that utilize *ent*-CPP as substrate.

Taken together, it can be inferred that (+)-abietane and pimarane diterpenene biosynthesis was acquired independently multiple times following relatively late duplication events of KSL enzymes in Rosids, Asterids, and Poaceae long after the monocot-dicot split. In monocots, this appears to have occurred after the split of the family Poaceae from other members of Poales, with episodes of independent (+)-abietane and pimarane diterpene evolution at the species level in wheat, and maize. Similarly, in the Lamiaceae, duplication events of KSL enzymes that gave rise to miltiradiene synthases appear to have arisen soon after this family split from other Asterids, giving rise to orthologous (+)-abietane and pimarane type di-TPSs across Lamiaceae genera. The presence of a similar miltiradiene synthase in *T. wilfordii* suggests an additional recent episode of convergence on this metabolite class.

II.III.II (+)-Abietane and pimarane synthases in gymnosperms

Gymnosperms have received the most attention in studies of the biosynthesis of (+)-abietanes and pimaranes and are frequently among focal points of reviews, particularly the family Pinaceae (Alicandri et al., 2020; Chen et al., 2011; Karunanithi & Zerbe, 2019; Keeling & Bohlmann, 2006; Zulak & Bohlmann, 2010). This is due to the central role of these metabolites in conifer resin acid biosynthesis, which is terminated via oxidation by P450s. Most gymnosperms utilize di-TPSs from the gymnosperm- specific TPS-d subfamily to form diterpene olefins. Similarly, the nonseed plant *Selaginella moellendorffii* possesses bifunctional class II/I di-TPSs from the lycophyte-specific TPS-h subfamily. However, barring horizontal gene transfer from a yet-unknown origin to lycophytes and gymnosperms, an explanation for the origin of these two subfamilies is lacking. A new phylogenetic reconstruction is therefore provided after mining the genome-predicted proteomes of 25 land plant species using OrthoFinder (Emms & Kelly, 2015; see Appendix Figure A2.1 for phylogeny of all species used). As shown in Figure 2.3, the orthogroup containing conifer TPS-d members also contains the *S. moellendorffii* TPS-h lineage, and phylogenetic reconstruction of this orthogroup resulted in an orthologous TPS-d/h relationship. Therefore, these two gene families will hereafter be referred to simply as TPS-d/h. Table 2.1 provides a summary of characterized enzymes retrieved by OrthoFinder.

In conifers belonging to the genus *Picea*, two paralogous di-TPSs belonging to the TPS-d/h subfamily are considered responsible for the production of at least seven of the eight known (+)-abietane and pimarane-type skeletal structures observed in bark and needle resins (Keeling et al., 2011a; Keeling et al., 2011b; Martin et al., 2004) and previous phylogenies suggest they arose as paralogs of each other following duplication early in the genus *Picea*, as both *Picea abies* and *Picea sitchensis* possess a copy (Alicandri et al., 2020; Hall et al., 2013). TPS-ISO and TPS-LAS from *Picea* spp. are both bifunctional di-TPSs that perform class II and class I cyclization reactions by first converting GGPP to normal-CPP, which then diffuses to a second active site that mediates class I diphosphate ionization (Peters et al., 2001). *in vitro* assays in the studies above show that TPS-ISO performs these two cyclizations, resulting in isopimaradiene (**13**) as either the sole product in *P. abies* or as a mixture with lower amounts of sandaracopimaradiene (**14**) in *P. sitchensis*.

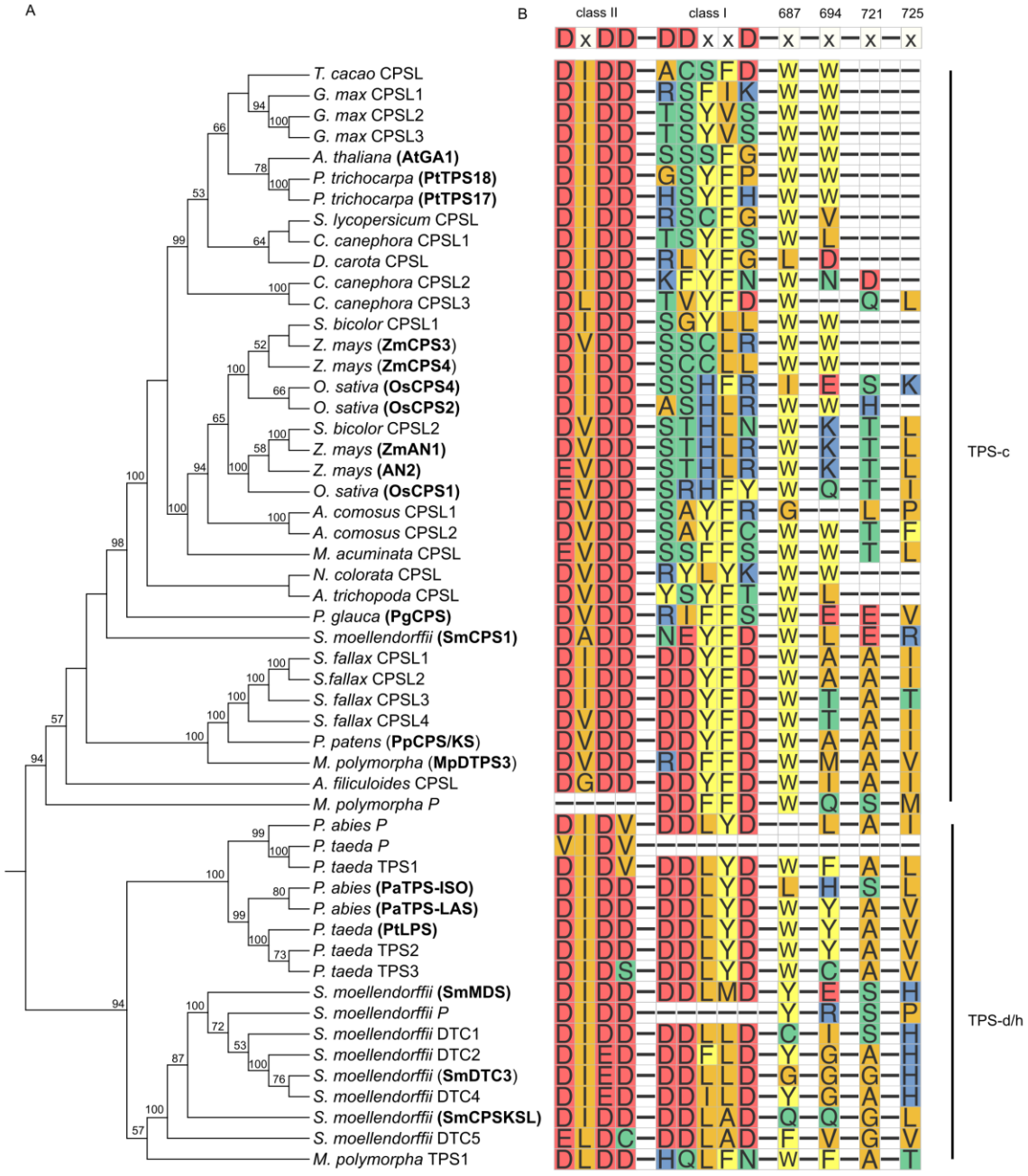


Figure 2.3: Ancient duplication of the bifunctional class II/I CPS lineage in the ancestor of land plants was followed by (+)-abietane and pimarane diversification in conifers and nonseed plants. A) Maximum-likelihood phylogeny of orthogroup containing the TPS-c and TPS-d/h subfamilies. Characterized sequences are in bold parentheses. Bootstrap support values of 50 or higher are indicated. Full binomials of species surveyed are given in Appendix Figure 2.1. B) Alignment segments of encoded protein products (showing the class II and class I binding motifs and four sites discussed that affect product outcome in disparate lineages). Predicted pseudogenes marked with "*P*". Amino acids colored according to side-chain chemistry (Taylor, 1997; Yu et al., 2017).

Table 2.1. Substrate and product summary of characterized CPS and CPS-like (subfamily TPS-c) and TPS-d and -h sequences retrieved from the proteomes examined here

Enzyme name	Organism	Substrate	Subfamily	Product(s)	Source
AtGA1	*Arabidopsis thaliana*	GGPP	TPS-c	*ent*-CPP	(Sun & Kamiya, 1994)
MpDTPS3	*Marchantia polymorpha*	GGPP	TPS-c	normal-CPP	(Kumar et al., 2016)
OsCPS1	*Oryza sativa*	GGPP	TPS-c	*ent*-CPP	(Prisic et al., 2004)
OsCPS2	*Oryza sativa*	GGPP	TPS-c	*ent*-CPP	(Prisic et al., 2004)
OsCPS4	*Oryza sativa*	GGPP	TPS-c	*syn*-CPP	(Xu et al., 2004)
PaTPS-ISO	*Picea abies*	GGPP	TPS-d	isopimaradiene	(Martin et al., 2004)
PaTPS-LAS	*Picea abies*	GGPP	TPS-d	13-hydroxy-8,14-abietene	(Keeling et al., 2011a; Martin et al., 2004)
PgCPS	*Picea glauca*	GGPP	TPC-c	*ent*-CPP	(Keeling et al., 2010)
PitLPS	*Pinus taeda*	GGPP	TPS-d	various diterpene olefins	(Ro & Bohlmann, 2006)
PtTPS17	*Populus trichocarpa*	GGPP	TPS-c	*ent*-CPP	(Irmisch et al., 2015)
PpCPS/KS	*Physcomitrella patens*	GGPP	TPS-c	*ent-kaurene,* 16α-hydroxy-*ent*-kaurane	(Hayashi et al., 2006)
SmCPS1	*Selaginella moellendorfii*	GGPP	TPS-c	normal-CPP	(G. Li et al., 2012)
SmMDS	*Selaginella moellendorffii*	GGPP	TPS-h	miltiradiene	(Sugai et al., 2011)
SmTPS7	*Selaginella moellendorffii*	GGPP	TPS-h	labda-7,13*E*-dien-ol	(Mafu et al., 2011)
ZmAN1	*Zea mays*	GGPP	TPS-c	*ent*-CPP	(Bensen et al., 1995)
ZmAN2	*Zea mays*	GGPP	TPS-c	*ent*-CPP	(Harris et al., 2005)
ZmCPS3	*Zea mays*	GGPP	TPS-c	normal-CPP	(Murphy et al., 2018)
ZmCPS4	*Zea mays*	GGPP	TPS-c	8,13-CPP	(Murphy et al., 2018)

TPS-LAS enzymes in *Picea* spp. yield the unstable abietane-type alcohol intermediate 13-hydroxy-8,14-abietene (**4**) that can be utilized directly as substrate by P450s or undergo non-enzymatic dehydration to four (+)-abietanes (**5-9)** (Geisler et al., 2016; Keeling et al., 2011a; Figures 1 & 2). Dehydroabietadiene is also frequently observed in the product profile of TPS-LAS in these studies, as most (+)-abietanes spontaneously deprotonate to yield the aromatic ring.

TPS-LAS from the genus *Pinus*, also from the family Pinaceae, is effectively identical and orthologous to those seen in *Picea* spp. (Hall et al., 2013; Ro & Bohlmann, 2006; Figure 2.2). However, congeners of *Pinus* possess monofunctional TPS-d/h (TPS-mPIM and TPS-mISO) enzymes in addition to TPS-LAS that show only low activity with GGPP and instead convert normal-CPP to pimaradiene (**11**) and isopimaradiene, respectively (Hall et al., 2013). The phylogenetic work from Hall et al. supports a *Pinus*-specific duplication event that gave rise to TPS-mPIM and TPS-mISO. It further implies that the monofunctional and bifunctional members of TPS-d/h subfamilies arose via a more ancient duplication in the common ancestor of all conifers, as TPS-mPIM and TPS-mISO are clearly paralogous to the *Pinus* and *Picea* TPS-LAS, as well as two di-TPSs from the *Abies* genus of conifers (See TPS-d/h phylogeny of Figure 2.2). This strongly implies that duplicate TPS-d/h family members (at least one monofunctional and one bifunctional) existed prior to speciation events leading to individual Pinaceae species, and that one copy was lost among *Picea* and *Abies* genera. Conifer gene duplication and subsequent generation of pseudogenes also appears to be supported by our analysis using OrthoFinder with the *Pinus taeda* and *P. abies* proteomes, as both species possess uncharacterized TPS-d/h like genes that met the screening criteria (see Materials and Methods) and are paralogous to TPS-ISO and TPS-LAS (Figure 2.3A). At least in *P. abies,* this uncharacterized paralog is not likely to be transcriptionally active, as the first intron is predicted to be over 7.5 kilobases in length (not shown) and we were additionally unable to clone it from cDNA of any tissue type. Additionally, both *P. abies* and *P. taeda* do not have an otherwise fairly conserved tryptophan at site 687, and overall lower similarity in this region than other tps-d/h members (Figure 2.3B).

A TPS-d/h enzyme from *Abies grandis* (family Pinaceae), the first abietane-type diterpene cyclase characterized, converts GGPP to five diterpene olefins (**5-9**) (Peters et al., 2000; Vogel et al., 1996). Later, TPS-d enzymes similar to TPS-LAS and TPS-ISO from *Abies* species were investigated in *Abies balsamea* (Zerbe et al., 2012b), a different species within the same genus, revealing that *A. balsamea* actually possesses its own set of TPS-LAS and TPS-ISO as paralogs

of each other (i.e. AbTPS-LAS is more closely related to AbTPS-ISO than it is to TPS-LAS from *Picea* and *Pinus* species). This is a striking example of parallel evolution, as it strongly implies that a single enzyme ancestor of the conifer TPS-d enzymes duplicated once independently in each lineage, resulting in identical sets of specialist enzymes. This would explain how *Abies* and *Picea* lost their monofunctional isopimaradiene synthase enzymes that are present in *Pinus*: they replaced them with specialized bifunctional isopimaradiene synthases (Figure 2.2).

The more distantly related conifer *Taiwania cryptomerioides* belonging to the family Cupressaceae was recently reported to express multiple monofunctional di-TPSs with novel divergence in catalytic properties (Ma et al., 2019). Among them, three were active with normal-CPP in single-product reactions. TcKSL1 dephosphorylated normal-CPP to produce biformene. Another (TcKSL5) cyclized normal-CPP to the kaurene-like diterpene alcohol phyllocladanol, and the third (TcKSL3) converted normal-CPP solely to the (+) abietane-type diterpene levopimaradiene (**6**). Interestingly, TcKSL3 is a paralog of TcKSL5, and both of these enzymes are more closely related to other TcKSL enzymes not involved in abietane or pimarane production than they are to the Pinaceae TPS-d subfamily (Ma et al., 2019; Figure 2.2). This partly explains the incongruence between the gymnosperm species phylogeny and the TPS-d/h gene tree. However, either unusually high substitution rates or ancient gene duplication followed by reciprocal gene loss in Ginkgoaceae and Pinaceae would be required to fully explain this placement.

The non-conifer gymnosperm *Ginkgo biloba* is the sole extant member of the Ginkgoaceae and also possesses a TPS-d/h (GbLPS), which converts GGPP solely to levopimaradiene (Schepmann et al., 2001). GbLPS is orthologous to the Pinaceae TPS-d/h members (Figure 2.2). However, as *T. cryptomerioides* has not yet been reported to possess an enzyme similar to the Ginkgoaceae and conifer TPS-d/h members, it is unclear whether the lineage harboring TcKSL existed as a paralog of conifer and *G. biloba* TPS-d/h prior to gymnosperm radiation. In any case, it appears that monofunctional levopimaradiene synthase activity evolved independently in *T. cryptomerioides* following duplication of the enzyme ancestral to TcKSL3 and TcKSL5. Reports of only levopimaradiene-specific synthases from *G. biloba* and *T. cryptomerioides*, together with the Pinaceae-specific duplications leading to isopimaradiene synthases, suggest that (+)-pimarane-type diterpene biosynthesis was acquired independently in the Pinaceae whereas (+)-abietane-type biosynthesis likely dates at least as far back as the gymnosperm ancestor.

II.III.III (+)-Abietane and pimarane synthases in nonseed plants

In comparison to angiosperms and gymnosperms, considerably less is known about specialized diterpenoid metabolism in nonseed plants. di-TPSs in nonseed plants are sometimes named by an abbreviation that identifies the main product, as is the case for miltiradiene synthase (MDS). Nonseed diTPS are also designated by subfamily homology (e.g. KSL, CPS, or both), or are referred to as "diterpene cyclases" (DTC). A bifunctional di-TPS in the lycophyte *Selaginella moellendorffii*, SmMDS, converts GGPP to miltiradiene, thereby catalyzing both class II and class I cyclization reactions to yield an abietane-type olefin (Sugai et al., 2011). SmMDS arose via lineage-specific duplications that gave rise to the expanded lycophyte TPS-d/h subfamily in lycophytes sometime after the divergence of *S. moellendorffii* from other plants (Figure 2.2). However, due to a lack of published lycophyte genomes, the distribution and relative ages of TPS-h family members in this lineage remains largely unexplored. Among the TPS-d/h paralogs of SmMDS are two other characterized di-TPSs with divergent catalytic properties. SmDTC3 converts *ent*- and *syn*-CPP to 16α-hydroxy-*ent*-kaurane and unidentified diterpene alcohols, respectively (Shimane et al., 2014). However, SmDTC3 also produces sandaracopimaradiene via normal-CPP as shown by Shimane and co-workers. A third TPS-h family member, SmCPSKSL1, is bifunctional in that it both performs the protonation-initiated cyclization of GGPP and ionizes the diphosphate group, but yields a labdadiene alcohol likely in the normal configuration (Mafu et al., 2011). As SmCPSKSL1 represents an anciently branching *S. moellendorffii* paralog (Figure 2.2) and SmDTC3 has the biochemical repertoire to produce pimarane-type diterpenoids, acquisition of miltiradiene biosynthesis by *S. moellendorffii* enzymes was likely an independent event that was contingent on an intrinsic ability of this lineage to synthesize normally configured diterpenes.

Investigation into diterpenoid biosynthesis in the moss *Hypnum plumaeforme* revealed that HpDTC1, which is orthologous to gymnosperm and lycophyte TPS-d/h subfamilies, respectively, converts GGPP to *syn*-pimaradiene (Okada et al., 2016). Although this enzyme uses a different stereoisomer of CPP than most characterized TPS-d/h members, activity and phylogenetic placement of HpDTC1 support three novel aspects of diterpenoid evolution: 1) an extremely ancient origin of the bifunctional TPS-d and TPS-h subfamilies dating to the ancestor of all land plants, 2) the presence of abietane and pimarane biosynthetic capabilities in all these di-TPS subfamilies regardless of stereochemistry, and 3) whole-gene loss of the TPS-d/h subfamily

among angiosperms and ferns (Figure 2.2). Together these factors explain the chemical expansion of normally configured diterpenes in conifers and why this class appears to be more limited in angiosperms.

II.IV Epistasis and constraint among diterpene synthases is prevalent

Angiosperm diterpenoids are highly diverse in large part via gene expansion of KSL members performing class I reactions with CPP substrates of various conformations to generate a variety of labdane and labdane-related products, as discussed previously. However, it seems that loss of the TPS-d/h subfamily among angiosperms, explainable by neutral mutational events, set flowering plants on a trajectory wherein early expansion of normally configured abietane and pimarane diterpenoids was absent. Not until monofunctional class II CPS enzymes capable of synthesizing normal-CPP appeared could angiosperms converge on some of the same diterpene scaffolds observed in gymnosperms and nonseed plants. However, the relevant mutations coupled with selective advantages conferred by normally configured diterpenes in angiosperms eventually drove the repeated acquisition of (+)-abietanes and pimaranes in limited angiosperm lineages as seen today. This scenario resembles the unusually large number of generations required for only a single *E. coli* population to evolve permissive mutations for citrate consumption, as mentioned previously.

While it appears that acquisition of the (+)-abietane/pimarane-type diterpene phenotype is repeatable, several questions remain regarding if and how a specific outcome will arise. On one hand, positive selection can be invoked. Indeed, one study showed an elevated rate of non-synonymous substitutions at sites expected to control product outcome of class II di-TPSs (Cui et al., 2015), which share common ancestry with the class II/I and class I di-TPSs discussed previously (Figure 2.2). Additionally, specific ecological roles in direct defense, defensive signaling, and allelopathy (Franceschi et al., 2005; Keeling & Bohlmann, 2006; Schmelz et al., 2014) undoubtedly provide fitness advantages that would subject to selection. On the other hand, class I reaction outcome *in planta* depends on which substrate conformation is provided. Thus, evolution of a novel (+)-abietane/pimarane class I cyclization step will be contingent on normal-CPP synthase machinery, which is available to maize and wheat following recent duplications of CPS enzymes (Murphy et al., 2018; Wu et al., 2012) as well as for the Lamiaceae species discussed above (Brückner et al., 2014; Gao et al., 2009; Pateraki et al., 2014). It follows that this

particular outcome of contingency between gene families (TPS-c/CPS and TPS-b/KSL enzymes involved in normally configured diterpene biosynthesis) is of greater consequence following the loss of genes encoding the bifunctional TPS-d/h lineage and following loss of bifunctionality in the TPC-c subfamily (Figure 2.3B).

This dependence on history is not confined to the presence, absence, or number of genes involved. Rather, layered genetic changes can be implicated in specific phenotypic outcomes. For instance, it has been established that domain loss plays a major role in di-TPS evolution and can lead to novel (+)-abietane biosynthetic phenotypes, as was shown in a KSL from *Sal*via *miltiorrhiza* (Hillwig et al., 2011). A deeper genetic layer would be the single mutations that were inherited by modern-day TPS, which will be discussed next.

II.IV.I Sites controlling diterpene synthase product outcomes are of large effect, limited, and pervasive across the broader diterpene synthase phylogeny

Mounting evidence from mutational analyses of di-TPSs also supports a model of constraint in the number of acceptable substitutions that can bring about phenotypic change and the order in which the substitutions must occur. In general, large changes in di-TPS catalysis have been demonstrated after just a few mutations. For example, the (+)-abietadiene synthase from *A. grandis* and TPS-LAS from *Picea abies* can switch from producing strictly (+)-abietanes (**5-8**) to (+)-pimaranes (**13** & **14**) simply by substituting serine for alanine at site 721 (as numbered in Figure 2.3B) (Keeling et al., 2008; Wilderman & Peters, 2007). For PaTPS-ISO, Keeling et al. found that four mutations that caused the amino acid sequence to more closely resemble that of PaTPS-LAS (sites 687, 694, 721, and 725 in Figure 2.3B) had additive effects increasing the proportion of (+)-abietanes in the product profile.

In mutational analyses of KSs and KSLs, which belong to a different orthogroup (Figure 2.4; see table 2.2 for summary of characterized enzymes retrieved by our proteome analysis), a similar trend of large-effect substitutions has been demonstrated that is constrained to homologous sites. In rice (*O. sativa indica*) and *Arabidopsis thaliana* KS and KSLs which normally produce *ent*-kaurene or a double-bond isomer thereof, a single replacement of isoleucine for threonine at site 725 (I725T) redirects carbocation intermediates to form *ent*-pimaradiene as the major product (Xu et al., 2007b).

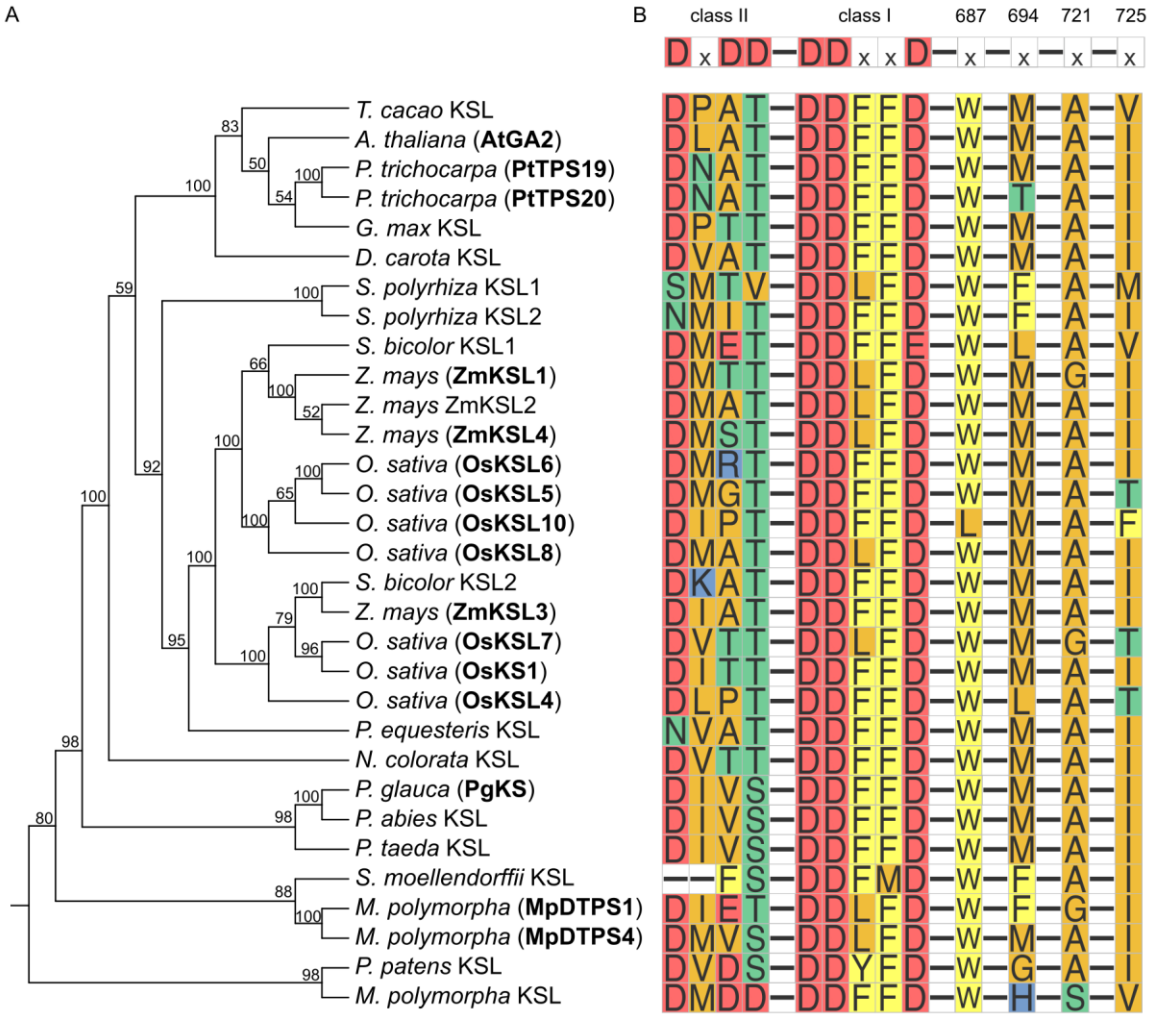


Figure 2.4: C-terminal domain of TPS-e/f subfamily retained a high degree of sequence similarity. A) Maximum-likelihood phylogeny of orthogroup containing the TPS-e/f gene family. Characterized sequences are in bold parentheses. Bootstrap support values of 50 and higher are indicated. Full binomials of species surveyed are given in Appendix Figure 2.1. B) Alignment segments of translated protein products, showing the class II and class I binding motifs and four sites discussed that affect product outcome in disparate lineages. Amino acids colored according to side-chain chemistry (Taylor, 1997; Yu et al., 2017).

Table 2.2. Substrate and product summary of characterized KS and KSL (subfamily TPS-e/f) sequences retrieved from the proteomes examined here

Name	Organism	Substrate	Product	Source
AtGA2	*Arabidopsis thaliana*	*ent*-CPP	*ent*-kaurene	(Yamaguchi et al., 1998)
MpDTPS1†	*Marchantia polymorpha*	*ent*-CPP	atiseran-16-ol*	(Kumar et al., 2016)
MpDTPS4†	*Marchantia polymorpha*	*ent*-CPP	*ent*-kaurene	(Kumar et al., 2016)
OsKS1	*Oryza sativa*	*ent*-CPP	*ent*-kaurene	(Xu et al., 2007a)
OsKSL10	*Oryza sativa*	*ent*-CPP	*ent*-sandaracopimaradiene	(Otomo et al., 2004)
OsKSL4	*Oryza sativa*	*syn*-CPP	*syn*-pimaradiene	(Otomo et al., 2004)
OsKSL5	*Oryza sativa*	*ent*-CPP	*ent*-isokaurene	(Xu et al., 2007a)
OsKSL6	*Oryza sativa*	*ent*-CPP	*ent*-isokaurene	(Xu et al., 2007a)
OsKSL7	*Oryza sativa*	*ent*-CPP	*ent*-cassadiene	(Cho et al., 2004)
OsKSL8	*Oryza sativa*	*syn*-CPP	stemar-13-ene	(Nemoto et al., 2004)
PgKS	*Picea glauca*	*ent*-CPP	*ent*-kaurene	(Keeling et al., 2010)
PtTPS19†	*Populus trichocarpa*	*ent-CPP*	*ent*-kaurene	(Irmisch et al., 2015)
PtTPS20†	*Populus trichocarpa*	*ent*-CPP	16α-hydroxy-*ent*-kaurene**	(Irmisch et al., 2015)
ZmKSL1	*Zea mays*	*ent*-CPP	*ent*-kaurene	(Fu et al., 2016)
ZmKSL2	*Zea mays*	*ent-CPP*	*ent*-kaurene	(Fu et al., 2016)
ZmKSL4	*Zea mays*	*ent*-CPP	dlabradiene	(Mafu et al., 2018)

* minor amounts of *ent*-kaurene and *ent*-atiserene also observed
** minor amounts of *ent*-kaurene and *ent*-isokaurene also observed

A different subspecies of rice (*japonica*) has a wild-type KSL that is 98% identical to one of the *indica* KSLs, which has a threonine instead of isoleucine at site 725 and produces *ent*-pimaradiene pimaradiene, a diastereomer of (**11**) (Kanno et al., 2006). Interestingly, in the same study by Xu et al., mutation of T725I (i.e. the reciprocal substitution) in the *japonica* KSL abolished *ent*-pimaradiene as a product and resulted in a more complex mixture of *ent*-kaurene and *ent*-kaurene-like products. Single amino acids at site 694 can likewise be swapped in *Populus trichocarpa* KSL (see site 694 of PtTPS19 and PtTPS20 in Figure 2.4B), controlling the formation of either hydroxylated or hydrocarbon *ent*-kaurene-like products (Irmisch et al., 2015). In the *ent*-kaurene synthase PgKS from the conifer *Picea glauca*, substituting either an alanine or leucine for methionine at site 694 resulted in cation quenching with water as seen for some of the above enzymes, yielding 16-hydroxy-*ent*-kaurane in addition to *ent*-kaurene in class I reactions with *ent*-CPP (Zerbe et al. 2012a). Further dissection of PgKS identified a substitution of isoleucine with alanine at site 725 (I725A) that nearly abolished production of *ent*-kaurene and resulted primarily in *ent*-pimaradiene.

Such a restriction in the number of sites controlling product outcome is predicted to limit the number of mutations that would result in a novel phenotype; if only few sites control product outcomes the probability decreases that the random process of mutation would produce an adaptive phenotype in the timeframe for which it is selectable (Kimura, 1983). This constraint and widespread gene loss across the land plant phylogeny provide two mechanistic layers that drive the direction of evolution. As will be discussed, it appears that the order in which mutation occurred historically also had a major impact on the phenotypes we see today.

II.IV.II Mutational studies of diterpene synthases suggest that mutational order matters

The paralogous TPS-d/h enzymes from *P. abies*, PaTPS-LAS and PaTPS-ISO, share over 90% sequence identity. As discussed, substituting A721 in PaTPS-LAS for serine abolished all PaTPS-LAS-like product outcomes when assayed with GGPP and resulted in a PaTPS-ISO-like product profile (Keeling et al., 2008; Wilderman & Peters, 2007, Figure 2.3B). Interestingly, the reciprocal substitution alone did not produce any PaTPS-LAS-like products in the mutant PaTPS-ISO (S721A). The serine to alanine substitution in PaTPS-ISO was found to affect product outcome, but only in the presence of a second mutation (H694Y). This site was also found to be consequential for product outcome in the mutational studies at homologous sites mentioned

previously. Thus, a mutational order effect can be discerned in which specific substitutions (a tyrosine at position 694, for example) must be fixed historically before (+)-abietane biosynthesis could evolve in conifer species.

While all plant di-TPSs are considered to have evolved from a bifunctional *ent*-kaurene synthase similar to a CPS found in the moss *P. patens* (Chen et al., 2011), the OrthoFinder result places at least three paralogous di-TPS clades in the ancestor of all land plants: Two which gave rise to the CPS and TPS-d/h subfamilies, and one from which modern-day KS and KSL sequences descended. Assuming a bifunctional *ent*-kaurene synthase ancestor and proper assignments of orthology, duplication of the CPS lineage just before the split of bryophytes from the rest of land plants gave rise to the TPS-d/h lineage, which would have then been redundantly producing *ent*-kaurene and was subsequently lost among most of the taxa investigated here. However, at least in conifers and *S. moellendorffii*, the TPS-d/h family duplicated and acquired novel function-changing substitutions in the regions discussed above (Figure 2.3B). The TPS-c subfamily in angiosperms, containing class II di-TPSs that are paralogous to the TPS-d/h family, appears to have undergone increased divergence in sequence similarity at these sites in angiosperms along with the class I DDxxD motif, in contrast to the orthogroup harboring KSL (Figures 2.3 and 2.4). The angiosperm CPS lineage may thus never have been a suitable candidate for convergent recruitment of genes towards the biosynthesis of specialized class I cyclization reactions in angiosperms. The relatively recent recruitment of KSL enzymes to this end may reflect the long timeframe required for evolution in a constrained pathway.

II.V Chapter Summary

This chapter has reviewed patterns in the recruitment of class I di-TPSs to diterpene biosynthetic pathways in light of the relationships between species harboring them and the repeated evolution of similar and identical diterpene olefins. It is argued that convergence on (+)-abietane and/or pimarane diterpene scaffolds in angiosperms is likely to be rarer than in gymnosperms or other groups possessing bifunctional class II/I di-TPSs already involved in specialized metabolism and have taken much longer due to chance events in the course of evolution. For instance, the fortuitous duplication of CPS genes capable of synthesizing normal-CPP in monocots and the Lamiaceae already provided ample substrate for downstream class I di-TPSs from multiple subfamilies. Secondly, since few mutations are capable of driving functional shifts in di-TPS

mutational analyses, this drastically reduces the probability that any given mutation would be promoted by natural selection. Furthermore, the evolutionary time scales for di-TPS recruitment might be expected to be extended by intramolecular epistasis, which has been shown to occur in closely related di-TPS paralogs reviewed here and may have restricted the order in which mutations arise.

Chapter III Conifers inherited a bifunctional (+)-abietane synthase that was later recruited in parallel towards biosynthesis of specific diterpene scaffolds

III.I Background

Historical context is fundamental to dissecting the molecular basis of modern enzyme catalysis and in understanding the conditions that precede the emergence of novel traits because varying evolutionary histories often lead to divergent outcomes (Hochberg & Thornton, 2017). Plant TPS convert a rather limited number of substrates to thousands of terpene skeletons, which are subsequently converted by other enzymes resulting in tens of thousands of known terpenoids that participate in primary plant growth, multi-trophic signaling networks, and plant defense (Degenhardt et al., 2009; Karunanithi & Zerbe, 2019; Pichersky & Raguso, 2018). It is often the case that specific terpenoid structures are confined to a limited number of taxa due in part to this enormous metabolic expansion (Zhou & Pichersky, 2020). In this study, an historical approach is taken to understand the ancestral conditions that gave rise to diversification of diterpene skeletal structures in the Pinaceae family of gymnosperms and in the nonseed plant *Selaginella moellendorffii*.

Gymnosperms possess duplicated class I and class II/I di-TPS from the TPS-d/h subfamily that synthesize an array of bi- and tri-cyclic diterpene skeletal structures, most of which are further oxidized by P450s (Alicandri et al., 2020). Among the class II/I di-TPSs of conifers are TPS-ISO and TPS-LAS enzymes that convert GGPP via normal-copalyl diphosphate (normal-CPP) to isopimaradiene and sandaracopimaradiene, and 13-hydroxy-(8)14-abietene, respectively, the latter becoming non-enzymatically dehydrated to form four diterpene olefins (Keeling et al., 2011a; Martin et al., 2004; Peters et al., 2000; Zerbe et al., 2012b). Prior to the discovery of the primary TPS-LAS product, TPS-d members from *A. grandis* were demonstrated to produce similar TPS-LAS-like products analyzed under standard GC-MS conditions (Peters et al., 2000). *Ginkgo biloba*, from a different gymnosperm family, possesses an orthologous TPS-d member

that converts GGPP specifically to levopimaradiene (Schepmann et al., 2001). *Taiwania cryptomerioides* (family Cupressaceae) utilizes a paralogous monofunctional TPS-d enzyme (TcKSL3) that converts normal-CPP specifically to levopimaradiene. The enzymes discussed so far facilitate reactions that proceed via normal-CPP. However, *Zerbe* and co-workers (2012b) additionally showed that bifunctional AbCAS from *Abies balsamea* uniquely catalyzes the production of *cis*-abienol, which is a bicyclic diterpene alcohol, resulting from dephosphorylation of labda-13-en-8-ol diphosphate.

Orthologous to the TPS-d lineage is the TPS-h subfamily possessed by some nonseed plants, of which the model lycophyte *S. moellendorffii* has been reported to produce a more limited number of similar diterpene structures. SmMDS and SmCPSKSL convert GGPP via normal-CPP to miltiradiene and labda-7,13*E*-dien-15-ol, respectively, whereas the monofunctional SmDTC3 converts either normal-CPP to sandaracopimaradiene or *syn*-CPP to so far unidentified diterpene alcohols (Mafu et al., 2011; Shimane et al., 2014; Sugai et al., 2011). Finally, the moss *Hypnum plumaeforme* utilizes HpDTC1 to produce *syn*-pimaradiene via *syn-CPP* (Okada et al., 2016).

Structural and functional plasticity, substrate promiscuity, flexibility in conformation of bound substrates, procession via carbocation intermediates, and domain acquisition and loss have all been implicated as causes of the large diversity in product phenotypes observed in modern-day TPSs (Cao et al., 2010; Christianson, 2008; Hillwig et al., 2011; Irmisch et al., 2015; Keeling et al., 2008; Zhou et al., 2012). While these factors are all well supported by the literature, little is known in general about ancestral conditions that preceded today's novel phenotypic outcomes. One possibility is that that the production of specific diterpene scaffolds such as levopimaradiene, *cis*-abienol, and isopimaradiene was inherited anciently by the Pinaceae lineage through an ancestral TPS-d of low product specificity, and that novel enzymes became more specialized. Alternatively, ancestral enzymes may have been more specific, while modern-day di-TPSs have added alternative diterpene products. Both scenarios are feasible as new phenotypes must evolve *de novo* at times, yet ancestral enzymes of low specificity have been shown empirically to lead to diverse sets of descendants (Olson-Manning, 2020; Thornton, 2001; Voordeckers et al., 2012).

As attempts to draw evolutionary conclusions based on modern diversity alone potentially leads to false inferences of ancestral phenotypes (Thornton & Carroll, 2011), we chose to directly test the biosynthetic capacity of an experimentally resurrected ancestral TPS-d. As a first step in

taking such an historical perspective, we estimated the amino acid sequence at the TPS-d node that gave rise to all modern-day characterized TPS-d members among the conifers. This provides a first empirical glimpse into the paleo-chemical past of conifers and allows more ancient ancestors to be similarly resurrected and characterized in the future to test hypotheses of enzyme evolution.

III.II Results

III.II.I Biochemical characterization of AncTPS-CON suggests repeated acquisition of specific diterpenoid scaffolds across specific land plant lineages

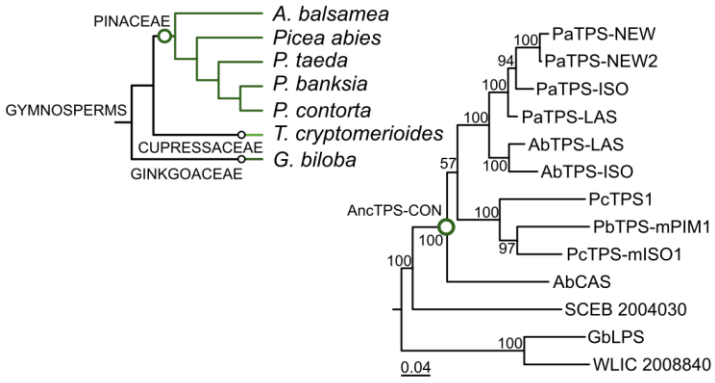


Figure 3.1: A 170-million-year-old enzyme (AncTPS-CON) represents the ancestral enzyme of Pinaceae (+)-abietane and pimarane-type diterpene synthases. Simplified species phylogeny of gymnosperms (Upper-left). Right) Phylogenetic tree of translated genes and SCEB 2004030 (uncharacterized, Podocarpaceae) and WLIC 2008840 (uncharacterized, Zamiaceae) used in estimation of AncTPS-CON.

Maximum-likelihood analysis of selected characterized and uncharacterized Pinaceae-specific TPS-d/h enzymes spanning the gymnosperm phylogeny was performed for ancestral sequence reconstruction of AncTPS-CON. In addition to published characterized sequences, uncharacterized full-length TPS-d members retrieved with BLAST against the oneKP transcriptome database of members outside the Pinaceae family were included. SCEB 2004030 from *Podocarpus coriaceus* (family Podocarpaceae) and WLIC 2008840 from *Dioon edule* (family Zamiaceae) each encode at least one predicted full-length TPS-d member and we included these in the ancestral state estimation (Figure 3.1 & Appendix Figure A3.1). Two full-length transcripts from an in-house transcriptome assembly of adult *P. abies* bark (PaTPS-NEW and PaTPS-NEW2) were also included.

Using the generated gene phylogeny and alignment of the encoded N-terminally truncated TPS-d/h, the sequence at the internal node predating the speciation event that gave rise to individual Pinaceae species was computationally estimated and named AncTPS-CON (Figure 3.2A). AncTPS-CON was estimated with overall high average confidence (average posterior probability > 0.95) and high per-site confidence in the class II and class I aspartate-rich binding motifs as

well as at the four sites previously determined to control product outcome in modern-day TPS-LAS enzymes. Heterologous co-expression of AncTPS-CON with a GGPP synthase resulted in a multi-product profile consisting of palustradiene, dehydroabietadiene, abietadiene, and neoabietadiene (Figure 3.2B). Although this profile closely resembles that of modern-day TPS-LAS enzymes from the genera *Abies, Picea*, and *Pinus* of the Pinaceae, it conspicuously lacks levopimaradiene, which is a distinct product formed by modern-day TPS-LAS enzymes following dehydration of 13-hydroxy-8,14-abietene (Hall et al., 2013; Keeling et al., 2011a; Zerbe et al., 2012, Figure 2.1) produced by modern-day TPS-LAS enzymes.

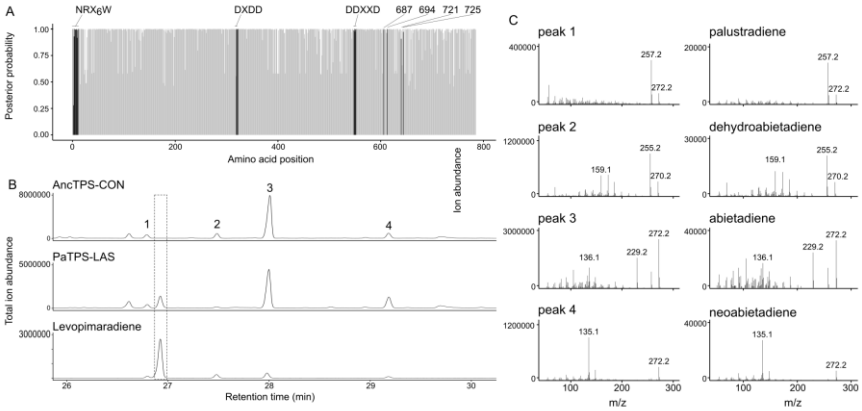


Figure 3.2: AncTPS-CON is a bifunctional, multi-product diterpene synthase that produces only four diterpene olefins. A) Per-site posterior probabilities of AncTPS-CON. The x-axis indicates each site along the N-terminally truncated ancestral conifer diterpene synthase and the y-axis shows confidence values of reconstructed ancestral amino acid states at each position. Black bars and labels indicate posterior probabilities for the NRX_8W motif, class II DxDD motif, class I DDxxD motif, and the four sites demonstrated by Keeling et al. (2008) and Wilderman & Peters (2007) as affecting product outcome (numbering follows usage of PaTPS-ISO by Keeling et al.). B) GC-MS chromatogram of N-terminally truncated AncTPS-CON co-expressed with a GGPP synthase with product peaks labeled 1-4. C) Mass spectra of peaks 1-4 and authentic standards of palustradiene, dehydroabietadiene, abietadiene, and neoabietadiene, respectively.

Previous TPS-LAS studies report that different handling conditions during sample preparation and instrumentation can lead to different ratios of the various double bond isomers of abietane-type diterpenes produced *in vitro*, the previously published PaTPS-LAS (Keeling et al., 2011a; Martin et al., 2004) was expressed alongside AncTPS-CON under similar conditions. PaTPS-LAS produced only the compounds previously reported in the expected ratios, supporting the present characterization of AncTPS-CON.

An inconsistency of conifer di-TPS studies is that diterpene resin acids with pimaradiene and sandaracopimaradiene structures are found in many conifers, but previous work and analyses using OrthoFinder presented here have not produced candidates for a di-TPS capable of

producing these products. To search for these products, the organic phase of PaTPS-LAS extracts was concentrated 250-fold and analyzed by high-resolution GC-MS. In addition to a prominent levopimaradiene peak and peaks for the other four products previously described for this enzyme (and produced by AncTPS-CON using current conditions), trace levels of sandaracopimaradiene were observed, potentially implicating PaTPS-LAS also in the biosynthesis of low amounts of sandaracopimaric acid in modern-day *Picea abies* (Appendix Figure 3.2).

To further investigate how far back in land plant history class I cyclization of normal-CPP might be traced, we attempted to clone and characterize four members of the *S. moellendorffii* TPS-h subfamily that have arisen via relatively recent gene duplications independently in the lycophyte lineage.

III.II.II Alternative diterpene synthases of high sequence identity in *Selaginella moellendorffii* are promiscuous and have different product profiles

Primers targeting SmDTC1, SmDTC2, SmDTC4, and SmDTC5 (see Figure 2.3A for phylogenetic relationship to Pinaceae TPSd/h members) from cDNA generated from aerial tissues, and MeJA-treated aerial, stem, and root *S. moellendorffii* tissue produced varied results. The previously described SmDTC3 was also cloned and assayed as a positive control. For SmDTC5, what appeared to be a full-length coding sequence encoding a class I di-TPS was amplified in PCR reactions. However, this required a large volume of cDNA and the amplified fragment had premature stop codons that appeared to be introduced by an intron. Although we treated all RNA samples with DNase prior to generation of cDNA, it appears that the high concentration of total cDNA may have included residual amounts of genomic DNA. Conversely, SmDTC1, SmDTC2, and SmDTC4 produced no amplification fragments from cDNA from both untreated aerial tissue and all three MeJA-treated tissues at either standard or high volumes of cDNA template. Nevertheless, SmDTC3 was amplified and subcloned, but while most *E. coli* colonies possessed the SmDTC3 construct, a smaller subset had a fragment that encoded a peptide ~97% identical to SmDTC3, which we named SmDTC3.1 (Figure 3.3).

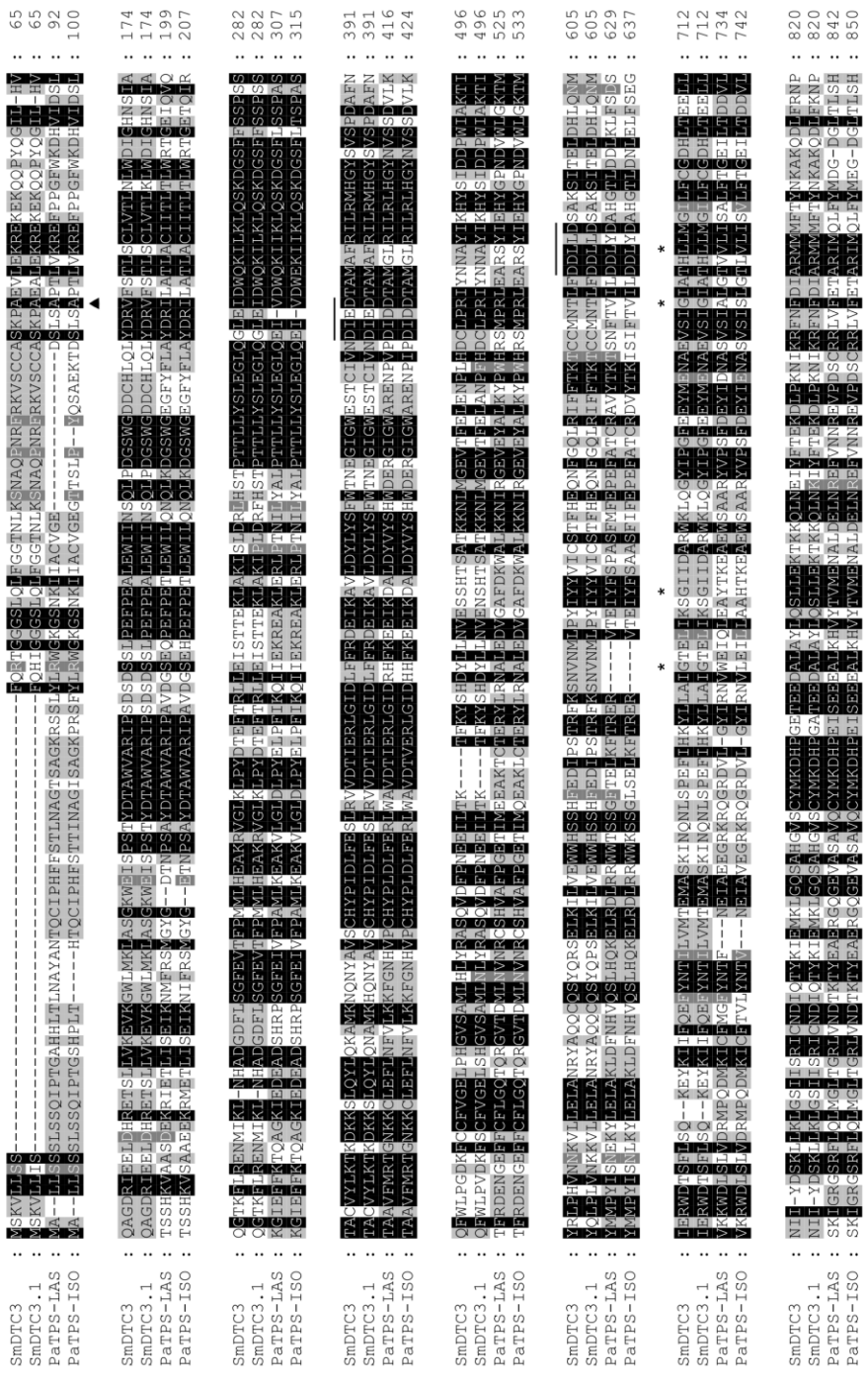


Figure 3.3: SmDTC3 and SmDTC3.1 share high sequence similarity with PaTPS-LAS and PaTPS-ISO. Multiple sequence alignment of SmDTC3, SmDTC3.1, PaTPS-LAS, and PaTPS-ISO. Sites demonstrated to influence product outcome in PaTPS marked by asterisks. Down- and upward-facing arrowheads show site of trunctations for heterologous expression for SmDTC and PaTPS enzymes, respectively. Horizontal bars indicate DxDD and DDxxD motifs.

Presuming this to be an extremely recent paralog or alternate allele of SmDTC3, N-terminally truncated SmDTC3.1 was subcloned with native codons and transformed into engineered *E. coli* cells, as was SmDTC3. SmDTC3 and SmDTC3.1 each possess a predicted C-terminal class I DDxxD motif, but not an N-terminal DxDD class II motif (Figure 3.3). Therefore, SmDTC3.1 was co-expressed with normal-, *syn*-, and *ent*-CPP synthases, which make the typical substrates of class I di-TPSs. SmDTC3.1 was able to convert normal-CPP to sandaracopimaradiene, similar to SmDTC3 (Shimane et al., 2014; Figure 3.4A). Surprisingly, SmDTC3.1 also converted normal-CPP to diterpenes with retention times identical to abietadiene and neoabietadiene, as well as to a fourth product resembling 16α-*ent*-hydroxykaurane. This was unexpected, as the highly similar SmDTC3 was not reported by Shimane et al. (2014) to produce abietane-type diterpene olefins or diterpene alcohols when assayed with normal-CPP, and diterpene alcohols were not detected in the SmDTC3 positive control (Figure 3.4A). The diterpene alcohol product

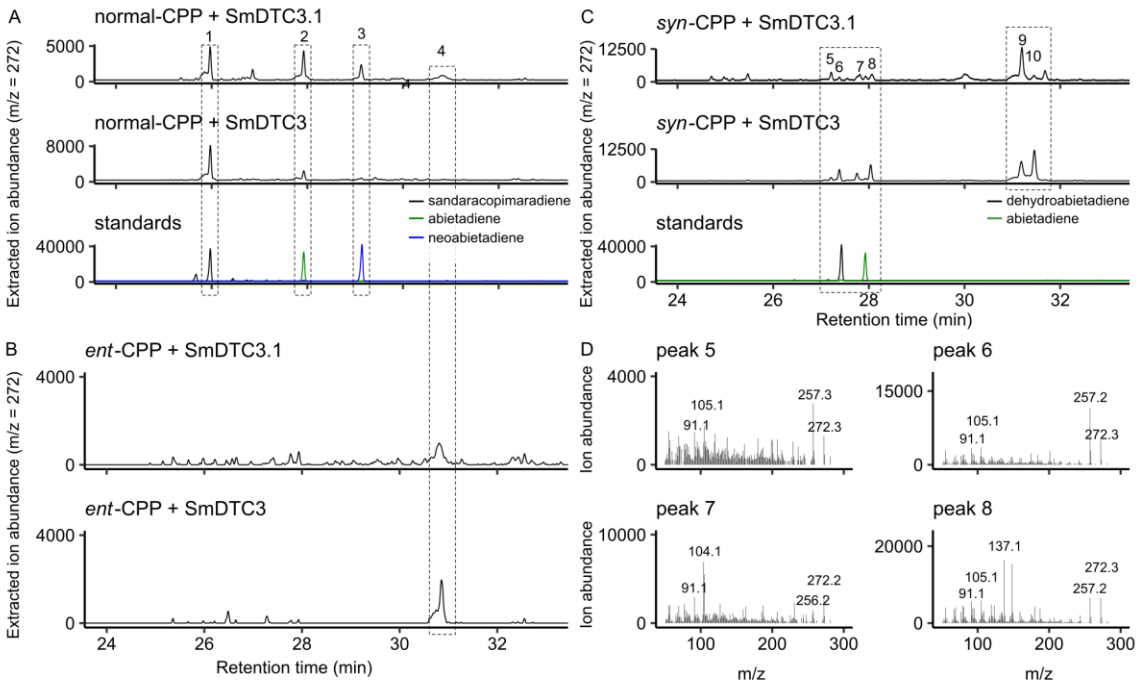


Figure 3.4: SmDTC3.1 is a class I diterpene synthase catalytically distinct form SmDTC3. GC-MS chromatograms of *Escherichia coli* extracts following heterologous expression of N-terminally truncated SmDTC3.1 and SmDTC3 with A) a normal-CPP synthase, and chromatograms of sandaracopimaradiene (black), abietadiene (green) and neoabietadiene (blue) as standards, B) an *ent*-CPP synthase, and C) a *syn*-CPP synthase and dehydroabietadiene (black) and abietadiene (green) as standards. Peak 1, sandaracopimaradiene; peaks 2 & 3, (+)-abietetane-like products; peak 4, 16-hydroxykuarane (putative); peaks 5-8, *syn*-abietane-like products; peaks 9 & 10, unknown diterpene alcohols. See Appendix Figure A3.3 for mass spectra of peaks 1-4. D) Mass spectra of peaks 5-8.

of SmDTC3.1 and normal-CPP was tentatively identified as 16α-hydroxykaurane based on retention time and mass spectrum.

In contrast, SmDTC3.1 resembled SmDTC3 when co-expressed with an *ent*-CPP synthase in producing 16α-*ent*-hydroxykaurane, albeit at only trace levels (Figure 3.4B). When SmDTC3.1 was co-expressed with a *syn*-CPP synthase, four unknown diterpenes and two unidentified diterpene alcohols were detected as was reported for SmDTC3 by Shimane and colleagues (2014) (Figure 3.4C). Interestingly, the retention times of dehydroabietadiene and abietadiene suggest that peaks 5-8 could be abietane-type diterpene olefins (Figure 3.4C, see peaks 5-8). Mass spectra of peaks 5-8 (Figure 3.4D) also resemble the mass spectrum for abietadiene shown in Figure 3.2C. It is also notable that, despite the high sequence identities between SmDTC3 and SmDTC3.1, these two enzymes appear to produce different ratios of the two diterpene alcohols reported for SmDTC3 (Figure 3.4C peaks 9 & 10). A summary of confirmed and putative conversion products is provided in Figure 3.5.

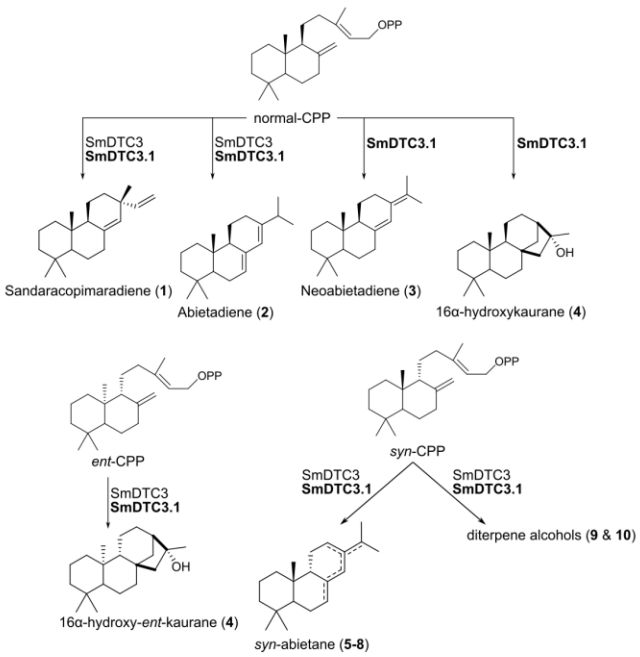


Figure 3.5: Scheme of verified and putative substrates and products of SmDTC3.1 characterized here (bold) and SmDTC3 (Shimane et al., 2014). Numbers correspond to peak numbers in Figure 3.4. Structures 2-8 are tentative assignments.

III.III Chapter Summary

In this chapter, it was demonstrated that the Pinaceae lineage inherited a multi-product class II/I bifunctional di-TPS (AncTPS-CON) active with GGPP as substrate. AncTPS-CON resembled modern-day PaTPS-LAS but had an incomplete set of products that are together produced by modern-day TPS-LAS, TPS-ISO, and AbCAS enzymes. Unfortunately, we were unable to determine whether AncTPS-CON also releases an unstable alcohol that is then dehydrated non-enzymatically to a group of olefins, as is the case for present day Pinaceae TPS-LAS. However, if the unstable alcohol was the sole product, as for the extant TPS-LAS enzymes, we would have expected a similar ratio of olefins to the PaTPS-LAS control, as we employed similar extraction and analytical conditions; but this was not the case. The phylogenetic placement and product profile of AncTPS-CON imply that the ability to form specific carbon skeletons, such as sandaracopimaradiene, evolved independently in Pinaceae and lycophytes such as *S. moellendorffii*, and that levopimaradiene biosynthesis evolved in parallel in the Pinaceae (e.g. *Abies*, *Pinus*, and *Picea* spp.), Ginkgoaceae (*G. biloba*), and Cupressaceae (*T. cryptomerioides*).

Further investigation of two other diterpene synthases (orthologous to the gymnosperm TPS-d/h lineage) from the lycophyte *S. moellendorffii* was also carried out. Although the previously reported SmDTC3 (Shimane et al., 2014) and our SmDTC3.1 are similar in terms of substrate promiscuity, these two enzymes are functionally distinct despite sharing 97% sequence identity. SmDTC3.1 was capable of producing the (+)-abietane-type diterpene olefin neoabietadiene from

normal-CPP, whereas no neoabietadiene was detected in the same assay with SmDTC3. SmDTC3.1 also produced additional hydroxylated diterpene products with normal-CPP, while supplying the enzymes with *syn*-CPP as substrate resulted in different ratios of diterpene alcohols and olefin-like products compared to SmDTC3.

Chapter IV Functional expansion, active site restructuring and gene loss of the CYP720 subfamily of cytochrome P450s shape diterpene diversity in gymnosperms and angiosperms

IV.I Background

P450s in land plants are structurally conserved (but functionally diverse) membrane-bound enzymes that have driven an enormous range of chemical diversity, primarily via NADPH-dependent monooxygenase activity (Bathe & Tissier, 2019; Werck-Reichhart & Feyereisen, 2000). P450s are among the largest plant gene superfamilies, ranging from several dozen to hundreds of genomic copies per species, spanning all land plant lineages (Nelson, 2006). The CYP85 clan of P450s harbors roughly 12 subfamilies of monooxygenases involved in the oxidation of intermediate sterols that constitute the early steps in brassinolide biosynthesis for plant primary growth and in the formation of diterpene intermediates for plant defense (Dejonghe et al., 2014; Hamberger & Bak, 2013; Nelson & Werck-Reichhart, 2011).

Among CYP85 subfamilies are CYP90Bs and CYP7204Bs, which are capable of hydroxylating multiple early and late intermediates of brassinosteroid biosynthesis. For instance, AtCYP90B1 from *Arabidopsis thaliana* and CYP724B2 and CYP724B3 from *Lycopersicum esculentum* oxidize the cholesterol-like steroid intermediate campesterol in the side chain to yield 22-hydroxycampesterol (Fujita et al., 2006; Ohnishi et al., 2006b; Figure 4.1). The x-ray crystal structure of AtCYP90B1 docked with cholesterol, which is also oxidized in the C-22 position, showed that cholesterol is coordinated by water molecules and polar contacts with active site residues such that the D-ring side chain is positioned directly above the reactive heme surface (Fujiyama et al., 2019). CYP90C1 and CYP90D1 from *A. thaliana* were shown to oxidize several sterols at the neighboring C-23 carbon to produce 6-deoxotyphasterol, among other intermediates (Ohnishi et al., 2006a). Conversely, CYP90A1 dehydrates a ring hydroxyl of sterol intermediates to a ketone in *A. thaliana* (Ohnishi et al., 2012). The CYP85A subfamily (also represented in the CYP85 clan) is responsible for the multi-step conversion of late sterol intermediates to biologically active brassinolide or other brassinosteroids by catalyzing the

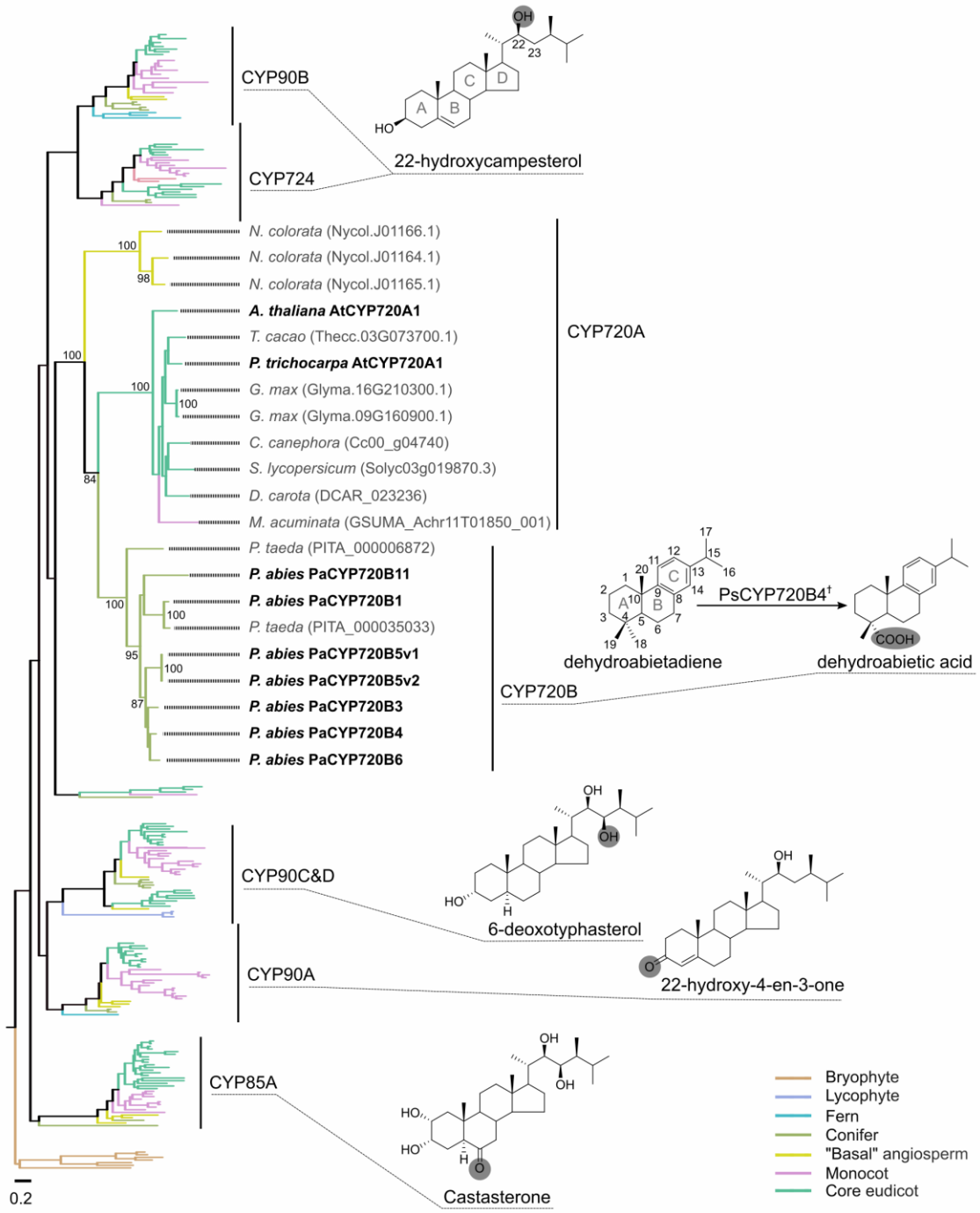


Figure 4.1: The specialized CYP720B family was recruited in conifers from cytochrome P450s involved in brassinolide metabolism. The tree depicts maximum-likelihood phylogeny of the orthogroup containing the land plant CYP85 clan with known sites of oxidation shown as gray circles. C-22 and C-23 of sterols are shown for 22-hydroxycampesterol and numbering for diterpenes is shown for dehydroabietadiene The CYP720 subfamily is expanded to show detail. Scale represents number of substitutions per site. Branch support values (n=1000 bootstrap replicates) above 80 are indicated for the CYP720 subfamily. Tree was rooted with a clade of uncharacterized cytochrome P450 from bryophytes. Color code in lower-right indicates lineages represented in the informatics analyses. "Bryophyte" classification consists of *Marchantia polymorpha*, *Physcomitrella patens*, and *Marchantia polymorpha*. Sequences in bold were cloned in this study. †, Enzyme characterized from *Picea sitchensis* by Hamberger et al. (2011). Full binomials of species surveyed are available in Appendix Figure 2.1.

insertion of a lactone function (Kim et al., 2005; Nomura et al., 2005). *Oryza sativa* CYP85 also oxidizes sterol intermediates, but terminates its reaction after generation of the steroid ketone castasterone rather than a lactone (Kim et al., 2008).

Some members from the CYP720B subfamily have also been characterized and shown to be responsible for the oxidation of diterpene olefins and tertiary alcohols which are products of class I diterpene cyclase activity in the biosynthesis of conifer diterpene resin acids in the Pinaceae. Early insight into this area came with the discovery of P450s capable of oxidizing the C-18 carbon of abietadiene and abietadienol in the conifer *Abies grandis*, yielding the alcohol and aldehyde precursors (respectively) of abietic acid (Funk & Croteau, 1994). PtCYP720B1 from *Pinus taeda* and PsCYP720B4 from *Picea sitchensis* can convert diterpene olefins, alcohols, and aldehydes to their corresponding resin acids (Hamberger et al., 2011; Ro et al., 2005). Several CYP720B2 and CYP720B12 copies from *Pinus* species have additionally been shown to perform C-18 conversion of a methyl group to a carboxylic acid function specifically with 13-hydroxy-(8),14-abietene (a tertiary alcohol) as a substrate, which can subsequently dehydrate non-enzymatically to form four (+)-abietane-type diterpene resin acids (Geisler et al., 2016).

As will be discussed in further detail, the CYP720B family duplicated multiple times independently in the genus *Picea* and at least once prior to the *Picea-Pinus* ancestor. All CYP720Bs are additionally orthologous to the uncharacterized CYP720A subfamily in angiosperms (Figure 4.1). However, angiosperms only rarely produce normally configured tricyclic diterpene scaffolds that could potentially be oxidized at the same C-18 position by angiosperm P450s as is seen in conifers. Possibly of greater interest is that both angiosperms and gymnosperms retained copies of the CYP720 subfamily since the ca. 350-million-year-old split of these two lineages (Morris et al., 2018) despite diterpene resin acid biosynthesis being unique to conifers. As no CYP720B family member has been biochemically characterized from Norway spruce, the functional space of CYP720B members from this species was further investigated to determine if the CYP720 subfamily is capable of a more diverse set of reactions. An expanded functional capacity of CYP720 enzymes beyond oxidation of diterpenes to resin acids can be postulated because the broader CYP85 clan appears to have a highly diverse range of substrates, position of oxidation, and resulting functional groups. Attempts to characterize CYP720A from the angiosperms *Populus trichocarpa* and *A. thaliana* are also described in this chapter.

IV.II Results

IV.II.I The CYP720 subfamily expanded in gymnosperms following recruitment from enzymes involved in brassinosteroid biosynthesis, but not in angiosperms

To determine the minimal set of CYP720 family members in *P. abies* and to resolve the origin, patterns of inheritance, and gene duplication history of this lineage, all P450 orthologs of previously characterized *P. abies* and *P. taeda* CYP720B members were first extracted from 25 genomes representing the land plant tree of life using OrthoFinder (Appendix Figure A2.1 contains species from which genomes were used). The CYP720 lineages from both angiosperms and gymnosperms were confined to a single orthogroup that was otherwise restricted to CYP85 members implicated in brassinosteroid biosynthesis (Figure 4.1; see Appendix Figure A4.1 for full bootstrap support values and sequence accessions). P450s involved in gibberellic acid biosynthesis (i.e. the CYP88 and CYP701 subfamilies) were absent within this orthogroup. This was expected for CYP701 as it comes from a different P450 clan than the one in focus here. However, the CYP88 subfamily is considered a member of the CYP85 clan (Nelson & Werck-Reichhart, 2011), suggesting that CYP88s are paralgous to the CYP85 clan. The remaining identified subfamilies from this clan were the CYP90Bs, represented in ferns and seed plants, and CYP724s, which appeared to be restricted to seed plants. These two subfamilies together appear to share a paralogous relationship, having duplicated prior to the separation of ferns and seed plants as inferred by fern representatives among the CYP90Bs (Figure 4.1). Members of the CYP90D and CYP90C subfamilies were also present, which share a paralogous relationship to each other and are represented in most lineages studied. CYP85A members were also present with representatives from vascular plants, while bryophyte sequences formed a single clade used to root the tree in Figure 4.1.

The CYP720 lineage harbors both the angiosperm CYP720A lineage (from which no enzyme has yet been characterized), and the conifer CYP720B lineage (Figure 4.1). CYP720s appears to have patterns of inheritance and duplication distinct from the other subfamilies mentioned above. Of the eight monocots and two basal angiosperms assessed, only one monocot (*Musa acuminata*) and one basal angiosperm (*Nymphaea colorata*) had representatives in the CYP720A subfamily. It should be noted that the placement of the three *N. colorata* paralogs as a neighbor taxon to the 720 subfamily is likely false; deep divergence times at the base of the angiosperm lineage could

contribute to the observed phylogenetic discordance. Nevertheless, we obtained a branch support value of 100 for the CYP720A lineage, which argues strongly for the exclusion of nonseed plants from the CYP720 families and suggests an origin at the base of seed plants.

IV.II.II Tissue expression and cloning of candidate CYP720 enzymes

Database searches resulted in seven *P. abies* CYP720B candidate genomic loci: MA_10427376g0010, MA_266173g0010, MA_46522g0010, MA_165926g0020, MA_178572g0010, MA_110777g0010, MA_131290g0010 (also identified but not characterized in (Geisler et al., 2016), which we renamed PaCYP720B1, PaCYP720B3, PaCYP720B4, PaCYP720B5v1, PaCYP720B5v2, PaCYP720B6, and PaCYP720B11, respectively. Note that protein products deduced by MA_165926g0020 and MA_178572g0010 are indistinguishable from each other, bringing the total number of transcripts and corresponding candidate protein products in *P. abies* to six.

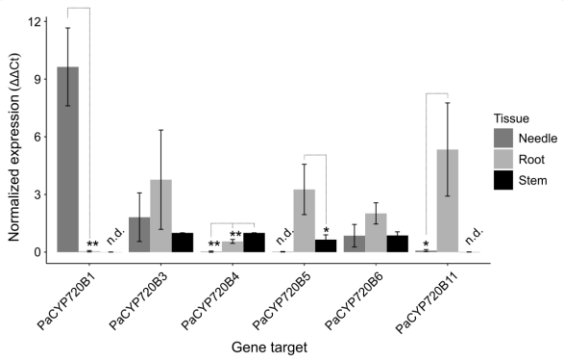


Figure 4.2: PaCYP720B enzymes have specialized tissue expression patterns. Depicted are expression values relative to zero (normalized to ubiquitin) of each cytochrome P450 gene in three tissues of 3-year-old *Picea abies* saplings. Legend provides tissue type from which expression is reported. Asterisks indicate adjusted p-value of means compared to the highest expression of each gene following TukeyHSD and one-way ANOVA tests (one asterisk, $p \leq 0.05$; two asterisks, $p \leq 0.005$).

Gene expression analysis in needles, roots, and stems of the six CYP720B candidates show that tissue expression patterns differed widely between paralogs (Figure 4.2). PaCYP720B1 and PaCYP720B11 appear to be highly specialized for expression in needle and roots, respectively, as their expression values were at or below our detection limit in other tissues. In needles, PaCYP720B1 expression was found at an abundance approximately 200 times greater than that found in roots ($p < 0.001$) whereas no expression was detected in stems. PaCYP720B11 expression was similarly not detected in stems but was ca. 65 times more abundant in roots than

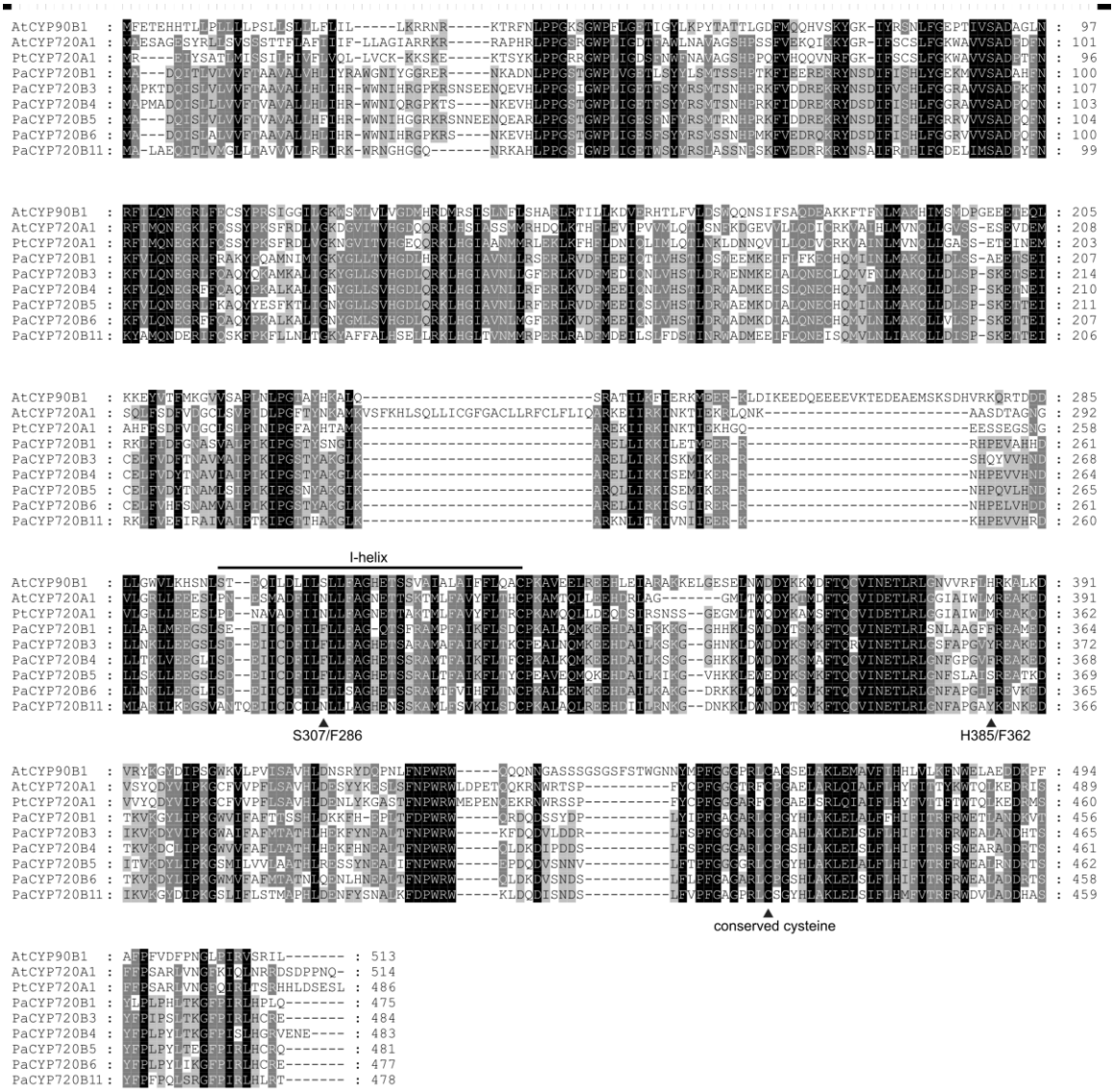


Figure 4.3: Multiple sequence alignment of AtCYP90B1 and CYP720 members examined here. I-helix is shown with horizontal bar. Conserved cysteine and discussed active site residues predicted to affect substrate binding indicated by arrowheads. Grayscale indicates level of amino acid residue conservation from not conserved (white) to highly-conserved (black).

in needles ($p = 0.022$). PaCYP720B3 and PaCYP720B6 appeared more broadly expressed, as we found no significant differences in transcript abundance across the three tissue types. For PaCYP720B4, the *P. abies* ortholog of the previously characterized PsCYP720B4 from *P. sitchensis*, its expression was just above our detection limit in needles but had normalized expression values of approximately 0.5 and 1 ($p < 0.001$) in roots and stems, respectively, with significant differences in abundance between all tissue types. PaCYP720B5 was not expressed in needles and was expressed approximately 5-fold higher in roots than in stems ($p = 0.034$), suggesting that this P450 also shares a degree of tissue specialization.

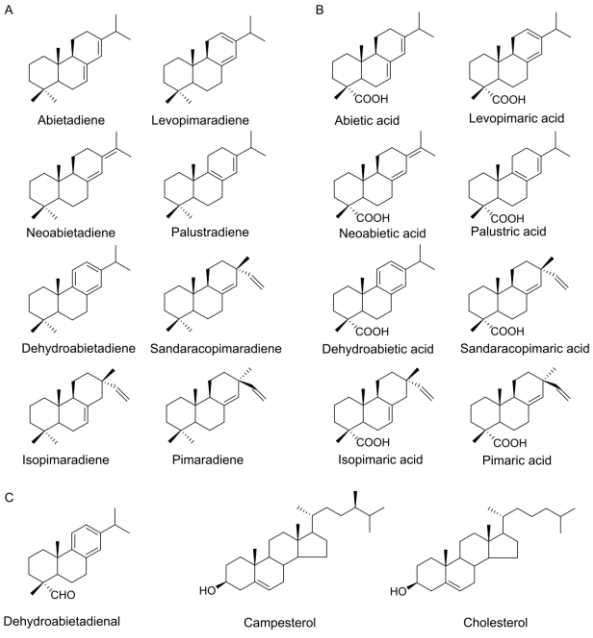


Figure 4.4: Chemical structures of potential CYP720 substrates and products. A) Diterpene olefin substrates used in assays with CYP720 enzymes. B) Diterpene resin acid products of PaCYP720B4. C) Chemical structures of dehydroabietadienal, campesterol and cholesterol.

We therefore cloned PaCYP720B3, PaCYP720B4, PaCYP720B5 and PaCYP720B6 from cDNA from *P. abies* sapling stems, and PaCYP720B1 and PaCYP720B11 from needle and root cDNA, respectively. The resulting sequences encoded full-length (~450 amino acid) P450s (Figure 4.3) and we expressed codon-optimized versions of each in the vector pESC-LEU and WAT11U *Saccharomyces cerevisiae* cells, the latter containing a genome-integrated P450 redox partner, cytochrome P450 reductase, from *Arabidopsis thaliana* (Urban et al., 1997). Isolated microsomes were then assayed with panels of diterpene olefins (Figure 4.4) that could potentially serve as substrates.

IV.II.III Characterization of P450s from the CYP720B subfamily

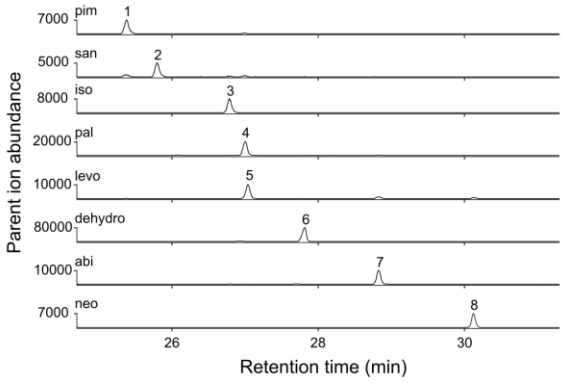


Figure 4.5: PaCYP720B4 converts eight diterpene olefins to their corresponding resin acids. Depicted are GC-MS extracted ion chromatograms of the methyl ester derivatives of the assay products following extraction of microsomes incubated with olefin substrates. Substrates used in each assay are indicated on individual chromatograms: pim, pimaradiene; san, sandaracopimaradiene; iso, isopimaradiene; pal, palustradiene; levo, levopimaradiene; dehydro, dehydroabietadiene; abi, abietadiene; neo, neoabietadiene. Peak numbers correspond to methyl ester derivatives of 1, pimaric acid; 2, sandaracopimaric acid; 3, isopimaric acid; 4, palustric acid; 5, levopimaric acid; 6, dehydroabietic acid; 7, abietic acid; 8, neoabietic acid.

After heterologous expression, *in vitro* assay of the PaCYP720B4 protein with eight abietane and pimarane-type diterpene olefins: pimaradiene, sandaracopimaradiene, isopimaradiene, palustradiene, levopimaradiene, dehydroabietadiene, abietadiene, and neoabietadiene (see Figure 4.4 for chemical structures of substrates tested and verified products) resulted in the conversion of each substrate to its corresponding C-18 carboxylic acid (Figure 4.5) as shown for PsCYP720B4 in previous work (Hamberger et al., 2011). Assays of PaCYP720B4 with sandaracopimaradiene and levopimaradiene as substrate contained minor amounts of other diterpene olefin impurities that were also oxidized in the same reactions. For instance, our sandaracopimaradiene standard comprised approximately 20% pimaradiene with additional trace levels of isopimaradiene and palustradiene. When assaying PaCYP720B4, pimaradiene was also oxidized in the same reaction vessel to yield a similar 4:1 ratio of sandaracopimaric to pimaric acid and trace levels of isopimaric and palustric acids. This is not surprising since previously characterized CYP720B enzymes had a $K_m \approx 14$ μM with diterpene olefin substrates (Hamberger et al., 2011), and assay was conducted with a 100 μM concentration of total olefin substrate for 1 h. A similar pattern was found when assaying PaCYP720B4 with levopimaradiene. Although the synthesized levopimaradiene was >99% pure, a minor amount appeared to spontaneously rearrange to abietadiene after storage time at -20°C such that approximately 5% of the levopimaradiene substrate delivered to PaCYP720B4 was abietadiene, accounting for the small peak corresponding to abietic acid in Figure 4.5.

Of the eight diterpene olefins assayed with PaCYP720B3, this enzyme was active only with dehydroabietadiene (Figure 4.6). Surprisingly, PaCYP720B3 converted dehydroabietadiene at

the C-18 position solely to the secondary alcohol dehydroabietadienol (Figure 4.6C), whereas dehydroabietic acid was not detected in derivatized samples. A secondary peak with trace-level ions characteristic of a diterpene alcohol was also present in the products of this reaction. Although it was not possible to identify this smaller peak using any of eight diterpene alcohol standards, it appears that PaCYP720B3 oxidizes dehydroabietadiene to yield multiple alcohol products.

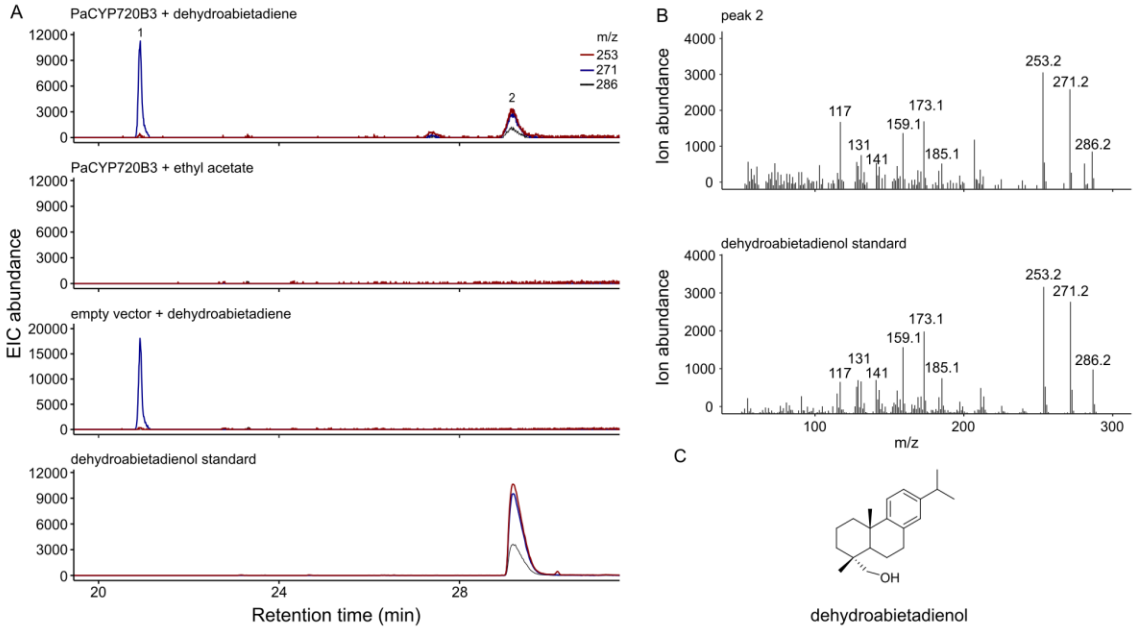


Figure 4.6: PaCYP720B3 converts dehydroabietadiene to dehydroabietadienol. A) GC-MS extracted ion chromatograms representing the molecular ion and characteristic fragments: m/z = 253 (dark red); m/z = 271 (blue); m/z = 286 (black). From top to bottom: PaCYP720B3 incubated with dehydroabietadiene (peak 1) and and resulting product (peak 2), PaCYP720B3 with no substrate added, empty vector yeast microsomal control incubated with dehydroabietadiene, and dehydroabietadienol standard. B) Mass spectrum of peak 2 and authentic dehydroabietadienol standard. C) Chemical structure of product dehydroabietadienol.

Similar to PaCYP720B3, PaCYP720B1 was found to be active only with dehydroabietadiene (Figure 4.7A). However, unlike the other Pinaceae CYP720 enzymes involved in diterpene resin acid biosynthesis, PaCYP720B1 did not oxidize dehydroabietadiene at the C-18 carbon. Instead, PaCYP720B1 converted dehydroabietadiene to a diterpenoid with a mass spectrum highly similar to that of the phenolic diterpene ferruginol (Figure 4.7B & C), compared to the NIST mass spectral library, in addition to at least one other unknown diterpene alcohol of approximately

equal abundance. This suggests oxidation at C-12 of dehydroabietadiene, implying an opposing orientation of the substrate within the active site of PaCYP720B1.

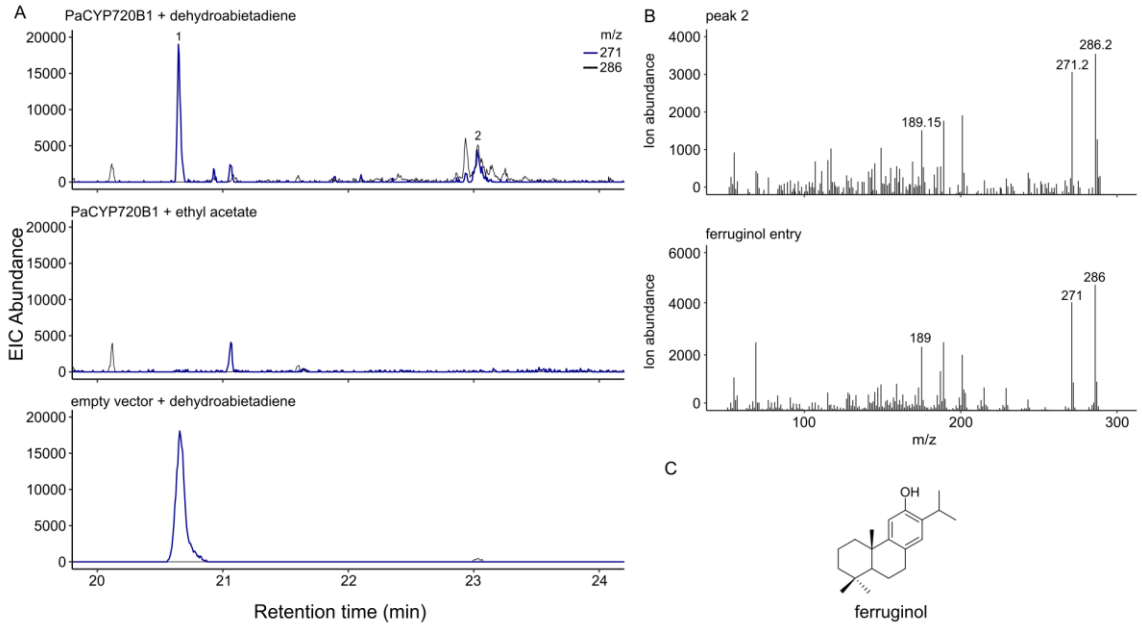


Figure 4.7: PaCYP720B1 converts dehydroabietadiene to multiple unknown diterpene alcohols. A) GC-MS extracted ion chromatograms representing the molecular ion and [M-15]* fragment: m/z = 271 (blue); m/z = 286 (black). From top to bottom: PaCYP720B1 incubated with dehydroabietadiene (peak 1) and and resulting product(s) (peak 2), PaCYP720B1 with no substrate added, empty vector yeast microsomal control incubated with dehydroabietadiene. B) Mass spectrum of peak 2 and NIST database entry for ferruginol. Note the small peak near 23 minutes in the empty vector control does not have a prominent m/z = 271 ion. C) Chemical structure of ferruginol.

PaCYP720B5, PaCYP720B6, and PaCYP720B11 were not active with any of the eight assayed diterpene olefins. Each enzyme was additionally screened for activity with eight diterpene alcohols representing intermediates of diterpene resin acid biosynthesis and the C-18 aldehyde dehydroabietadienal (See Figure 4.4C for chemical structure), but yielded no products. These three enzymes also had no activity with campesterol or cholesterol (structures given in Figure 4.4C), both of which are oxidized by members of the paralogous CYP90B and CYP724B subfamilies (Fujita et al., 2006; Ohnishi, et al., 2006b; Figure 4.1).

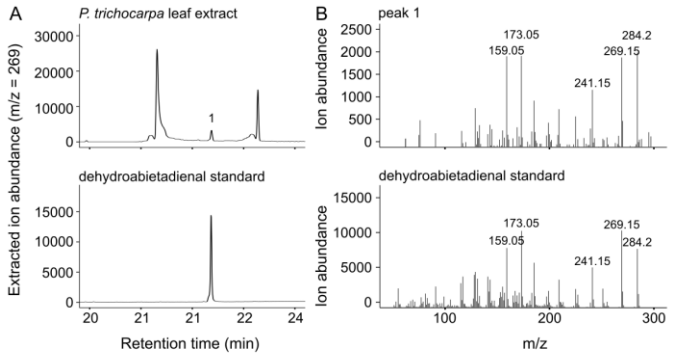


Figure 4.8: Dehydroabietadienal (DA) is present in leaves of *Populus trichocarpa* seedlings. A) GC-MS chromatogram of purified DA from leaf tissue and authentic standard. B) Mass spectra of peak 1 and dehydroabietadienal standard. Due to the low abundance of DA and its co-retention with a short-chain hydrocarbon present in *P. trichocarpa* seedlings, data are presented as an extracted ion rather than total ion chromatogram. Structure for DA given in Figure 4.4.

As dehydroabietadienal (DA) has been reported from the leaves of *A. thaliana* (Chaturvedi et al., 2012) and is structurally similar to diterpene olefins present in conifers (Figure 4.4A & C), a methanol extract of *P. trichocarpa* leaves was extracted and analyzed on GC-MS to search for the occurrence of DA in other angiosperms and two potential DA precursors were assayed with angiosperm CYP720As. The presence of DA in *P. trichocarpa* leaves was confirmed at a low level: 70.5 ± 11.6 ng g^{-1} dry weight (Figure 4.8). However, the angiosperm CYP720A members from *A. thaliana* and *P. trichocarpa* were inactive with both dehydroabietadiene and dehydroabietadienol, both of which could be further oxidized to DA. These two angiosperm CYP720A enzymes were also inactive with campesterol and cholesterol. *A. thaliana* has not been reported to produce normally configured diterpenes. Thus, if an angiosperm CYP720A is involved in DA accumulation, it could be oxidizing *ent*-miltiradiene, or dehydroabietadiene (10*R*) following non-enzymatic deprotonation of *ent*-miltiradiene, as *A. thaliana* and *P. trichocarpa* both produce *ent*-CPP (Irmisch et al., 2015; Sun & Kamiya, 1994).

In order to provide structural insight into novel defense metabolism of CYP720B enzymes, homology modeling of PaCYP720B4 was performed in relation to the paralogous CYP90B1 of *A. thaliana*, whose structure bound to cholesterol has recently been resolved (Fujiyama et al., 2019).

IV.II.IV Homology modeling and molecular docking simulation with PaCYP720B4

PaCYP720B4 and AtCYP90B1 share approximately 39% sequence identity. However, even highly divergent P450s are known to have high rates of structural conservation (Hasemann et al., 1995). Homology modeling of PaCYP720B4 to AtCYP90B1 yielded a 3D structure for PaCYP720B4 with a global quality model estimation of 0.69 and a QMEAN of -2.21, which fall within range of acceptable quality metrics. Figure 4.9A shows the I-helix of AtCYP90B1, which is known to span the active site proximal to the catalytic surface of the heme group and to be involved in proton transfer (Fujiyama et al., 2019; Hasemann et al., 1995; Werck-Reichhart & Feyereisen, 2000). Also depicted is the heme with coordinating cysteine, serine 307 (S307) and histidine 385 (H385), as well as all I-helix residues within 5Å of cholesterol as resolved by (Fujiyama et al., 2019). The active site of PaCYP720B4 replaces S307 of AtCYP90B1 with phenylalanine 286 (F286), drastically reducing the estimated distance of this site to the side chain of cholesterol (Figure 4.9B). H385 of AtCYP90B1 was replaced by phenylalanine 362 (F362) in PaCYP720B4, resulting in the loss of a polar contact between the hydroxyl group of cholesterol and histidine at this site (Fujiyama et al., 2019).

Docking simulation with dehydroabietadiene, which is oxidized at the C-18 methyl position (see Figure 4.4 for numbering), resulted in dehydroabietadiene being oriented such that the C-18 and -19 A-ring methyl groups were proximal to the PaCYP720B4 heme surface (Figure 7C). Tracing amino acid substitutions at the sites mentioned above revealed that all CYP720 members in angiosperms and gymnosperms have aromatic amino acids (either a tryptophan or phenylalanine) in their I helices at site 307/286, which historically replaced the smaller serine found at the same position of AtCYP90B1 (Figure 4.9C). This replacement would likely be incompatible with the binding of steroids oriented in the same manner as found in AtCYP90B1. Additionally, five out of the six PaCYP720Bs, as well as both angiosperm CYP720As, have amino acids with hydrophobic side chains at a position critical for coordination of cholesterol and campesterol in AtCYP90B1 via polar contact with the A-ring hydroxyl group (lettering shown in Figure 4.1).

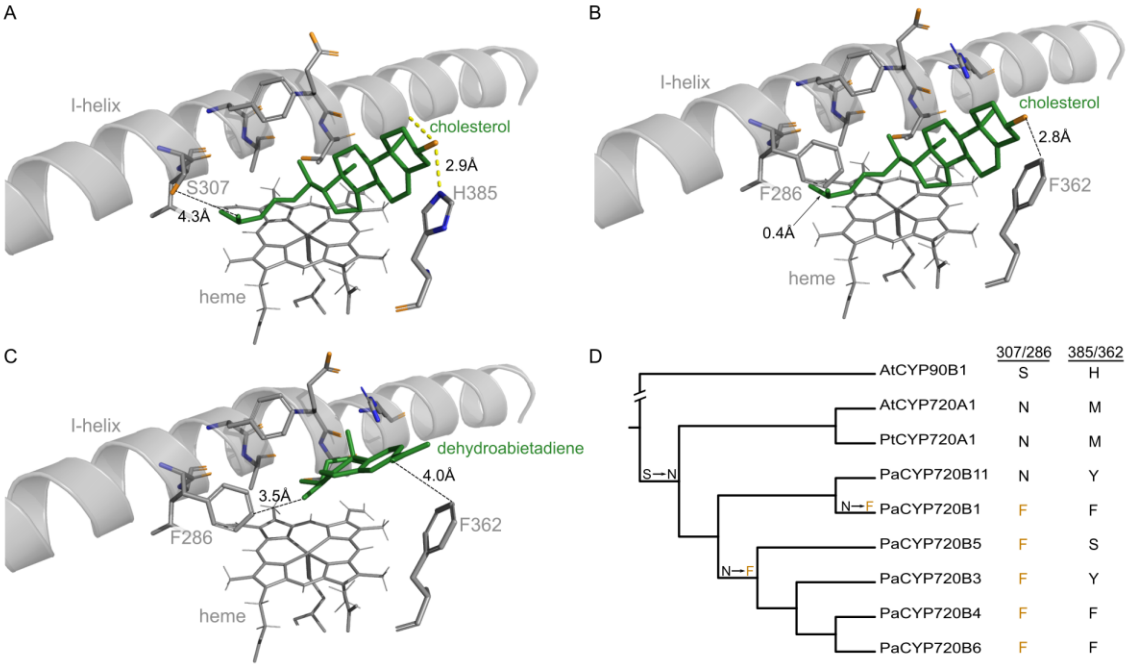


Figure 4.9: Novel mutations in *Picea abies* CYP720B enzymes are predicted to have introduced steric clashes and loss of polar contacts with the steroid substrates of their putative ancestors. A) Empirically-derived binding position of cholesterol with *Arabidopsis thaliana* AtCYP90B1 as previously reported (Fujiyama et al., 2019). B) PaCYP720B4 with cholesterol super-imposed from (A). C) PaCYP720B4 after docking simulation with its native substrate, dehydroabietadiene. Substrates are shown in dark green. I-helix amino acids within 5.0Å of cholesterol are shown in gray. S307 and H385 of AtCYP90B1, F286 and F362 of PaCYP720B4, and heme with coordinating cysteine are also shown in gray. Polar contacts are depicted by yellow dotted line and dotted lines with measurements denote distances in angstroms. Water molecules are not shown. Nitrogen and oxygen are shown in blue and orange, respectively. D) Simplified phylogeny of Pinaceae and dicot CYP720 subfamily rooted with *A. thaliana* AtCYP720A1 and amino acid states at sites 307/286 and 285/360 (numbered according to AtCYP90B1 and PaCYP720B4, respectively). Orange amino acids are predicted to introduce a steric clash with steroids and inferred branches of mutational transitions are indicated with arrows.

IV.III Chapter Summary

This chapter investigated diversification of the CYP720 family, which was found to have emerged via duplication of genes encoding brassinosteroid oxidation enzymes in seed plants just prior to the angiosperm-gymnosperm speciation event. The gene copy number of the CYP720 lineage was found to be constrained among dicots whereas monocots appear to have lost their CYP720 orthologs. Norway spruce expresses at least six CYP720B genes that arose as paralogs of each other via more recent gene duplications both prior to and following radiation of the Pinaceae family, leading to at least four *Picea*-specific CYP720B copies. Some CYP720B

paralogs in Norway spruce also show signs of expression tailored to specific tissues, whereas others are expressed more broadly.

Biochemical characterization of three of these paralogs indicate that Norway spruce CYP720B representatives might have narrow substrate ranges, as both the anciently-branching PaCYP720B1 and PaCYP720B3 (the latter of more recent origin) were only active with dehydroabietadiene. Interestingly, the orthologous CYP720A lineage in angiosperms were not active with any normally configured diterpene olefins tested, nor were any CYP720A proteins active with campesterol or cholesterol. PaCYP720B4, which is among the youngest of Norway spruce CYP720B paralogs, was conversely active with a wide panel of diterpene olefins. Homology modeling using the resolved structure of AtCYP90B1 as a template indicated active site restructuring that is predicted to block the binding of the larger steroids that may have been substrates for the ancestors of this lineage.

Chapter V Manipulation of biosynthetic enzymes provides insight into the regulation of conifer diterpene formation

V.I Background

A hallmark of Norway spruce chemical defense common to other conifers from the family Pinaceae as well, is the abundance of monoterpenes, sesquiterpenes and diterpene resin acids in oleoresins, which exude in large quantities from damaged bark and needles (Keeling & Bohlmann, 2006; Zulak & Bohlmann, 2010). The terpenoid content of resin provides physical and chemical barriers against pests and microbial pathogens. The abundance of oleoresin in conifer tissue is achieved in part by its deposition in specialized secretory structures referred to as resin ducts, which act as reservoirs for resin storage prior to herbivore or pathogen attack. In addition to constitutive ducts in needles and stems, traumatic resin ducts in the secondary xylem of stems can also be induced by herbivore attack, pathogen infection, or mechanical wounding via a jasmonate signaling cascade, thereby providing additional resin sources in response to enemy attack (Mageroy et al., 2020; Martin et al., 2002; Schmidt et al., 2011).

Several terpenoids from spruce species have been shown to exert biological effects on attacking and/or colonizing insects and their fungal symbionts. The volatile monoterpenes from spruce bark are toxic to adult bark beetles (Franceschi et al., 2005). These compounds are also proposed to act as a form of solvent to lower the viscosity of diterpene resin acids and allow for their outflow from damaged tissue, thereby forming a physical and chemical barrier to prevent further feeding and to immobilize natural enemies (Keeling & Bohlmann, 2006). High diterpene resin acid concentrations in *Picea sitchensis* are associated with resistance to weevil feeding and appear to have feeding-deterrent properties against *Lymantria dispar* (Byun-McKay et al., 2006; Powell & Raffa, 1999; Tomlin et al., 1996, 1998). A small number of diterpene resin acids have also been shown to inhibit fungal growth and prevent sporulation (Kopper et al., 2005).

The biosynthesis of all plant diterpenoids begins with the condensation of DMAPP and IPP under catalysis by IDS enzymes, which was reviewed in Chapter I (Cordoba et al., 2009; Nagel et al.,

2019; Zulak & Bohlmann, 2010). Some IDSs are known to catalyze a single condensation of IPP and DMAPP, yielding the 10-carbon isoprenyl diphosphate geranyl diphosphate (GPP), as is the case for PaIDS2 (Schmidt & Gershenzon, 2007b). However, PaIDS3 from Norway spruce condenses DMAPP successively with either two or three molecules of IPP too yield the C_{15} and C_{20} isoprenyl diphosphates farnesyl diphosphate (FPP) and geranylgeranyl diphosphate (GGPP), respectively. Thus, IDS enzymes can have high or low product specificity, and also vary in their substrate specificities. For example, PaIDS1 can accept DMAPP and IPP to yield GPP *in vitro* (Schmidt et al., 2010). PaIDS1 can also condense GPP with one or two more units of IPP to produce FPP and GGPP, respectively or FPP with one more unit of IPP to produce GGPP. Some IDSs are also inducible via jasmonate spraying, as is the case for the GGPP synthase PaIDS5 (Schmidt & Gershenzon, 2007a).

GGPP is formed in plastids and is a substrate for plastid-localized class I diterpene synthases (di-TPSs) that catalyze the Mg^{2+}-dependent ionization of the phosphate group and the conversion of the resulting carbocation to an immense range of cyclic and acyclic diterpene hydrocarbons and oxygenated compounds (Chen et al., 2011; Degenhardt et al., 2009; Peters, 2010). As mentioned in previous chapters, many of these olefins undergo further rearrangements and oxidative steps catalyzed by cytochromes p450s (P450s) that ultimately yield bioactive diterpenes utilized in plant chemical defenses as described above.

Due to the economic importance of spruce species and interest in the ecological roles of conifer resins, efforts have been made to create transgenic lines modified at various points in the diterpene biosynthetic pathway to understand *in planta* regulation of diterpene formation and factors controlling their accumulation. For instance, overexpression of PaIDS1 (a bifunctional GPP/GGPP synthase *in vitro*) in Norway spruce did not lead to an increase in either mono- or diterpenoids *in planta*, but instead to an overabundance of geranylgeranyl fatty acid esters (Nagel et al., 2014). Conversely, RNA interference (RNAi) experiments that targeted PgCYP720B4 in *Picea glauca*, which is biochemically analogous to PaCYP720B4 of Norway spruce, led to slight decreases in only some diterpene resin acids despite a strong interference effect (Hamberger et al., 2011). As much remains unknown about *in planta* control over terpenoid abundance in conifers, a combination of transgenic experiments and phytohormone treatments were chosen to investigate three potential regulatory steps in terpene biosynthesis in Norway spruce: an isoprenoid synthase with predicted *in planta* GPP synthase activity (PaIDS2), bifunctional class

II/I di-TPSs that form (+)-abietane and pimarane olefins *in vitro* (PaTPS-LAS and PaTPS-ISO), and conifer-specific cytochromes p450 from the CYP720B subfamily.

V.II Results

V.II.I PaIDS2 has unexpected influence on diterpene resin acid abundance

Overexpression of PaIDS2 in Norway spruce seedlings showed that whole-stem mono- sesqui- and diterpene contents were largely unaffected in overexpression lines as compared to empty vector and wild-type controls (Figure 5.1A-C). Likewise, needle mono- and sesquiterpene content were unaffected in overexpression lines (Figure 5.1D & E). As PaIDS2 does not appear to be expressed normally in wild-type needles, unlike PaIDS1 and PaIDS5 (Figure 5.1F), this phenotype was not considered unusual. However, we did predict a decrease in stem monoterpene abundance, as PaIDS2 was originally cloned and characterized as a GPP synthase from this tissue (Schmidt & Gershenzon, 2007b).

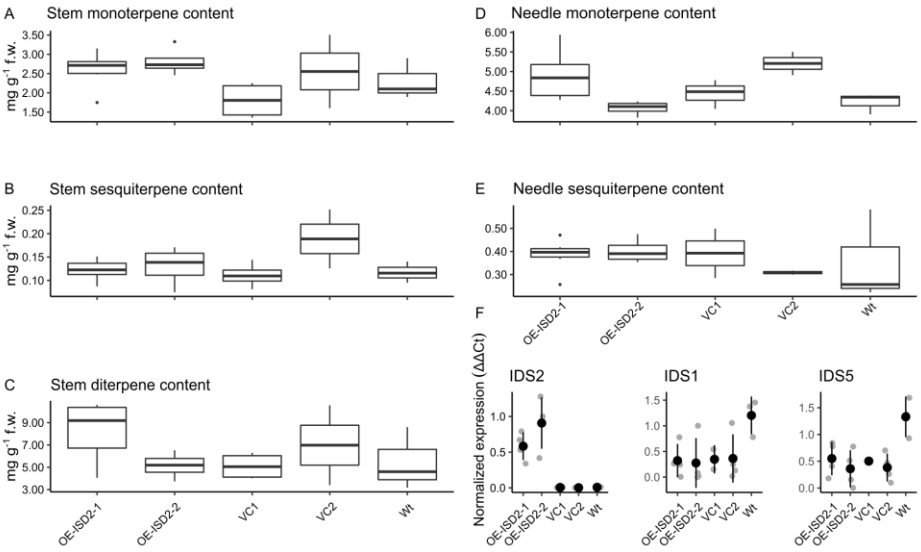


Figure 5.1: Overexpression of PaIDS2 does not affect monoterpene or sesquiterpene content in *Picea abies* bark or needles. A-C) Total mono-, sesqui- and diterpenoid content as determined by quantitative GC-MS, respectively, of transgenic *P. abies* seedling stems. D & E) Total needle mono- and sesquiterpene content as determined by quantitative GC-MS of transgenic *P. abies* seedling needles, respectively. Black dots represent outliers. F) Relative expression (normalized to ubiquitin) of PaIDS2 in overexpression, vector control, and wild-type lines. Gray dots represent individual samples. OE-IDS2-1 and OE-IDS2-2, PaIDS2 overexpression lines; VC1 and VC2, empty vector control (pCAMGW) lines; Wt, wild-type control line.

Characterization of needle tissues from line OE-IDS2-2, harboring the overexpressed PaIDS2 construct, unexpectedly revealed a 3-fold increase in diterpene resin acid content ($p < 0.001$) (Figure 5.2). A second overexpression line, OE-IDS2-4, similarly showed a 2-fold increase in total diterpene resin acid content compared to empty vector and wild-type controls ($p < 0.01$) and had slightly lower diterpene resin acid abundance than that of OE-IDS2-2 ($p = 0.009$), despite both of these lines appearing to have similar PaIDS2 overexpression levels (Figure 5.1F). However, quantification of individual diterpene resin acids show that this discrepancy could be due in part to one individual OE-IDS2-4 plant that had a low abundance of palustric and/or levopimaric acid (these two metabolites have the same retention time and therefore cannot be individually quantified). OE-IDS2-2 nevertheless had higher levels of individual resin acids overall compared to OE-IDS2-4 when each metabolite was quantified (Figure 5.2).

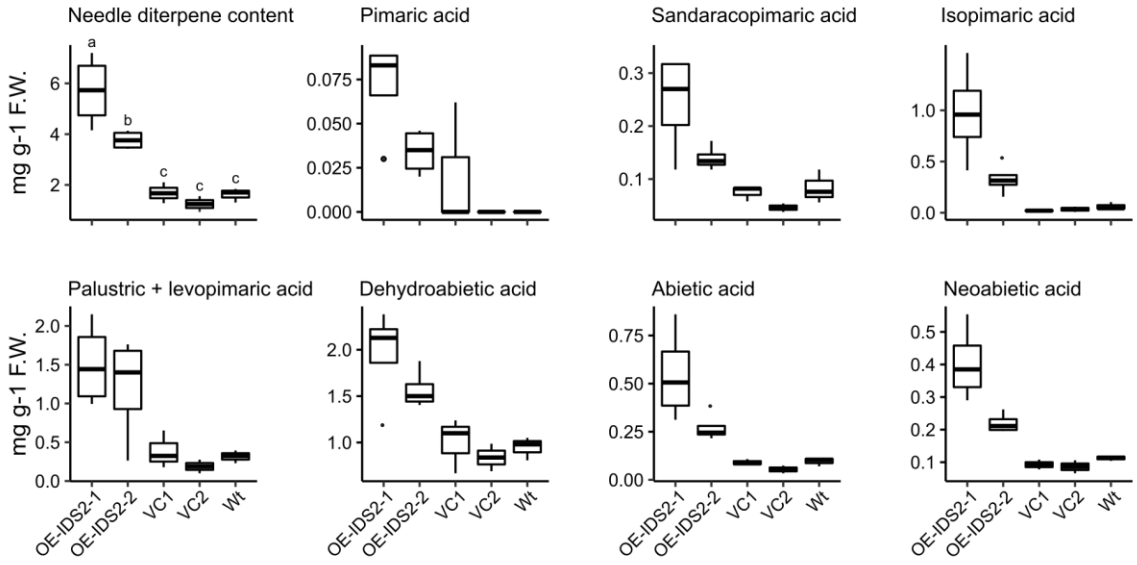


Figure 5.2: PaIDS2 overexpression leads to elevated abundance of diterpene resin acids in *Picea abies* sapling needles. Depicted is total needle content of individual diterpene resin acids as determined by quantitative GC-MS. Note that palustric and levopimaric acids have identical GC-MS retention times. OE-IDS2-1 and OE-IDS2-2, PaIDS2 overexpression lines; VC1 and VC2, empty vector control lines; Wt, wild-type control line. Dots represent outliers. Different letters indicate global significant differences ($p \leq 0.05$) between lines and same letters indicate a p-value above 0.05 following one-way ANOVA and TukeyHSD tests.

PaIDS2 overexpression did not appear to influence expression of PaIDS1 or PaIDS5 (Figure 5.2), PaIDS2 therefore must be participating *in planta* in the biosynthesis of GGPP. The GPP produced by PaIDS2 may be used as a substrate for further condensations with IPP catalyzed by other PaIDS enzymes rather than being a substrate for monoterpene biosynthesis. Or, PaIDS2 may act *in planta* as a GGPP synthase although it was characterized as a GPP synthase *in vitro*.

RNAi experiments of PaIDS2 produced no changes in the content of any terpene class in Norway spruce needles, as was expected because this gene is not expressed in needles. However, it was surprising that PaIDS2 RNAi knockdown lines also did not show significant changes in any of the three terpene classes in stems. To further investigate whether later enzymes of diterpene biosynthesis are likely to affect *in planta* resin acid abundance, expression of PaTPS-LAS and PaTPS-ISO was induced with methyl jasmonate (MeJA), and RNAi and overexpression experiments performed with CYP720B family members.

V.II.II Manipulation of later steps in diterpene biosynthesis does not elicit changes in diterpene resin acid abundance

Expression of the PaTPS-ISO and PaTPS-LAS genes was increased up to 10-20-fold in Norway spruce roots ($p < 0.001$) and 2-7-fold in needles over a 15-day time course following MeJA spraying (Figure 5.3). Interestingly, both PaTPS-LAS and PaTPS-ISO appeared to be increased in tandem in roots following phytohormone treatment, and total root diterpene resin acid content was also increased up to 2.5-fold ($p = 0.045$) (Figure 5.3A & B). Quantification of abundance of individual resin acids in roots shows that this increase is due to metabolites synthesized by both enzymes. However, while the highest expression of PaTPS-LAS and PaTPS-ISO transcripts occurred after 15 days, diterpene resin acid induction appeared following just 24 hours of MeJA spraying. This suggests that additional mechanisms are operating to modulate resin acid accumulation than just variation in the transcript level of di-TPS genes, possibly at earlier steps in biosynthesis. Further experiments are needed to measure the transcript levels and enzyme activities of additional steps in the pathway.

We additionally investigated the effects of MeJA treatment on di-TPS expression and resin acid content in stems. Gene expression was increased more rapidly than in roots, but never reached as high of a level relative to untreated controls as in roots. (Figure 5.3C). Nevertheless, the median resin content increased at later points in the time course, but this did not appear to be significant ($p = 0.09$) (Figure 5.3D). These results also suggest that the transcript levels of di-TPS genes are not major factors influencing diterpene accumulation. We also attempted to reduce PaTPS-ISO and PaTPS-LAS expression using RNAi but did not observe a reduction in constitutive resin acid content (Table 5.1).

Consistent with the results from di-TPS induction, several overexpression and RNAi experiments that targeted CYP720B members, some of which were characterized in Chapter IV, did not yield consistent results on either diterpene resin acid abundance or accumulation of predicted alcohol or aldehyde intermediates in any tissue (Table 5.1).

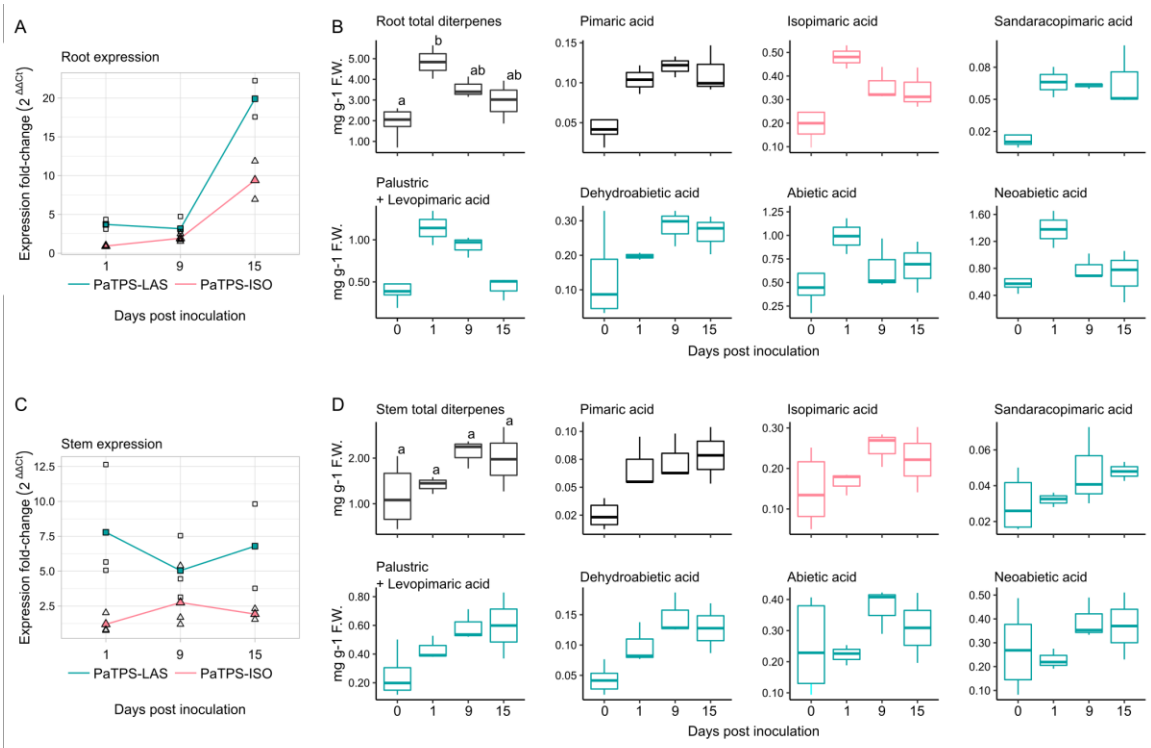


Figure 5.3: Induction of PaTPS-LAS and PaTPS-ISO gene transcripts is not tightly correlated with *in planta* diterpene resin acid abundance. Induction of PaTPS-ISO and PaTPS-LAS expression, and total and individual diterpene resin acid abundance in *Picea abies* seedling roots (A & B) and stems (C & D) as determined by quantitative GC-MS. Wild type plants were sprayed with methyl jasmonate, and expression of PaTPS-ISO (triangles) and PaTPS-LAS (squares) was measured 1, 9 and 15 days post induction (dpi), as were total and individual abundances of diterpene resin acids. Note that palustric and levopimaric acids have identical retention times. The average fold-change relative to time zero is shown in magenta for TPS-LAS and cyan for TPS-ISO. Different letters indicate global significant differences ($p \leq 0.05$) between timepoints and same letters indicate indicate a p-value above 0.05 following one-way ANOVA and TukeyHSD tests.

Table 5.1. Experiments with transgenic *Picea abies* lines showing no significant changes in diterpene abundance compared with wild-type and empty vector control lines

Gene	Enzyme class	Manipulation	Tissue(s)	Analyses
PaIDS2	IDS	RNA-interference	Needles & stem	Ether fraction on GC-MS
PaTPS-d*	di-TPS	RNA-interference	Needle & stem	Ether fraction on GC-MS
PaCYP720B1	P450	Overexpression	Needle, root & stem	Ether fraction on GC-MS
PaCYP720B3	P450	RNA-interference	Stem	Ether fraction on GC-MS
PaCYP720B4	P450	RNA-interference & Overexpression	Needle & stem	Ether fraction on GC-MS
PaCYP720B11	P450	RNA-interference & Overexpression	Root & needle	Ether fraction on GC-MS & MeOH fraction on LCMS

* Due to high homology between PaTPS-ISO and PaTPS-LAS open reading frames, primers targeting conserved regions of open reading frames and gene-specific primers targeting 5' UTR were attempted

V.III Chapter Summary

The focus of this chapter was on the role of different biosynthetic steps in the regulation of terpenoid accumulation in the model conifer Norway spruce. Various methods were employed to manipulate biosynthesis, including MeJA treatment, RNA-interference, and gene overexpression, and the targets included IDS, di-TPS and P450 enzymes. Overexpression of an IDS enzymes, PaIDS2 with a ubiquitous *Zea mays* promoter, increased the diterpene content in needles up to 3-fold, but had no effect on monoterpene level. Curiously, this enzyme had previously been demonstrated to catalyze the conversion of IPP and DMAPP to GPP *in vitro*. When the later steps of diterpene biosynthesis, including di-TPS and P450, were manipulated by RNAi or induction, there were no effects on diterpene resin content.

Chapter VI Discussion

This thesis addresses two questions about the evolutionary process, with emphasis on diterpenoid metabolism in the model conifer Norway spruce. The first question raised is to what extent specific phenotypic outcomes depend on the past. This question is justified because similar diterpenoid structures appear to have repeatedly evolved across the land plants but only a handful of times. If selection were strong for (+)-abietane and pimarane-type diterpene olefins, then a high frequency of convergent events would be expected in the absence of lineage-specific contingency. A second question asked is how novel biosynthetic pathways evolve given that one enzyme family often provides products that are substrates for enzymes of a second unrelated family. Driving this question is the observation that the diterpene olefins of focus here, which are synthesized by diterpene synthases (di-TPSs), do not accumulate in high abundance in plants but are instead further converted by disparate cytochrome P450s (P450s) in different lineages. If ancestral enzymes from each lineage were multi-product or highly promiscuous with substrates, patchwork assembly is predicted to occur and novel pathway steps would have been encoded for anciently. Alternatively, evolution might be directional such that either early or late pathway steps precede the complete assembly of modern-day biosynthetic routes.

In addition to questions regarding evolution and biosynthesis of diterpenoids, this thesis includes investigation into the regulation of terpenoid biosynthesis in Norway spruce, as this compound class serves defensive roles against insect and pathogen attack. Three potential metabolic control points were investigated in attempts to dissect some of the genetic elements underlying *in planta* terpenoid abundance. Often, manipulative studies of plant metabolic systems produce surprising results. To investigate this and to answer the evolutionary questions stated above, the results described in thesis were derived using a mix of comparative plant genomics, phylogenetics, ancestral sequence resurrection, molecular tools and comparative biochemistry.

VI.I Contingency and determinism shape evolutionary patterns in prototypical land plant diterpene synthases

The idea that chance events direct evolutionary outcomes is a major theme in our modern understanding of evolution (Blount et al., 2018). However, it is far from certain to what extent

specific phenotypic outcomes are dependent on lineage history as opposed to adaptive positive selection. The repeated acquisition of the same or very similar enzymes and products provides insight into not only the mechanistic basis of convergence, but also the role of contingency in such repeated natural experiments. Chapter II employed the bioinformatics tool OrthoFinder on 25 land plant proteomes to determine at which points in the land plant phylogeny the encoding gene families of di-TPSs originated and were inherited. OrthoFinder effectively searches for a complete set of orthologous sequences across the species phylogeny and proteomes provided, resulting in so-called "orthogroups" that contain a single ancestor and all of its descendants. Enzyme families that existed as multiple copies in the common ancestor of all species investigated (i.e. families that descended from multiple ancestral gene copies) are split into separate orthogroups. It was reported in Chapter II that an orthogroup containing the TPS-d subfamily of di-TPSs involved in (+)-abietane and pimarane biosynthesis in the Pinaceae unexpectedly contained the TPS-h subfamily from *Selaginella moellendorffii*, which is also involved in normally configured diterpene biosynthesis. It was also found that the Pinaceae TPS-d lineage was orthologous to the *S. moellendorffii* TPS-h lineage (Figure 2.3), thus being renamed the TPS-d/h lineage. Additionally, this orthogroup contained the TPS-c subfamily, also known as copalyl diphosphate synthases (CPSs), which catalyze the formation of *ent*-copalyl diphosphate (*ent*-CPP) in plant primary metabolism.

The most conspicuous implication in the phylogenetic analyses provided after application of OrthoFinder in Chapter II (Figure 2.3 and Figure 2.4) is that there were likely at least three paralogous di-TPSs in the common ancestor of the land plants investigated here. One paralog gave rise to the TPS-e/f subfamily of di-TPSs, which harbors kaurene synthase and kaurene-synthase-like (KSL) enzymes and was inherited by all major lineages with the exception of ferns (Figure 2.4). The second two paralogs dated to the common land plant ancestor are the TPS-c and TPS-d/h subfamilies mentioned above. Presence of TPS-c and TPS-d/h paralogues in the land plant ancestor is made apparent by the polyphyletic placement of gymnosperms, lycophytes and liverworts (represented by two Pinaceae members, *Marchantia polymorpha*, and *Selaginella moellendorffii*, respectively), in the TPS-c and TPS-d/h phylogeny in Figure 2.3. The TPS-d/h subfamily appears to have been lost by most lineages studied, as this lineage is represented exclusively by gymnosperms, lycophytes and liverworts.

OrthoFinder utilizes reciprocal BLAST searches with a supplied species tree to infer orthology rather than phylogenetic methods. Thus, OrthoFinder would not have taken subsequent tree rooting into account. The most parsimonious (i.e. simplest) root inference for the TPS-c and TPS-d/h subfamilies was to consider the gymnosperm TPS-d and lycophyte TPS-h subfamilies as orthologs of one another and specifying TPS-d/h as the tree root. Subsequent loss of the paralog that gave rise to the TPS-d/h subfamily (with the exception of gymnosperms, liverworts, and lycophytes) seems to be the most likely scenario for the origin of this subfamily. Widespread loss of the TPS-d/h subfamily among the other species investigated without gene phylogeny context would also explain how ancestral paralogs would end up in the same orthogroup. In other words, OrthoFinder is not likely to place the TPS-c and TPS-d/h subfamilies in separate orthogroups when the latter is present in only three out of twenty-five investigated species. Two less likely explanations for the TPS-d/h origin are horizontal gene transfer events from an as of yet unknown source into gymnosperms and lycophytes independently, and independent gene fusions in both gymnosperms and lycophytes. Therefore, it appears that of the three ancestral di-TPS paralogs, most lineages retained only two.

Loss of an ancient di-TPS paralog indicates a first level of dependency on past events in shaping future evolutionary outcomes for diterpene metabolism. One possibility is that the early copies of the TPS-c-like ancestor following duplication were functionally redundant, and loss due to incomplete lineage sorting or pseudogenation in angiosperms was of little selective consequence. One elegant study of fungal V-ATPases demonstrated that evolution of heterodimerism from a two- to three-ring system, and specific positions between ring subunits, occurred through neutral loss-of-function mutations, and evolution therefore need not implicate stepwise stages of increasing fitness (Finnigan et al., 2012). As these mutations were inherited by fungi early in their lineage, the phenotypic fate of fungal ring subunits appears to have been sealed near their onset by chance. Similarly, it could be that early neutral events of loss sent angiosperms on an evolutionary path in which phenotypic space, or chemical niches, were made available to alternative configurations of diterpene olefins catalyzed by KSLs, leading to the reported expansions of *ent*- and *syn*- configurations of diterpene olefins as previously reviewed (Murphy & Zerbe, 2020; Schmelz et al., 2014). Use of the alternative KSLs, which were historically maintained for production of crucial plant hormones (Zi et al., 2014), would be required for class I di-TPS expansion in cases where evolutionary reversions in other enzyme candidates are extremely unlikely. Indeed, it appears from alignment of the C-terminal domain of the TPS-c

subfamily that sequence similarity (defined here as the number of editing events required to align two or more sequences) decreased rapidly among angiosperms (Figure 2.3B). Additionally, loss of the C-terminal class I DDxxD motif in angiosperm TPS-c members would make this lineage highly unlikely ever to revert to class I reactions.

The patterns noted above in angiosperm TPS-c members strongly contrast with what is observed in the TPS-e/f phylogeny, represented primarily by *ent*-kaurene synthases (KSs) and KSLs, in which case duplication events are rather common (and recent) among angiosperms (Figure 2.4). High divergence between the ancestral di-TPS paralogs would lead to proportionally different genetic backgrounds from which recruitment could occur convergently in the repeated evolution of identical diterpene olefins. As reviewed in Chapter II, highly insightful studies demonstrated that the same sites across orthogroups control divergent product outcomes, albeit via replacements with different amino acids. This suggests that the number of sites available to control product outcomes may be limited, although additional sites have been shown to have catalytic effects and will be discussed later. Furthermore, angiosperms do not possess bifunctional diterpene synthases capable of both supplying normal-CPP and performing class I TPS cyclizations, and that catalysis of novel abietane and pimarane configurations by the angiosperm TPS-e/f subfamily are therefore dependent on fortuitous mutations in the class II angiosperm TPS-c subfamily.

Finally, as also reviewed in Chapter II, the epistatic interactions between amino acid positions in at least some di-TPSs is evident, as is the case for PaTPS-LAS and PaTPS-ISO. Although PaTPS-LAS and PaTPS-ISO are from the TPS-d/h subfamily, terpene synthases from TPS-c and TPS-e/f subfamilies use similar catalytic mechanisms and so might similarly require multi-site substitutions in the evolution of selectable phenotypes, which could explain some of the surprising results from mutational studies discussed in Chapter II. In fact, it was recently reported that two neighboring threonine residues are universally conserved among land plant KSs from the TPS-e/f subfamily, proximal to the class I DDxxD motif (Brown et al., 2021). Although the authors report that these are not strictly required for *ent*-kaurene biosynthesis and have only minimal effects on catalytic efficiency, such conservation indicates strong purifying selection and suggests that these sites may be interacting with each other or with distal sites. Moreover, catalytic diads (allelic linkages between two sites affecting catalysis) appear likely in a TPS-d/h

member involved in taxadiene synthase from the conifer *Taxus brevifolia* (family Taxaceae, not covered by the genomic analyses here) (Schrepfer et al., 2016).

Taken together, it appears that in the case of (+)-abietane and pimarane diterpene biosynthesis, factors that constrain the direction that evolution can take are 1) a limit to the number of available mutations that can bring about phenotypic change, 2) contingency on mutational order, 3) gene loss, which is explainable by neutral events such as incomplete lineage sorting or pseudogenation, and 4) contingency on novel early pathway steps that supply normally configured copalyl diphosphate (CPP) in the divergent TPS-c subfamily. These genetic and historical factors together are consistent with the reported "gatekeeping" effect that terpene synthases have on plant chemical evolution in general (Karunanithi & Zerbe, 2019) and also suggest that extremely large evolutionary timescales might be needed to re-evolve the same traits under these conditions. Thus, in addition to the facile nature of divergent trait evolution among di-TPSs (Christianson, 2008), the above studies provide mechanistic bases underlying slow convergence in this enzyme family, reminiscent of the large number of generations required for adaptive evolution in the experimental *E. coli* populations mentioned in Chapter II. Although natural selection might favor similar adaptive solutions to ecological pressures in multiple lineages, it seems that the divergent diterpene chemistry we see today may be highly contingent on early chance events.

VI.II Origin and evolution of the TPS-d/h subfamily

In the biosynthesis of most diterpenes, the products of di-TPSs become substrates for P450 enzymes. However, the specific P450 lineage recruited is not predictable. This raises a second major evolutionary question in this thesis: how are novel biochemical pathways assembled. Results of ancestral sequence resurrection of AncTPS-CON and characterization of SmDTC3.1 in Chapter III provide the first set of answers to this question and will be discussed next. In this chapter, it was found that AncTPS-CON produced an incomplete set of diterpene olefins relative to that of its modern-day descendants. SmDTC3.1 from the orthologous TPS-d/h lineage in *S. moellendorffii* produced multiple diterpene hydrocarbons and alcohols from *ent*-, *syn*-, and normal-CPP.

VI.II.I Convergent evolution of (+)-pimarane biosynthesis from ancestral TPS-d/h members

Gymnosperms appear to have independently acquired the biosynthesis of one or more diterpene olefins using TPS-d/h family members at least three times since the origin of their lineage, given the highly varied product sets catalyzed by di-TPSs across gymnosperm families. One occurrence is evident in the Pinaceae, in which the genera *Pinus, Picea*, and *Abies* each possess their own duplicated copies of TPS-ISO and TPS-LAS enzymes. A second occurrence is evident in the monofunctional *T. cryptomerioides* TPS-d/h member reviewed in Chapter II (TcKSL3, Cupressaceae), and a third appears in Ginkgoaceae represented by GbLPS in *Ginkgo biloba* (Figure 2.2). The ancestor of all gymnosperm TPS-d/h members, which was not resurrected in this study, must have diverged in *G. biloba* following speciation, leading to the levopimaradiene-specific GbLPS, and in the Pinaceae leading to AncTPS-CON (Figure 3.1). AncTPS-CON in turn duplicated in the Pinaceae ancestor and gave rise to 1) AbCAS, which catalyzes the conversion of GGPP to the normally configured, bicyclic labdane alcohol *cis*-abienol, and 2) the (+)-abietane and pimarane synthases of the Pinaceae. In conifers, levopimaradiene is one of several products formed following dehydration of the unstable intermediate 13-hydroxy-8,14-abietene by TPS-LAS enzymes, as reviewed in Chapter II. Isopimaradiene and pimaradiene are products of TPS-ISO, which arose as a paralog of TPS-LAS independently in each Pinaceae genus. Sandaracopimaradiene can additionally be produced by either abietane- or pimarane-type TPS-d/h enzymes.

Intriguingly, AncTPS-CON was found not to produce any of these known modern-day products when coexpressed with a GGPP synthase but rather a mixture of four (+)-abietane-type olefins, not including levopimaradiene (Figure 3.1). This product profile appears to occur via direct deprotonation of alternative hydrogen atoms of the abietenyl cation depicted in Figure 2.1, rather than dehydration of 13-hydroxy-8,14-abietene, which should yield an identical set of olefins as found for the Pinaceae TPS-LAS enzymes. It therefore appears that Pinaceae inherited a limited means to forming diterpene olefins, which would later be recruited into resin acid biosynthesis. Interestingly, AncTPS-CON possesses TPS-LAS-like amino acid residues estimated with high confidence at the four sites determined to affect product outcome of modern-day TPS-LAS and TPS-ISO enzymes (Figure 3.1). Detailed mutagenic analysis of PaTPS-LAS and PaTPS-ISO shows that introducing these four TPS-LAS-like sites (W687, Y694, A721 and V725) to PaTPS-ISO resulted in a product profile similar, but not identical, to PaTPS-LAS, as notable levels of

PaTPS-ISO products remained (Keeling et al., 2008). Moreover, the authors report that individual substitutions had cumulative effects.

Although it remains unknown to what extent carbocation quenching with water occurs in these mutant enzymes, it seems likely that these sites are not controlling formation of 13-hydroxy-8,14-abietene production in lieu of the AncTPS-CON results reported here. Indeed, a prior study of the AncTPS-CON descendant, AbAS (the *Abies grandis* ortholog of *Abies balsamea* TPS-LAS), showed that a single mutation of large effect not covered by the sites discussed here is enough to switch the product profile of AbAS from resembling other TPS-LAS enzymes to producing neoabietadiene almost exclusively (Peters & Croteau, 2002). Although this AbAS mutant is unlike any wild-type TPS-d/h member in terms of product profile reported to date, it does suggest that there remains a large unexplored set of mutational combinations that control specific and non-specific (+)-abietane or pimarane product outcomes, as well as alcohol formation.

It appears likely that, following duplication of AncTPS-CON, the individual TPS-ISO and TPS-LAS paralogs that arise in each lineage represent innovations independently acquired within the three Pinaceae genera discussed that led to a total of eight diterpene olefins and a bicyclic labdane alcohol in modern day conifers. As it is unclear why TPS-d/h members in gymnosperms would lose levopimaradiene biosynthesis between the ancestral gymnosperm TPS-d/h and AncTPS-CON, only to later converge on this product in TPS-LAS, GbLPS, and TcKSL3, it is concluded here that production of levopimaradiene also arose independently at least three times among gymnosperms, utilizing three different enzymatic strategies (bifunctional conversion of GGPP to levopimaradiene by GbLPS, monofunctional conversion of normal-CPP by TcKSL3 to levopimaradiene, and via 13-hydroxy-8,14-abietene by TPS-LAS.

VI.II.II Lycophyte TPS-d/h members demonstrate lineage specific expansion of substrate promiscuity and relaxed product specificity

Investigation into TPS-d/h members in *S. moellendorffii* produced some results consistent with a previous report on SmDTC3 and some surprising features of the newly described SmDTC3.1, which shares 97% sequence identity with SmDTC3 (Shimane et al., 2014; Figure 3.3 and Figure 3.4). SmDTC3.1 was assayed with normal-, *ent*-, and *syn*-CPP and compared to SmDTC3. Several SmDTC members are predicted to be monofunctional, using only one CPP stereoisomer, and this is due to independent substitutions in the class II DxDD motif in about half of the

predicted SmDTC paralogs (Figure 2.3). However, SmDTC3.1 was active with all three configurations of CPP, as was SmDTC3, but produced no detectable products with GGPP (negative result chromatograms not shown). SmDTC3.1 additionally produced a diterpene alcohol with normal-CPP with identical retention time and a similar mass spectrum to 16-hydroxy-*ent*-kaurane, which was not observed in SmDTC3 (Shimane et al., 2014; Figure 3.4A & B; Appendix Figure A3.3). SmDTC3.1 additionally produced several (+)-abietane-like diterpene olefins with normal and *syn*-CPP not present in SmDTC3 assays, and sandaracopimaradiene was produced by both enzymes. Notably, SmDTC3 and SmDTC3.1 had identical amino acid states in the four sites discussed previously that affect catalytic outcome of di-TPSs in both the Pinaceae TPS-d/h and the broader TPS-e/f subfamily (Figure 3.3). This lends support to the notion that additional deterministic residues exist among di-TPSs, and that TPS-d/h members in *S. moellendorffii* may exhibit widespread plasticity unique to the lycophyte lineages.

A SmDTC paralog, SmMDS, which performs both class II and class I reactions with GGPP to yield miltiradiene arose more anciently than SmDTC3 and SmDTC3.1 (Sugai et al., 2011; Figure 2.3). However, the two most anciently branching paralogs in the *S. moellendorffii* TPS-d/h subfamily - the uncharacterized SmDTC6 and the characterized SmCPSKSL (Mafu et al., 2011) - either lack an intact class II motif or result in bicyclic labdane alcohols, respectively. The apparent late acquisition of normally configured (+)-abietanes and pimaranes among SmDTC paralogs lends further credence to the argument above that gymnosperm and lycophyte lineages did not encode normally configured diterpene olefins at the time of their origin, and further that specific structures such as sandaracopimaradiene repeatedly evolved in parallel in gymnosperms and lycophytes. Thus, the redundant nature of early land plant di-TPS paralogs, likely resembling the modern-day bifunctional KS from the moss *Physcomitrella patens* (Hayashi et al., 2007), has the surprising implication that this redundancy was maintained for at least 50 and 150 million years between the origin of land plants and the appearance of lycophytes and gymnosperms, respectively (Morris et al., 2018). Thereafter, the paralog that gave rise to the TPS-d/h subfamily in lycophytes and gymnosperms must have independently acquired normal-CPP synthase activity in each lineage.

VI.III Origin and evolution of CYP720 reaction diversity

It is not yet possible to infer a mechanism by which a paralogous *ent*-kaurene synthase would be maintained during the ca. 50- and 150-million-year spans of di-TPS evolution before expansion

of the TPS-d/h subfamily. However, it may be speculated that an advantageous gene dosage effect on *ent*-kaurene abundance became either useless or deleterious in angiosperms after the advent of more complex hormone signaling networks. In any case, how a duplicate di-TPS might have been assembled into novel biochemical pathways together with P450s is discussed in this section, as well as broader evolutionary patterns in the recently duplicated CYP720 subfamily in Norway spruce. The following conclusions are based on results from phylogenetic analysis using OrthoFinder on 25 land plant genome-predicted proteomes, and biochemical assays and homology modeling of PaCYP720B enzymes from Norway spruce in Chapter IV. It was reported that the orthogroup containing PaCYP720Bs also contained encoded enzymes involved in brassinosteroid biosynthesis in land plants. The highly duplicated PaCYP720Bs from Norway spruce and the conifer *Pinus taeda*, also from the Pinaceae family, was orthologous to the single-copy dicot CYP720A subfamily. Biochemical analysis showed that some PaCYP720B enzymes oxidized a range of diterpene olefins, converting them to diterpene resin acids. Other PaCYP720Bs were only active with one substrate, adding hydroxyl functions to alternative carbon atoms.

VI.III.I The CYP720 family arose in the angiosperm-gymnosperm ancestor from brassinosteroid oxidases but was widely lost among monocots

The results of our ortholog searching and phylogenetic analysis of P450s suggest that CYP720 arose via gene duplication of an ancient brassinosteroid oxidative enzyme just prior to seed plant divergence (Figure 4.1). This duplication would have occurred in the same time frame that duplicate bifunctional di-TPSs existed, just prior to loss of the TPS-d/h subfamily in angiosperms. The subsequent duplications of CYP720 members in *P. abies* are in stark contrast to the stability of the CYP720A lineage in angiosperms, suggesting that CYP720 duplicates were either deleterious or not utilized in lineages lacking TPS-d/h members. Conversely, we found that angiosperm CYP720A members are conserved as a single genomic copy in all dicots represented, barring one recent duplication leading to paralogs in *Glycine max* (the latter duplication interpreted to be very recent, as both copies possess comparatively short branch lengths). Nevertheless, we conclude that the CYP720 lineage inherited by angiosperms and gymnosperms, and later recruited in conifers for secondary metabolite biosynthesis, was lost widely among monocots and conserved but as just one genomic copy number in dicots. Notably, PaCYP720B4, which is orthologous to PsCYP720B4 from *Picea sitchensis*, appears to have arisen from the most recent of duplications within the CYP720B subfamily.

VI.III.II CYP720B enzyme activities further expand known functional space via degrees of substrate specificity and substrate orientation

PaCYP720B4, which shares 98.6% sequence identity to the *P. sitchensis* ortholog PsCYP720B4 (Hamberger et al., 2011), oxidized all eight olefin substrates supplied as was reported for PsCYP720B4. Of additional interest, however, was that PaCYP720B4 was able to oxidize multiple olefin impurities in our assays in the presence of high amounts of sandaracopimaradiene and levopimaradiene (Figure 4.5). This finding indicates that CYP720B4 is likely unable to discriminate between structurally similar diterpene olefin substrates. Conversely, the finding that PaCYP720B3 was active only with dehydroabietadiene in contrast to the promiscuous PaCYP720B4, establishes that closely related paralogs of CYP720 enzymes can have very different specializations (Figure 4.6). On the product side, the fact that PaCYP720B3 produced a diterpene alcohol rather than carboxylic acids like PaCYP720B4 demonstrates a striking level of control over product formation, considering PaCYP720B3 is ca. 85% identical to PaCYP720B4.

PaCYP720B1, which is less than 60% identical to both PaCYP720B3 and PaCYP720B4, appears to further expand the known biochemical space occupied by the gymnosperm CYP720B subfamily as dehydroabietadiene was converted to ferruginol or a similar diterpene alcohol with functionality at C-12, a position unusual for this family of enzymes (Figure 4.7). This suggests a different substrate orientation in the active site placing the C-ring proximal to the heme surface, leading to an alternative product outcome. Although ferruginol has not yet been reported in *P. abies* and a role for it in diterpene biosynthesis in this species has therefore not been proposed, ferruginol has been detected in the roots of the angiosperm *Tripterygium wilfordii*, produced biosynthetically by unrelated Lamiaceae P450s for tanshinone biosynthesis (Guo et al., 2013; Zi & Peters, 2013), and is a major metabolite in species of the Podocarpaceae and Cupressaceae conifer families (Cox et al., 2007; Hansen et al., 2017). Thus, regardless of its precise *in planta* role, ferruginol biosynthesis in conifers via P450 oxidation has precedence and even appears to have evolved independently in conifers and Lamiaceae.

VI.III.III Homology modeling supports restructuring of PaCYP720B active sites for accommodation of novel substrates and the occlusion of putative ancestral substrates

Historically, it is likely that the CYP720 enzymes catalyzed oxidation steps in the biosynthesis of brassinosteroids or other steroids, as inferred by our phylogenetic analysis (Figure 4.1).

However, the enzyme PaCYP720B4, a product of recent gene duplication in the *Picea* lineage, which we investigated here, was shown to be capable of oxidizing dehydroabietadiene as well as seven additional diterpene olefins. PaCYP720B4 possess only about 40% amino acid identity to the CYP720 enzymes involved in sterol oxidation. Among the amino acid substitutions are some inherited uniquely through the CYP720 lineage and some recently derived following divergence of the conifer CYP720B family. The replacements observed in CYP720B4 appear not only to facilitate binding and proper orientation of the C_{20} olefin substrate tested, which is significantly smaller than a steroid (Figure 4.9C & D), but also include larger hydrophobic substitutions that would lead to steric interference with the side chain of cholesterol or campesterol and a loss of polar contact with the A-ring hydroxyl group. Interestingly, mapping the inferred amino acid changes to the CYP720B phylogeny shows parallel replacement of asparagine with phenylalanine in PaCYP720B1 and in the enzyme ancestor of PaCYP720B3, PaCYP720B4, PaCYP720B5 and PaCYP720B6 (Figure 4.9D). Thus, the substitutions observed suggest that, concomitant with the acquisition of diterpene olefin oxidation, the novel CYP720B enzymes would have lost their capacity to bind phytosterols and appear to have done so repeatedly. Such a loss in substrate recognition would presumably be important if duplication of brassinosteroid oxidases resulted historically in a detrimental gene dosage effect by altering the titers of steroid hormones.

As a result of the amino acid replacements mentioned above, F286 of PaCYP720B4 is also phenylalanine in all *P. abies* CYP720B enzymes with the exception of PaCYP720B11. In PaCYP720B11, as well as in CYP720A1 in *A. thaliana* and *P. trichocarpa,* this site is an asparagine, which more closely resembles in size and polarity the serine found at this site in AtCYP90B1. An intriguing avenue of further research would be to screen a wider panel of substrates with angiosperm CYP720A members including phytosterol intermediates. As it is unclear why the novel CYP720B members would lose ancestral activities inherited by the CYP720 lineage, it could be that angiosperm CYP720A and conifer CYP720B11 are performing analogous oxidative roles. Nevertheless, it is concluded here that functional expansion of specialized P450s in diterpene oxidation in Norway spruce, and likely other Pinaceae, followed early amino acid replacements that would have likely blocked ancestral substrates from the active site and facilitated novel substrate orientations. Furthermore, it appears that broad substrate acceptance by PaCYP720B4 occurred long after inheritance of AncTPS-CON, following late duplications within the CYP720B subfamily, as other PaCYP720B enzymes appear to employ dehydroabietadiene as a substrate.

These results together indicate that, despite a high degree of promiscuity observed in CYP720B4, the P450 enzymes of the CYP720 subfamily also exhibit high substrate selectivity. This would explain why the three paralogous CYP720 candidates investigated here were unable to oxidize any of the supplied substrates. It should be mentioned that other Pinaceae species harbor CYP720B12 and CYP720B2 enzymes that selectively oxidize the hydroxylated product of TPS-LAS enzymes (Geisler et al., 2016). CYP720B12 and CYP720B2 have orthologs that could be identified in Norway spruce. However, our screening suggests that direct oxidation of 13-hydroxy-8,14-abietene could be restricted to certain Pinaceae species, as PaCYP720B12 and PaCYP720B2 from Norway spruce appear truncated in the genome and therefore did not meet the screening criteria imposed here. Although cloning of partial fragments of PaCYP720B2 and PaCYP720B12 was possible, 5'- and 3'-RACE performed for each failed to yield full-length transcripts (data not shown). In *Picea*, these could represent orthologous pseudogenes that are nevertheless still transcribed as no introns were detected in PCR fragments over a span of approximately 850 nucleotides.

VI.III.IV Multiple genetic, biochemical and structural layers primed the CYP720 subfamily in Pinaceae for divergent pathway recruitment

The placement of the seed plant CYP720 subfamily in the orthogroup containing land plant steroid oxidases strongly implicates gene recruitment from growth regulation towards specialized defense metabolism just prior to the angiosperm-gymnosperm speciation event. This is of interest because the expanded biochemical capabilities of CYP720B members in conifers, in conjunction with the various products produced by PaCYP720B1, PaCYP720B3, and PaCYP720B4, open the possibility that this subfamily is inherently plastic, and might have historically performed a wide range of oxidative reactions on substrates. It is of additional interest that the orthologous CYP720A subfamily was not capable of performing oxidation reactions on any olefin or alcohol diterpene intermediates; it could be that permissive mutations that accrued in the conifer CYP720B family never occurred historically in the angiosperm CYP720A family, thereby limiting the range of evolutionary options for this angiosperm gene lineage. The CYP720A subfamily might therefore have retained historical selectivity for its preferred substrate, which could be plant sterols as mentioned previously, or some yet unidentified enantiomer of normally configured diterpene olefins.

In conifers, it appears that gene recruitment towards defense metabolism involved more than acceptance of new substrates. Substrate orientation, sites and degrees of oxidation and occlusion of ancient substrates from the active site appear to have played roles in fine-tuning the catalytic capabilities of the CYP720B lineage. However, in angiosperms, the highly constrained gene copy number in dicots and loss of the CYP720A representatives in monocots suggest an alternative fate for the CYP720 lineage. Interestingly, the monocot *O. sativa* utilizes *syn*-pimaradiene, which is structurally similar to the diterpene olefins tested here, as an intermediate in the formation of the highly oxidized and chemically-divergent momilactone A (Otomo et al., 2004). *O. sativa* also appears to utilize a member of the CYP701A family that oxidizes *ent*-sandaracopimaradiene for oryzalexin biosynthesis (Otomo et al., 2004; Wang et al., 2012). As mentioned previously, Lamiaceae members conversely use the CYP76AH subfamily for diterpene oxidation in tanshinone biosynthesis (Guo et al., 2013; Zi & Peters, 2013). Loss of catalytic flexibility in the CYP720 lineage might actually have facilitated alternative solutions in recruiting enzymes for diterpene oxidation.

VI.IV Assembly of modern-day pathways in Norway spruce diterpenoid metabolism

Early diterpenoid routes in Pinaceae, represented here by Norway spruce resin acid biosynthesis, appear to be explained by the patchwork hypothesis of biochemical pathway assembly. It was argued above that TPS-d/h members were highly unlikely of catalyzing the formation of (+)-abietane and pimarane-type diterpene olefins ancestrally in both lycophyte and gymnosperm lineages. However, the 150-million-year timeframe between the land plant ancestor and appearance of gymnosperms coincided with the maintenance of an ancient bifunctional di-TPS paralog, presumed here to be a duplicate *ent*-kaurene synthase. Just prior to its duplication, AncTPS-CON was capable of synthesizing an incomplete set of (+)-abietane-type diterpene hydrocarbons, and this was likely unique to gymnosperms following loss of an orthologous copy in angiosperms. This strongly suggests that the functional role of the ancestral gymnosperm TPS-d/h was lost by the time the Pinaceae appeared and before the first duplications of PaCYP720B enzymes.

The most common substrate across the CYP720B lineage in Pinaceae appears to be dehydroabietadiene, with PaCYP720B4 arising later in *Picea* and catalyzing the formation of the eight known diterpene resin acids seen today. Given that miltiradiene and levopimaradiene

rearrange to dehydroabietadiene in other known diterpenoid pathways, and abietadiene can rearrange to dehydroabietadiene as seen in experiments here, it is possible that AncTPS-CON was maintained for rearrangement of its products to dehydroabietadiene, which was ancestrally oxidized by paralogous CYP720B enzymes. As no activity was seen with angiosperm CYP720A members when assayed with diterpene olefins or steroids tested here, it is not yet possible to determine for which role CYP720 enzymes were maintained prior to recruitment into resin acid biosynthesis in Norway spruce and other Pinaceae members. However, the conservation in copy number in angiosperms, specifically the core eudicots, suggests that dosage of its product is tightly regulated. Nevertheless, via parallel duplications leading to TPS-LAS and TPS-ISO paralogs in the genera *Abies, Picea,* and *Pinus*, resulting in one copy capable of producing 13-hydroxy-8,14-abietene, each lineage acquired its own means of levopimaradiene biosynthesis following dehydration. This duplication in Norway spruce occurred in the same time frame as the late duplication events of PaCYP720B that gave rise to PaCYP720B4, implying that direct oxidation of the eight individual olefins in Norway spruce, and likely other Pinaceae species, surprisingly evolved backwards (i.e. in retrograde fashion). Without a CYP720B ancestral enzyme oxidizing dehydroabietadiene but with fortuitous secondary oxidative capacity of a broad range of both (+)-abietane and pimarane-type diterpene olefins, there would be no explanatory mechanism behind the parallel evolution of identical TPS-LAS and TPS-ISO paralogs in individual Pinaceae genera.

The patchwork hypothesis of biochemical pathway evolution appears to have the largest amount of empirical support (Copley, 2000; Fani & Fondi, 2009). For instance, the patchwork hypothesis explains the evolution of degradation of the aromatic pesticide pentachlorophenol (PCP) in the bacteria *Sphingomonas chlorophenolica*, which likely evolved in the past few decades (Copley, 2000). The pathway consists of three reaction types (NADPH-dependent hydroxylation, dehalogenation, and oxidative ring cleavage) catalyzed by enzymes from divergent gene families. The PCP degradation pathway appears to be limited to only a few strains in nature. However, the dehalogenation step is performed by an enzyme related to the ubiquitous maleylacetoacetate isomerases that degrade tyrosine, which were likely recruited to the PCP degradation pathway. Copley reports that broad specificity of the participating hydroxylase in PCP degradation facilitated pathway assembly via a patchwork of enzymes performing alternative roles. Similarly, it appears likely that the CYP720 subfamily was ancestrally capable of hydroxylating dehydroabietadiene in various positions for as yet unknown downstream metabolism. Already

accepting abietane-type compounds as substrates, it was likely primed in the Pinaceae for the retrograde assembly of specific diterpene oxidative steps together with novel TPS-LAS and TPS-ISO paralogs.

VI.V Regulation of defensive terpenes in Norway spruce

The results presented in Chapter V support a model of control in which flux in terpene biosynthesis is regulated upstream of the enzymes dedicated to functionalization of biologically active terpenoids. Overexpression of PaIDS2 in needles provided an opportunity to investigate *in planta* control on terpene content more directly because PaIDS2 does not appear to be localized to this tissue naturally. Increasing the proportion of transgenic PaIDS to the other PaIDS proteins present led to a ~2-3 fold increase in diterpenoid products. It is notable that overexpression of PaIDS2, which functions *in vitro* as a GPP synthase, increases the content of diterpene resin acids, but not monoterpenes, suggesting that any GPP released by this enzyme is not preferentially channeled to monoterpene synthases as substrate. One possible mechanism underlying this phenotype is the presence of metabolons between PaIDS2 and a GGPP synthase, the latter condensing GPP from PaIDS2 with IPP to yield GGPP. Acceptance of GPP and IPP as substrates by other PaIDS has precedence in the literature, as discussed above with PaIDS1 (Schmidt et al., 2010). Any additional GGPP formed in this way could then supply additional substrate for diterpene resin acid biosynthesis via conversion by di-TPSs and P450s.

For PaTPS-ISO and PaTPS-LAS, increased transcript abundance generated by methyl jasmonate treatment did not correlate well with increases in diterpene content. Samples taken at much later time points may be necessary to observe changes in total terpenoid abundances. Although we appeared to induce PaTPS-ISO and PaTPS-LAS expression soon after MeJA treatment, fully formed resin ducts are usually reported 15 days after treatment and a previous study on MeJA-treated Norway spruce reported increased traumatic resin accumulation two months after treatment (Martin et al., 2002). Martin and co-workers also separated stems into bark and wood contents, potentially enriching the per-gram content of induced diterpene content in the wood. Additionally, the negative results of the various CYP720 transgenic experiments (Table 5.1) suggest that terminal pathway enzymes do not represent limiting steps in pathway flux.

Conclusions

This thesis reports on the first instance of ancestral sequence resurrection applied to di-TPSs, which produced results that could not have been predicted by examining modern-day enzymes alone. Additionally, the catalytic properties of P450 enzymes (which evolved increasing substrate preference and additional active site restructuring within the same timeframe as the diverging di-TPSs) were investigated and reported here. In extant Pinaceae lineages, these di-TPSs and P450s together complete defensive diterpenoid biosynthetic pathways. Control points of pathway flux were additionally examined.

There are a number of limitations to the methods employed here that require certain caveats on data interpretation. For the evolutionary studies, it is not possible to infer the tissue of expression of ancestral enzymes as ancestral sequence estimation relies on either codon usage or amino acid substitution models. Therefore, non-coding DNA regions such as gene promoters cannot be reconstructed and examined. Secondly, the fraction of modern-day genes and encoded proteins available for study represent only a fraction of all genes and proteins that existed in the past. Although it is possible to reconstruct ancestors with high confidence from a set of modern-day enzymes, for instance, there are likely additional clades of di-TPSs that have long been extinct. Additionally, there may have been additional CYP720 members that have participated in ancient biochemical pathways that are now obscure due to extinction, as with any family of genes and encoded protein products. For these reasons, the evolutionary patterns and processes observed in this thesis likely represent only a fraction of the body of forces driving diterpenoid evolution. Finally, although tests of positive selection can shed light on sites important in gene and protein evolution, false inferences may be frequent (Venkat et al., 2018) and therefore ancestral inferences of fitness and ecological roles remain beyond the scope of the studies here.

The results presented in Chapter III and Chapter IV provide initial insight into the acquisition of unique phenotypes in a restricted lineage of conifers. To gain more complete information about the genetic mechanisms responsible for these processes, it is recommended to reconstruct and resurrect the remaining ancestral di-TPS enzymes of the gymnosperm lineage. Specifically, the ancestral TPS-d/h that existed shortly after the split of gymnosperms from other vascular plants could be estimated using any number of TPS-d/h sequences from Cycadales, Ginkgoales,

Coniferales, and Gnetales orders of gymnosperms, using additional orders if necessary. This could similarly be done for the ancestral P450 of the CYP720 subfamily, with inclusion of CYP720A sequences in the estimate from angiosperms that possess them. Mutagenic analysis of ancestors by introducing amino acid substitutions in the order in which they occurred historically to re-evolve modern biochemical pathway steps would shed light on the critical mutations (relatively free of epistasis) that resulted in di-TPS products that were compatible with early CYP720 ancestors. Extending this further, applying such methods to other areas of diterpene metabolism would determine whether the patterns reported here explain the repeated compatibility between novel class II and class I di-TPS, and which P450 lineages were recruited into diterpene pathways. Ideally, such experiments would focus on combinatorial assays between ancestral di-TPS and P450s whose descendants participate together in completing specialized diterpene routes. Experiments using ancestral sequence resurrection to examine co-evolution of enzymes encoded by different gene families have not been conducted before. However, results from these studies could help inform better hypotheses of protein evolution, and be applied to, for example, ancestral enzymes involved in plant defense and ancestral receptors from insect pest that have evolved resistances to those plants.

Evolutionary questions aside, the precise nature of control over terpenoid pathway flux remains elusive. Future efforts could confirm or refute existence of potential metabolons by generating CRISPR-CAS background lines (thereby knocking out potential metabolon partners) in which PaIDS2 overexpression could be repeated. A primary target for a metabolon partner in such an experiment would be PaIDS1, which we showed to be present in the needles and therefore likely capable of converting GPP and IPP *in planta* to GGPP. Alternatively, PaIDS2 may be acting *in planta* as a GGPP synthase rather than as an intermediate enzyme. This would predict that the same knockout lines would have a similar phenotype as reported here, and also predict that feeding radiolabeled DMAPP and IPP would lead to the accumulation of labeled GGPP rather than GPP.

Materials & Methods

Bioinformatics

Obtaining proteomes for phylo-genomic assessment of land plant diterpene synthases and P450 evolution

The primary (longest) amino acid translations of predicted gene models from 25 land plants were obtained and are hereafter referred to as "proteomes". The latest proteome versions as of the time of this work for *Theobroma cacao*, *Arabidopsis thaliana*, *Populus trichocarpa*, *Glycine max*, *Solanum lycopersicum*, *Daucus carota*, *Spirodela polyrhiza*, *Zostera marina*, *Sorghum bicolor*, *Zea mays*, *Oryza sativa japonica*, *Ananas comosus*, *Musa acuminata, Nymphaea colorata*, *Amborella trichopoda*, *Selaginella moellendorffii, Physcomitrella patens*, *Sphagnum fallax* and *Marchantia polymorpha* were obtained from Phytozome 12 (Goodstein et al., 2012). The *Coffea canephora* proteome was obtained from the Coffee Genome Hub (Dereeper et al., 2015). *Phalaenopsis equestris* was obtained from Plaza 4.0 (Van Bel et al., 2018). *Azolla filiculoides* and *Saccostrea cucullata* proteomes were obtained from FernBase (Li et al., 2018) and *Pinus taeda* (version 1.0) and *Picea abies* (version 1.0) were acquired from ConGenIE (Sundell et al., 2015). Although *Picea glauca* has a publically available genome sequence (Warren et al., 2015), this species was omitted from screening because this work focuses on *P. abies* and the latter sufficiently represents the *Picea* genus. As PaTPS-ISO, PaTPS-LAS, and a putative *ent*-kaurene synthase are not represented by the current Norway spruce genome release, the amino acid sequence of PaTPS-ISO, PaTPS-LAS, and the characterized *ent*-kaurene synthase from *Picea glauca* were each appended to the Norway spruce proteome prior to analysis. To reduce downstream computational load while retaining all potential land plant di-TPSs and P450s, all proteomes were reduced to only those peptides predicted to be between 450 and 900 amino acids in length.

Orthology searching in land plant genomes with OrthoFinder

To obtain single orthogroups that contained open reading frames of interest, reduced proteomes from above were used in two independent analyses using OrthoFinder version 2.3.3 using either

"diamond" or "BLAST" search options and a user-supplied species tree (Figure A2.1). Moss and liverwort species were considered sister clades as a viable alternative to other topologies (Morris et al., 2018). Two orthogroups harbored di-TPSs of interest, one represented primarily by *ent*- and *syn*-CPP synthases and TPS-d/h members, and the other orthogroup represented largely by the TPS-e/f subfamily. These three orthogroups were then extracted and used for downstream analysis. The orthogroup found to contain the CYP720B subfamily also contained other members of the CYP85 clan in both analyses mentioned above and was further concatenated to transcripts between 450 and 550 amino acids in length.

Maximum-likelihood analysis of orthogroups

For the di-TPS orthogroups, structure-guided alignment was performed using the PROMALS webserver (Pei & Grishin, 2007) to better facilitate alignment of long divergent regions. Manual adjustments were performed by aligning start methionines and sequences were removed based on apparent truncations, resulting in 53 full-length amino acid sequences for the orthogroup containing TPS-c and TPS-d/h members, and 31 full-length peptides in the orthogroup containing TPS-e/f. The di-TPS orthogroups were then analyzed with RAxML v8.2.12 (Stamatakis, 2014) with 1000 bootstrap replicates and the JTT substitution model (automatically chosen by RAxML). The best-scoring trees were used to generate the gene phylogenies in Figure 2.3 and Figure 2.4. The most parsimonious tree rooting option, which minimized the number of inferred gene loss events for the TPS-c and TPS-d/h orthogroup, was chosen using NOTUNG 2.9 (Chen et al., 2000) by supplying the generated gene phylogeny and species tree.

The CYP85 orthogroup was aligned using the MAFFT (Katoh & Standley, 2013) l-ins-i algorithm (v7.310) and default parameters, followed by manual alignment as above. Sequences that did not possess a start methionine or an absolutely conserved cysteine residue for heme coordination, and sequences that appeared truncated were removed from the resulting orthogroup. This resulted in 189 full-length genome-encoded CYP85 protein sequences identified across the 25 land plants surveyed. Phylogenetic analysis was performed on the 189 sequences with RAxML exactly as above for di-TPSs. The best-scoring trees were used to generate the gene phylogenies in Figure 4.1 and Figure A4.1.

Nucleotide sequence collection for use in ancestral sequence reconstruction

Nucleotide sequences encoding selected characterized conifer TPS-d/h enzymes were obtained from GenBank. PaTPS-NEW and PaTPS-NEW2 were generated *in silico* following *de novo* single-end Illumina assembly from three biological replicates of adult *Picea abies* bark. Due to low coverage of di-TPS coding genes in available published *Picea* spp. genomes, a genome-guided approach was not selected. An uncharacterized full-length TPS-d representative from Podocarpaceae (SCEB 2004030) and Zamiaceae (WLIC 2008840) were retrieved by running tBLASTn against the available *Podocarpus coriaceus* and *Dioon edule* transcriptomes available on OneKP. As the N-terminal portion of TPS-d enzymes largely contains a highly divergent chloroplast transit peptide presumably irrelevant to ancestral catalytic activity, this region was removed from all sequences prior to further analysis at sites designated as a transit peptide on the ChloroP server.

Computational reconstruction of AncTPS-CON

Amino acid translations (with an appended start methionine) of nucleotide sequences lacking predicted N-terminal transit region and were aligned using the l-ins-i algorithm in MAFFT setting the maximum allowed iterations to 1000 and leaving the remaining parameters at the default setting. The resulting alignment was adjusted manually and used in maximum-likelihood analysis (100 replicates) in PhyML version 3.0 (Guindon et al., 2010) using the JTT substitution model as selected by ProtTest version 2.4 (Abascal et al., 2005) and rooted along the branch containing *Ginkgo biloba* and WLIC 2008840. This tree and a nucleotide alignment (Appendix Figure A3.1) guided by the amino acid alignment from above was supplied to PAML (Yang, 2007). Using the codeml package with empirical codon models, ancestral states were reconstructed for all 12 internal nodes of the supplied phylogenetic tree and per-site posterior probabilities were retrieved. The node inferred as the ancestral di-TPS that gave rise to all TPS-d/h sequences in conifers was extracted, named "AncTPS-CON", and processed further. As PAML is unreliable in gapped regions, AncTPS-CON was translated to its deduced amino acid sequence, aligned against the amino acid sequence alignment of extant TPS-d/h members from above. Insertions in the estimated ancestor that were inferred to be more recent than the Pinaceae ancestral species were removed.

Molecular cloning and quantitative real-time PCR

Total RNA isolation and cDNA synthesis

A modified phenol-chloroform extraction procedure was used (Schmidt et al., 2010) to extract RNA from *P.* abies needles, as most commercially available RNA extraction kits are not suitable for extraction of total RNA from *P. abies* needles. Briefly, the protocol was modified with the use of 500 mg of homogenized tissue, and following precipitation of RNA and discarding the supernatant, the resulting pellet of purified RNA was resuspended in 0.75 mL 70% ethanol and purified using an InviTrap® spin plant RNA mini kit according to the manufacturer's instructions. Stem and root RNA from *P. abies* and *S. moellendorffii* and RNA from leaves of *A. thaliana* and *P. trichocarpa*, as well as *S. moellendorffii* aerial tissue, was isolated using a Zymoclean™ gel RNA recovery kit according to the manufacturer's instructions. cDNA was synthesized from purified template RNA using the SuperScript™ III first-strand synthesis kit from Invitrogen™ according to the manufacturer's instructions, and the resulting cDNA was diluted two-fold with sterilized H_2O.

Cloning of full-length coding DNA sequences and qPCR

Primers for TPS-LAS (accession no. AY473621) were designed based on the reported mRNA coding sequence (Martin et al., 2004). Primers for SmDTC3 and SmDTC3.1 were designed based on the nucleotide sequence of the *S. moellendorffii* draft genome transcript no. 407280 (see bioinformatics section above for genome information). For amplification of P450 candidate sequences, primers were designed based on AtCYP720A, PaCYP720A and PaCYP720B-coding DNA sequences from *A. thaliana*, *P. trichocarpa* and *P. abies* genomes, respectively. All attempted primer sequences are shown in Appendix Table 4.1. All amplified sequences were subcloned using the Zero Blunt™ TOPO™ PCR Cloning Kit from Invitrogen™. For *E. coli* transformation, ligation products were transformed into One Shot™ TOP10 chemically competent *E. coli* cells from Invitrogen™, miniprepped using Invisorb® Spin Plasmid Mini Two from STRATEC, and sequenced in-house with a 3130*xl* Genetic Analyzer from Applied Biosystems™ using pUC/M13 forward and reverse sequencing primers.

Real-time qPCR reactions for PaTPS-LAS, PaTPS-ISO and six PaCYP720B candidates were prepared using SYBR® Green Supermix according to the manufacturer's protocol and analyzed using CFX Connect™ Real-Time System from Bio-Rad™. C_t values were measured using each set of target primers with three biological cDNA replicate samples and three technical replicates for each sample. A three-minute initial denaturation step at 95°C was used, followed by 50 cycles of a 10 second denaturation step at 95°C and annealing and elongation step for 20 seconds at 60°C. A final denaturation step of two minutes at 95°C was performed. Amplicon melting temperatures were determined by a 60-95°C ramp at a rate of 0.1°C/sec for five minutes and 50 seconds. Amplified fragments were ligated directly into TOPO™ Zero Blunt™ and sequenced with M13 primers to verify fragment identities.

Enzymatic characterization of diterpene synthases and cytochrome P450s

Biochemical characterization of AncTPS-CON and SmDTC3.1

A nucleotide sequence encoding AncTPS-CON was synthesized (codon-optimized for expression in *Escherichia coli*, kindly provided by the Joint Genome Institute) with Gateway™-modified ends and was cloned into vector pGG (Cyr et al., 2007), which encodes a GGPP synthase that supplies free GGPP, using LR Clonase™ II thus creating pGG-AncTPS-CON. As a positive control and for product comparisons, PaTPS-LAS was amplified from sapling bark cDNA from *P. abies* with Gateway™-modified primers and cloned into pGG as above to yield pGG-LAS. SmDTC3 and SmDTC3.1 were subcloned into pGG-eC, pGG-sC, and pGG-nC, each containing an *ent*-, *syn*-, and normal-CPP synthase, respectively (Cyr et al., 2007).

Heterologous expression was performed following Cyr et al. (2007) with only minimal changes. pGG constructs were individually co-transformed into C41 OverExpress™ cells along with vector pIRS, which increases free IPP and DMAPP in *E. coli* by heterologously expressing upstream enzymes of the isoprenoid pathway (Morrone et al., 2010). 50 mL cultures were grown to an $O.D._{600} \approx 0.6$ in terrific broth containing 10% Tris-HCl buffer (pH = 7.5) in the presence of spectinomycin (34 µg/ml) and chloramphenicol (50 µg/ml). Cell cultures were cooled for one hour to 16°C, induced with IPTG at a final concentration of 1 mM, and shaken at 180 RPM for 72 hours. pGG, pGG-eC, pGG-sC, and pGG-nC were kindly provided by Reuben Peters.

Cell cultures were extracted once with equal volumes of hexane in a 100 ml separatory funnel, adding saturated NaCl solution when necessary to lower surface tension of emulsified layers. Organic phases were then further dried with 1 g Na_2SO_4 and evaporated under a gentle N_2 flow. Extracts of AncTPS-CON and modern-day PaTPS-LAS were resuspended in either 1 mL hexane for analysis. A coexpression assay with PaTPS-LAS was also replicated as above, but resuspended in 50 µL hexane.

All GC-MS analyses were carried out on an Agilent 6890 GC system containing a Machery-Nagel™ Optima® 30 m by 0.25 mm fused silica capillary column and an Agilent 5973 Mass Selective Detector. 1µL of each sample was injected at 270°C with an initial oven temperature of 150°C and held for 3 minutes. Oven temperature was ramped at a rate of 3.5°C/min to 280°C and held at 280°C for 4 minutes. Product identity was confirmed by comparing retention factors and mass spectra to olefin standards synthesized from corresponding individual resin acids. The highly-concentrated PaTPS-LAS extract from above was additionally run using the same parameters, but with a 10°C/min oven heating rate to 280°C.

Heterologous expression and yeast microsomal preparation of P450

Coding sequences for PaCYP720B1, PaCYP720B3, PaCYP720B4, PaCYP720B5, PaCYP720B6, AtCYP720A1 and PtCYP720A1 were synthesized by Thermo Fisher Scientific, codon-optimized for expression in *Saccharomyces cerevisiae*, and inserted between the NotI and SacI cloning sites of pESC-LEU from Agilent Technologies. Each insert contained an in-frame stop codon to remove the pESC-LEU FLAG epitope. Wat11 yeast, an engineered strain of *S. cerevisiae* containing a genome-integrated cytochrome P450 reductase from *A. thaliana*, was transformed with each pESC-LEU construct and plated on leucine dropout SC plates. A single colony of yeast containing each construct was grown in 30 mL leucine dropout SC media for 18 hours at 28°C under continuous shaking at 180 RPM. 500 µL of each culture was transferred to 100 mL YPA media containing 100 g/L filter-sterilized glucose in 500 mL vented glass flasks and shaken continuously at 180 RPM at 28°C for a further 32 hours. After reaching an $O.D._{600}$1:10 ≈ 0.6, cultures were centrifuged at 5000 RPM at 16°C for five minutes. Supernatant was removed and each pellet was washed three times by resuspending in sterile H_2O, centrifuging again, and clearing the supernatant. Washed pellets were then resuspended in 100 mL YPA media

containing 100 g/L filter-sterilized galactose and shaken continuously at 160 RPM for a further 18 hours at 25°C.

Yeast cells were pelleted by centrifugation at 4°C for 10 minutes at 7500 RPM in 50 mL falcon tubes. After removal of supernatant, pellets were rinsed briefly twice with 1 mL TEK buffer (50 mM TRIS-HCl, 1 mM EDTA, 100 mM KCl, pH = 7.5) and resuspended in 5 mL TES buffer (50 mM TRIS-HCl, 1 mM EDTA, 600 mM sorbitol, 10 g/L bovine serum albumin, pH = 7.5). Cells were disrupted by vigorous shaking with 2 mm glass beads for 10 minutes and lysate was transferred to a new 50 mL falcon tube. Beads were washed three times with 5 mL TES buffer and supernatant was transferred to the falcon tube containing yeast lysate. Lysates were subjected to ultracentrifugation at 35000 RPM at 4°C for 90 minutes and supernatant was cleared. Pellets containing microsomes were washed once with 1 mL TES buffer, followed by 1 mL TEG buffer (50 mM TRIS-HCl, 1 mM EDTA, 30% glycerol, pH = 7.5), and resuspended in 2 mL TEG buffer. Suspended microsomes were then homogenized in sterile potters, flash-frozen in liquid nitrogen and stored at -80°C until used.

In vitro microsomal CYP720 assays and GC-MS analysis

Microsomal preparations containing heterologously expressed CYP720 and empty pESC-LEU vector were assayed according to a modified version of (Ro et al., 2005). 75 µL of microsomes were added to 225 µL 100 mM potassium phosphate buffer containing 1µl of 100 mM diterpene olefin substrate, 0.25 mg glucose-6-phosphate, 1 µL (1 unit) of glucose-6-phosphate dehydrogenase, and 143 µL H_2O in a 2 mL Eppendorf Tube®. Reactions were initiated and brought to a final volume of 450 µL with 5 µl 90 mM NADPH and tubes were gently rocked horizontally at 28°C for one hour. Each assay was extracted three times with 500 µL ethyl acetate, transferred to a new tube and evaporated at room temperature under a gentle flow of N_2 gas. Extracts were resuspended in 100 µl ethyl acetate and either derivitized with 5 µL Trimethylsulfonium hydroxide (TMSH) for methylation of carboxy groups at room temperature for 30 minutes or used directly in GC-MS analysis. CYP720 microsomes were also incubated with cholesterol and campesterol and extracted with ethyl acetates exactly as above. Ethyl acetate was then completely evaporated in 1 mL HPLC vials, and residues were respsunded in 200 µL TMS-HT derivatization reagent (Tokyo Chemical Industry) in anhydrous pyridine for

silation of hydroxylated compounds, heated at 80°C for 30 minutes, and transferred to HPLC vial inserts.

For P450 assays with olefin substrates, GC-MS analyses were carried out on the same system described above for di-TPSs. 1 µL of each concentrated assay extract was injected at 270°C with an initial oven temperature of 150°C and held for 3 minutes with a column pressure and flow rate of 13 psi and 1 mL/min, respectively. Oven temperature was ramped at a rate of 3.5°C/min to 280°C and held at 280°C for 4 minutes. Parent ions were extracted for visualization of product peaks in derivatized samples. For methylated dehydroabietic acid, m/z = 314 was extracted and m/z = 316 was extracted for all other methylated products. Derivatized resin acids were identified by comparison of mass spectra and retention times to derivatized commercial standards. 1 µL of each assay extract after incubation with cholesterol or campesterol were injected on the same GC-MS system at 270°C with an initial oven temperature of 180°C and held for one minute, followed by a 20°C/min ramp to 300°C and held for 16 minutes. Negative results were determined by comparison to pESC-LEU empty vector assay extracts.

Heterologous expression and in vivo characterization of diterpene synthase genes in E. coli

di-TPS activity was assessed in *E. coli* using a modular di-TPS coexpression system that supplies substrates *in vivo* to di-TPSs of interest (cite Reuben 2007). Briefly, an engineered vector (pIRS, cite Reuben 2010) containing a spectinomycin resistance marker expresses the MEP pathway genes *idi*, *dxr*, and *dxs* to allow accumulation of free DMAPP and IPP in *E. coli*. A second vector (pGG) derived from pACYCDuet™-1 containing chloramphenicol resistance and a pseudo-mature GGPP synthase from the conifer *Abies grandis* makes use of the free isoprenyl diphosphates to supply free GGPP in *E. coli* . A suspected bifunctional di-TPS can then be inserted into a second cloning site within this vector using standard gateway cloning, and expression of this system provides the putative bifunctional di-TPS of interest with access to GGPP as a substrate.

N-terminally truncated versions of TPS-LAS and TPS-ISO were inserted into pGG from pDONR207 with Gateway™ LR Clonase™ II following the manufacturer's instructions. This created the constructs pGG-LAS and pGG-ISO. pGG-LAS and pGG-ISO were cotransformed into OverExpress™ C41 (DE3) cells. Three colonies containing each construct combination were

grown in 50 mL Terrific Broth containing 20µg/mL chloramphenicol and 25µg/mL spectinomycin at 37°C until reaching an OD_{600} of 0.6-0.8. Cultures were then cooled to 16°C for one hour prior to induction. Culture pH was then stabilized by the addition of 5.6 mL of TRIS-HCl buffer (pH = 7.5) and induced with IPTG at a final concentration of 1mM. Induction cultures were shaken for 72 hours and extracted once with equal volumes of hexane. Hexane was removed using a rotary evaporator and extracts were resuspended in 200 µL diethyl ether.

Live plant material, treatments and genetic manipulation

***P. abies* sapling treatment with methyl jasmonate**

For MeJA treatment and PaTPS expression analysis, three-year-old monoclonal wild-type Norway spruce saplings (genotype 3369-Schongau) were cultivated outdoors and an experimental group (n = 9) was treated via aerosol on aboveground tissues with 1mM methyl jasmonate solution containing Tween 20 (0.05% v/v) until runoff. A control group (n = 12) was not sprayed. Root and stem tissues were excised from three untreated plants, representing a "time zero" time point, and root and stem tissues were subsequently harvested from three treated and untreated plants each at 1, 9, and 15 days post induction (dpi). All tissues were immediately flash-frozen and later homogenized with mortar and pestle under liquid nitrogen.

A similar treatment was applied to six additional plants of the same Norway spruce line for induction of P450s involved in resin acid biosynthesis. For this set of experiments, four control plants were left untreated, and six saplings were sprayed with methyl jasmonate and roots, stems and needles were harvested separately from three plants at two and six dpi.

Generation of transgenic *P. abies* lines

For generating overexpression constructs of IDSs, di-TPSs, and P450s, open reading frames of each were cloned into pCAMGW, which carries a *Zea mays* ubiquitin promoter (Schmidt et al., 2010). Each construct, or pCAMBIA2301 as a control, was transformed into *Agrobacterium tumefaciens* strain C58/pMP90 according to Schmidt et al. (2010). Transcript knockdown was achieved through RNA-mediated suppression. For RNAi constructs, approximately 300-base-pair regions within open reading frames of low homology to other transcripts of the same gene/protein class were amplified by PCR and cloned in sense and antisense orientations into an

RNAi cassette flanking an intervening sequence, the latter providing stability to the inverted repeat DNA via a spacer region between the arms of the hairpin loop. This cassette carried the same *Z. mays* promoter described above and 35s terminator and was subcloned into pCAMBIA1305.2 (www.cambia.org) and transformed into *A. tumefaciens* as above.

A. tumefaciens-mediated transformation of *P. abies* embryogenic tissue culture was also performed according to Schmidt et al. (2010). Briefly, one volume of each transformed *A. tumefaciens* (OD_{600} = 0.6) cell culture in MLV media was added to an equal volume of embryogenic *P. abies* cells alongside negative controls where *A. tumefaciens* was not added. Cells were separated from liquid via paper filtration and a five-second vacuum pulse, and filter paper was placed on semi-solid MLV media containing 50 µM acetosyringone in petri dishes for 48 hours. Six replicate dishes were used for each transformant and controls. Filter paper was then transferred to 250 mL Erlenmeyer flasks containing 50 mL liquid MLV media, and cells were harvested following manual shaking and funneling the media through new filter paper. Cells on filter paper were then transferred to new MLV media containing 250 mg/L cefotaxime from Duchefa. After 48 hours, filter paper containing cells were transferred to MLV containing cefotaxime and 35 mg/L kanamycin.

Additional methods

Diterpene Standards and Substrates

Dehydroabietic acid (90%) was purchased from FUJIFILM Wako Pure Chemical Corporation. Abietic acid (85%) was purchased from Acros Organics. Palustric acid (99%) was purchased from Toronto Research Chemicals. Neoabietic acid (99%) and levopimaric acid (99%) were purchased from Santa Cruz Biotech. Isopimaric acid (99%) and pimaric acid (99% as a standard, 84% used for synthesis) were purchased from Alomone labs. Sandaracopimaric acid was purchased from CHEMOS GmbH as a standard. Campesterol (65%) and cholesterol (99%) were purchased from Sigma Aldrich. The eight diterpene olefins, dehydroabietadiene, abietadiene, levopimaradiene, palustradiene, neoabietadiene, pimaradiene, sandaracopimaradiene, isopimaradiene and their respective diterpene alcohols dehydroabietadienol, abietadienol, levopimaradienol, palustradienol, neoabietadienol, pimaradienol (84% purity, containing 13% sandaracopimaradienol) and isopimaradienol were synthesized following established methods (Lee et al., 2001). As starting molecules, the corresponding diterpene acids dehydroabietic acid,

abietic acid, levopimaric acid, palustric acid, neoabietic acid, pimaric acid (84% purity, contained 13% sandaracopimaric acid), and isopimaric acid were used. Dehydroabietadiene, abietadiene, neoabietadiene and isopimaradiene were obtained as pure compounds using this method. In contrast, further purification by preparative-TLC (SiO_2-$AgNO_3$, developed twice with 100% hexane) was needed to obtain pure palustradiene, levopimaradiene, pimaradiene and sandaracopimaradiene (79% purity, containing 21% pimaradiene). Dehydroabietadienal was synthesized from dehydroabietadienol according to published methods (Chaturvedi et al., 2012). The structures and purities of all synthesized compounds were confirmed by NMR spectroscopy and GC-MS analysis.

Purification and detection of dehydroabietadienal from *P. trichocarpa* leaves

For purification of dehydroabietadienal (DA) from leaves of *P. trichocarpa,* 20 mg of pulverized freeze-dried tissue was extracted once with 20 mL of MeOH at room temperature. Vials were sonicated in a water bath for three minutes at 25°C and agitated until all lipophilic residue was suspended. A 6 mL Chromabond® HR-X solid phase extraction column was washed once with one volume H_2O and equilibrated with MeOH prior to loading leaf extract. All leaf extract was allowed to pass through the column resin layer via gravity flow. The column was then subjected to airflow with a vacuum pump for two hours until the resin was dry. Dried resin was equilibrated again with 2 mL MeOH, eluted twice with 1 mL of MeOH, and three times with 1 mL acetone. The second, third and fourth acetone fractions contained DA, which were combined and concentrated to 100 µL using a gentle N_2 stream. As per Chaturvedi et al. (2012), 2 µL of concentrated acetone eluate was injected on the same GC-MS system described above at 265°C at an initial oven temperature of 40°C and held for one minute with a column pressure and flow rate of 8 psi and 1 mL/min, respectively. Oven temperature was ramped at a rate of 10°C/min to 300°C and held for 5 minutes. Characteristic ions for DA of m/z = 284 and 269 were extracted for visualization of standard and corresponding peak from leaf extract

Extraction and GC-MS analysis of Norway spruce terpene content

All plant tissues were excised, immediately frozen in liquid nitrogen and stored at -80°C until further processing. Plant tissues were pulverized to a fine powder under a constant supply of liquid nitrogen. 100 mg frozen tissue was then extracted once overnight at room temperature following addition of 2 ml ether containing 50 µg ml/L 1,9-decadiene (Sigma) and 47.3 µg ml/L

dichlordehydroabietic acid as internal standards and a 30-second vortexing step. 1.6 ml of ether extract was used for monoterpene and sesquiterpene analyses. 50 µl 0.2M N-trimethylsulfonium hydroxide was added to the remaining 400 µl to derivatize the carboxy group of diterpene resin acids, briefly vortexed, and allowed to stand at room temperature for one hour prior to GC injection.

All GC-MS analyses were carried out on the GC-MS system described above. For monoterpene analysis, samples were injected at 270°C with an initial oven temperature of 40°C and held for three minutes. An oven ramp step then followed at a rate of 5°C/min to 200°C, followed by a second ramp at 60°C/min to 300°C and a 4 minute bake out. Derivatize diterpene fractions were injected at 270°C with an initial oven temperature of 150°C and held for three minutes, followed by a 3.5°C/min ramp to 300 degrees and held for five minutes. Peak areas of analytes and internal standards were used to approximate absolute abundances and normalized to amount of tissue used for extraction.

PaCYP720B4 homology modeling

CYP720B4 was modeled using the SWISS-MODEL server (Schwede, 2003) and the resolved crystal structure of *A. thaliana* CYP90B1 bound to cholesterol (PDB accession 6a15.1.A) as template. Hydrogens were added and energy minimalization was performed using ModRefiner (Xu & Zhang, 2011) and docking simulations with energy-minimalized dehydroabietadiene were performed using AutoDock Vina (Trott & Olson, 2010) using a rigid ligand, selecting docking coordinates immediately adjacent to the apical surface of the heme group. The best-scoring pose of dehydroabietadiene was chosen for visualization.

Bibliography

Abascal, F., Zardoya, R., & Posada, D. (2005). ProtTest: selection of best-fit models of protein evolution. *Bioinformatics*, *21*(9), 2104–2105. https://doi.org/10.1093/bioinformatics/bti263

Aharoni, A., Gaidukov, L., Khersonsky, O., Gould, S. M. Q., Roodveldt, C., & Tawfik, D. S. (2005). The "evolvability" of promiscuous protein functions. *Nature Genetics*. https://doi.org/10.1038/ng1482

Alicandri, E., Paolacci, A. R., Osadolor, S., Sorgonà, A., Badiani, M., & Ciaffi, M. (2020). On the evolution and functional diversity of terpene synthases in the *Pinus* species: a review. *Journal of Molecular Evolution 88*(3), 253–283. https://doi.org/10.1007/s00239-020-09930-8

Bathe, U., & Tissier, A. (2019). Cytochrome P450 enzymes: a driving force of plant diterpene diversity. *Phytochemistry*, *161*, 149–162. https://doi.org/10.1016/j.phytochem.2018.12.003

Bensen, R. J., Johal, G. S., Crane, V. C., Tossberg, J. T., Schnable, P. S., Meeley, R. B., & Briggs, S. P. (1995). Cloning and characterization of the maize An1 gene. *Plant Cell*, *7*(1), 75–84. https://doi.org/10.1105/tpc.7.1.75

Bergthorsson, U., Andersson, D. I., & Roth, J. R. (2007). Ohno's dilemma: evolution of new genes under continuous selection. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(43), 17004–17009. https://doi.org/10.1073/pnas.0707158104

Bhat, P. J., & Murthy, T. V. S. (2001). Transcriptional control of the GAL/MEL regulon of yeast *Saccharomyces cerevisiae*: mechanism of galactose-mediated signal transduction. *Molecular Microbiology*, *40*(5), 1059–1066. https://doi.org/10.1046/j.1365-2958.2001.02421.x

Biedermann, P. H. W., Müller, J., Grégoire, J. C., Gruppe, A., Hagge, J., Hammerbacher, A., Hofstetter, R. W., Kandasamy, D., Kolarik, M., Kostovcik, M., Krokene, P., Sallé, A., Six, D. L., Turrini, T., Vanderpool, D., Wingfield, M. J., & Bässler, C. (2019). Bark beetle population dynamics in the anthropocene: challenges and solutions. *Trends in Ecology & Evolution*, *34*(10), 914–924. https://doi.org/10.1016/j.tree.2019.06.002

Blount, Z. D., Borland, C. Z., & Lenski, R. E. (2008). Historical contingency and the evolution of a key innovation in an experimental population of *Escherichia coli*. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(23), 7899–7906. https://doi.org/10.1073/pnas.0803151105

Blount, Z. D., Lenski, R. E., & Losos, J. B. (2018). Contingency and determinism in evolution: replaying life's tape. In *Science*. https://doi.org/10.1126/science.aam5979

Bohlmann, J., Steele, C. L., & Croteau, R. (1997). Monoterpene synthases from grand fir (*Abies grandis*). *Journal of Biological Chemistry*, *272*(35), 21784–21792. https://doi.org/10.1074/jbc.272.35.21784

Boutanaev, A. M., Moses, T., Zi, J., Nelson, D. R., Mugford, S. T., Peters, R. J., & Osbourn, A. (2015). Investigation of terpene diversification across multiple sequenced plant genomes. *Proceedings of the National Academy of Sciences*, *112*(1), E81–E88. https://doi.org/10.1073/pnas.1419547112

Božić, D., Papaefthimiou, D., Brückner, K., de Vos, R. C. H., Tsoleridis, C. A., Katsarou, D., Papanikolaou, A., Pateraki, I., Chatzopoulou, F. M., Dimitriadou, E., Kostas, S., Manzano, D., Scheler, U., Ferrer, A., Tissier, A., Makris, A. M., Kampranis, S. C., & Kanellis, A. K. (2015). Towards elucidating carnosic acid biosynthesis in Lamiaceae: functional characterization of the three first steps of the pathway in *Salvia fruticosa* and *Rosmarinus officinalis*. *PLOS ONE*, *10*(5), e0124106. https://doi.org/10.1371/journal.pone.0124106

Breen, M. S., Kemena, C., Vlasov, P. K., Notredame, C., & Kondrashov, F. A. (2012). Epistasis as the primary factor in molecular evolution. *Nature*, *490*(7421), 535–538. https://doi.org/10.1038/nature11510

Bridgham, J. T., Ortlund, E. A., & Thornton, J. W. (2009). An epistatic ratchet constrains the direction of glucocorticoid receptor evolution. *Nature*. https://doi.org/10.1038/nature08249

Brown, R., Jia, M., & Peters, R. J. (2021). A pair of threonines mark ent-kaurene synthases for phytohormone biosynthesis. *Phytochemistry*, *184*. https://doi.org/10.1016/j.phytochem.2021.112672

Brückner, K., Božić, D., Manzano, D., Papaefthimiou, D., Pateraki, I., Scheler, U., Ferrer, A., De Vos, R. C. H., Kanellis, A. K., & Tissier, A. (2014). Characterization of two genes for the biosynthesis of abietane-type diterpenes in rosemary (*Rosmarinus officinalis*) glandular trichomes. *Phytochemistry*, *101*, 52–64. https://doi.org/10.1016/j.phytochem.2014.01.021

Byun-McKay, A., Godard, K. A., Toudefallah, M., Martin, D. M., Alfaro, R., King, J., Bohlmann, J., & Plant, A. L. (2006). Wound-induced terpene synthase gene expression in Sitka spruce that exhibit resistance or susceptibility to attack by the white pine weevil. *Plant Physiology*, *140*(3), 1009–1021. https://doi.org/10.1104/pp.105.071803

Cao, R., Zhang, Y., Mann, F. M., Huang, C., Mukkamala, D., Hudock, M. P., Mead, M. E., Prisic, S., Wang, K., Lin, F. Y., Chang, T. K., Peters, R. J., & Oldfield, E. (2010). Diterpene cyclases and the nature of the isoprene fold. *Proteins: Structure, Function and Bioinformatics*, *78*(11), 2417–2432. https://doi.org/10.1002/prot.22751

Chaturvedi, R., Venables, B., Petros, R. A., Nalam, V., Li, M., Wang, X., Takemoto, L. J., & Shah, J. (2012). An abietane diterpenoid is a potent activator of systemic acquired resistance. *The Plant Journal*, *71*(1), 161–172. https://doi.org/10.1111/j.1365-313X.2012.04981.x

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

Chen, K., Durand, D., & Farach-Colton, M. (2000). NOTUNG: A program for dating gene duplications and optimizing gene family trees. *Journal of Computational Biology*, *7*(3–4), 429–447. https://doi.org/10.1089/106652700750050871

Chen, L., Tong, H., Wang, M., Zhu, J., Zi, J., Song, L., & Yu, R. (2015). Effect of enzyme inhibitors on terpene trilactones biosynthesis and gene expression profiling in ginkgo biloba cultured cells. *Natural Product Communications*. https://doi.org/10.1177/1934578X1501001205

Chiu, C. C., Keeling, C. I., & Bohlmann, J. (2017). Toxicity of pine monoterpenes to mountain pine beetle. *Scientific Reports*, *7*(1), 8858. https://doi.org/10.1038/s41598-017-08983-y

Cho, E. M., Okada, A., Kenmoku, H., Otomo, K., Toyomasu, T., Mitsuhashi, W., Sassa, T., Yajima, A., Yabuta, G., Mori, K., Oikawa, H., Toshima, H., Shibuya, N., Nojiri, H., Omori, T., Nishiyama, M., & Yamane, H. (2004). Molecular cloning and characterization of a cDNA encoding *ent*-cassa-12,15-diene synthase, a putative diterpenoid phytoalexin biosynthetic enzyme, from suspension-cultured rice cells treated with a chitin elicitor. *Plant Journal*, *37*(1), 1–8. https://doi.org/10.1046/j.1365-313X.2003.01926.x

Christianson, D. W. (2008). Unearthing the roots of the terpenome. *Current Opinion in Chemical Biology*, *12*(2), 141–150. https://doi.org/10.1016/j.cbpa.2007.12.008

Copley, S. D. (2000). Evolution of a metabolic pathway for degradation of a toxic xenobiotic: the patchwork approach. *Trends in Biochemical Sciences*, *25*(6), 261–265. https://doi.org/10.1016/S0968-0004(00)01562-0

Cordoba, E., Salmi, M., & Leon, P. (2009). Unravelling the regulatory mechanisms that modulate the MEP pathway in higher plants. *Journal of Experimental Botany*, *60*(10), 2933–2943. https://doi.org/10.1093/jxb/erp190

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

Cui, G., Duan, L., Jin, B., Qian, J., Xue, Z., Shen, G., Snyder, J. H., Song, J., Chen, S., Huang, L., Peters, R. J., & Qi, X. (2015). Functional divergence of diterpene syntheses in the medicinal plant *Salvia miltiorrhiza*. *Plant Physiology*, *169*(3), 1607–1618. https://doi.org/10.1104/pp.15.00695

Cyr, A., Wilderman, P. R., Determan, M., & Peters, R. J. (2007). A modular approach for facile biosynthesis of labdane-related diterpenes. *Journal of the American Chemical Society*, *129*(21), 6684–6685. https://doi.org/10.1021/ja071158n

Dalla, S., & Dobler, S. (2016). Gene duplications circumvent trade-offs in enzyme function: insect adaptation to toxic host plants. *Evolution*. https://doi.org/10.1111/evo.13077

Danner, H., Boeckler, G. A., Irmisch, S., Yuan, J. S., Chen, F., Gershenzon, J., Unsicker, S. B., & Köllner, T. G. (2011). Four terpene synthases produce major compounds of the gypsy moth feeding-induced volatile blend of *Populus trichocarpa*. *Phytochemistry*, *72*(9), 897–908. https://doi.org/10.1016/j.phytochem.2011.03.014

Degenhardt, J., Köllner, T. G., & Gershenzon, J. (2009). Monoterpene and sesquiterpene synthases and the origin of terpene skeletal diversity in plants. In *Phytochemistry* (Vol. 70, Issues 15–16, pp. 1621–1637). https://doi.org/10.1016/j.phytochem.2009.07.030

Dejonghe, W., Mishev, K., & Russinova, E. (2014). The brassinosteroid chemical toolbox. *Current Opinion in Plant Biology*, *22*, 48–55. https://doi.org/10.1016/j.pbi.2014.09.002

Dereeper, A., Bocs, S., Rouard, M., Guignon, V., Ravel, S., Tranchant-Dubreuil, C., Poncet, V., Garsmeur, O., Lashermes, P., & Droc, G. (2015). The coffee genome hub: a resource for coffee genomes. *Nucleic Acids Research*, *43*(D1), D1028–D1035. https://doi.org/10.1093/nar/gku1108

Des Marais, D. L., & Rausher, M. D. (2008). Escape from adaptive conflict after duplication in an anthocyanin pathway gene. *Nature*, *454*(7205), 762–765. https://doi.org/10.1038/nature07092

Eick, G. N., Bridgham, J. T., Anderson, D. P., Harms, M. J., & Thornton, J. W. (2017). Robustness of reconstructed ancestral protein functions to statistical uncertainty. *Molecular Biology and Evolution*, *34*(2), 247–261. https://doi.org/10.1093/molbev/msw223

Emms, D. M., & Kelly, S. (2015). OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. *Genome Biology*, *16*(1), 157. https://doi.org/10.1186/s13059-015-0721-2

Erwin, D. H. (2006). Evolutionary contingency. *Current Biology*, *16*(19), R825–R826. https://doi.org/10.1016/j.cub.2006.08.076

Espinosa-Cantú, A., Ascencio, D., Barona-Gómez, F., & De Luna, A. (2015). Gene duplication and the evolution of moonlighting proteins. *Frontiers in Genetics* (Vol. 6, Issue JUL). https://doi.org/10.3389/fgene.2015.00227

Faccoli, M., Blaženec, M., & Schlyter, F. (2005). Feeding response to host and nonhost compounds by males and females of the spruce bark beetle *Ips typographus* in a tunneling microassay. *Journal of Chemical Ecology*, *31*(4), 745–759. https://doi.org/10.1007/s10886-005-3542-z

Fani, R., & Fondi, M. (2009). Origin and evolution of metabolic pathways. *Physics of Life Reviews*, *6*(1), 23–52. https://doi.org/10.1016/j.plrev.2008.12.003

Finnigan, G. C., Hanson-Smith, V., Stevens, T. H., & Thornton, J. W. (2012). Evolution of increased complexity in a molecular machine. *Nature*, *481*(7381), 360–364. https://doi.org/10.1038/nature10724

Franceschi, V. R., Krokene, P., Christiansen, E., & Krekling, T. (2005). Anatomical and chemical defenses of conifer bark against bark beetles and other pests. *New Phytologist*, *167*(2), 353–376. https://doi.org/10.1111/j.1469-8137.2005.01436.x

Fu, J., Ren, F., Lu, X., Mao, H., Xu, M., Degenhardt, J., Peters, R. J., & Wang, Q. (2016). A tandem array of *ent*-kaurene synthases in maize with roles in gibberellin and more specialized metabolism. *Plant Physiology*, *170*(2), 742–751. https://doi.org/10.1104/pp.15.01727

Fujita, S., Ohnishi, T., Watanabe, B., Yokota, T., Takatsuto, S., Fujioka, S., Yoshida, S., Sakata, K., & Mizutani, M. (2006). *Arabidopsis* CYP90B1 catalyses the early C-22 hydroxylation of C_{27}, C_{28} and C_{29} sterols. *Plant Journal*, *45*(5), 765–774. https://doi.org/10.1111/j.1365-313X.2005.02639.x

Fujiyama, K., Hino, T., Kanadani, M., Watanabe, B., Jae Lee, H., Mizutani, M., & Nagano, S. (2019). Structural insights into a key step of brassinosteroid biosynthesis and its inhibition. *Nature Plants*, *5*(6), 589–594. https://doi.org/10.1038/s41477-019-0436-6

Funk, C., & Croteau, R. (1994). Diterpenoid resin acid biosynthesis in conifers: characterization of two cytochrome P450-dependent monooxygenases and an aldehyde dehydrogenase involved in abietic acid biosynthesis. *Archives of Biochemistry and Biophysics*, *308*(1), 258–266. https://doi.org/10.1006/abbi.1994.1036

Gao, W., Hillwig, M. L., Huango, L., Cui, G., Wang, X., Kong, J., Yang, B., & Peters, R. J. (2009). A functional genomics approach to tanshinone biosynthesis provides stereochemical insights. *Organic Letters*, *11*(22), 5170–5173. https://doi.org/10.1021/ol902051v

Gaucher, E. A., Thomson, J. M., Burgan, M. F., & Benner, S. A. (2003). Inferring the palaeoenvironment of ancient bacteria on the basis of resurrected proteins. *Nature*, *425*(6955), 285–288. https://doi.org/10.1038/nature01977

Geisler, K., Jensen, N. B., Yuen, M. M. S., Madilao, L., & Bohlmann, J. (2016). Modularity of conifer diterpene resin acid biosynthesis: P450 enzymes of different CYP720B clades use alternative substrates and converge on the same products. *Plant Physiology*, *171*(1), 152–164. https://doi.org/10.1104/pp.16.00180

Gong, L. I., Suchard, M. A., & Bloom, J. D. (2013). Stability-mediated epistasis constrains the evolution of an influenza protein. *ELife*, *2*(2). https://doi.org/10.7554/eLife.00631

Goodstein, D. M., Shu, S., Howson, R., Neupane, R., Hayes, R. D., Fazo, J., Mitros, T., Dirks, W., Hellsten, U., Putnam, N., & Rokhsar, D. S. (2012). Phytozome: a comparative platform for green plant genomics. *Nucleic Acids Research*, *40*(D1), D1178–D1186. https://doi.org/10.1093/nar/gkr944

Granick, S. (1965). Evolution of heme and chlorophyll. *Evolving Genes and Proteins* (pp. 67–88). Elsevier. https://doi.org/10.1016/b978-1-4832-2734-4.50014-0

Guindon, S., Dufayard, J.-F., Lefort, V., Anisimova, M., Hordijk, W., & Gascuel, O. (2010). New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Systematic Biology*, *59*(3), 307–321. https://doi.org/10.1093/sysbio/syq010

Guo, J., Ma, X., Cai, Y., Ma, Y., Zhan, Z., Zhou, Y. J., Liu, W., Guan, M., Yang, J., Cui, G., Kang, L., Yang, L., Shen, Y., Tang, J., Lin, H., Ma, X., Jin, B., Liu, Z., Peters, R. J., Zhao, Z.K., Huang, L. (2016). Cytochrome P450 promiscuity leads to a bifurcating biosynthetic pathway for tanshinones. *New Phytologist*, *210*(2), 525–534. https://doi.org/10.1111/nph.13790

Guo, J., Zhou, Y. J., Hillwig, M. L., Shen, Y., Yang, L., Wang, Y., Zhang, X., Liu, W., Peters, R. J., Chen, X., Zhao, Z. K., & Huang, L. (2013). CYP76AH1 catalyzes turnover of miltiradiene in tanshinones biosynthesis and enables heterologous production of ferruginol in yeasts. *Proceedings of the National Academy of Sciences*, *110*(29), 12108–12113. https://doi.org/10.1073/pnas.1218061110

Hall, D. E., Zerbe, P., Jancsik, S., Quesada, A. L., Dullat, H., Madilao, L. L., Yuen, M., & Bohlmann, J. (2013). Evolution of conifer diterpene synthases: diterpene resin acid biosynthesis in lodgepole pine and jack pine involves monofunctional and bifunctional diterpene synthases. *Plant Physiology*, *161*(2), 600–616. https://doi.org/10.1104/pp.112.208546

Hamberger, B., & Bak, S. (2013). Plant P450s as versatile drivers for evolution of species-specific chemical diversity. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *368*(1612), 20120426. https://doi.org/10.1098/rstb.2012.0426

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

Hansen, N. L., Heskes, A. M., Hamberger, B., Olsen, C. E., Hallström, B. M., Andersen-Ranberg, J., & Hamberger, B. (2017). The terpene synthase gene family in *Tripterygium wilfordii* harbors a labdane-type diterpene synthase among the monoterpene synthase TPS-b subfamily. *The Plant Journal*, *89*(3), 429–441. https://doi.org/10.1111/tpj.13410

Harris, L. J., Saparno, A., Johnston, A., Prisic, S., Xu, M., Allard, S., Kathiresan, A., Ouellet, T., & Peters, R. J. (2005). The maize An2 gene is induced by *Fusarium* attack and encodes an *ent*-copalyl diphosphate synthase. *Plant Molecular Biology*, *59*(6), 881–894. https://doi.org/10.1007/s11103-005-1674-8

Hasemann, C. A., Kurumbail, R. G., Boddupalli, S. S., Peterson, J. A., & Deisenhofer, J. (1995). Structure and function of cytochromes P450: a comparative analysis of three crystal structures. *Structure*, *3*(1), 41–62. https://doi.org/10.1016/S0969-2126(01)00134-4

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

Hedden, P., & Sponsel, V. (2015). A century of gibberellin research. *Journal of Plant Growth Regulation*, *34*(4), 740–760. https://doi.org/10.1007/s00344-015-9546-1

Helliwell, C. A., Poole, A., Peacock, W. J., & Dennis, E. S. (1999). *Arabidopsis ent*-kaurene oxidase catalyzes three steps of gibberellin biosynthesis. *Plant Physiology*, *119*(2), 507–510. https://doi.org/10.1104/pp.119.2.507

Hemmerlin, A., Hoeffler, J. F., Meyer, O., Tritsch, D., Kagan, I. A., Grosdemange-Billiard, C., Rohmer, M., & Bach, T. J. (2003). Cross-talk between the cytosolic mevalonate and the plastidial methylerythritol phosphate pathways in tobacco bright yellow-2 cells. *Journal of Biological Chemistry*, *278*(29), 26666–26676. https://doi.org/10.1074/jbc.M302526200

Heskes, A. M., Sundram, T. C. M., Boughton, B. A., Jensen, N. B., Hansen, N. L., Crocoll, C., Cozzi, F., Rasmussen, S., Hamberger, B., Hamberger, B., Staerk, D., Møller, B. L., & Pateraki, I. (2018). Biosynthesis of bioactive diterpenoids in the medicinal plant *Vitex agnus-castus*. *Plant Journal*, *93*(5), 943–958. https://doi.org/10.1111/tpj.13822

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

Hittinger, C. T., & Carroll, S. B. (2007). Gene duplication and the adaptive evolution of a classic genetic switch. *Nature*, *449*(7163), 677–681. https://doi.org/10.1038/nature06151

Hochberg, G. K. A., & Thornton, J. W. (2017). Reconstructing ancient proteins to understand the causes of structure and function. *Annual Review of Biophysics*. https://doi.org/10.1146/annurev-biophys-070816-033631

Horowitz, N. H. (1945). On the evolution of biochemical syntheses. *Proceedings of the National Academy of Sciences*, *31*(6), 153–157. https://doi.org/10.1073/pnas.31.6.153

Hou, X., Tai, Y., Ling, S., Li, D., Guo, C., Sun, J., Pan, G., & Yuan, Y. (2018). Molecular cloning and characterization of squalene synthase from *Matricaria recutita* L. *Acta Physiologiae Plantarum*, *40*(6), 103. https://doi.org/10.1007/s11738-018-2675-y

Hughes, A. L. (1994). The evolution of functionally novel proteins after gene duplication. *Proceedings of the Royal Society B: Biological Sciences*, *256*(1346), 119–124. https://doi.org/10.1098/rspb.1994.0058

Innan, H., & Kondrashov, F. (2010). The evolution of gene duplications: classifying and distinguishing between models. *Nature Reviews Genetics*, *11*(2), 97–108. https://doi.org/10.1038/nrg2689

Irmisch, S., Müller, A. T., Schmidt, L., Günther, J., Gershenzon, J., & Köllner, T. G. (2015). One amino acid makes the difference: the formation of *ent*-kaurene and 16α-hydroxy-*ent*-kaurane by diterpene synthases in poplar. *BMC Plant Biology*, *15*(1), 262. https://doi.org/10.1186/s12870-015-0647-6

Jensen, R. A. (1976). Enzyme recruitment in evolution of new function. *Annual Review of Microbiology*, *30*(1), 409–425. https://doi.org/10.1146/annurev.mi.30.100176.002205

Jermann, T. M., Opitz, J. G., Stackhouse, J., & Benner, S. A. (1995). Reconstructing the evolutionary history of the artiodactyl ribonuclease superfamily. *Nature*, *374*(6517), 57–59. https://doi.org/10.1038/374057a0

Jia, Q., Li, G., Köllner, T. G., Fu, J., Chen, X., Xiong, W., Crandall-Stotler, B. J., Bowman, J. L., Weston, D. J., Zhang, Y., Chen, L., Xie, Y., Li, F.W., Rothfels, C. J., Larsson, A., Graham, S. W., Stevenson, D. W., Wong, G. K.S., Gershenzon, J., & Chen, F. (2016). Microbial-type terpene synthase genes occur widely in nonseed land plants, but not in seed plants. *Proceedings of the National Academy of Sciences*, *113*(43), 12328–12333. https://doi.org/10.1073/pnas.1607973113

Jin, B., Cui, G., Guo, J., Tang, J., Duan, L., Lin, H., Shen, Y., Chen, T., Zhang, H., & Huang, L. (2017). Functional diversification of kaurene synthase-like genes in *Isodon rubescens*. *Plant Physiology*, *174*(2), 943–955. https://doi.org/10.1104/pp.17.00202

Johnston, M. (1987). A model fungal gene regulatory mechanism: the GAL genes of *Saccharomyces cerevisiae*. *Microbiological Reviews*, *51*(4), 458–476. https://doi.org/10.1128/MMBR.51.4.458-476.1987

Kanno, Y., Otomo, K., Kenmoku, H., Mitsuhashi, W., Yamane, H., Oikawa, H., Toshima, H., Matsuoka, M., Sassa, T., & Toyomasu, T. (2006). Characterization of a rice gene family encoding type-A diterpene cyclases. *Bioscience, Biotechnology and Biochemistry*, *70*(7), 1702–1710. https://doi.org/10.1271/bbb.60044

Karunanithi, P. S., & Zerbe, P. (2019). Terpene synthases as metabolic gatekeepers in the evolution of plant terpenoid chemical diversity. *Frontiers in Plant Science*, *10*, 1166. https://doi.org/10.3389/fpls.2019.01166

Kato-Noguchi, H., Ino, T., Sata, N., & Yamamura, S. (2002). Isolation and identification of a potent allelopathic substance in rice root exudates. *Physiologia Plantarum*, *115*(3), 401–405. https://doi.org/10.1034/j.1399-3054.2002.1150310.x

Katoh, K., & Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Molecular Biology and Evolution*, *30*(4), 772–780. https://doi.org/10.1093/molbev/mst010

Keeling, C. I., & Bohlmann, J. (2006). Diterpene resin acids in conifers. *Phytochemistry*, *67*(22), 2415–2423. https://doi.org/10.1016/j.phytochem.2006.08.019

Keeling, C. I., Dullat, H. K., Yuen, M., Ralph, S. G., Jancsik, S., & Bohlmann, J. (2010). Identification and functional characterization of monofunctional *ent*-copalyl diphosphate and *ent*-kaurene synthases in white spruce reveal different patterns for diterpene synthase evolution for primary and secondary metabolism in gymnosperms. *Plant Physiology*, *152*(3), 1197–1208. https://doi.org/10.1104/pp.109.151456

Keeling, C. I., Madilao, L. L., Zerbe, P., Dullat, H. K., & Bohlmann, J. (2011a). The primary diterpene synthase products of *Picea abies* levopimaradiene/abietadiene synthase (PaLAS) are epimers of a thermally unstable diterpenol. *Journal of Biological Chemistry*, *286*(24), 21145–21153. https://doi.org/10.1074/jbc.M111.245951

Keeling, C. I., Weisshaar, S., Lin, R. P. C., & Bohlmann, J. (2008). Functional plasticity of paralogous diterpene synthases involved in conifer defense. *Proceedings of the National Academy of Sciences*, *105*(3), 1085–1090. https://doi.org/10.1073/pnas.0709466105

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

Kim, B. K., Fujioka, S., Takatsuto, S., Tsujimoto, M., & Choe, S. (2008). Castasterone is a likely end product of brassinosteroid biosynthetic pathway in rice. *Biochemical and Biophysical Research Communications*, *374*(4), 614–619. https://doi.org/10.1016/j.bbrc.2008.07.073

Kim, T.W., Hwang, J.Y., Kim, Y.S., Joo, S.H., Chang, S. C., Lee, J. S., Takatsuto, S., & Kim, S.K. (2005). *Arabidopsis* CYP85A2, a cytochrome P450, mediates the baeyer-villiger oxidation of castasterone to brassinolide in brassinosteroid biosynthesis. *The Plant Cell*, *17*(8), 2397–2412. https://doi.org/10.1105/tpc.105.033738

Kimura, M. (1983). *The Neutral Theory of Molecular Evolution*. Cambridge University Press. https://doi.org/10.1017/CBO9780511623486

Köksal, M., Zimmer, I., Schnitzler, J. P., & Christianson, D. W. (2010). Structure of isoprene synthase illuminates the chemical mechanism of teragram atmospheric carbon emission. *Journal of Molecular Biology*, *402*(2), 363–373. https://doi.org/10.1016/j.jmb.2010.07.009

Kopper, B. J., Illman, B. L., Kersten, P. J., Klepzig, K. D., & Raffa, K. F. (2005). Effects of diterpene acids on components of a conifer bark beetle-fungal interaction: tolerance by *Ips pini* and sensitivity by its associate *Ophiostoma ips*. *Environmental Entomology*, *34*(2), 486–493. https://doi.org/10.1603/0046-225X-34.2.486

Krause, T., Reichelt, M., Gershenzon, J., & Schmidt, A. (2020). Analysis of the isoprenoid pathway intermediates, dimethylallyl diphosphate and isopentenyl diphosphate, from crude plant extracts by liquid chromatography tandem mass spectrometry. *Phytochemical Analysis*, *31*(6), 770–777. https://doi.org/10.1002/pca.2941

Kumar, S., Kempinski, C., Zhuang, X., Norris, A., Mafu, S., Zi, J., Bell, S. A., Nybo, S. E., Kinison, S. E., Jiang, Z., Goklany, S., Linscott, K. B., Chen, X., Jia, Q., Brown, S. D., Bowman, J. L., Babbitt, P. C., Peters, R. J., Chen, F., & Chappell, J. (2016). Molecular diversity of terpene synthases in the liverwort *Marchantia polymorpha*. *Plant Cell*, *28*(10), 2632–2650. https://doi.org/10.1105/tpc.16.00062

Laule, O., Furholz, A., Chang, H.S., Zhu, T., Wang, X., Heifetz, P. B., Gruissem, W., & Lange, M. (2003). Crosstalk between cytosolic and plastidial pathways of isoprenoid biosynthesis in *Arabidopsis thaliana*. *Proceedings of the National Academy of Sciences*, *100*(11), 6866–6871. https://doi.org/10.1073/pnas.1031755100

Lee, H.J., Ravn, M. M., & Coates, R. M. (2001). Synthesis and characterization of abietadiene, levopimaradiene, palustradiene, and neoabietadiene: hydrocarbon precursors of the abietane diterpene resin acids. *Tetrahedron*, *57*(29), 6155–6167. https://doi.org/10.1016/S0040-4020(01)00605-6

Li, F.W., Brouwer, P., Carretero-Paulet, L., Cheng, S., de Vries, J., Delaux, P.M., Eily, A., Koppers, N., Kuo, L.Y., Li, Z., Simenc, M., Small, I., Wafula, E., Angarita, S., Barker, M. S., Bräutigam, A., DePamphilis, C., Gould, S., Hosmani, P. S., et al. (2018). Fern genomes elucidate land plant evolution and cyanobacterial symbioses. *Nature Plants*, *4*(7), 460–472. https://doi.org/10.1038/s41477-018-0188-8

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

Liang, P.H., Ko, T.P., & Wang, A. H. J. (2002). Structure, mechanism and function of prenyltransferases. *European Journal of Biochemistry*, *269*(14), 3339–3354. https://doi.org/10.1046/j.1432-1033.2002.03014.x

Lu, S., Xu, R., Jia, J.W., Pang, J., Matsuda, S. P. T., & Chen, X.Y. (2002). Cloning and functional characterization of a β-pinene synthase from *Artemisia annua* that shows a circadian pattern of expression. *Plant Physiology*, *130*(1), 477–486. https://doi.org/10.1104/pp.006544

Lücker, J., El Tamer, M. K., Schwab, W., Verstappen, F. W. A., van der Plas, L. H. W., Bouwmeester, H. J., & Verhoeven, H. A. (2002). Monoterpene biosynthesis in lemon (*Citrus limon*) - cDNA isolation and functional analysis of four monoterpene synthases. *European Journal of Biochemistry*, *269*(13), 3160–3171. https://doi.org/10.1046/j.1432-1033.2002.02985.x

Lynch, M., & Force, A. (2000). The probability of duplicate gene preservation by subfunctionalization. *Genetics, 154*(1), 459–473. www.genetics-gsa.org

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

Mafu, S., Ding, Y., Murphy, K. M., Yaacoobi, O., Addison, J. B., Wang, Q., Shen, Z., Briggs, S. P., Bohlmann, J., Castro-Falcon, G., Hughes, C. C., Betsiashvili, M., Huffaker, A., Schmelz, E. A., & Zerbe, P. (2018). Discovery, biosynthesis and stress-related accumulation of dolabradiene-derived defenses in maize. *Plant Physiology*, *176*(4), 2677–2690. https://doi.org/10.1104/pp.17.01351

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

Mageroy, M. H., Christiansen, E., Långström, B., Borg-Karlson, A. K., Solheim, H., Björklund, N., Zhao, T., Schmidt, A., Fossdal, C. G., & Krokene, P. (2020). Priming of inducible defenses protects Norway spruce against tree-killing bark beetles. *Plant Cell and Environment*, *43*(2), 420–430. https://doi.org/10.1111/pce.13661

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

Martin, D., Tholl, D., Gershenzon, J., & Bohlmann, J. (2002). Methyl jasmonate induces traumatic resin ducts, terpenoid resin biosynthesis, and terpenoid accumulation in developing xylem of Norway spruce stems. *Plant Physiology*, *129*(3), 1003–1018. https://doi.org/10.1104/pp.011001

Meyer, J., Walker-Jonah, A., & Hollenberg, C. P. (1991). Galactokinase encoded by GAL1 is a bifunctional protein required for induction of the GAL genes in *Kluyveromyces lactis* and is able to suppress the gal3 phenotype in *Saccharomyces cerevisiae*. *Molecular and Cellular Biology*, *11*(11), 5454–5461. https://doi.org/10.1128/MCB.11.11.5454

Miller, B., Oschinski, C., & Zimmer, W. (2001). First isolation of an isoprene synthase gene from poplar and successful expression of the gene in *Escherichia coli*. *Planta*, *213*(3), 483–487. https://doi.org/10.1007/s004250100557

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

Morris, J. L., Puttick, M. N., Clark, J. W., Edwards, D., Kenrick, P., Pressel, S., Wellman, C. H., Yang, Z., Schneider, H., & Donoghue, P. C. J. (2018). The timescale of early land plant evolution. *Proceedings of the National Academy of Sciences*, *115*(10), E2274–E2283. https://doi.org/10.1073/pnas.1719588115

Morrone, D., Lowry, L., Determan, M. K., Hershey, D. M., Xu, M., & Peters, R. J. (2010). Increasing diterpene yield with a modular metabolic engineering system in E. coli: comparison of MEV and MEP isoprenoid precursor pathway engineering. *Applied Microbiology and Biotechnology*, *85*(6), 1893–1906. https://doi.org/10.1007/s00253-009-2219-x

Murphy, K. M., Ma, L. T., Ding, Y., Schmelz, E. A., & Zerbe, P. (2018). Functional characterization of two class II diterpene synthases indicates additional specialized diterpenoid pathways in maize (*Zea mays*). *Frontiers in Plant Science*, *871*. https://doi.org/10.3389/fpls.2018.01542

Murphy, K. M., & Zerbe, P. (2020). Specialized diterpenoid metabolism in monocot crops: biosynthesis and chemical diversity. *Phytochemistry*, *172*, 112289. https://doi.org/10.1016/j.phytochem.2020.112289

Nagel, R., Berasategui, A., Paetz, C., Gershenzon, J., & Schmidt, A. (2014). Overexpression of an isoprenyl diphosphate synthase in spruce leads to unexpected terpene diversion products that function in plant defense. *Plant Physiology*, *164*(2), 555–569. https://doi.org/10.1104/pp.113.228940

Nagel, R., Schmidt, A., & Peters, R. J. (2019). Isoprenyl diphosphate synthases: the chain length determining step in terpene biosynthesis. *Planta*, *249*(1), 9–20. https://doi.org/10.1007/s00425-018-3052-1

Nelson, D. R. (2006). Plant cytochrome P450s from moss to poplar. *Phytochemistry Reviews*, 5, 193–204. https://doi.org/10.1007/s11101-006-9015-3

Nelson, D., & Werck-Reichhart, D. (2011). A P450-centric view of plant evolution. *The Plant Journal*, *66*(1), 194–211. https://doi.org/10.1111/j.1365-313X.2011.04529.x

Nemoto, T., Cho, E. M., Okada, A., Okada, K., Otomo, K., Kanno, Y., Toyomasu, T., Mitsuhashi, W., Sassa, T., Minami, E., Shibuya, N., Nishiyama, M., Nojiri, H., & Yamane, H. (2004). Stemar-13-ene synthase, a diterpene cyclase involved in the biosynthesis of the phytoalexin oryzalexin S in rice. *FEBS Letters*, *571*(1–3), 182–186. https://doi.org/10.1016/j.febslet.2004.07.002

Nomura, T., Kushiro, T., Yokota, T., Kamiya, Y., Bishop, G. J., & Yamaguchi, S. (2005). The last reaction producing brassinolide is catalyzed by cytochrome P450s, CYP85A3 in tomato and CYP85A2 in *Arabidopsis*. *Journal of Biological Chemistry*, *280*(18), 17873–17879. https://doi.org/10.1074/jbc.M414592200

Nozaki, H., Hayashi, K., Nishimura, N., Kawaide, H., Matsuo, A., & Takaoka, D. (2007). Momilactone A and B as allelochemicals from moss *Hypnum plumaeforme*: first occurrence in bryophytes. *Bioscience, Biotechnology, and Biochemistry*, *71*(12), 3127–3130. https://doi.org/10.1271/bbb.70625

Ohnishi, T., Godza, B., Watanabe, B., Fujioka, S., Hategan, L., Ide, K., Shibata, K., Yokota, T., Szekeres, M., & Mizutani, M. (2012). CYP90A1/CPD, a brassinosteroid biosynthetic cytochrome P450 of *Arabidopsis*, catalyzes C-3 oxidation. *Journal of Biological Chemistry*, *287*(37), 31551–31560. https://doi.org/10.1074/jbc.M112.392720

Ohnishi, T., Szatmari, A. M., Watanabe, B., Fujita, S., Bancos, S., Koncz, C., Lafos, M., Shibata, K., Yokota, T., Sakata, K., Szekeres, M., & Mizutani, M. (2006a). C-23 hydroxylation by *Arabidopsis* CYP90C1 and CYP90D1 reveals a novel shortcut in brassinosteroid biosynthesis. *Plant Cell*, *18*(11), 3275–3288. https://doi.org/10.1105/tpc.106.045443

Ohnishi, T., Watanabe, B., Sakata, K., & Mizutani, M. (2006b). CYP724B2 and CYP90B3 function in the early C-22 hydroxylation steps of brassinosteroid biosynthetic pathway in tomato. *Bioscience, Biotechnology and Biochemistry*, *70*(9), 2071–2080. https://doi.org/10.1271/bbb.60034

Ohno, S. (1970). *Evolution by Gene Duplication*. Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-86659-3

Okada, K., Kawaide, H., Miyamoto, K., Miyazaki, S., Kainuma, R., Kimura, H., Fujiwara, K., Natsume, M., Nojiri, H., Nakajima, M., Yamane, H., Hatano, Y., Nozaki, H., & Hayashi, K. I. (2016). HpDTC1, a stress-inducible bifunctional diterpene cyclase involved in momilactone biosynthesis, functions in chemical defence in the moss *Hypnum plumaeforme*. *Scientific Reports*, *6*(1), 25316. https://doi.org/10.1038/srep25316

Olson-Manning, C. F. (2020). Elaboration of the corticosteroid synthesis pathway in primates through a multistep enzyme. *Molecular Biology and Evolution*, *37*(8), 2257–2267. https://doi.org/10.1093/molbev/msaa080

Ortlund, E. A., Bridgham, J. T., Redinbo, M. R., & Thornton, J. W. (2007). Crystal structure of an ancient protein: evolution by conformational epistasis. *Science*, *317*(5844), 1544–1548. https://doi.org/10.1126/science.1142819

Otomo, K., Kanno, Y., Motegi, A., Kenmoku, H., Yamane, H., Mitsuhashi, W., Oikawa, H., Toshima, H., Itoh, H., Matsuoka, M., Sassa, T., & Toyomasu, T. (2004). Diterpene cyclases responsible for the biosynthesis of phytoalexins, momilactones A, B, and oryzalexins A-F in rice. *Bioscience, Biotechnology and Biochemistry*, *68*(9), 2001–2006. https://doi.org/10.1271/bbb.68.2001

Pateraki, I., Andersen-Ranberg, J., Hamberger, B., Heskes, A. M., Martens, H. J., Zerbe, P., Bach, S. S., Møller, B. L., Bohlmann, J., & Hamberger, B. (2014). Manoyl oxide (13R), the biosynthetic precursor of forskolin, is synthesized in specialized root cork cells in *Coleus forskohlii*. *Plant Physiology*, *164*(3), 1222–1236. https://doi.org/10.1104/pp.113.228429

Pei, J., & Grishin, N. V. (2007). PROMALS: Towards accurate multiple sequence alignments of distantly related proteins. *Bioinformatics*, *23*(7), 802–808. https://doi.org/10.1093/bioinformatics/btm017

Pelot, K. A., Hagelthorn, D. M., Addison, J. B., & Zerbe, P. (2017). Biosynthesis of the oxygenated diterpene nezukol in the medicinal plant *Isodon rubescens* is catalyzed by a pair of diterpene synthases. *PLoS ONE*, *12*(4). https://doi.org/10.1371/journal.pone.0176507

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

Peters, R. J., & Croteau, R. B. (2002). Abietadiene synthase catalysis: Mutational analysis of a prenyl diphosphate ionization-initiated cyclization and rearrangement. *Proceedings of the National Academy of Sciences*, *99*(2), 580–584. https://doi.org/10.1073/pnas.022627099

Peters, R. J., Flory, J. E., Jetter, R., Ravn, M. M., Lee, H. J., Coates, R. M., & Croteau, R. B. (2000). Abietadiene synthase from grand fir (*Abies grandis*): dharacterization and mechanism of action of the "pseudomature" recombinant enzyme. *Biochemistry*, *39*(50), 15592–15602. https://doi.org/10.1021/bi001997l

Peters, R. J., Ravn, M. M., Coates, R. M., & Croteau, R. B. (2001). Bifunctional abietadiene synthase: free diffusive transfer of the (+)-copalyl diphosphate intermediate between two distinct active sites. *Journal of the American Chemical Society*, *123*(37), 8974–8978. https://doi.org/10.1021/ja010670k

Piatigorsky, J. (1991). The recruitment of crystallins: new functions precede gene duplication. *Science*, *252*(5009), 1078–1079. https://doi.org/10.1126/science.252.5009.1078

Pichersky, E., & Raguso, R. A. (2018). Why do plants produce so many terpenoid compounds? *New Phytologist*, *220*(3), 692–702. https://doi.org/10.1111/nph.14178

Powell, J. S., & Raffa, K. F. (1999). Effects of selected *Larix laricina* terpenoids on *Lymantria dispar* (Lepidoptera: Lymantriidae) development and behavior. *Environmental Entomology*, *28*(2), 148–154. https://doi.org/10.1093/ee/28.2.148

Prisic, S., Xu, M., Wilderman, P. R., & Peters, R. J. (2004). Rice contains two disparate *ent*-copalyl diphosphate synthases with distinct metabolic functions. *Plant Physiology*, *136*(4), 4228–4236. https://doi.org/10.1104/pp.104.050567

Projecto-Garcia, J., Natarajan, C., Moriyama, H., Weber, R. E., Fago, A., Cheviron, Z. A., Dudley, R., McGuire, J. A., Witt, C. C., & Storz, J. F. (2013). Repeated elevational transitions in hemoglobin function during the evolution of Andean hummingbirds. *Proceedings of the National Academy of Sciences of the United States of America*. https://doi.org/10.1073/pnas.1315456110

Raffa, K. F., Berryman, A. A., Simasko, J., Teal, W., & Wong, B. L. (1985). Effects of grand fir monoterpenes on the fir engraver, *Scolytus ventralis* (Coleoptera: Scolytidae), and its symbiotic fungus. *Environmental Entomology*, *14*(6), 552–556. https://doi.org/10.1093/ee/14.5.552

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

Ro, D. K., Arimura, G. I., Lau, S. Y. W., Piers, E., & Bohlmann, J. (2005). Loblolly pine abietadienol/abietadienal oxidase PtAO (CYP720B1) is a multifunctional, multisubstrate cytochrome P450 monooxygenase. *Proceedings of the National Academy of Sciences*, *102*(22), 8060–8065. https://doi.org/10.1073/pnas.0500825102

Sallaud, C., Giacalone, C., Töpfer, R., Goepfert, S., Bakaher, N., Rösti, S., & Tissier, A. (2012). Characterization of two genes for the biosynthesis of the labdane diterpene *Z*-abienol in tobacco (*Nicotiana tabacum*) glandular trichomes. *Plant Journal*, *72*(1), 1–17. https://doi.org/10.1111/j.1365-313X.2012.05068.x

Sallaud, C., Rontein, D., Onillon, S., Jabès, F., Duffé, P., Giacalone, C., Thoraval, S., Escoffier, C., Herbette, G., Leonhardt, N., Causse, M., & Tissiera, A. (2009). A novel pathway for sesquiterpene biosynthesis from *Z,Z*-farnesyl pyrophosphate in the wild tomato *Solanum habrochaites*. *Plant Cell*, *21*(1), 301–317. https://doi.org/10.1105/tpc.107.057885

Schepmann, H. G., Pang, J., & Matsuda, S. P. T. (2001). Cloning and characterization of *Ginkgo biloba* levopimaradiene synthase, which catalyzes the first committed step in ginkgolide biosynthesis. *Archives of Biochemistry and Biophysics*, *392*(2), 263–269. https://doi.org/10.1006/abbi.2001.2438

Schilmiller, A. L., Schauvinhold, I., Larson, M., Xu, R., Charbonneau, A. L., Schmidt, A., Wilkerson, C., Last, R. L., & Pichersky, E. (2009). Monoterpenes in the glandular trichomes of tomato are synthesized from a neryl diphosphate precursor rather than geranyl diphosphate. *Proceedings of the National Academy of Sciences*, *106*(26), 10865–10870. https://doi.org/10.1073/pnas.0904113106

Schmelz, E. A., Huffaker, A., Sims, J. W., Christensen, S. A., Lu, X., Okada, K., & Peters, R. J. (2014). Biosynthesis, elicitation and roles of monocot terpenoid phytoalexins. *The Plant Journal*, *79*(4), 659–678. https://doi.org/10.1111/tpj.12436

Schmidt, A., & Gershenzon, J. (2007a). Cloning and characterization of isoprenyl diphosphate synthases with farnesyl diphosphate and geranylgeranyl diphosphate synthase activity from Norway spruce (*Picea abies*) and their relation to induced oleoresin formation. *Phytochemistry*, *68*(21), 2649–2659. https://doi.org/10.1016/j.phytochem.2007.05.037

Schmidt, A., & Gershenzon, J. (2007b). Cloning and characterization of two different types of geranyl diphosphate synthases from Norway spruce *(Picea abies*). *Phytochemistry*, *69*(1), 49–57. https://doi.org/10.1016/j.phytochem.2007.06.022

Schmidt, A., Nagel, R., Krekling, T., Christiansen, E., Gershenzon, J., & Krokene, P. (2011). Induction of isoprenyl diphosphate synthases, plant hormones and defense signalling genes correlates with traumatic resin duct formation in Norway spruce (*Picea abies*). *Plant Molecular Biology*, *77*(6), 577–590. https://doi.org/10.1007/s11103-011-9832-7

Schmidt, A., Wächtler, B., Temp, U., Krekling, T., Séguin, A., & Gershenzon, J. (2010). A bifunctional geranyl and geranylgeranyl diphosphate synthase is involved in terpene oleoresin formation in *Picea abies*. *Plant Physiology*, *152*(2), 639–655. https://doi.org/10.1104/pp.109.144691

Schrepfer, P., Buettner, A., Goerner, C., Hertel, M., van Rijn, J., Wallrapp, F., Eisenreich, W., Sieber, V., Kourist, R., & Brück, T. (2016). Identification of amino acid networks governing catalysis in the closed complex of class I terpene synthases. *Proceedings of the National Academy of Sciences*, *113*(8), E958–E967. https://doi.org/10.1073/pnas.1519680113

Schwede, T. (2003). SWISS-MODEL: an automated protein homology-modeling server. *Nucleic Acids Research*, *31*(13), 3381–3385. https://doi.org/10.1093/nar/gkg520

Semiz, A., & Sen, A. (2015). Cloning and expression of a CYP720B orthologue involved in the biosynthesis of diterpene resin acids in *Pinus brutia*. *Molecular Biology Reports*, *42*(3), 737–744. https://doi.org/10.1007/s11033-014-3822-1

Shah, P., McCandlish, D. M., & Plotkin, J. B. (2015). Contingency and entrenchment in protein evolution under purifying selection. *Proceedings of the National Academy of Sciences of the United States of America*. https://doi.org/10.1073/pnas.1412933112

Shi, M., Luo, X., Ju, G., Li, L., Huang, S., Zhang, T., Wang, H., & Kai, G. (2016). Enhanced diterpene tanshinone accumulation and bioactivity of transgenic *Salvia miltiorrhiza* hairy roots by pathway engineering. *Journal of Agricultural and Food Chemistry*, *64*(12), 2523–2530. https://doi.org/10.1021/acs.jafc.5b04697

Shimada, T., Endo, T., Fujii, H., Hara, M., Ueda, T., Kita, M., & Omura, M. (2004). Molecular cloning and functional characterization of four monoterpene synthase genes from *Citrus unshiu* Marc. *Plant Science*, *166*(1), 49–58. https://doi.org/10.1016/j.plantsci.2003.07.006

Shimane, M., Ueno, Y., Morisaki, K., Oogami, S., Natsume, M., Hayashi, K., Nozaki, H., & Kawaide, H. (2014). Molecular evolution of the substrate specificity of *ent*-kaurene synthases to adapt to gibberellin biosynthesis in land plants. *Biochemical Journal*, *462*(3), 539–546. https://doi.org/10.1042/BJ20140134

Shimura, K., Okada, A., Okada, K., Jikumaru, Y., Ko, K. W., Toyomasu, T., Sassa, T., Hasegawa, M., Kodama, O., Shibuya, N., Koga, J., Nojiri, H., & Yamane, H. (2007). Identification of a biosynthetic gene cluster in rice for momilactones. *Journal of Biological Chemistry*, *282*(47), 34013–34018. https://doi.org/10.1074/jbc.M703344200

Stamatakis, A. (2014). RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*, *30*(9), 1312–1313. https://doi.org/10.1093/bioinformatics/btu033

Starr, T. N., Flynn, J. M., Mishra, P., Bolon, D. N. A., & Thornton, J. W. (2018). Pervasive contingency and entrenchment in a billion years of Hsp90 evolution. *Proceedings of the National Academy of Sciences*, *115*(17), 4453–4458. https://doi.org/10.1073/pnas.1718133115

Storz, J. F. (2016a). Causes of molecular convergence and parallelism in protein evolution. *Nature Reviews Genetics*. https://doi.org/10.1038/nrg.2016.11

Storz, J. F. (2016b). Gene duplication and evolutionary innovations in hemoglobin-oxygen transport. *Physiology, 31*(3), 223–232. https://doi.org/10.1152/physiol.00060.2015

Sugai, Y., Ueno, Y., Hayashi, K. I., Oogami, S., Toyomasu, T., Matsumoto, S., Natsume, M., Nozaki, H., & Kawaide, H. (2011). Enzymatic ^{13}C labeling and multidimensional NMR analysis of miltiradiene synthesized by bifunctional diterpene cyclase in *Selaginella moellendorffii*. *Journal of Biological Chemistry*, *286*(50), 42840–42847. https://doi.org/10.1074/jbc.M111.302703

Sun T. P., & Kamiya, Y. (1994). The *Arabidopsis* GA1 locus encodes the cyclase *ent*-kaurene synthetase A of gibberellin biosynthesis. *Plant Cell*, *6*(10), 1509–1518. https://doi.org/10.1105/tpc.6.10.1509

Sundell, D., Mannapperuma, C., Netotea, S., Delhomme, N., Lin, Y., Sjödin, A., Van de Peer, Y., Jansson, S., Hvidsten, T. R., & Street, N. R. (2015). The plant genome integrative explorer resource: PlantGenIE.org. *New Phytologist*, *208*(4), 1149–1156. https://doi.org/10.1111/nph.13557

Swaminathan, S., Morrone, D., Wang, Q., Bruce Fulton, D., & Peters, R. J. (2009). CYP76M7 is an *ent*-cassadiene C11 α-hydroxylase defining a second multifunctional diterpenoid biosynthetic gene cluster in rice. *Plant Cell*, *21*(10), 3315–3325. https://doi.org/10.1105/tpc.108.063677

Tata, S. K., Jung, J., Kim, Y.H., Choi, J. Y., Jung, J.Y., Lee, I.J., Shin, J. S., & Ryu, S. B. (2016). Heterologous expression of chloroplast-localized geranylgeranyl pyrophosphate synthase confers fast plant growth, early flowering and increased seed yield. *Plant Biotechnology Journal*, *14*(1), 29–39. https://doi.org/10.1111/pbi.12333

Tawfik, O. K. and D. S. (2010). Enzyme promiscuity: a mechanistic and evolutionary perspective. *Annual Review of Biochemistry*, *79*(1), 471–505. https://doi.org/10.1146/annurev-biochem-030409-143718

Thornton, J. W. (2001). Evolution of vertebrate steroid receptors from an ancestral estrogen receptor by ligand exploitation and serial genome expansions. *Proceedings of the National Academy of Sciences*, *98*(10), 5671–5676. https://doi.org/10.1073/pnas.091553298

Thornton, J. W. (2004). Resurrecting ancient genes: experimental analysis of extinct molecules. *Nature Reviews Genetics*, *5*(5), 366–375. https://doi.org/10.1038/nrg1324

Thornton, J. W., & Carroll, S. M. (2011). Lamprey endocrinology is not ancestral. *Proceedings of the National Academy of Sciences*, *108*(2), E5–E5. https://doi.org/10.1073/pnas.1014896108

Tomlin, E. S., Alfaro, R. I., Borden, J. H., & He, F. (1998). Histological response of resistant and susceptible white spruce to simulated white pine weevil damage. *Tree Physiology*, *18*(1), 21–28. https://doi.org/10.1093/treephys/18.1.21

Tomlin, E. S., Borden, J. H., & Pierce, H. D. (1996). Relationship between cortical resin acids and resistance of Sitka spruce to the white pine weevil. *Canadian Journal of Botany*, *74*(4), 599–606. https://doi.org/10.1139/b96-076

Trott, O., & Olson, A. J. (2010). AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *Journal of Computational Chemistry*, *31*(2), 455–461. https://doi.org/10.1002/jcc.21334

Unland, K., Pütter, K. M., Vorwerk, K., van Deenen, N., Twyman, R. M., Prüfer, D., & Schulze Gronover, C. (2018). Functional characterization of squalene synthase and squalene epoxidase in Taraxacum koksaghyz. *Plant Direct*, *2*(6), e00063. https://doi.org/10.1002/pld3.63

Urban, P., Mignotte, C., Kazmaier, M., Delorme, F., & Pompon, D. (1997). Cloning, yeast expression, and characterization of the coupling of two distantly related *Arabidopsis thaliana* NADPH-cytochrome P450 reductases with P450 CYP73A5. *Journal of Biological Chemistry*, *272*(31), 19176–19186. https://doi.org/10.1074/jbc.272.31.19176

Van Bel, M., Diels, T., Vancaester, E., Kreft, L., Botzki, A., Van de Peer, Y., Coppens, F., & Vandepoele, K. (2018). PLAZA 4.0: an integrative resource for functional, evolutionary and comparative plant genomics. *Nucleic Acids Research*, *46*(D1), D1190–D1196. https://doi.org/10.1093/nar/gkx1002

Venkat, A., Hahn, M. W., & Thornton, J. W. (2018). Multinucleotide mutations cause false inferences of lineage-specific positive selection. *Nature Ecology & Evolution*, *2*(8). https://doi.org/10.1038/s41559-018-0584-5

Vogel, B. S., Wildung, M. R., Vogel, G., & Croteau, R. (1996). Abietadiene synthase from grand fir (*Abies grandis*). *Journal of Biological Chemistry*, *271*(38), 23262–23268.

Voordeckers, K., Brown, C. A., Vanneste, K., van der Zande, E., Voet, A., Maere, S., & Verstrepen, K. J. (2012). Reconstruction of ancestral metabolic enzymes reveals molecular mechanisms underlying evolutionary innovation through gene duplication. *PLoS Biology*, *10*(12), e1001446. https://doi.org/10.1371/journal.pbio.1001446

Vranová, E., Coman, D., & Gruissem, W. (2012). Structure and dynamics of the isoprenoid pathway network. *Molecular Plant*, *5*(2), 318–333. https://doi.org/10.1093/mp/sss015

Wang, Q., Hillwig, M. L., & Peters, R. J. (2011). CYP99A3: functional identification of a diterpene oxidase from the momilactone biosynthetic gene cluster in rice. *The Plant Journal*, *65*(1), 87–95. https://doi.org/10.1111/j.1365-313X.2010.04408.x

Wang, Q., Hillwig, M. L., Wu, Y., & Peters, R. J. (2012). CYP701A8: A rice *ent*-kaurene oxidase paralog diverted to more specialized diterpenoid metabolism. *Plant Physiology*, *158*(3), 1418–1425. https://doi.org/10.1104/pp.111.187518

Warren, R. L., Keeling, C. I., Yuen, M. M. Saint, Raymond, A., Taylor, G. A., Vandervalk, B. P., Mohamadi, H., Paulino, D., Chiu, R., Jackman, S. D., Robertson, G., Yang, C., Boyle, B., Hoffmann, M., Weigel, D., Nelson, D. R., Ritland, C., Isabel, N., Jaquish, B., et al. (2015). Improved white spruce (*Picea glauca*) genome assemblies and annotation of large gene families of conifer terpenoid and phenolic defense metabolism. *Plant Journal*, *83*(2), 189–212. https://doi.org/10.1111/tpj.12886

Werck-Reichhart, D., & Feyereisen, R. (2000). Cytochromes P450: a success story. *Genome Biology*, *1*(6). https://doi.org/10.1186/gb-2000-1-6-reviews3003

Wilderman, P. R., & Peters, R. J. (2007). A single residue switch converts abietadiene synthase into a pimaradiene specific cyclase. *Journal of the American Chemical Society*, *129*(51), 15736–15737. https://doi.org/10.1021/ja074977g

Wilderman, P. R., Xu, M., Jin, Y., Coates, R. M., & Peters, R. J. (2004). Identification of *syn*-pimara-7,15-diene synthase reveals functional clustering of terpene synthases involved in rice phytoalexin/allelochemical biosynthesis. *Plant Physiology*, *135*(4), 2098–2105. https://doi.org/10.1104/pp.104.045971

Wu, Y., Zhou, K., Toyomasu, T., Sugawara, C., Oku, M., Abe, S., Usui, M., Mitsuhashi, W., Chono, M., Chandler, P. M., & Peters, R. J. (2012). Functional characterization of wheat copalyl diphosphate synthases sheds light on the early evolution of labdane-related diterpenoid metabolism in the cereals. *Phytochemistry*, *84*, 40–46. https://doi.org/10.1016/j.phytochem.2012.08.022

Xu, D., & Zhang, Y. (2011). Improving the physical realism and structural accuracy of protein models by a two-step atomic-level energy minimization. *Biophysical Journal*, *101*(10), 2525–2534. https://doi.org/10.1016/j.bpj.2011.10.024

Xu, M., Hillwig, M. L., Prisic, S., Coates, R. M., & Peters, R. J. (2004). Functional identification of rice *syn*-copalyl diphosphate synthase and its role in initiating biosynthesis of diterpenoid phytoalexin/allelopathic natural products. *Plant Journal*, *39*(3), 309–318. https://doi.org/10.1111/j.1365-313X.2004.02137.x

Xu, M., Ross Wilderman, P., Morrone, D., Xu, J., Roy, A., Margis-Pinheiro, M., Upadhyaya, N. M., Coates, R. M., & Peters, R. J. (2007a). Functional characterization of the rice kaurene synthase-like gene family. *Phytochemistry*, *68*(3), 312–326. https://doi.org/10.1016/j.phytochem.2006.10.016

Xu, M., Wilderman, P. R., & Peters, R. J. (2007b). Following evolution's lead to a single residue switch for diterpene synthase product outcome. *Proceedings of the National Academy of Sciences*, *104*(18), 7397–7401. https://doi.org/10.1073/pnas.0611454104

Yamaguchi, S., Sun, T. P., Kawaide, H., & Kamiya, Y. (1998). The GA2 locus of Arabidopsis thaliana Encodes *ent*-kaurene synthase of gibberellin biosynthesis. *Plant Physiology*, *116*(4), 1271–1278. https://doi.org/10.1104/pp.116.4.1271

Yang, Z. (2007). PAML 4: Phylogenetic analysis by maximum likelihood. *Molecular Biology and Evolution*, *24*(8), 1586–1591. https://doi.org/10.1093/molbev/msm088

Yčas, M. (1974). On earlier states of the biochemical system. *Journal of Theoretical Biology*, *44*(1), 145–160. https://doi.org/10.1016/S0022-5193(74)80035-4

Yu, F., & Utsumi, R. (2009). Diversity, regulation, and genetic manipulation of plant mono- and sesquiterpenoid biosynthesis. *Cellular and Molecular Life Sciences*, *66*(18), 3043–3052. https://doi.org/10.1007/s00018-009-0066-7

Zerbe, P., Chiang, A., & Bohlmann, J. (2012a). Mutational analysis of white spruce (*Picea glauca*) *ent*-kaurene synthase (PgKS) reveals common and distinct mechanisms of conifer diterpene synthases of general and specialized metabolism. *Phytochemistry*, *74*, 30–39. https://doi.org/10.1016/j.phytochem.2011.11.004

Zerbe, P., Chiang, A., Dullat, H., O'Neil-Johnson, M., Starks, C., Hamberger, B., & Bohlmann, J. (2014). Diterpene synthases of the biosynthetic system of medicinally active diterpenoids in *Marrubium vulgare*. *Plant Journal*, *79*(6), 914–927. https://doi.org/10.1111/tpj.12589

Zerbe, P., Chiang, A., Yuen, M., Hamberger, B., Hamberger, B., Draper, J. A., Britton, R., & Bohlmann, J. (2012b). Bifunctional *cis*-abienol synthase from *Abies balsamea* discovered by transcriptome sequencing and its implications for diterpenoid fragrance production. *Journal of Biological Chemistry*, *287*(15), 12121–12131. https://doi.org/10.1074/jbc.M111.317669

Zhang, B., Liu, Y., Chen, M., Feng, J., Ma, Z., Zhang, X., & Zhu, C. (2018). Cloning, expression analysis and functional characterization of squalene synthase (SQS) from *Tripterygium wilfordii*. *Molecules*, *23*(2), 269. https://doi.org/10.3390/molecules23020269

Zhang, C., Yang, D., Liang, Z., Liu, J., Yan, K., Zhu, Y., & Yang, S. (2019). Climatic factors control the geospatial distribution of active ingredients in *Salvia miltiorrhiza* Bunge in China. *Scientific Reports*, *9*(1), 904. https://doi.org/10.1038/s41598-018-36729-x

Zhang, J. (2003). Evolution by gene duplication: an update. *Trends in Ecology & Evolution*, *18*(6), 292–298. https://doi.org/10.1016/S0169-5347(03)00033-8

Zhang, J. (2006). Parallel adaptive origins of digestive RNases in Asian and African leaf monkeys. *Nature Genetics*. https://doi.org/10.1038/ng1812

Zhang, X.D., Yu, Y.G., Yang, D.F., Qi, Z.C., Liu, R.Z., Deng, F.T., Cai, Z.X., Li, Y., Sun, Y.F., & Liang, Z.S. (2018). Chemotaxonomic variation in secondary metabolites contents and their correlation between environmental factors in *Salvia miltiorrhiza* Bunge from natural habitat of China. *Industrial Crops and Products*, *113*, 335–347. https://doi.org/10.1016/j.indcrop.2018.01.043

Zhen, Y., Aardema, M. L., Medina, E. M., Schumer, M., & Andolfatto, P. (2012). Parallel molecular evolution in an herbivore community. *Science*. https://doi.org/10.1126/science.1226630

Zhou, F., & Pichersky, E. (2020). More is better: the diversity of terpene metabolism in plants. *Current Opinion in Plant Biology 55*, 1–10. https://doi.org/10.1016/j.pbi.2020.01.005

Zhou, K., Xu, M., Tiernan, M., Xie, Q., Toyomasu, T., Sugawara, C., Oku, M., Usui, M., Mitsuhashi, W., Chono, M., Chandler, P. M., & Peters, R. J. (2012). Functional characterization of wheat *ent*-kaurene(-like) synthases indicates continuing evolution of labdane-related diterpenoid metabolism in the cereals. *Phytochemistry*, *84*, 47–55. https://doi.org/10.1016/j.phytochem.2012.08.021

Zi, J., Mafu, S., & Peters, R. J. (2014). To gibberellins and beyond! Surveying the evolution of (di)terpenoid metabolism. *Annual Review of Plant Biology*, *65*(1), 259–286. https://doi.org/10.1146/annurev-arplant-050213-035705

Zi, J., & Peters, R. J. (2013). Characterization of CYP76AH4 clarifies phenolic diterpenoid biosynthesis in the Lamiaceae. *Organic & Biomolecular Chemistry*, *11*(44), 7650. https://doi.org/10.1039/c3ob41885e

Zulak, K. G., & Bohlmann, J. (2010). Terpenoid biosynthesis and specialized vascular cells of conifer defense. In *Journal of Integrative Plant Biology* (Vol. 52, Issue 1, pp. 86–97). https://doi.org/10.1111/j.1744-7909.2010.00910.x

Appendix

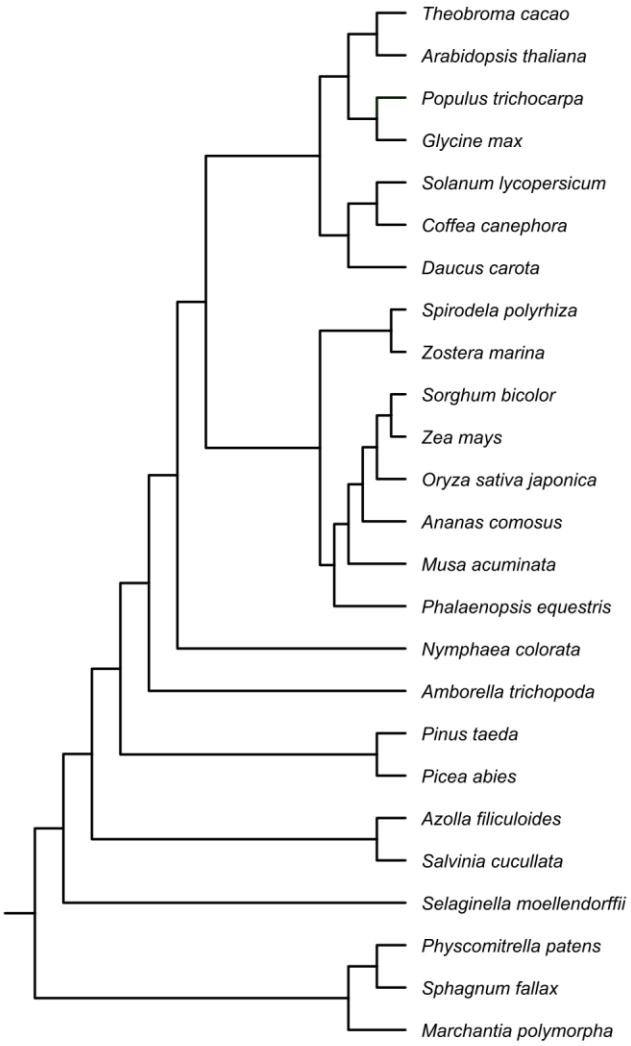
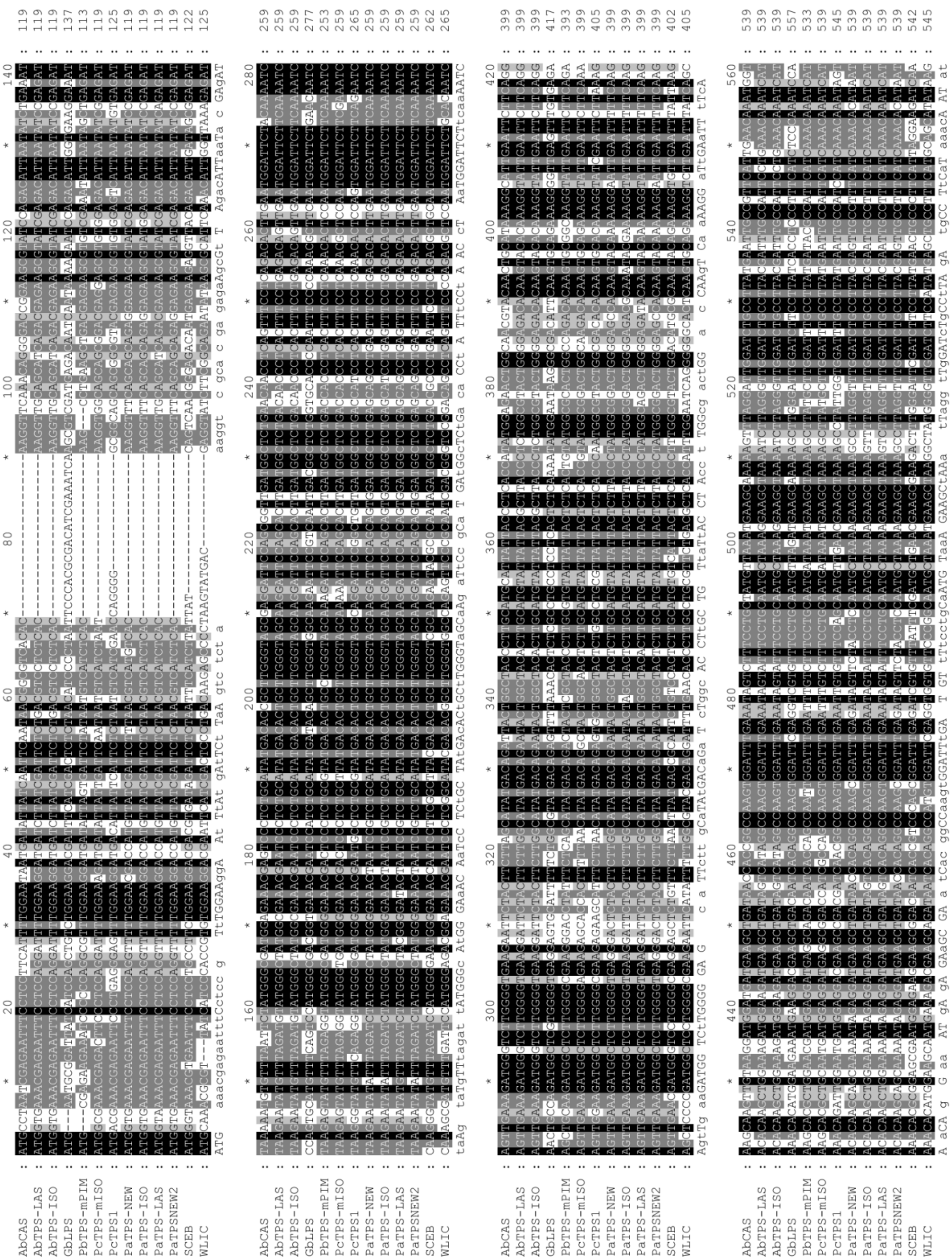
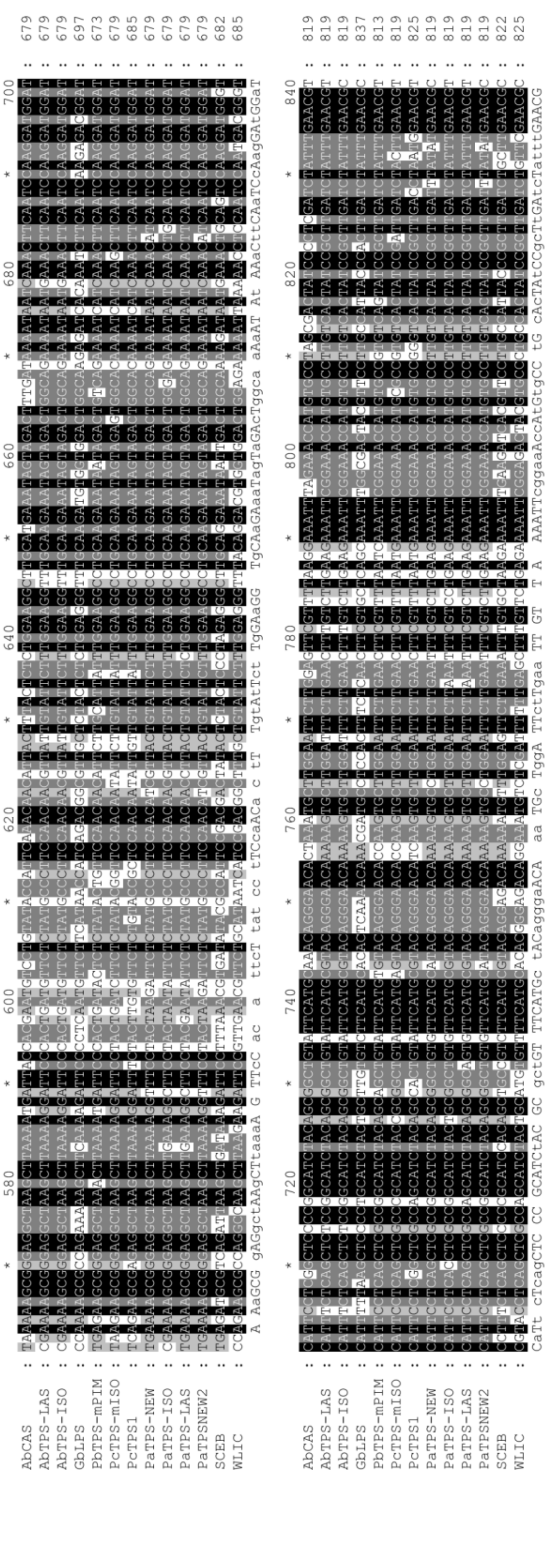
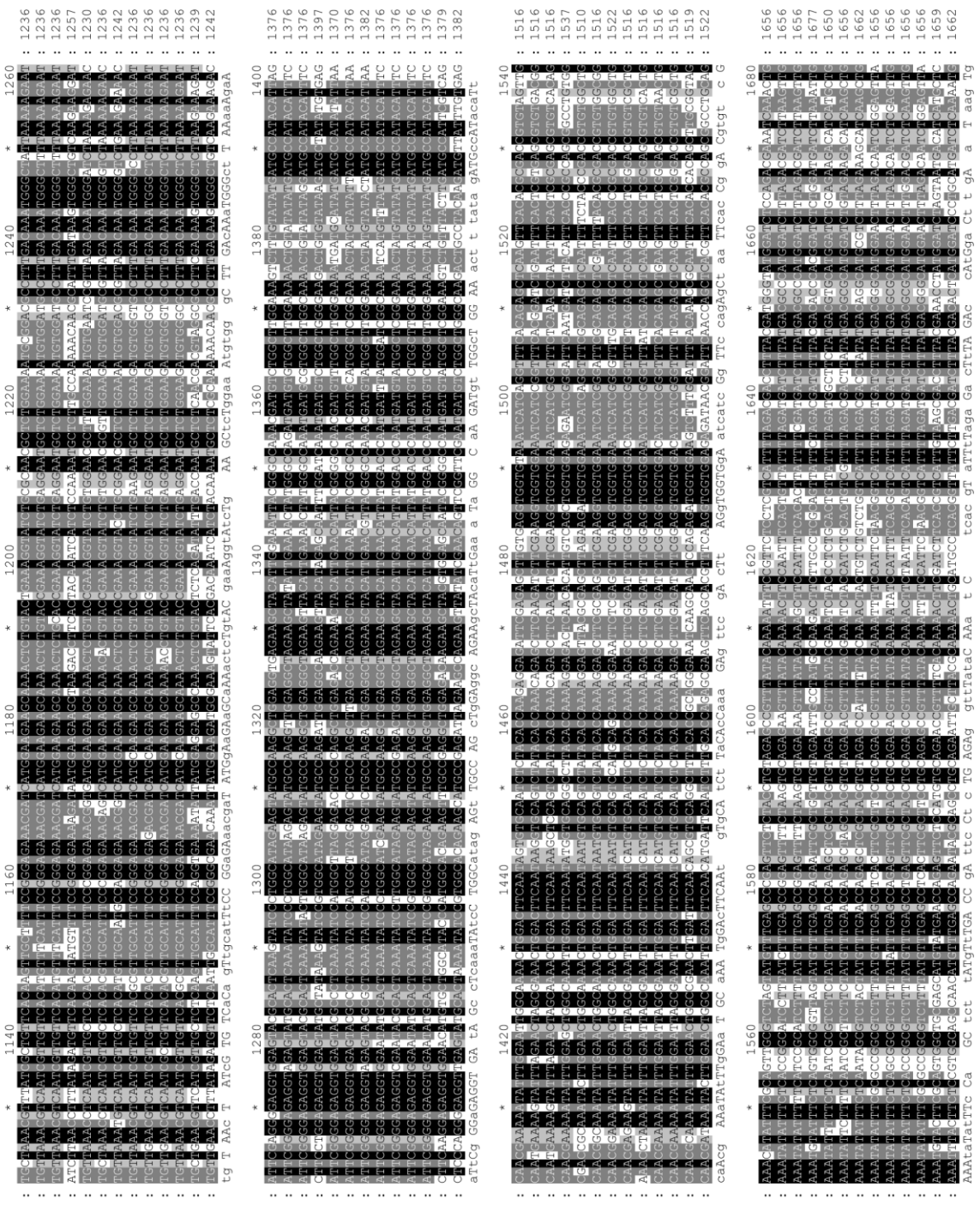
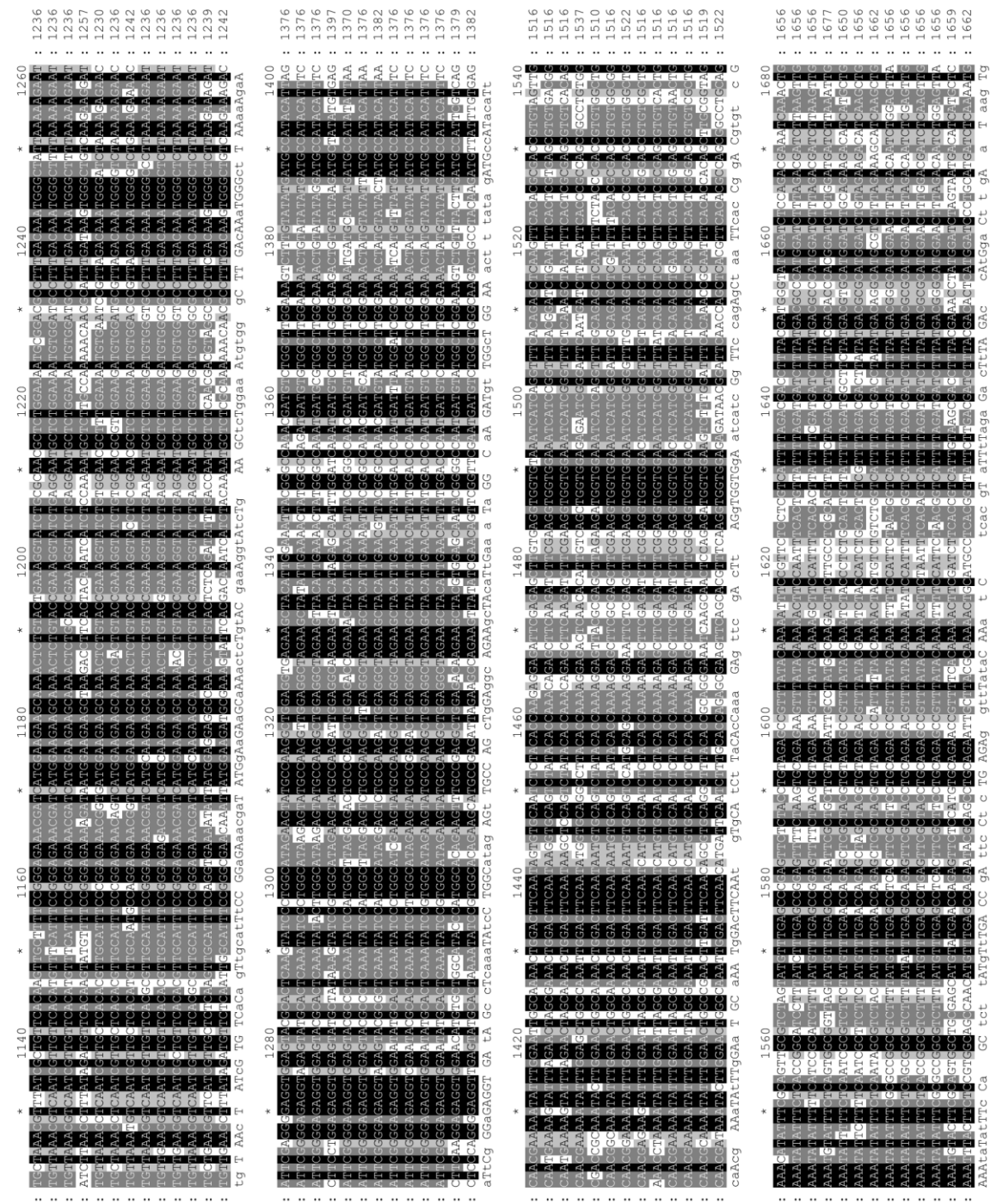


Figure A2.1: Species tree provided to OrthoFinder for proteome analysis in Chapter II and Chapter IV. Classification of "bryophytes" (*P. patens, S. fallax, and M. polymorpha*) follows one of seven alternatives by Morris et al. (2018)









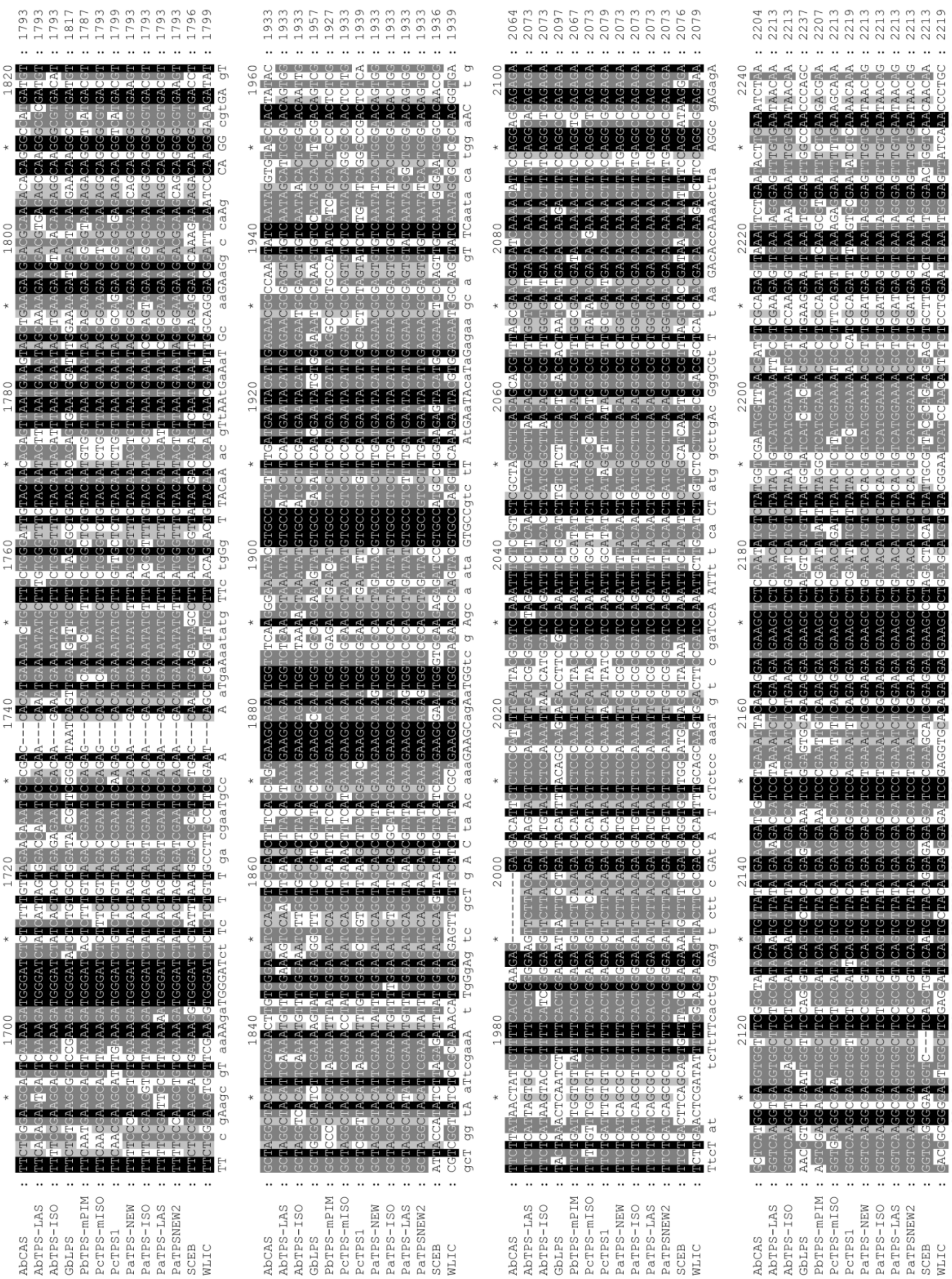


Figure A3.1: Multiple sequence alignment of nucleotide sequences provided to codeml for ancestral sequence reconstruction.

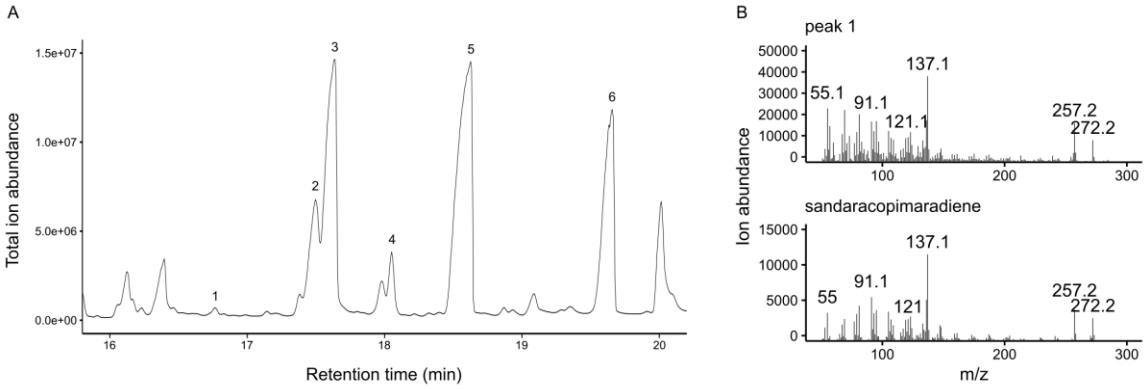


Figure A3.2: PaTPS-LAS produces trace amounts of sandaracopimaradiene in addition to five previously-published diterepene olefins. A) GC-MS chromatogram of N-terminally truncated PaTPS-LAS when coexpressed in pACYC-DUET with a GGPP synthase and highly concentration (see Materials & Methods) prior to injection. 1, sandaracopimaradiene; 2, palustradiene; 3, levopimaradiene; 4, dehydroabietadiene; 5, abietadiene; 6, neoabietadiene. B) Mass spectra of peak 1 and authentic sandaracopimaradiene (distinguished from pimaradiene by retention factors relative to levopimaraidne, as determined using standards. Remaining peaks were identified by retention factor and mass spectra of published PaTPS-LAS products. Note that retention times differ from previous runs due to a shorter method and slight column shortening.

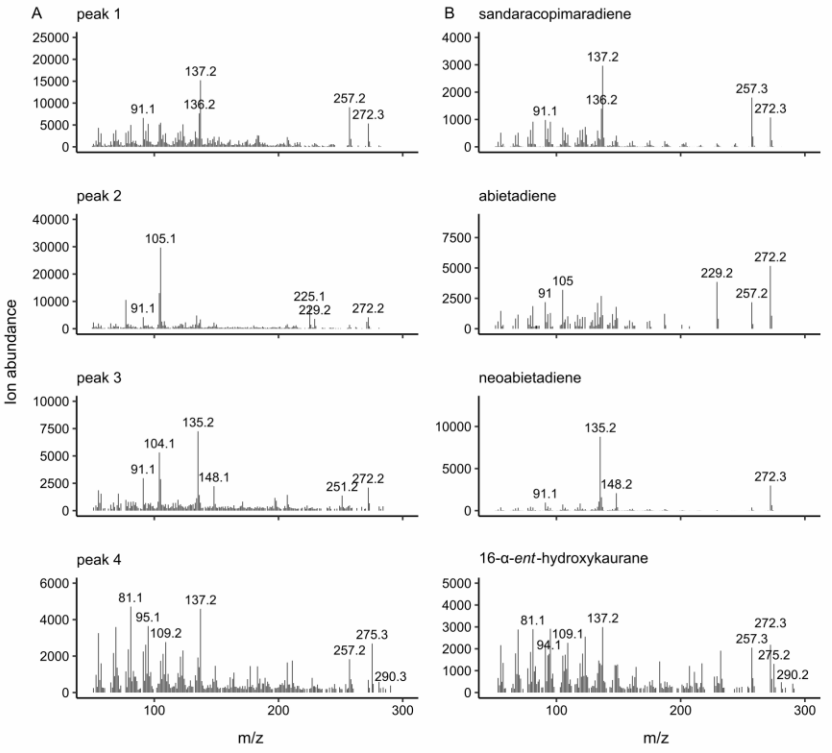


Figure A3.3: Supporting mass spectra for DTC3.1 assays with normal-CPP in Chapter III. A) Mass spectra of peaks 1-4 after coexpression of SmDTC3.1 with a normal-CPP synthase. B) Mass spectra of authentic standards sandaracopimaradiene, abietadiene, neoabietadiene, and mass spectrum of 16-α-*ent*-hydroxykaurane produced by SmDTC3, as reported by Shimane et al., (2014). Ion abundance differences may reflect alternative double bond arrangement, low signal-to-noise, or analyte co-retention.

Appendix Table A4.1. Gene names, accession numbers, and primer sequences used for cloning and expression of CYP720s

Gene name	Genome accession	Primer sequence
PaCYP720B1	MA_10427376g0010	5'-ATGGCAGACCAAATAACTCTAGTGTTGG -3' 3'-TCATTGCAGAGGATGGAGGCGAATG-5'
PaCYP720B3	MA_266173g0010	5'-ATGGCGCCAAAGACAGACCAAATATC -3' 3'-TCATTCTCTACAATGAAGGCGAATGGG -5'
PaCYP720B4	MA_46522g0010	5'-ATGGCGCCCATGGCAGAGC-3' 3'-TTATTCATTCTCTACTCTACCATGAAGGC-5'
PaCYP720B5	MA_165926g0020 & MA_178572g0010	5'-ATGGCAGATCAAATATCATTAGTGCTGG-3' 3'-TCATTGTCTACAATGAAGACGAATGGG-5'
PaCYP720B6	MA_110777g0010	5'-ATGGCAGATCAAATATCATTAGCGTTGG-3' 3'-TCATTCTCTGCAATGAAGACGGATAGG-5'
PaCYP720B11	MA_131290g0010	5'-ATGGCACTGGCCGAGCAAATAAC-3' 3'-TCATGTTCTTAAATGAAGACGAATGGG-5'
AtCYP720A1	AT1G73340	5'-ATGGCGGAATCTGCAGGAG-3' 3'-TTATTGATTCGGAGGATCACTG-5'
PtCYP720A1	Potri.017G154600	5'-ATGAGAGAAATTTATTCGGCAACACTTATG-3' 3'-TTACAACGATTCTGAGTCGAGATGATG-5'

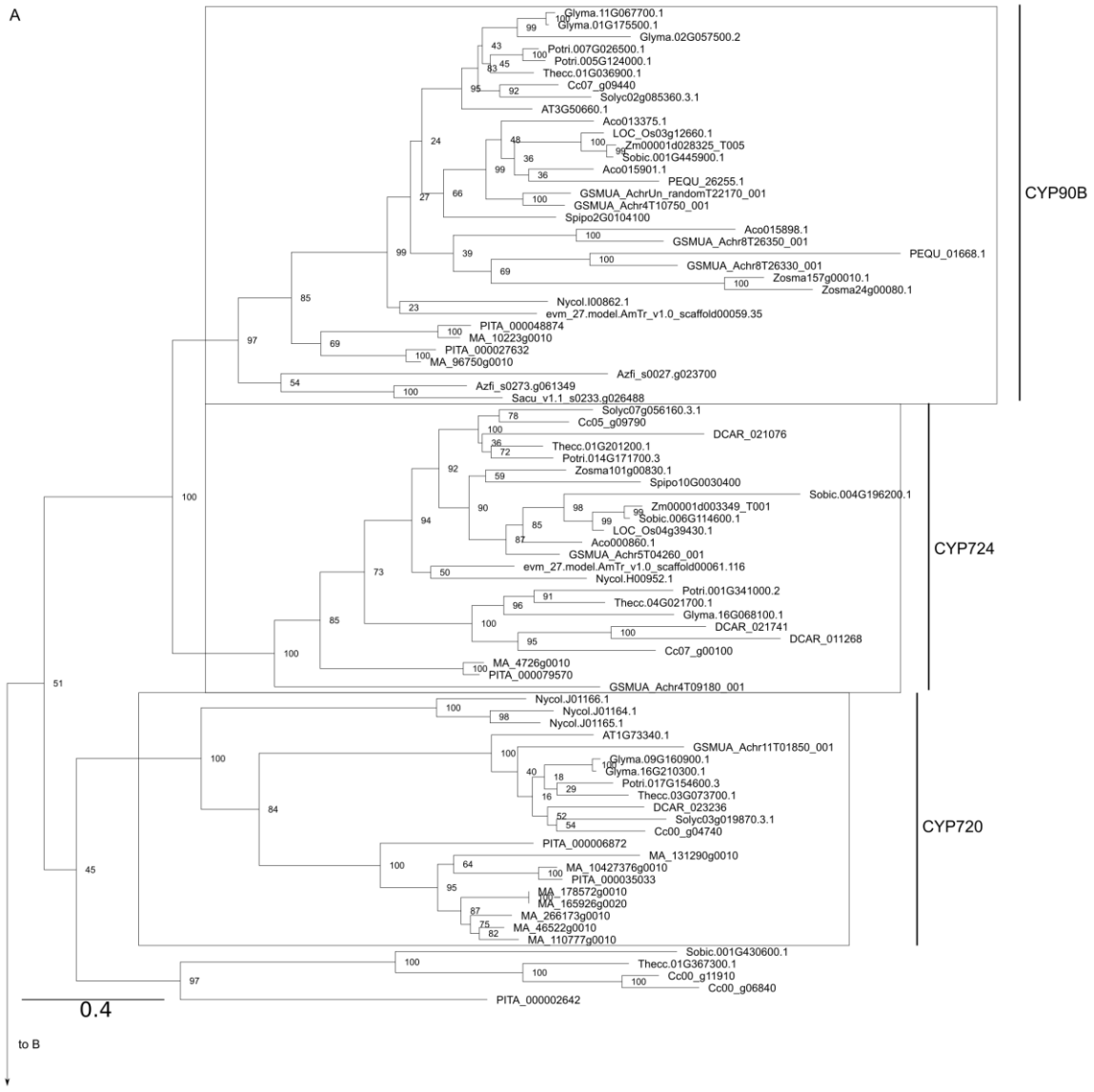

A
Glyma.11G067700.1
Glyma.01G175500.1
Glyma.02G057500.2
Potri.007G026500.1
Potri.005G124000.1
Thecc.01G036900.1
Cc07_g09440
Solyc02g085360.3.1
AT3G50660.1
Aco013375.1
LOC_Os03g12660.1
Zm00001d028325_T005
Sobic.001G445900.1
Aco015901.1
PEQU_26255.1
GSMUA_AchrUn_randomT22170_001
GSMUA_Achr4T10750_001
Spipo2G0104100
Aco015898.1
GSMUA_Achr8T26350_001
PEQU_01668.1
GSMUA_Achr8T26330_001
Zosma157g00010.1
Zosma24g00080.1
Nycol.I00862.1
evm_27.model.AmTr_v1.0_scaffold00059.35
PITA_000048874
MA_10223g0010
PITA_000027632
MA_96750g0010
Azfi_s0027.g023700
Azfi_s0273.g061349
Sacu_v1.1_s0233.g026488
CYP90B
Solyc07g056160.3.1
Cc05_g09790
DCAR_021076
Thecc.01G201200.1
Potri.014G171700.3
Zosma101g00830.1
Spipo10G0030400
Sobic.004G196200.1
Zm00001d003349_T001
Sobic.006G114600.1
LOC_Os04g39430.1
Aco000860.1
GSMUA_Achr5T04260_001
evm_27.model.AmTr_v1.0_scaffold00061.116
Nycol.H00952.1
Potri.001G341000.2
Thecc.04G021700.1
Glyma.16G068100.1
DCAR_021741
DCAR_011268
Cc07_g00100
MA_4726g0010
PITA_000079570
GSMUA_Achr4T09180_001
CYP724
Nycol.J01166.1
Nycol.J01164.1
Nycol.J01165.1
AT1G73340.1
GSMUA_Achr11T01850_001
Glyma.09G160900.1
Glyma.16G210300.1
Potri.017G154600.3
Thecc.03G073700.1
DCAR_023236
Solyc03g019870.3.1
Cc00_g04740
PITA_000006872
MA_131290g0010
MA_10427376g0010
PITA_000035033
MA_178572g0010
MA_165926g0020
MA_266173g0010
MA_46522g0010
MA_110777g0010
CYP720
Sobic.001G430600.1
Thecc.01G367300.1
Cc00_g11910
Cc00_g06840
PITA_000002642
0.4
to B

Figure A4.1: Full genome accession names and bootsrap support values of CYP85 clan of cytochrome P450s. The tree depicted was used to infer gene relationships between CYP85 members in Chapter IV. The CYP90B, CYP724, and CYP720 subfamilies are indicated. Scale indicates number of substitutions per site.

Abstract

An area of high interest in biology is the underlying causes of diversity. From an ecological standpoint, central questions are how particular traits, such as physiological or chemical phenotypes, improve an organism's fitness within a population and the effects of varying traits on organisms between populations. From molecular and biochemical perspectives, studies often attempt to dissect regulatory, structural, and chemical differences to infer mechanistic causes of diversity between species lineages. However, the complexity of biological systems and large timescales of evolutionary divergence often hamper comparative approaches to dissecting modern-day phenotypes in dissimilar genetic contexts. One solution to this problem is to consider extant traits in an historical light and to investigate the causes of diversity as they developed during the course of evolution.

Multi-step biochemical pathways in plants is a challenging area of trait dissection as plants rapidly evolve novel specialized metabolites, often resulting in genus- or species-specific chemical compounds that are synthesized by unrelated gene families. Although the evolution of multi-step biochemical pathways catalyzed by proteins encoded by paralogous (duplicate) gene copies is understood to some extent, assembly of novel pathway steps across gene families is rarely investigated. Diterpenoids comprise one such area of divergent chemical traits in plants that are often restricted to limited taxa. As diterpenoid biosynthesis is catalyzed by families of diterpene synthases and cytochrome P450s, natural experiments in biochemical pathway assembly have occurred repeatedly during evolution.

The conifer Norway spruce (*Picea abies*) is an industrially and ecologically important tree in northern, central, and eastern Europe that produces a rich mixture of defensive diterpenoids in high abundance. Conifers, particularly the family Pinaceae, uniquely produce a mixture of C_{20} diterpene resin acids that can flush out and/or suffocate invading pests. As Norway spruce dominate large forests in Europe, and is used as an ornamental tree and for lumber, beetle invasions that decimate populations of Norway spruce are devastating. Due to the availability of molecular tools for genetic dissection and a published genome, this thesis attempts to answer questions about the evolution of novel diterpenoid pathways in Norway spruce using a combination of bioinformatics and biochemical analyses. Additionally, this work investigates regulation of terpene pathways and attempts to identify pathway steps that control pathway flux using transgenic manipulation of diterpene synthases and cytochrome P450s.

Zusammenfassung

Bei der Erforschung über die Entstehung biologischer Vielfalt stellt sich aus ökologischer Sicht die zentralen Fragen, auf welche Weise bestimmte Merkmale (wie physiologische oder chemische Phänotypen) die Fitness eines Organismus innerhalb einer Population erhöhen und welche Auswirkungen unterschiedliche Merkmale auf Organismen zwischen Populationen haben. Molekulare und biochemische Studien versuchen häufig, aus regulatorischen, strukturellen und chemischen Unterschieden mechanistische Ursachen für die Diversität zwischen Artenlinien abzuleiten. Die Komplexität biologischer Systeme und die großen Zeiträume der evolutionären Divergenz erschweren jedoch häufig vergleichende Ansätze zur Analyse moderner Phänotypen in unterschiedlichen Kontexten. Hierbei hilft es, die heutigen Merkmale in einem historischen Licht zu betrachten und zu untersuchen, wie sich die Ursachen der Vielfalt im Laufe der Evolution entwickelten.

Mehrstufige biochemische Synthesewege sind in Pflanzen aufgrund der schnellen Entwicklung neuer spezialisierter Metaboliten ein komplexes Gebiet der Merkmalsektion. Dies führt häufig zu gattungs- oder artspezifischen chemischen Verbindungen, die von unabhängigen Genfamilien synthetisiert werden. Obwohl die Entwicklung solcher mehrstufigen biochemischen Synthesewege, die durch von paralogen (doppelten) Genkopien codierten Enzyme katalysiert werden, bis zu einem gewissen Grad verstanden ist, wurde die Anordnung einzelner Syntheseschritte über Genfamilien hinweg nur selten untersucht. Diterpenoide stellen einen solchen Bereich unterschiedlicher chemischer Merkmale in Pflanzen dar, die häufig auf begrenzte Taxa beschränkt sind. Die Synthese von Diterpenoiden wird durch Enzyme der Diterpensynthase- und Cytochrom P450-Familien katalysiert und im Laufe der Evolution wurden natürlicherweise unterschiedlichste Variationen dieser biochemischen Wege durchlaufen.

Die Gemeine Fichte (*Picea abies*) ist die vorherrschende Baumart in den Wäldern Nord-, Mittel- und Osteuropas. Da sie sowohl Lebensraum bietet und als auch als Nutzholz verwendet wird, ist ein Befall mit Schädlingen (z.B. Käfer), die den Baumbestand dezimieren, ökologisch und ökonomisch verheerend. Nadelbäume, insbesondere die Familie der Pinaceae, produzieren große Mengen verschiedener defensiver Diterpenoide, darunter eine Mischung aus C20-Diterpenharzsäuren. Diese auf einzigartige Weise produzierten Diterpenharzsäuren können in das Holz eindringende Schädlinge mechanisch ausspülen und/oder ersticken. Mit der Veröffentlichung des Genoms der Gemeinen Fichte und der Verfügbarkeit molekularer Werkzeuge für genetische Manipulation bestehen ideale Voraussetzungen, um die Synthesewege von Diterpenharzsäuren in Nadelbäumen zu entschlüsseln. Diese Arbeit nutzt eine Kombination aus Bioinformatik und biochemischen Analysen, um die Entstehung und Entwicklung dieser einzigartigen Diterpenoidwege zu erforschen. Zusätzlich werden einzelne Syntheseschritte charakterisiert und durch genetische Manipulation von Diterpensynthasen und Cytochrom P450-Enzymen untersucht, welchen Einfluss diese auf die Konzentrationen und Flüsse von Metaboliten in Terpenbiosynthesewegen haben.

Acknowledgements

The work carried out in this thesis would not have been possible without the support of my supervisors Dr. Axel Schmidt and Prof. Dr. Jonathan Gershenzon, which I value to no end. Their wisdom and experience has positively impacted my views on biochemistry, chemical ecology, plant physiology and molecular biology through daily personal communications, project and group meetings, and through their contributions to the literature. I am extremely grateful for their advice, guidance, project input, facilitation of fruitful collaborations, and also material support. Additionally, I am thankful for the opportunities they have given me to grow as a scientist outside the boundaries of this thesis. Supporting my attendance at scientific conferences and workshops over the years has expanded both my expertise and professional network. I hope to carry what they have taught me throughout my future endeavors.

I would like to express my gratitude to my committee supervisor Prof. Dr. Günter Theißen for his thoughtful feedback and scientific questions regarding my PhD project, as well as for guiding me during my time as a PhD student. I would additionally like to thank Prof. Dr. Reuben J Peters for valuable discussions and materials, as well as for pushing highly insightful questions regarding terpene and terpene synthase chemistry. Many experiments in this thesis would not be possible without his support. I am also grateful to Prof. Dr. Jörg Bohlmann for providing insight, materials, and for taking the time to discuss my project at conferences.

Gratitude also goes to Dr. Yoko Nakamura, who provided guidance, helpful discussions, and terpene substrates used in this thesis. Marion Stäger, Dr. Michael Reichelt, and Bettina Raguschke have also been extremely helpful in providing transgenic, analytical, and molecular materials and support. I also thank the Max Planck Society for providing funding and professional training, as well as everyone at the Max-Planck-Institute for Chemical Ecology who has supported me over the years, and to the IMPRS for professional development.

I would also like to thank my peers Michael Easson, Benjamin Bartels, Wiebke Häger, Anton Shekhov, and the Stößel family for making Germany feel like home. Also, my peers and office mates Tine Uhe and Pam Medina have been wonderful support and company. I am also grateful for my mom and brothers, who have put up with me over the years.

Eigenständigkeitserklärung

Hiermit erkläre ich, dass mir die geltende Promotionsordnung der Biologisch-Pharmazeutischen Fakultät der Friedrich-Schiller-Universität Jena bekannt ist. Entsprechend § 5 Abs. 4 der Promotionsordnung bestätige ich, dass ich diese Dissertation selbst angefertigt habe und keine Textabschnitte eines Dritten oder eigener Prüfungsarbeiten ohne Kennzeichnung übernommen habe. Weiterhin habe ich alle benutzten Hilfsmittel und Quellen angegeben. Personen, die mich bei der Erhebung und Auswahl des Materials sowie bei der Erstellung der Manuskripte unterstützt haben, sind in der Danksagung genannt. Ich habe keine Hilfe eines Promotionsberaters in Anspruch genommen und es wurden im Zusammenhang mit dem Inhalt der Dissertation keine Geldwerte oder Leistungen unmittelbar oder mittelbar an Dritte weitergegeben. Die Dissertation wurde nicht bereits zuvor als Prüfungsarbeit für eine staatliche oder andere wissenschaftliche Prüfung eingereicht. Weiterhin wurde keine gleiche, in wesentlichen Teilen ähnliche oder andere Abhandlung als Dissertation bei einer anderen Hochschule eingereicht.

Andrew John O'Donnell Jena, den