THE PLANT GENOME: AN EVOLUTIONARY VIEW ON STRUCTURE AND FUNCTION

The family of terpene synthases in plants: a mid-size family of genes for specialized metabolism that is highly diversified throughout the kingdom

Feng Chen[1,*], Dorothea Tholl[2], Jörg Bohlmann[3] and Eran Pichersky[4]

[1]*Department of Plant Sciences, University of Tennessee, Knoxville, TN 37996, USA,*
[2]*Department of Biological Sciences, 408 Latham Hall, Virginia Polytechnic Institute and State University, Blacksburg, VA 24061, USA,*
[3]*Michael Smith Laboratories, 2185 East Mall, University of British Columbia, Vancouver, BC V6T 1Z4, Canada, and*
[4]*Department of Molecular, Cellular and Developmental Biology, University of Michigan, Ann Arbor, MI 48109, USA*

SUMMARY

Some plant terpenes such as sterols and carotenes are part of primary metabolism and found essentially in all plants. However, the majority of the terpenes found in plants are classified as 'secondary' compounds, those chemicals whose synthesis has evolved in plants as a result of selection for increased fitness via better adaptation to the local ecological niche of each species. Thousands of such terpenes have been found in the plant kingdom, but each species is capable of synthesizing only a small fraction of this total. In plants, a family of terpene synthases (TPSs) is responsible for the synthesis of the various terpene molecules from two isomeric 5-carbon precursor 'building blocks', leading to 5-carbon isoprene, 10-carbon monoterpenes, 15-carbon sesquiterpenes and 20-carbon diterpenes. The bryophyte *Physcomitrella patens* has a single *TPS* gene, copalyl synthase/kaurene synthase (CPS/KS), encoding a bifunctional enzyme producing *ent*-kaurene, which is a precursor of gibberellins. The genome of the lycophyte *Selaginella moellendorffii* contains 18 *TPS* genes, and the genomes of some model angiosperms and gymnosperms contain 40–152 *TPS* genes, not all of them functional and most of the functional ones having lost activity in either the CPS- or KS-type domains. *TPS* genes are generally divided into seven clades, with some plant lineages having a majority of their *TPS* genes in one or two clades, indicating lineage-specific expansion of specific types of genes. Evolutionary plasticity is evident in the TPS family, with closely related enzymes differing in their product profiles, subcellular localization, or the *in planta* substrates they use.

Keywords: gene family evolution, terpene synthase subfamily, transit peptide, neofunctionalization, comparative genomics, molecular phylogeny.

INTRODUCTION

Many plants are truly autotrophic organisms. Using only carbon dioxide, minerals, water, and sunlight, they are able to synthesize all of the metabolites necessary for their growth and development, often referred to as 'primary' metabolites. The very existence of autotrophic plants has opened up ecological niches for organisms – animals, fungi, bacteria, and even parasitic plants – that do not need to synthesize all the components of their own primary metab-

olism but instead obtain them by consuming plants (or other organisms further up the food chain). Conversely, plants often depend on other organisms for optimal growth and completion of their life cycle. Selection on plants for greater fitness is likely to have played a major role in the evolution of a suite of biochemical pathways in plants, the so-called 'secondary' metabolism, whereby plants synthesize chemicals that can be harmful or beneficial to other organisms

(Fraenkel, 1959). For example, plants produce an arsenal of toxic compounds that serve as defense against pathogens or herbivores, and plants also produce an array of compounds that attract other organisms which, while feeding on the plants, provide benefits to the plants (e.g., pollinators, nitrogen-fixing bacteria, and mycorrhizal fungi) (Pichersky and Gang, 2000; Pichersky and Gershenzon, 2002; Hartmann, 2007).

Since each plant species occupies a unique ecological niche, it is not surprising that it typically evolves the ability to synthesize a different set of such compounds to help it interact with its biotic and even its abiotic environment. This part of plant metabolism, now referred to as 'specialized' metabolism because it represents specific solutions (i.e., adaptations) to specific ecological problems, is therefore not wholly shared by all plant lineages. By definition, specialized metabolism is the most diverse part of plant metabolism. Furthermore, because other organisms evolve new responses to plant specialized metabolites in the process of adaptive selection, the same process of selection for higher fitness in the plant leads in turn to the appearance of yet new specialized metabolites, a veritable arms race. The result is that plants, in the aggregate, are known to be able to synthesize several hundred thousand different specialized metabolites, and the actual number might be much higher (Hartmann, 2007). However, it is clear that any given plant species can synthesize only a small fraction of such compounds.

Often, plant genomes contain sets of related genes (i.e., 'gene families') that encode enzymes that use similar substrates and give similar products, but have clearly diverged in different lineages. The terpene synthase (TPS) family in plants is such an example (Bohlmann *et al.*, 1998a). Some estimates suggest that more than 25 000 terpene structures may exist in plants (McCaskill and Croteau, 1997).

Analysis of the several plant genomes that have been sequenced and annotated indicates that, with the exception of the moss *Physcomitrella patens*, which has a single functional *TPS* gene, the *TPS* gene family is a mid-size family, with gene numbers ranging from approximately 20 to 150 in sequenced plant genomes analyzed in this article (Table 1). Here we discuss present knowledge of the *TPS* family in plants–the number and organization of the genes in each genome, function of specific genes, and the evolution of novel biochemical functions in different lineages.

SUBSTRATES AND PRODUCTS OF TERPENE SYNTHASES

Terpenoids (isoprenoids) are any compounds that are derived from the isomeric 5-carbon building blocks isopentenyl diphosphate (IPP) and dimethylallyl diphosphate (DMAPP). Two independent pathways in plants, the MEP pathway operating in plastids and the mevalonic acid pathway operating in the cytosol/peroxisomes, lead to the synthesis of both IPP and DMAPP (Lichtenthaler, 1999; Sapir-Mir *et al.*, 2008). These building blocks are combined with each other by the activity of prenyl transferases, which are also known as isoprenyl diphosphate synthases (Wang and Ohnuma, 2000). The condensation of one DMAPP molecule with one IPP molecule in what is called a head-to-tail condensation gives geranyl diphosphate (GPP), if in the *trans* configuration, or neryl diphosphate (NPP), if in the *cis*-configuration. Adding another IPP unit to GPP will give farnesyl diphosphate (FPP) in the *trans* configuration or in the *cis* configuration, and another IPP unit will give geranylgeranyl diphosphate (GGPP) (Takahashi and Koyama, 2006). Longer prenyl diphosphates are also synthesized and some are ultimately added (with the loss of the diphosphate moiety) as components to form such compounds as chlorophylls, phylloquinones, and ubiquinones (McGarvey and Croteau, 1995).

Table 1 Sizes of the TPS family and subfamilies in seven model plant genomes

Species	Genome size (Mb)	Chromosome number (1N)	Total *TPS* gene models[a]	Putative full length *TPSs*[b]	TPS subfamily[c]								
					a	b	c	d	(e)	(f)	e/f	g	h
P. patens	480	27	4	1	0	0	1	0	(0)	(0)	0	0	0
S. moellendorffii	106	27	18[d]	14[d]	0	0	3	0	(3)	(0)	3	0	8
V. vinifera	487	19	152[e]	69[e]	30	19	2	0	(1)	(0)	1	17	0
P. trichocarpa	410	19	68	32	13	12	2	0	(2)	(1)	3	2	0
A. thaliana	125	5	40[f]	32[f]	22	6	1	0	(1)	(1)	2	1	0
O. sativa	389	12	57	34	18	0	3	0	(9)	(0)	9	2	0
S. bicolor	730	10	48	24	15	2	1	0	(3)	(0)	3	3	0

[a]Total number of *TPS* gene models identified in this analysis as described in Data S1 unless noted otherwise.
[b]The number of gene models encoding proteins larger than 500 amino acids and therefore likely to encode full length TPS proteins.
[c]The number of putative full-length *TPS* genes in each subfamily. The columns '(e)' and '(f)' list the number of *TPS* genes in each species according to the original classification. The subfamilies TPS-e and TPS-f are merged to form a new subfamily designated as TPS-e/f.
[d]These may contain allelic variants of same genes.
[e]These numbers were based on the analysis presented by Martin *et al.* (2010).
[f]These numbers were based on the analysis presented by Aubourg *et al.* (2002).

Some terpenoids produced by the activity of prenyl transferases and TPSs are found in all plants and therefore belong to 'primary,' or general, metabolism. For example, head-to-head condensation of two FPP molecules leads to the production of squalene (with the loss of both diphosphate groups), the precursor of sterols. A similar condensation of two GGPP molecules gives phytoene, the precursor of carotenoids. GGPP is also a precursor of gibberellins; a pair of structurally related TPS enzymes in angiosperms and gymnosperms, copalyl diphosphate synthase (CPS) and kaurene synthase (KS), convert GGPP first to copalyl diphosphate (CPP), then to *ent*-kaurene, the precursor of all plant gibberellins (Yamaguchi, 2008) (Figure 1).

In addition to CPS and KS, the plant TPS family consists of enzymes that use DMAPP to make isoprene (C5), GPP or NPP to make various cyclic and acyclic monoterpene hydrocarbons (C10) and monoterpene alcohols, FPP to make various cyclic and acyclic sesquiterpene hydrocarbons (C15) and sesquiterpene alcohols, and GGPP, CPP, or hydroxyl-CPP to make various cyclic and acyclic diterpene hydrocarbons (C20) and diterpene alcohols (Tholl, 2006) (Figure 1). Based on the reaction mechanism and products formed, plant TPSs can be classified into two groups: class I and class II. CPS is a representative of class II TPSs. It catalyzes the formation of CPP through protonation-induced cyclization of GGPP. However, most known plant TPSs are class I TPSs. In the
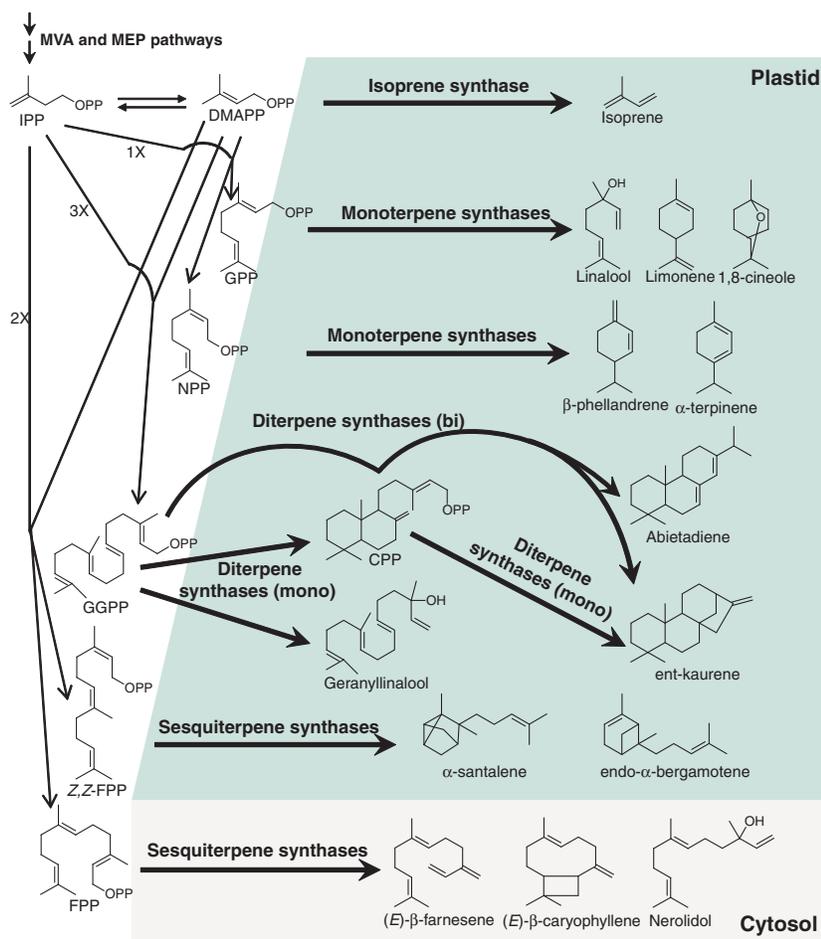
Figure 1. An outline on the formation of terpenoids catalyzed by various types of terpene synthases.
Isopentenyl diphosphate (IPP) is the common precursor of all terpenes. It is synthesized by both the cytosol-localized mevalonic acid (MVA) pathway and the MEP pathway in plastids. IPP is isomerized to give dimethylallyl pyrophosphate (DMAPP). DMAPP either serves as the substrate for hemiterpene biosynthesis or fuses with one IPP unit to form geranyl diphosphate (GPP). The condensation of one GPP molecule with one IPP molecule gives farnesyl diphosphate (FPP), and the condensation of one FPP molecule with one IPP molecule will give geranylgeranyl diphosphate (GGPP). GPP, FPP and GGPP are the precursor for monoterpenes, sesquiterpenes and diterpenes, respectively. While these prenyl diphosphates in the *trans*-configuration have been believed to be the ubiquitous natural substrates for terpene synthases, recent studies showed that two prenyl diphosphates in the *cis*-configuration, neryl diphosphate (NPP) and *Z,Z*-FPP, are also the naturally occurring substrates of terpene synthases. Isoprene synthase, monoterpene synthases, sesquiterpene synthases, and diterpene synthases convert DMAPP, GPP (or NPP), FPP (or *Z,Z*-FPP), and GGPP to isoprene, monoterpenes, sesquiterpenes and diterpenes, respectively. In general, the biosynthesis of isoprene, monoterpenes, and diterpenes occurs in the plastid and the biosynthesis of sesquiterpenes occurs in the cytosol. One exception exists for sesquiterpene synthases: one plastid-localized TPS in tomato produces sesquiterpenes derived from *Z,Z*-FPP (Sallaud *et al.*, 2009). Diterpene synthases can be divided into two groups: bifunctional (bi) and monofunctional (mono). Bifunctional diterpene synthases catalyze the consecutive conversion of GGPP to copalyl diphospate (CPP) then CPP to stable diterpenes. Monofunctional diterpene synthases catalyze the formation of diterpenes from CPP or directly from GGPP without a CPP intermediate.

initial step of the enzymatic reactions catalyzed by class I TPSs, the prenyl diphosphate is ionized and carbocation intermediates are formed. A striking feature of class I TPS enzymes is that, because of the stochastic nature of bond rearrangements that follow the creation of the unusual carbocation intermediates, a single TPS enzyme using a single substrate often gives rise to multiple products (Steele *et al.*, 1998; Christianson, 2006; Degenhardt *et al.*, 2009a). For example, AtTPS-Cin from *Arabidopsis thaliana* catalyzes the formation of 10 monoterpenes with 1,8-cineole being the most abundant product (Chen *et al.*, 2004). The reaction that converts GPP to the final terpene products catalyzed by AtTPS-Cin starts with the metal ion-dependent ionization of GPP. The resulting carbocation can undergo cyclizations, hydride shifts and other rearrangement before the reaction is terminated via deprotonation or nucleophile capture (often water) (Figure 2).

Many of the direct products of TPS enzymes in specialized metabolism, namely isoprene, mono- and sesquiterpenes, and some diterpenes, are volatile compounds under the temperature and atmospheric conditions of the environment in which plants live. Modification of the TPS products by oxidation, peroxidation, methylation, acylation, or cleavage changes their physical properties and may alter their biological activities. The terpenes produced directly by TPS enzymes and those obtained after additional modification reactions may serve various ecological roles such as pollinator attraction (Pichersky and Gershenzon, 2002), attraction of insect predators of herbivores (Unsicker *et al.*, 2009), and chemical and physical barriers to feeding or ovipositing insects (Keeling and Bohlmann, 2006a; Heiling *et al.*, 2010), among others. In addition, specialized products of TPSs and their derivatives are widely used as pharmaceuticals (e.g., taxol derived from taxadiene), flavors (e.g., menthol derived from limonene) and fragrances (e.g., santalols derived from santalene), or as industrial materials (e.g., diterpene resin acids), and some are being explored as biofuel precursors (e.g., farnesene) (Bohlmann and Keeling, 2008).

BASIC STRUCTURE OF *TPS* GENES AND ENZYMES: *TPS* GENES ARE LIKELY DERIVED FROM A DUPLICATION OF AN ANCESTRAL GENE ENCODING A BIFUNCTIONAL KAURENE SYNTHASE

The genome of the moss *Physcomitrella patens* contains a single functional *TPS* gene, a bifunctional *CPS/KS*, with homology to *CPS*, *KS*, and *TPS* genes from gymnosperms and angiosperms (Hayashi *et al.*, 2006). This TPS catalyzes the formation of *ent*-kaurene and 16-hydroxykaurene *in vitro* (Hayashi *et al.*, 2006; Anterola *et al.*, 2009). The N-terminal domain of the protein has features of the active site of a CPS, (i.e., features of a class II diterpene synthase), while the KS activity (a class I diterpene synthase) seems to be located in the C-terminal half of the protein (Hayashi *et al.*, 2006). *P. patens* utilizes a diterpene metabolite derived from *ent*-kaurene as an endogenous developmental regulator, which provides insights into the evolution of gibberellin functions in land plants (Hayashi *et al.*, 2010). Sequence analysis indicates that *CPS* and *KS* in both gymnosperms and angiosperms are derived from a duplication of an ancestral *CPS/KS* gene followed by subfunctionalization that involved the loss of activity in one of these domains in one gene and the loss of the other domain in the other duplicate gene (Keeling *et al.*, 2010). In addition to plants, some fungi and bacteria are known to produce gibberellins. In fungi, as in the moss *P. patens*, the biosynthesis of *ent*-kaurene, the gibberellin precursor from GGPP, is catalyzed by bifunctional enzymes (Toyomasu *et al.*, 2000). In contrast, in the legume-associated bacterium *Bradyrhizobium japonicum*, the production of gibberellin precursor *ent*-kaurene from GGPP is catalyzed by two separate monofunctional diterpene synthases, CPS and KS (Morrone *et al.*, 2009), resembling gibberellin biosynthesis in higher plants. Sequence analysis suggests that the diterpene synthases associated with gibberellin biosynthesis, as well as those of related diterpenoids, in plants, fungi, and bacteria may share a common evolutionary origin (Morrone *et al.*, 2009).

Similar scenarios of gene duplication and subfunctionalization, as well as neofunctionalization, can be invoked for the evolution of *TPS* genes and enzymes of specialized

Figure 2. The reaction mechanism of a representative monoterpene synthase, AtTPS-Cin from Arabidopsis (Chen *et al.*, 2004).
The multiple products of AtTPS-Cin are enclosed in dashed boxes. In the reaction, the production of the initial cyclic species, the α-terpinyl cation, opens the door for the secondary cyclizations. The production of acyclic monoterpenes may proceed via either the geranyl cation or the linalyl cation. OPP represents the diphosphate moiety.

metabolism from an ancestral *CPS/KS* gene prototype (Trapp and Croteau, 2001; Keeling *et al.*, 2010). Some gymnosperm diterpene synthases, such as abietadiene synthase from grand fir (*Abies grandis*), retain both class I (KS-type) and class II (CPS-type) functional domains and bifunctional properties, catalyzing the formation of an enzyme-bound CPP from GGPP and then converting CPP to a diterpene (Peters *et al.*, 2000). In angiosperms, all diterpene synthases that have been characterized to date are monofunctional, with loss of activity in one domain or the other. Some of these monofunctional diterpene synthases (class II) catalyze the formation of *ent*-CPP, *syn*-CPP, or hydroxy-CPP from GGPP, while others (class I) use one of these intermediates to make a diterpene hydrocarbon (Cho *et al.*, 2004; Otomo *et al.*, 2004a; Prisic *et al.*, 2004; Falara *et al.*, 2010). However, in both gymnosperms and angiosperms some diterpene synthases have been found that have retained class I activity only (with or without the actual loss of the N-terminal class II domain) and use GGPP to directly produce a diterpene without a CPP intermediate (Mau and West, 1994; Wildung and Croteau, 1996; Herde *et al.*, 2008; Köksal *et al.*, 2010b).

In contrast to the diterpene synthases, which exist both in the form of bifunctional and monofunctional enzymes, all known mono- and sesquiterpene synthases are believed to be monofunctional, having retained only one active site corresponding to the KS-type domain or class I activity of the *CPS/KS* gene prototype. In most cases, the actual CPS-type domain has been substantially deleted (and not just rendered inactive), but some exceptions have been found, which probably represent recent evolution from diterpene synthases (Dudareva *et al.*, 1996; Sallaud *et al.*, 2009; Schilmiller *et al.*, 2009; and see below).

The topic of the structure and function of plant TPS proteins has been examined in a number of excellent recent reviews (e.g., Christianson, 2006, 2008; Degenhardt *et al.*, 2009a). For details on the three-dimensional structures, reaction mechanisms, catalytic plasticity, and diversity of product profiles of plant TPSs, with emphasis on class I TPS from angiosperms and gymnopserms, we refer the reader to the representative work by Starks *et al.* (1997), Whittington *et al.* (2002), Greenhagen *et al.* (2006), Yoshikuni *et al.* (2006), Hyatt *et al.* (2007), Kampranis *et al.* (2007), O'Maille *et al.* (2008), and Tantillo (2010). X-ray structural analyses of plant TPSs have been performed with angiosperm and gymnosperm class I TPS proteins (see below). Although a three-dimensional structure of a functional class II plant diterpene synthase has not yet been reported, work on the gymnosperm class I/class II multi-product abietadiene synthase (Peters *et al.*, 2001, 2003) provided substantial insights into the structural organization and interactions of the two active sites and reaction mechanisms of a plant TPS that closely resembles an ancestral bifunctional enzyme. Of particular interest in the context of the evolution of the plant TPS family is a model for the origin of class I/class II diterpene synthase structural and functional domains recently presented by Cao *et al.* (2010) and supported by X-ray structure analysis of taxadiene synthase, a gymnosperm class I diterpene synthase (Köksal *et al.*, 2010b). Since there is much evidence that gymnosperm and angiosperm mono- and sesquiterpene synthases evolved from an ancestral diterpene synthase, the model presented by Cao *et al.* (2010) serves as a unifying, general framework, which can explain the evolution of protein domain structures of all plant TPSs. Briefly, in agreement with their common origin, all gymnosperm and angiosperm TPSs share a common overall structure derived from a three-domain ancestral TPS resembled by the gymnosperm abietadiene synthase (which in turn resembles *P. patens* PpCPS/KS). The three domains of an ancestral plant diterpene synthase can be related further back to domains of bacterial and fungal TPSs and *trans* prenyl transferases (Cao *et al.*, 2010). Variations of the three-domain structure in many of the well characterized enzymes of the plant TPS family can be explained by loss of a particular domain such as the KS- or CPS-type domain or by loss of activity associated with a particular domain in different lineages of TPS evolution. These variations also account for the variable length of plant TPSs of approximately 600–900 amino acids.

The class I activity, which is found in all plant TPSs, except for monofunctional CPS, resides in the C-terminal domain and includes the 'DDXXD' and 'NSE/DTE' motifs for metal-dependent ionization of the prenyl diphosphate substrate. The class II activity resides in a separate domain and contains, in its active form, a 'DXDD' motif for the protonation-initiated cyclization of GGPP to CPP. The class II active site domain is non-functional in those TPSs for which three-dimensional structures have been obtained (Starks *et al.*, 1997; Whittington *et al.*, 2002; Hyatt *et al.*, 2007; Kampranis *et al.*, 2007; Gennadios *et al.*, 2009; Köksal *et al.*, 2010a,b). A third, functionally less well characterized TPS domain (Cao *et al.*, 2010; Köksal *et al.*, 2010b), was first identified in plants with a sequence of approximately 200 amino acids near the N-terminus (Bohlmann *et al.*, 1998a). This third domain contains variations of an 'EDXXD' motif, which in its active form contributes to class II diterpene synthase activity (Cao *et al.*, 2010; Köksal *et al.*, 2010b). All TPSs containing this third domain are members of the TPS-d3, -c, -e, -f clades of the TPS family (Martin *et al.*, 2004; also see below). In agreement with an ancestral origin of this domain (Bohlmann *et al.*, 1998a; Cao *et al.*, 2010; Köksal *et al.*, 2010b), these clades represent a diverse spectrum of angiosperm and gymnosperm TPSs including diterpene, sesquiterpene, and monoterpene synthases.

In contrast to the cytosolic sequiterpene synthases, most mono- and diterpene synthases have obvious N-terminal plastid transit peptides. Transit peptides are presumably cleaved off in the mature TPS upstream of the 'RRX$_8$W'

motif, which is essential for catalysis of monoterpene cyclization (Whittington *et al.*, 2002; Hyatt *et al.*, 2007) and is also conserved with variations in most sesquiterpene- and diterpene synthases. The model of common ancestry of all plant TPSs, as reconstructed on the level protein domain structure by Cao *et al.* (2010) and supported by recent x-ray structure analysis of a gymnosperm diterpene synthase (Köksal *et al.*, 2010b), is also supported on the level of genomic TPS sequences with patterns of conserved intron positions across many TPSs from bryophytes, gymnosperms and angiosperms representing all major TPS subfamilies (Trapp and Croteau, 2001; Aubourg *et al.*, 2002; Keeling *et al.*, 2010; Martin *et al.*, 2010), although exceptions from conserved intron-exon structures have been reported (Lee and Chappell, 2008).

COMPARATIVE GENOMICS OF THE *TPS* FAMILY AMONG DIVERSE PLANT TAXA

We present here an analysis of the *TPS* family (which formally includes *bona fide CPS* and *KS* genes) in five angiosperm species whose genomes have been most extensively sequenced and annotated – the dicot species grapevine (*Vitis vinifera*) (Jaillon *et al.*, 2007), poplar (*Populus trichocarpa*) (Tuskan *et al.*, 2006) and *A. thaliana* (Arabidopsis Genome Initiative 2000), and the monocot species rice (*Oryza sativa*) (Goff *et al.*, 2002) and sorghum (*Sorghum bicolor*) (Paterson *et al.*, 2009). We also include in the analysis, for comparison purposes, the genomes of the bryophyte *Physcomitrella patens* (Rensing *et al.*, 2008) and the lycophyte *Selaginella moellendorffii* (http://genome.jgi-psf.org/Selmo1/Selmo1.home.html). As terpenes play an important role in the physiology of various gymnosperms but there is currently no complete sequence of a gymnosperm genome, we also included, in some comparisons, data on *TPS* gene composition from various EST and full-length cDNA databases from a number of gymnosperm species, with the majority of the annotated *TPS* sequences coming from several spruce (*Picea*) species (Martin *et al.*, 2004; Ralph *et al.*, 2008) and grand fir (*Abies grandis*) (Bohlmann *et al.*, 1999).

TPS family size

The *P. patens* genome contains four *TPS* gene models (Table 1), one of which has been shown to be the bifunctional PpCPS/KS (Hayashi *et al.*, 2006). A second *TPS* gene encodes a protein that aligns well with PpCPS/KS, except that a number of stop codons are present in frame, suggesting it is a pseudogene. The other two *TPS* gene models are short gene fragments. While these shorter sequences may be due to the poor genome sequence and/or automated annotation, our manual annotation suggests that they are more likely to be true pseudogenes, i.e., the degraded products of duplicated *PpCPS/KS* genes.

The number of *TPS* gene models in the five angiosperm plant species ranges from 40 in Arabidopsis (Aubourg *et al.*,

2002) to 152 in *V. vinifera* (Martin *et al.*, 2010), counting both full-length genes, clearly non-functional genes, and, in the less well annotated genomes, short gene fragments (i.e., incompletely sequenced) (Table 1). Arabidopsis, with the smallest number of total *TPS* gene models, has 32 full-length *TPS* genes (Aubourg *et al.*, 2002), and it is likely that the genomes of the other four angiosperm plants analyzed here have at least as many functional *TPS* genes and probably considerably more. Because of the lack of a completely sequenced and annotated gymnosperm genome, it is not currently possible to estimate the number of *TPS* genes in such genomes, but analysis of data from EST and full length cDNA databases for conifers suggests that the number of expressed functional *TPS* genes – at least 70 in white spruce (*Picea glauca*) – is similar to the number of *TPS* genes found in angiosperms (Keeling *et al.*, 2011). The *S. moellendorffii* genome contains 18 *TPS* gene models (Table 1), but to the best of our knowledge none of them has been functionally characterized. Among these 18 TPS proteins, six contain both the class II 'DXDD' and the class I 'DDXXD' motifs; four contain only the 'DXDD' motif and five contain only the 'DXXDD' motif. Neither motif can be found in the remaining *S. moellendorffii* TPSs, which may represent pseudogenes. The majority of the *S. moellendorffii TPS* genes contain more than 10 introns, similar to the diterpene synthases in gymnosperms and in the moss *P. patens* (Hayashi *et al.*, 2006; Keeling *et al.*, 2010). Several pairs of *TPS* genes in *S. moellendorffii* are highly homologous, implying recent gene duplication. In our analysis of the *S. moellendorffii* genome sequence we noted eight TPS gene models that encode proteins with very limited homology to plant TPSs and high similarity to microbial TPSs. Most of these genes do not contain introns. The proteins they encode are approximately 200–400 amino acids and contain the conserved 'DDXXD' and 'NSE' motifs. While it is possible that these TPSs are authentic *S. moellendorffii* genes, it is also possible that they represent DNA from endophytic organisms.

TPS gene arrangement on chromosomes

The space occupied by *TPS* genes in the *P. patens* genome is about 0.008 gene/Mb. In the other six plants analyzed here, the content of *TPS* genes ranges from 0.07 gene/Mb in sorghum to approximately 0.3 gene/Mb in Arabidopsis and grapevine. These data suggest that the TPS family has undergone significant expansion during the evolution of land plants. Chromosomal locations of the *TPS* genes in *P. patens* and *S. moellendorffii* have not yet been determined. *TPS* genes are found on all five chromosomes in Arabidopsis, and in all but one chromosome in both rice and sorghum, whose chromosome number is 12 and 10, respectively (Table 1). In poplar, with 19 chromosomes, 12 chromosomes contain *TPS* genes. However, this number needs to be viewed with caution because 11 *TPS* genes in

poplar have not yet been mapped to any specific chromosomes. In grapevine, a recent analysis of the 12-fold coverage genome sequence of a highly inbred Pinot noir variety identified 69 putatively functional *VvTPS* genes, 20 partial *VvTPS*, and 63 *VvTPS* probable pseudogenes (Martin *et al.*, 2010). These *VvTPS* gene models are localized on seven of the 19 grapevine chromosomes, but 18 *VvTPSs* remained unmapped.

A significant number of *TPS* genes in the genome of the five angiosperm plants occur in tandem arrays of two or more genes (sometime separated by one or a few unrelated genes). In Arabidopsis, rice, poplar, grapevine and sorghum, 42, 64, 59, 85 and 66% of *TPS* genes, respectively, occur in such tandem arrays. These tandem arrays are likely the consequence of local gene duplication by unequal crossover. Consistent with this hypothesis, the genes in the tandem arrays are typically highly homologous to each other. For example, *AtTPS23* and *AtTPS27* are two Arabidopsis *TPS* genes located in such a cluster, and they are identical to each other both in the coding region and intron sequences, and thus represent a very recent *TPS* gene duplication (Chen *et al.*, 2004). The tandem arrays of *TPS* genes are in some cases quite extensive, as in grapevine, where 45 *VvTPS* genes (almost one third of all *VvTPSs*) are organized as an extremely dense *TPS* gene cluster across a stretch of 690 kb on chromosome 18 (Martin *et al.*, 2010) and in rice, where 14 *TPS* genes (approximately one-quarter of the total rice *TPS* genes) occur in a 480-kb stretch on chromosome 4.

Functions of *TPS* genes

The enzyme encoded by the *P. patens PpCPS/KS* gene catalyzes the formation of *ent*-kaurene and 16-hydroxy-*ent*-kaurene (Hayashi *et al.*, 2006). *ent*-Kaurene is a precursor to *ent*-kaurenoic acid, which in *P. patens* serves the general physiological role in development that the gibberellins play in angiosperms (Hayashi *et al.*, 2010). A specialized ecological role for PpCPS/KS is not presently known; however, the 16-hydroxy-*ent*-kaurene product of PpCPS/KS accumulates in *P. patens* with substantial concentrations of up to 0.5–1 mM and is also being released as a volatile compound (von Schwartzenberg *et al.*, 2004). These levels of accumulation and release might be indicative of a role of 16-hydroxy-*ent*-kaurene in interactions with other organisms, but this hypothesis remains to be tested. No terpenes have yet been reported in *S. moellendorffii*, so possible ecological roles for *TPS* genes in this lycophyte species are not yet known either. In contrast, the ecological roles of terpenes in gymnosperms and angiosperms are many (Gershenzon and Dudareva, 2007), and much has been learned about the genes and enzymes that specify their production. In reviewing the information on the function of the *TPS* genes in the model systems discussed here, it is worth noting that the gymnosperm species studied are extremely long-lived trees with

wind-pollinated cones, the two monocot angiosperm species, rice and sorghum, are grasses whose flowers are wind-pollinated, and among the three dicot angiosperm species, Arabidopsis is a small annual with scented flowers, poplar is a tree with wind-pollinated flowers, and grapevine is a cultivated perennial plant with scented flowers and highly aromatic fruits.

In gymnosperms, the largest number of *TPS* genes of specialized metabolism has been identified and functionally characterized in species of spruce (*Picea*) and in grand fir (*Abies grandis*), with a few individual *TPS* genes identified and characterized in other species such as loblolly pine (*Pinus taeda*), Douglas fir (*Pseudotsuga menziesii*), *Ginkgo biloba* and *Taxus* (Keeling and Bohlmann, 2006a; Keeling *et al.*, 2011). The interest in diterpene synthases from *Ginkgo* and *Taxus* is due to their importance for the biosynthesis of medicinally useful metabolites, of which the anticancer drug taxol is perhaps the most prominent example of a high value terpenoid (Croteau *et al.*, 2006). For the exploration of the *TPS* gene family in conifers (order Coniferales), and in particular members of the pine family (*Pinacea*), the major driving force has been to understand the biochemical, molecular, genomic and evolutionary underpinnings of the great chemical diversity of terpenoids in conifer defense against insects and pathogens, and to apply this knowledge for tree breeding and forest health protection. Conifers produce large quantities of oleoresin composed of complex mixtures of dozens of different acyclic and cyclic monoterpenes, sesquiterpenes, and diterpenes, the latter predominantly in the form of diterpene resin acids (Keeling and Bohlmann, 2006a,b). In addition, conifers emit volatile terpenoids in response to insect attack (Miller *et al.*, 2005; Mumm and Hilker, 2006). Oleoresin terpenoid defenses act directly against invading insects and pathogens, as well as larger herbivores, while induced volatile emissions may function in indirect defense to attract the natural enemies of the attacking herbivores. The size of the *TPS* gene family and the diversity of biochemical functions of TPSs in any one conifer species mirror the complexity of oleoresin composition and volatile emissions. Published reports describe functions for 10 different *TPS* genes in Norway spruce (*P. abies*) (Fäldt *et al.*, 2003a; Martin *et al.*, 2004) and 11 different *TPS* genes in grand fir (Vogel *et al.*, 1996; Bohlmann *et al.*, 1997, 1998b, 1999; Steele *et al.*, 1998). A recent large-scale transcriptome analysis identified 69 *TPS* genes in white spruce (*P. glauca*), 55 *TPS* genes in Sitka spruce (*P. sitchensis*), and 20 *TPS* genes in a hybrid white spruce (*P. glauca* × *P. engelmannii*), of which 21 have been functionally characterized as mono-, sesqui- and diterpene synthases of specialized metabolism (Keeling *et al.*, 2011).

The majority of the conifer TPSs are multi-product enzymes, and it was in grand fir that the phenomenon of highly multi-product TPS enzymes was first highlighted with the discovery of the sequiterpene synthase genes encoding

δ-selinene synthase (34 products) and γ-humulene synthase (52 products) (Steele *et al.*, 1998). However, a few single-product TPSs have also been described in conifers. In the case of the Norway spruce ditepene synthases levopima-radiene/abietadiene synthase (PaTPS-LAS) and isopimard-iene synthase (PaTPS-Iso), multi-product and single-product enzymes exist as closely related (>90% identical) paralogous TPSs in the same species (Martin *et al.*, 2004). Reciprocal site-directed mutagenesis revealed that the functional bifur-cation of these two enzymes is due to only four amino acid residues (Keeling *et al.*, 2008), thus providing an example of *TPS* gene duplication and neofunctionalization and the plasticity of TPS function. Recent studies confirmed that a critical mutation in the isopimaradiene synthase leading to change of product profiles in site-directed mutagenesis also occurs as a natural TPS variation in Sitka spruce (Keeling *et al.*, 2011). This example is perhaps representative of the many events of *TPS* gene duplication, followed by sub- and neofunctionalization that have shaped the functional diver-sity of large families of active *TPS* genes of specialized metabolism in conifers. In addition to the more than 30 functionally characterized *TPS* genes of specialized metab-olism functionally characterized in spruce, primary metab-olism *CPS* and *KS* genes have also been characterized in white spruce and in Sitka spruce, where these genes appear to be actively expressed as single-copy genes (Keeling *et al.*, 2010).

Expression of many of the conifer *TPS* genes involved in oleoresin defenses and volatile emissions is induced by herbivore or pathogen attack, and this effect can be mimicked to some extend by mechanical wounding of trees or treatment with methyl jasmonate (Miller *et al.*, 2005; Zulak *et al.*, 2009). Analysis of TPSs at the proteome and transcriptome levels in Norway spruce and Sitka spruce showed that methyl jasmonate-induced *TPS* transcription lead to increased protein abundance and enzyme activity and correlated with changes in the terpenoid metabolite profiles (Zulak *et al.*, 2009; Hall *et al.*, 2011). Spatially, the constitutive and induced expression of *TPS* genes and proteins in spruce is localized, at least in part, to the epithelial cells of resin ducts in stem tissues as determined by immunofluorescence localization (Zulak *et al.*, 2010) and laser-capture microdissection (Abbott *et al.*, 2010). In most conifers, constitutive and induced traumatic resin ducts are the primary site of massive accumulation of TPS products. Most of the products of mono- and sesquiterpene synthase activity accumulate without further modification. In contrast, most of the products of the conifer diterpene synthases are oxidized to diterpene resin acids by the activity of CYP720B cytochrome P450 enzymes, before being sequestered into resin ducts (Ro *et al.*, 2005).

In Arabidopsis, 14 *TPS* genes have been functionally characterized, which include two diterpene synthase genes that encode CPS and KS respectively for the biosynthesis of gibberellins (Sun and Kamiya, 1994; Yamaguchi *et al.*, 1998) and 12 for the biosynthesis of specialized metabolites. Because of its comprehensive molecular and genetic re-sources, Arabidopsis has become a particularly useful model for in depth studies on the spatial organization and temporal regulation of terpene metabolism in relation to the various biological functions of terpenes. Also, comparisons between ecotype- and genus-specific genomes allow investigating basic molecular mechanisms that control the intra- and inter-specific natural variation of terpene biosynthesis. Volatile terpenoids are emitted from several Arabidopsis tissues under constitutive and/or induced conditions (van Poecke *et al.*, 2001; Aharoni *et al.*, 2003; Chen *et al.*, 2003; Fäldt *et al.*, 2003b; Rohloff and Bones, 2005; Snoeren *et al.*, 2010).

Most of the terpenoids detected so far in Arabidopsis are synthesized by approximately one-third of the enzymes of the Arabidopsis TPS family, with the enzymatic activities and products of the others yet to be determined. Microarray and RT-PCR analyses have indicated distinct or overlapping tissue-specificity of *AtTPS* gene expression, which has been investigated in greater detail for 10 genes (Figure 3). Flowers of the Arabidopsis ecotype Columbia (Col) emit a complex mixture of monoterpenes and over 20 sesquiterpene hydro-carbons with (*E*)-β-caryophyllene as the predominant com-pound (Chen *et al.*, 2003; Tholl *et al.*, 2005). Nearly all of the sesquiterpenes are produced by two flower-specific enzymes, the (*E*)-β-caryophyllene/α-humulene synthase (At5g23960, AtTPS21) and the multi-product sesquiterpene synthase AtTPS11 (At5g44630), and the other terpene com-pounds are synthesized by four additional enzymes (At4g16740, AtTPS03; At2g24210, AtTPS10; At1g61680, AtT-PS14; At3g25810, AtTPS24) (Figure 3a). None of the floral terpenes is synthesized in flower petals, instead their formation occurs particularly in the stigma and sepals (AtTPS21), intrafloral nectaries and ovules (AtTPS11), and in pollen (AtTPS03) (Chen *et al.*, 2003; Tholl *et al.*, 2005; Huang *et al.*, 2010). The terpenes that are emitted from these floral tissues at comparatively low rates can serve multiple functions: they may attract small pollinating insects at short range that contribute to cross-fertilization in natural popu-lations. More recently, floral volatiles have been associated with defensive activities (Raguso, 2009) and Arabidopsis floral terpenes are likely to exhibit such functions to specif-ically protect floral organs such as the stigma against pathogen invasion. Differences in life histories and breeding systems influence floral terpene volatile emissions in the genus Arabidopsis. For example, in the highly scented flowers of the outcrossing perennial *A. lyrata*, formation of (*E*)-β-caryophyllene is substituted by the biosynthesis of benzenoid volatiles but instead has evolved as an insect-inducible trait in leaves catalyzed by a *TPS21* orthologue (Abel *et al.*, 2009).

Similar defensive functions as suggested for flower-specific *TPS* genes are predicted for 14 *AtTPS* genes with

(a)

3, 21, 24
3, 24
11
3, 21, 24
3, 21, 24
3, 24
11

(b)

Endodermis — Stele — Cortex/Endodermis
Epidermis

23, 27
(Chen *et al.*, 2004
Birnbaum *et al.*, 2003)

12, 13
(mature root, Ro *et al.*, 2006)

12, 13
(young root, Ro *et al.*, 2006)

23, 27
(elongation zone with
transition to maturation
zone, Chen *et al.*, 2004)

8 (Birnbaum *et al.*, 2003)

22 (Birnbaum *et al.*, 2003)
Lateral root cap

Root cap

(c)

3, 4, 10
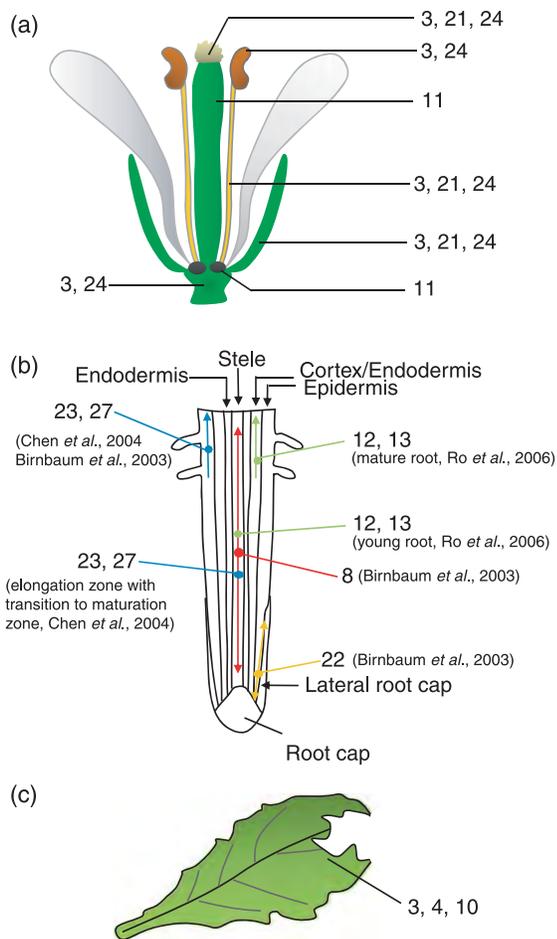
Figure 3. Tissue- and cell type-specific expression patterns of Arabidopsis *TPS* genes in flowers (a), roots (b), and leaves (c) according to transcript and/ or promoter-reporter gene analyses.
Numbers indicate different *TPS* genes. Arrows and dots in (b) indicate root-specific expression. Expression profiles of two root-specific *TPS* genes with predicted function as sesquiterpene and diterpene synthases are shown. In (c), herbivore-induced *AtTPSs* are shown. 3, AtTPS03, α-farnesene synthase; 4, AtTPS04, (*E,E*)-geranyllinalool synthase; 8, AtTPS08, putative diterpene synthase; 11, AtTPS11, multi-sesquiterpene synthase; 12 and 13, AtTPS12, AtTPS13, (*Z*)-γ-bisabolene synthases; 21, AtTPS21, (*E*)-β-caryophyllene synthase; 22, AtTPS22, putative sesquiterpene synthase; 24, AtTPS24, multi-product monoterpene synthase; 23 and 27, AtTPS23, AtTPS27, 1,8-cineole synthases.

primary or exclusive expression in Arabidopsis roots. Four of these genes have been functionally characterized to encode monoterpene and sesquiterpene synthases (Chen *et al.*, 2004; Ro *et al.*, 2005) and diterpene synthase or sesquiterpene synthase activities have been predicted for others. According to high resolution transcriptome maps (Birnbaum *et al.*, 2003; Brady *et al.*, 2007) and promoter-reporter gene studies, many root-expressed genes exhibit cell type- and root zone-specific expression patterns (Figure 3b) that may emerge in defense against soil-borne, root-attacking organisms with different feeding or infection strategies. For example, promoter-GUS analyses demon-

strated that the identical, duplicated genes *AtTPS23* and *AtTPS27* (At3g25820 and At3g25830), which are responsible for the formation of the monoterpene 1,8-cineole (Chen *et al.*, 2004), are expressed in the vascular tissue of the root elongation zone and in epidermal cells and root hairs of older root-growth zones (Chen *et al.*, 2004) (Figure 3b), where 1,8-cineole can directly be released into the rhizosphere. Similarly, the two (*Z*)-γ-bisabolene synthase genes *AtTPS12* (At4g13280) and *AtTPS13* (At4g13300) are expressed in the stele of younger roots and in the cortex and endodermis of mature roots (Ro *et al.*, 2006) (Figure 3b). Both genes have emerged by gene duplication and are found in pairs together with the co-expressed P450 genes *CYP71A19* and *CYP71A20* suggesting that these gene pairs form cell type-specific biosynthetic modules.

Leaves of the Arabidopsis ecotype Col release a mixture of volatiles containing the C_{16}-homoterpene TMTT, the sesquiterpene (*E,E*)-α-farnesene, and the monoterpene β-myrcene in an induced response to herbivory, elicitor treatment, and the application of jasmonic acid or jasmonate mimics (van Poecke *et al.*, 2001; Chen *et al.*, 2003; Herde *et al.*, 2008; Huang *et al.*, 2010; Snoeren *et al.*, 2010). AtTPS04 (At1g61120) catalyzes the formation of (*E,E*)-geranyllinalool as the first dedicated step in the biosynthesis of TMTT (Herde *et al.*, 2008; Lee *et al.*, 2010), while (*E,E*)-α-farnesene and β-myrcene are produced by the (*E,E*)-α-farnesene synthase AtTPS03 and presumably the β-myrcene/ocimene synthase AtTPS10, respectively (Bohlmann *et al.*, 2000; Huang *et al.*, 2010). Both *AtTPS03* and *AtTPS04* genes are induced locally at wound sites (Herde *et al.*, 2008; Huang *et al.*, 2010) (Figure 3c). The herbivore-induced volatile mixture was shown to attract parasitoids of herbivores (van Poecke *et al.*, 2001; Loivamäki *et al.*, 2008) and contributes to increased plant fitness (van Loon *et al.*, 2000). Parasitoid attraction varies depending on the volatile blend emitted from different Arabidopsis ecotypes (Snoeren *et al.*, 2010). For example, the ecotypes Col and Ws differ in their emission of (*E*)-β-ocimene, and this difference is due to a mutation that inactivates the (*E*)-β-ocimene-producing *AtTPS02* in the Col ecotype (Huang *et al.*, 2010).

Although the genome sequence for *Populus* was published in 2006, the functions of only three poplar *TPS* cDNAs have been characterized so far, including two encoding isoprene synthase and one encoding a sesquiterpene synthase. All of these genes are involved in the formation of volatile terpenes. Genes encoding isoprene synthase have been isolated from *P. alba* × *tremula* (Miller *et al.*, 2001) and *P. alba* (Sasaki *et al.*, 2005). The proteins encoded by these two isoprene synthase genes are 99% identical, suggesting that they are variants of the same gene. The release of volatile isoprene, produced by isoprene synthase, from *Populus* accounts for a large amount of the biogenic carbon found in the atmosphere. It is believed that isoprene emission has a role in the ecological response of *Populus*

to abiotic stress (Sharkey and Yeh, 2001; Behnke *et al.*, 2007). A full-length cDNA (PtdTPS1) was isolated from *P. tricho-carpa × deltoides* and shown to encode a sesquiterpene synthase responsible for the synthesis of insect-induced (−)-germacrene D (Arimura *et al.*, 2004). The herbivore-induced local and systemic mono- and sequiterpenoid emissions produced by poplar *TPS* genes may serve functions in multitrophic defence (Arimura *et al.*, 2004) and in within-plant signaling (Frost *et al.*, 2008).

In contrast to the poplar system, a substantial effort has been made to comprehensively analyze the genome organization and functions of the *TPS* gene family in grapevine (Martin *et al.*, 2010). Terpenoids, in the form of free volatiles and as glycoside conjugates of monoterpene alcohols, are among the most important flavor compounds of grape berries and wine bouquet (Swiegers *et al.*, 2005; Lund and Bohlmann, 2006). In addition, grapevine flowers produce a terpenoid-rich floral scent (Martin *et al.*, 2009). In total, 43 *VvTPS* FLcDNAs from several grapevine varieties have been functionally characterized, representing mostly monoterpene synthases and sesquiterpene synthases (Lücker *et al.*, 2004; Martin and Bohlmann, 2004; Martin *et al.*, 2010). The majority of the functionally characterized VvTPS are multi-product enzymes. The products of these TPSs account for many of the major and minor terpenoids of berry and wine flavor, including a suite of monoterpene alcohols. A functionally characterized sesquiterpene synthase from Cabernet Sauvignon (VvValCS) and Gewürztraminer (VvValGw) produces (−)-valencene and (−)-7-epi-α-selinene as the main products (Lücker *et al.*, 2004; Martin *et al.*, 2009). This *TPS* gene is expressed in anthers and developing pollen grains and contributes to diurnal floral scent emission (Martin *et al.*, 2009).

The rice *TPS* family has also been relatively well studied. Rice plants produce and emit a mixture of volatile terpenoids, including monoterpenes and sesquiterpenes, after insect herbivory (Yuan *et al.*, 2008). By correlating terpenoid emission and expression of *TPS* genes analyzed by microarrays, three *TPS* genes were identified to be responsible for the production of the majority of insect-induced volatile terpenoids in rice. One of these genes encodes a monoterpene synthase making linalool as the single product, which is the most abundant insect-induced volatile from nipponbare rice plants (Yuan *et al.*, 2008). The other two characterized *TPS* genes encode sesquiterpene synthases. Both of them make multiple sesquiterpenoids (Yuan *et al.*, 2008).

Rice plants also produce a large number of labdane type diterpenoids. The known rice diterpenoids fall into five structurally related families: gibberellins, phytocassanes A–E, oryzalexins A–F, momilactones A and B, and oryzalexin S, which are derived from *ent*-kaurene, *ent*-cassa-12,15-diene, *ent*-sandaracopimaradiene, syn-pimara-7,15-diene and syn-stemar-13-ene, respectively (Peters, 2006;

Toyomasu, 2008). Three rice TPSs have been shown to function as CPS, including OsCPS1 involved in gibberellin biosynthesis, OsCPS2 involved in the biosynthesis of non-gibberellin diterpenoids (Prisic *et al.*, 2004) and OsCPSsyn producing *syn*-CPP for the biosynthesis of syn-labdane-related diterpenoids (Xu *et al.*, 2004). A number of rice TPSs that use *ent*-CPP or *syn*-CPP as the direct substrate have also been characterized. These include OsKS1, which is a *bona fide* KS involved in gibberellin biosynthesis to produce *ent*-kaurene (Sakamoto *et al.*, 2004), OsKSL4 (OsDTS2) for producing *syn*-pimaradiene (Wilderman *et al.*, 2004), OsKS5 for producing *ent*-pimara-8(14),15-diene (Kanno *et al.*, 2006), OsKS6 for producing *ent*-kaur-15-ene (*ent*-isokaurene) (Kanno *et al.*, 2006; Xu *et al.*, 2007), OsKSL7 (OsDTC1) for producing *ent*-cassadiene (Cho *et al.*, 2004), OsKSL8 (OsDTC2) for producing syn-stemarene (Nemoto *et al.*, 2004), OsKSL10 for producing *ent*-sandaracopimaradiene (Otomo *et al.*, 2004a), and OsKSL11 for producing stemod-13(17)-ene (Morrone *et al.*, 2006). These non-gibberellin diterpenoids function as phytoalexins to defend rice plants against microbial infection (Kato *et al.*, 1994). Some of these compounds are produced by the rice plants then released into the environment to inhibit the germination and growth of neighboring plants, therefore acting as allelochemicals (Kato-Noguchi and Ino, 2003).

In contrast to rice, little is known about terpenes and *TPS* genes in sorghum. A recent study found that like rice sorghum plants emit volatile terpenoids, predominated by sesquiterpenes, upon insect herbivory. Colinearity analysis based on the synteny between the rice and sorghum genomes led to the identification of sorghum *TPS* genes, which are the orthologues of key rice *TPS* genes for producing insect-induced volatile terpenoids (X. Zhuang and F. Chen, unpublished results). Although not included in the analysis in this review, maize has served as another important monocot model for investigating terpenoid metabolism, especially in the context of plant-insect inter-actions. The majority of the *TPS* genes isolated from maize are involved in making herbivory-induced volatile terpenoids (Schnee *et al.*, 2002, 2006; Köllner *et al.*, 2008, 2009). Transgenic studies using some of these maize *TPS* genes convincingly showed that their products have a role in attracting parasitoids (Schnee *et al.*, 2006) and entomopath-ogenic nematodes (Degenhardt *et al.*, 2009b), the natural enemies of maize herbivores.

EVOLUTION OF TERPENE SYNTHASES: MOLECULAR PHYLOGENY VERSUS FUNCTION

Topography of the TPS family tree

Previous phylogentic analyses of TPS protein sequences from gymnosperms and angiosperms recognized seven major clades (or subfamilies), designated TPS-a through TPS-g (Bohlmann *et al.*, 1998a; Dudareva *et al.*, 2003; Martin

et al., 2004). We have now substantially extended this analysis to include new TPS sequences from the sequenced genomes of several plant species. The new analysis also recognizes seven TPS subfamilies – the original a, b, c, d, and g, a merged clade of the original e and f subfamilies now designated as e/f, and a new subfamily h (Figure 4). Function and taxonomic distribution of plant TPS subfamilies are summarized in Table 2.

PpCPS/KS, the single member of the TPS family in *P. patens*, is placed in the TPS-c clade. Other members of this clade are all *bona fide* CPS proteins from gymnosperms (spruce PgCPS and PsCPS) and angiosperms (Arabidopsis CPS, and rice OsCPS1, OsCPS2 and OsCPSsyn) as well as three *S. moellendorffii* TPSs that contain only the 'DXDD' motif but not the 'DDXXD' motif, suggesting that they are monofunctional CPS. Since, as discussed above, evidence suggests that all other *TPS* genes evolved from a prototype bifunctional CS/KS gene, it is likely that the TPS-c clade represents the base of the tree (the tree in Figure 4 is unrooted, see Figure S1 for a rooted tree using bacteria CPS as the outgroup), and, furthermore, that mono-functional-

ization occurred in the TPS family very early in land plant evolution.

Closely related to the TPS-c subfamily are the TPS-e and TPS-f subfamilies. TPS-e contains all *bona fide* KS proteins from gymnosperms (PgKS and PsKS) and angiosperms (Arabidopsis KS and rice OsKS1). As in the TPS-c subfamily, several species analyzed have more than one TPS belonging to the TPS-e subfamily. Three *S. moellendorffii* TPSs form a subclade that is located near the bifurcation node of the TPS-c and the TPS-e clades (Figure 4). The presence of the 'DDXXD' motif but not the 'DXDD' motif in the proteins encoded by these *S. moellendorffii* genes suggests they function as class I TPS, probably KS. Therefore, this branch of *S. moellendorffii* TPSs are placed in the TPS-e subfamily. When the TPS-f subfamily (Figure 4), which contains AtTPS04, a diterpene synthase making geranyllinalool in Arabidopsis (Herde *et al.*, 2008), uncharacterized TPSs from poplar and grapevine (Martin *et al.*, 2010) as well as CbLIS, a monoterpene synthase producing *S*-linalool from *C. breweri* flowers (Dudareva *et al.*, 1996) and two unusual TPSs from *Solanum* (see below), was first defined, it split from the base
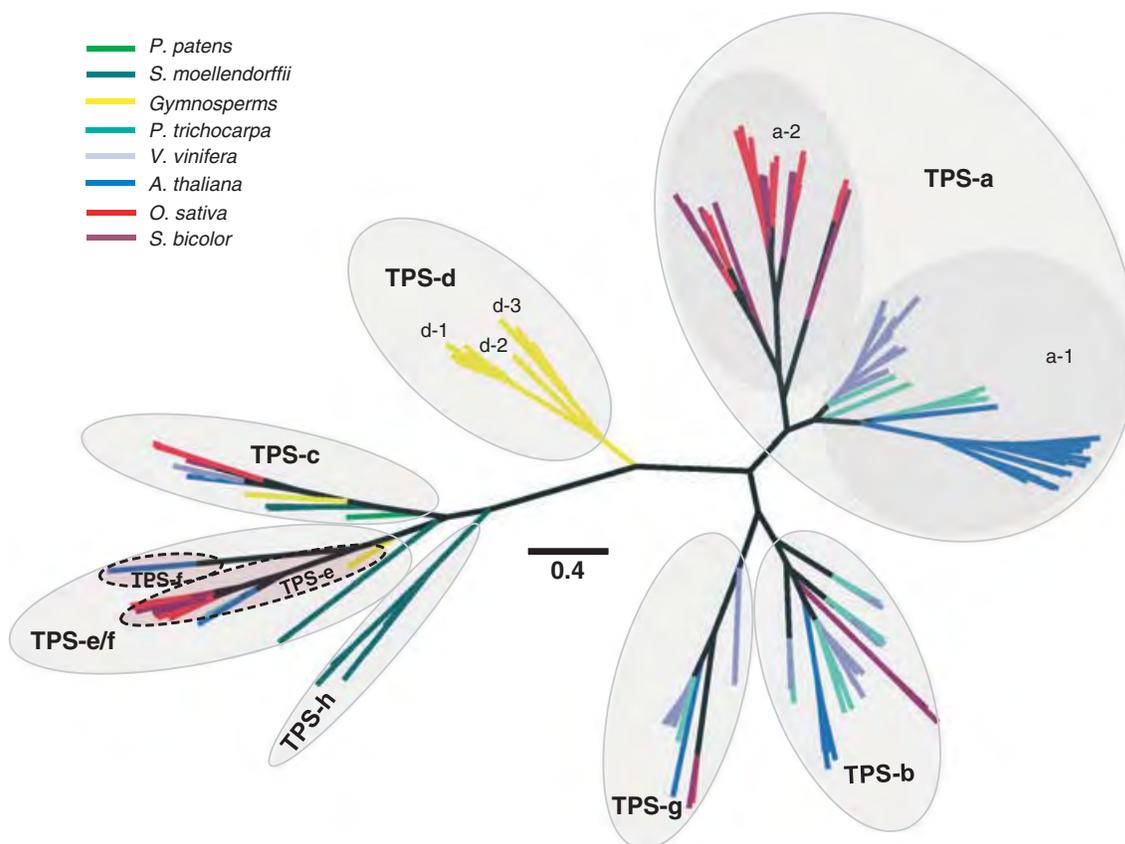
Figure 4. Phylogeny of putative full-length TPSs from the seven sequenced plant genomes and representative characterized TPSs from gymnosperms (Table S1). Based on the phylogeny and functions of known TPSs, seven subfamilies of TPSs are recognized. These include subfamily TPS-c (most conserved among land plants), subfamily TPS-e/f (conserved among vascular plants), subfamily TPS-h (*Selaginella moellendorffii* specific), subfamily TPS-d (gymnosperm specific), and three angiosperm-specific subfamilies TPS-b, TPS-g and TPS-a. The TPS-a subfamily is further divided into two groups, a-1 being dicot-specifc and a-2 being monocot-specific. The TPS-d subfamily is further divided into three groups, d-1, d-2 and d-3, which show distinction in function of TPSs in each group. The TPS-e/f subfamily is merged from the previously separate TPS-e and TPS-f subfamilies, which are also shown on the phylogenetic tree.

Table 2 Function and taxonomic distribution of plant TPS subfamilies

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

[a]DiTPS, diterpene synthase; IspS, isoprene synthase; MonoTPS, monoterpene synthase; SesquiTPS, sesquiterpene synthase.

of the TPS-e subfamily (i.e., it appeared to be a sister clade). However, with the additional sequences in the present analysis, it is clear that TPS-f is derived from TPS-e (Figure 4), and we have therefore combined the two subfamilies into one clade, designated as the TPS-e/f subfamily. Evidently, genes in subclade f represent evolutionary novelties, but we note that none of the rice and sorghum TPSs belongs to the TPS-f subclade, indicating that it is probably dicot specific.

Eight *S. moellendorffii* TPSs that do not belong to the previously defined subfamilies of TPS-c and TPS-e/f form a new clade, which is introduced here as the new TPS-h subfamily (Figure 4). Except for one TPS that is presently missing the sequence at the region containing the 'DDXXD' motif, all *S. moellendorffii* TPSs in the TPS-h subfamily contain both 'DXDD' and 'DDXXD' motifs. Interestingly, among the more than 300 *TPS* genes identified from the five angiosperm species analyzed in the work described here, none contains both 'DXDD' and 'DDXXD' motifs. Conversely, all functionally characterized bifunctional diterpene synthases of specialized metabolism in the gymnosperms contain both the 'DXDD' and the 'DDXXD' motif, similar to the bifunctional PpCPS/KS (Keeling *et al.*, 2010). These gymnosperm bifunctional diterpene synthases belong to the TPS-d subfamily (Figure 4). It is likely that gymnosperm bifunctional diterpene synthases evolved from a CPS/KS prototype TPS (probably before the gymnosperm-angiosperm split, since neither lineage appears to contain a CPS/KS) (Keeling *et al.*, 2010). Similarly, it is likely that the putative bifunctional TPSs in the newly defined subfamily TPS-h evolved from PpCPS/KS and may be involved in specialized metabolism in *S. moellendorffii*.

In the gymnosperms we see a clear phylogenetic separation of *TPS* genes of specialized metabolism and *TPS* genes of primary gibberellin metabolism. Known gymnosperm CPS and KS of gibberellin metabolism belong, respectively, to the TPS-c and TPS-e/f clades like their counterparts in the angiosperms. Consistent with previous analyses, all gymnosperm TPSs for specialized metabolism belong to the gymnosperm-specific subfamily TPS-d. The TPS-d subfamily can be further divided into TPS-d-1, TPS-d-2 and TPSd-3 (Martin *et al.*, 2004). TPS-d-1 contains all known gymnosperm monoterpene synthases for the production of a large array of conifer defense compounds, in addition to a few TPSs which produce the simple acyclic sesquiterpene (*E,E*)-α-farnesene. Most known gymnosperm sesquiterpene synthases, including those enzymes that produce large arrays of multiple terpenoids (Steele *et al.*, 1998) belong to the TPS-d-2 group. The TPS-d-3 contains primarily diterpene synthases and several known sesquiterpene synthases. Members of the gymnosperm group TPS-d-3 contain an ancestral domain of approximately 200 amino acids which is also present in TPSs of the clades TPS-c, TPS-e/f, and TPS-h. Well characterized diterpene synthases in the TPS-d-3 group are the single-product and multi-product enzymes of conifer diterpene resin acid biosynthesis (Peters *et al.*, 2000; Keeling *et al.*, 2008) and taxadiene synthase from Taxus (Wildung and Croteau, 1996). Although sesquiterpene synthases are found in all three groups, they can be distinguished by gene structures: the sesquiterpene synthases in TPS-d-1 and TPS-d-2 are approximately 600 amino acids in length and those in TPS-d-3 are approximately 800 amino acids.

The angiosperm-specific TPS-a, TPS-b and TPS-g clades have substantially diverged from the other TPS clades (Figure 4). Many of the genes in these three clades have been functionally characterized in model- and non-model systems. Based on current knowledge, these three clades comprise entirely of genes of specialized mono-, sesqui- or diterpene biosynthesis with roles in ecological plant interactions, rather than roles in primary plant metabolism. For the five flowering plants for which genome sequences have been analyzed in this review, the *TPS* genes in TPS-a account for more than half of their *TPS* genes. TPS-a can be further divided into two groups, TPS-a-1 and TPS-a-2, with the former being monocot-specific and the latter dicot-specific. Most of the characterized TPSs in the TPS-a clade are sesquiterpene synthases. The TPS-a clade includes the first ever published *TPS* gene of specialized plant metabolism, the 5-*epi*-aristolochene synthase from tobacco which has since served as a model TPS for much of our understanding of *TPS* genes and enzymes in plants (Facchini and Chappell, 1992; Starks *et al.*, 1997; Greenhagen *et al.*, 2006; O'Maille *et al.*, 2008). Overall, the sesquiterpene synthase TPS-a clade appears to be highly divergent in all seed plants charatcterized to date.

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

TPS-g is a clade closely related to TPS-b and was first defined by monoterpene synthases that produce the acyclic floral scent compounds myrcene and ocimene in snapdragon (Dudareva *et al.*, 2003). Two members of the TPS-g group identified and characterized in Arabidopsis (At1g61680; Chen *et al.*, 2003) and rice (Os02g02930; Yuan *et al.*, 2008), also produce an acyclic monoterpene, namely linalool synthases. Similarly, all members of the TPS-g group identified in grapevine also only produce acyclic mono-, sesqui-, and diterpene alcohols (Martin *et al.*, 2010). Thus a prominent feature of the members of the TPS-g group is the prevalence of acyclic products. A common structural feature of the members of the TPS-g subfamily is the lack of the RRX_8W motif which is highly conserved near the N-terminus of monoterpene synthases (mostly cyclases) of the angiosperm TPS-b clade and the gymnosperm TPS-d-1 clade.

Lineage-specific expansion in the TPS family and changes in subcellular localization and substrate specificity

The presence of the *S. moellendorffii*-specific TPS-h subfamily, the gymnosperm-specific TPS-d subfamily and the angiosperm-specific subfamilies of TPS-a, TPS-d and TPS-g indicates lineage-specific expansion of the TPS family (Figure 4). In angiosperms, the TPS-a subfamily is the major determinant of the size of the TPS family of individual species. Apparently, the expansion of the TPS-a family occurred after the split of the monocot and dicot lineages (Figure 4). Moreover, the positions of Arabidopsis *TPS* genes on the branches of clade TPS-a-1 indicate that many of them arose by gene duplications that occurred after the divergence of the *Arabidopsis* lineage from the *V. vinifera* and *P. trichocarpa* lineages. Another example of species-specific expansion of the *TPS* genes can be observed within the TPS-e/f subfamily. Among the six plant species that contain TPS-e/f genes, the rice genome contains nine members while the other five species have 1–3 genes (Table 1). The relatively large number of TPS-e/f genes in rice leads to the production of a suite of labdane-type diterpenoids that provide defense against pathogens (Otomo *et al.*, 2004b).

Building on the pioneering work of Joe Chappell, Rodney Croteau, and Charles A. West since the early nineties, the last decade has witnessed an enormous proliferation of functional characterization of *TPS* genes of plant specialized metabolism. The most recent acceleration of *TPS* gene discovery in well-travelled and remote places of the plant kingdom has been fueled by access to comprehensive genome and transcriptome sequence resources for a number of model and non-model systems. Just over 10 years ago our understanding of the plant *TPS* gene family was based on just over 30 genes, without a comprehensive coverage of a representative set of *TPS* genes in any one species (Bohlmann *et al.*, 1998a). However, even with these few genes, a general model for the description of the TPS family became apparent. For example, all monoterpene synthases were thought to be localized to plastids and to use GPP as substrate. Sesquiterpene synthases were thought to be generally localized to the cytosol where they use all-*trans* FPP as their susbtrate with the use of GPP occurring only *in vitro.* And finally, the larger diterpene synthases had their place in plastids with all-*trans* GGPP as their substrate.

While much of this general model still holds true, new discoveries continue to reveal new facets of the *TPS* gene family. Even with this initial model, it was clear that the TPS family possesses a remarkable flexibility to evolve enzymes with new subcellular localization and substrate specificity. For example, the topology of the phylogenetic tree (Figure 4) suggests that mono- and sesquiterpene synthases evolved from diterpene synthases independently in gymnosperms and angiosperms. A change from diterpene synthase to sesquiterpene synthase involves not only a change in substrate specificity but also in subcellular localization (by the loss of a transit peptide). A relative recent example of such an event may exist with the closely related diterpene synthases and sesquiterpene synthases of the gymnosperm TPS-d-3 clade. Here, the *bona fide* sesquiterpene synthase genes encoding (*E*)-α-bisabolene synthases in grand fir and Norway spruce (Bohlmann *et al.*, 1998b; Martin *et al.*, 2004) still share with the conifer diterpene synthases the sequence encoding the ancestral 200 amino acid motif, but have apparently lost the region encoding the N-terminal transit peptide. Similary, *Clarkia breweri* linalool synthase (Dudareva *et al.*, 1996), a monoterpene synthase that belongs to the TPS-f subfamily, evolved from a diterpene synthase, possibly a geranyllinalool synthase such as Arabidopsis AtTPS04, a change that in this case involved only a switch in substrate specificity. This enzyme also retained the 200 amino acid motif, linking it closely to a diterpene synthase ancestor.

Another good example of neofunctionalization of duplicated *TPS* genes involving a change in subcellular localization comes from the two *TPS* genes of the TPS-g subfamily that were isolated from snapdragon (Nagegowda *et al.*, 2008). These two genes, *AmNES/LIS-1* and *AmNES/LIS-2*, encode two TPS proteins that are nearly identical,

suggesting that they are the consequence of recent gene duplication. The two proteins showed similar catalytic properties *in vitro*, both synthesize linalool and nerolidol as specific products using GPP and FPP as substrate, respectively. However, AmNES/LIS-2 protein has an additional 30 amino acids at the N-terminus, causing the protein to be transported into the plastid, where it uses the available GPP to make linalool. In contrast, AmNES/LIS-1 lacks a transit peptide, and is localized in the cytosol where it uses the available FPP to make nerolidol. These results indicate that *in planta* AmNES/LIS-1 functions as a sesquiterpene synthase while AmNES/LIS-2 acts as a monoterpene synthase, suggesting that subcellular targeting of TPSs controls the *in planta* substrate they use and the type of terpene products formed. As described above, the recently duplicated Arabidopsis AtTPS02 and AtTPS03 genes represent a very similar case, where the former produces (*E*)-β-ocimene in the plastid and the transit peptide-lacking AtTPS03 produces (*E,E*)-α-farnesene in the cytosol, although both enzymes can synthesize both compounds (from the respective substrates) *in vitro* (Huang *et al.*, 2010).

Recent work based on the characterization of trichome transcriptomes in tomato revealed other surprising evolutionary diversifications in the TPS family. A *Solanum lycopersicum* (cultivated tomato) TPS enzyme, β-phellandrene synthase (PHS1), belongs to the diterpene synthase clade TPS-e/f but it uses the unusual substrate neryl diphosphate, the *cis*-isomer of GPP, to make mostly β-phellandrene as well as a few other monoterpenes in the plastids of the glandular trichomes on the leaf and stem surfaces (Schilmiller *et al.*, 2009). Even more surprisingly, its closely related enzyme in *Solanum habrochaites* (89% identical) uses *Z,Z*-FPP to make two sesquiterpenes, bergamotene and santalene, in the plastids of the trichomes (Sallaud *et al.*, 2009).

The TPS-a clade also appears to be more diverse with regard to subcellular targeting and substrates of its enzymes than what was initially thought. Several angiosperm species have substantial proportions of their *TPS* genes in this subfamily, some with apparent transit peptide (including 18 uncharacterized Arabidopsis TPS proteins) and some without. Most biochemically tested enzymes of this subfamily are sesquiterpene synthases but the diterpene synthase casbene synthase (Mau and West, 1994) was an early recognized member, suggesting that alternative substrate specificities may have evolved within the TPS-a subfamily (Bohlmann *et al.*, 1998a). This pattern has been substantiated with the molecular characterization of several *Nicotiana sylvestris TPS* genes of the TPS-a subfamily. While direct biochemical characterization for the *N. sylvestris* proteins remains to be done, Ennajdaoui *et al.* (2010) showed that when the trichome-specific expression of these genes is blocked, the production of the diterpene cembratrien-ol is diminished. Consistent with a role in diterpene formation,

the predicted TPS proteins appear to have plastid transit peptides.

The above mentioned new findings highlight the fascinating potential of the *TPS* gene family to evolve surprising variations of biochemical functions and subcellular localization, and strongly establish that substrate specificity, product profiles and localization of a *TPS* gene cannot be predicted based on association with a specific TPS subfamily (i.e., gross structure indicating evolutionary relationships) or general sequence similarity. As different subfamilies expand in different lineages by gene duplication and divergence, as has happened for example in the TPS-a, TPS-b, and TPS-g in angiosperms, TPS-d in gymnosperms, and TPS-h in *S. moellendorffii*, it can be expected that proteins with altered subcellular localization and new substrate specificities would have evolved.

FUTURE DIRECTIONS

Along with other types of plant specialized metabolites, terpenoids play important roles in plant interactions with the environment (Gershenzon and Dudareva, 2007). Understanding the function and evolution of the TPS family can provide important insights into species-specific adaptations to unique niches. To better understand the physiological and ecological roles of specific *TPS* genes and enzymes will require research in several areas. Roles of specific terpenes or general roles of classes of terpenes, a topic only slightly covered here, must be examined *in planta* and ideally in the natural environments of the plants that produce these terpenes. The manipulation of *TPS* gene expression in model and non-model plants will be critical to this end. Despite the many major accomplishments since the first elucidation of the function of a plant *TPS* gene (Facchini and Chappell, 1992) and the three-dimensional structure of a plant TPS protein (Starks *et al.*, 1997), the ongoing and future structural and biochemical investigation of TPS enzymes will continue to be a field of exciting new discoveries. At present, functional characterization has been completed only for subsets of the TPS families in any one species, including Arabidopsis, grapevine and spruce, the three species for which, relative to other species, the biochemical functions of a substantial number of TPS proteins are already known, and there are only a few experimentally determined three-dimensional structures for TPS of plant origin (Gennadios *et al.*, 2009; Hyatt *et al.*, 2007; Kampranis *et al.*, 2007; Köksal *et al.*, 2010a,b; Starks *et al.*, 1997; Whittington *et al.*, 2002). Functional and structural characterization of complete TPS families in selected model plant species such as Arabidopsis and rice, as well as plants occupying key position in the plant phylogeny, such as the conifers and *S. moellendorffii*, will provide insights into the array of terpenes that a single species can synthesize, and how its enzymes evolved the capacity to do so. At the same time, large transcriptome sequencing projects targeted at plant species that produce

interesting terpenoid metabolites will accelerate the identi-fication of comprehensive sets of *TPS* genes in a large variety of non-model systems. An improved knowledge base on plant terpenoid metabolism will facilitate the manipulation of the terpene pathway for changing agro-nomic traits such as fruit flavors (Lewinsohn *et al.*, 2001; Davidovich-Rikanati *et al.*, 2007), floral scent (Lücker *et al.*, 2001), plant defense against insects (Schnee *et al.*, 2006) and high level production of known and novel biochemicals (Bohlmann and Keeling, 2008; Kirby and Keasling, 2009).

ACKNOWLEDGEMENTS

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article:
Data S1. Experimental procedures.
Figure S1. A rooted phylogenetic tree of plant TPSs.
Table S1. TPS proteins represented in phylogenetic analysis.
Please note: As a service to our authors and readers, this journal provides supporting information supplied by the authors. Such materials are peer-reviewed and may be re-organized for online delivery, but are not copy-edited or typeset. Technical support issues arising from supporting information (other than missing files) should be addressed to the authors.

REFERENCES

Abbott, E., Hall, D., Hamberger, B. and Bohlmann, J. (2010) Laser microdis-section of conifer stem tissues: isolation and analysis of high quality RNA, terpene synthase enzyme activity and terpenoid metabolites from resin ducts and cambial zone tissue of white spruce (*Picea glauca*). *BMC Plant Biol.* **10**, 106.

Abel, C., Clauss, M., Schaub, A., Gershenzon, J. and Tholl, D. (2009) Floral and insect-induced volatile formation in *Arabidopsis lyrata* ssp. *petraea*, a perennial, outcrossing relative of *A. thaliana*. *Planta*, **230**, 1–11.

Aharoni, A., Giri, A.P., Deuerlein, S. *et al.* (2003) Terpenoid metabolism in wild-type and transgenic *Arabidopsis* plants. *Plant Cell*, **15**, 2866–2884.

Anterola, A., Shanle, E., Mansouri, K., Schuette, S. and Renzaglia, K. (2009) Gibberellin precursor is involved in spore germination in the moss *Physcomitrella patens*. *Planta*, **229**, 1003–1007.

Arabidopsis Genome Initiative (2000) Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature*, **408**, 796–815.

Arimura, G., Huber, D.P.W. and Bohlmann, J. (2004) Forest tent caterpillars (*Malacosoma disstria*) induce local and systemic diurnal emissions of ter-penoid volatiles in hybrid poplar (*Populus trichocarpa* × deltoides): cDNA cloning, functional characterization, and patterns of gene expression of (−)-germacrene D synthase PtdTPS1. *Plant J.* **37**, 603–616.

Aubourg, S., Lecharny, A. and Bohlmann, J. (2002) Genomic analysis of the terpenoid synthase (AtTPS) gene family of *Arabidopsis thaliana*. *Mol. Genet. Genomics*, **267**, 730–745.

Behnke, K., Ehlting, B., Teuber, M. *et al.* (2007) Transgenic, non-isoprene emitting poplars don't like it hot. *Plant J.* **51**, 485–499.

Birnbaum, K., Shasha, D.E., Wang, J.Y., Jung, J.W., Lambert, G.M., Galbraith, D.W. and Benfey, P.N. (2003) A gene expression map of the *Arabidopsis* root. *Science*, **302**, 1956–1960.

Bohlmann, J. and Keeling, C.I. (2008) Terpenoid biomaterials. *Plant J.* **54**, 656–669.

Bohlmann, J., Steele, C.L. and Croteau, R. (1997) Monoterpene synthases from grand fir (*Abies grandis*). cDNA isolation, characterization, and functional expression of myrcene synthase, (−)-(4S)-limonene synthase, and (−)-(1S,5S)-pinene synthase. *J. Biol. Chem.* **272**, 21784–21792.

Bohlmann, J., Meyer-Gauen, G. and Croteau, R. (1998a) Plant terpenoid synthases: molecular biology and phylogenetic analysis. *Proc. Natl Acad. Sci. USA*, **95**, 4126–4133.

Bohlmann, J., Crock, J., Jetter, R. and Croteau, R. (1998b) Terpenoid-based defenses in conifers: cDNA cloning, characterization, and functional expression of wound-inducible (E)-alpha-bisabolene synthase from grand fir (*Abies grandis*). *Proc. Natl Acad. Sci. USA*, **95**, 6756–6761.

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

Bohlmann, J., Martin, D., Oldham, N.J. and Gershenzon, J. (2000) Terpenoid secondary metabolism in *Arabidopsis thaliana*: cDNA cloning, character-ization, and functional expression of a myrcene/(E)-beta-ocimene synthase. *Arch. Biochem. Biophys.* **375**, 261–269.

Brady, S.M., Orlando, D.A., Lee, J.Y. *et al.* (2007) A high-resolution root spa-tiotemporal map reveals dominant expression patterns. *Science*, **318**, 801–806.

Cao, R., Zhang, Y., Mann, F.M. *et al.* (2010) Diterpene cyclases and the nature of the isoprene fold. *Proteins*, **78**, 2417–2432.

Chen, F., Tholl, D., D'Auria, J.C., Farooq, A., Pichersky, E. and Gershenzon, J. (2003) Biosynthesis and emission of terpenoid volatiles from *Arabidopsis* flowers. *Plant Cell*, **15**, 481–494.

Chen, F., Ro, D.K., Petri, J., Gershenzon, J., Bohlmann, J., Pichersky, E. and Tholl, D. (2004) Characterization of a root-specific *Arabidopsis* terpene synthase responsible for the formation of the volatile monoterpene 1,8-cineole. *Plant Physiol.* **135**, 1956–1966.

Cho, E.M., Okada, A., Kenmoku, H. *et al.* (2004) Molecular cloning and char-acterization 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 J.* **37**, 1–8.

Christianson, D.W. (2006) Structural biology and chemistry of the terpenoid cyclases. *Chem. Rev.* **106**, 3412–3442.

Christianson, D.W. (2008) Unearthing the roots of the terpenome. *Curr. Opin. Chem. Biol.* **12**, 141–150.

Colby, S.M., Alonso, W.R., Katahira, E.J., McGarvey, D.J. and Croteau, R. (1993) 4S-Limonene synthase from the oil glands of spearmint (*Mentha spicata*). cDNA isolation, characterization, and bacterial expression of the catalytically active monoterpene cyclase. *J. Biol. Chem.* **268**, 23016–23024.

Croteau, R., Ketchum, R.E., Long, R.M., Kaspera, R. and Wildung, M.R. (2006) Taxol biosynthesis and molecular genetics. *Phytochem. Rev.* **5**, 75–97.

Davidovich-Rikanati, R., Sitrit, Y., Tadmor, Y. *et al.* (2007) Enrichment of tomato flavor by diversion of the early plastidial terpenoid pathway. *Nat. Biotechnol.* **25**, 899–901.

Degenhardt, J., Köllner, T.G. and Gershenzon, J. (2009a) Monoterpene and sesquiterpene synthases and the origin of terpene skeletal diversity in plants. *Phytochemistry*, **70**, 1621–1637.

Degenhardt, J., Hiltpold, I., Köllner, T.G. *et al.* (2009b) Restoring a maize root signal that attracts insect-killing nematodes to control a major pest. *Proc. Natl Acad. Sci. USA*, **106**, 13213–13218.

Dudareva, N., Cseke, L., Blanc, V.M. and Pichersky, E. (1996) Evolution of floral scent in Clarkia: novel patterns of *S*-linalool synthase gene expression in the *C. breweri* flower. *Plant Cell*, **8**, 1137–1148.

Dudareva, N., Martin, D., Kish, C.M., Kolosova, N., Gorenstein, N., Fäldt, J., Miller, B. and Bohlmann, J. (2003) (E)-beta-ocimene and myrcene synthase genes of floral scent biosynthesis in snapdragon: function and expression

of three terpene synthase genes of a new terpene synthase subfamily. *Plant Cell*, **15**, 1227–1241.

Ennajdaoui, H., Vachon, G., Giacalone, C., Besse, I., Sallaud, C., Herzog, M. and Tissiers, A. (2010) Trichome specific expression of the tobacco (*Nicotiana sylvestris*) cembratrien-ol synthase genes is controlled by both activating and repressing *cis*-regions. *Plant Mol. Biol.* **73**, 673–685.

Facchini, P.J. and Chappell, J. (1992) Gene family for an elicitor-induced sesquiterpene cyclase in tobacco. *Proc. Natl Acad. Sci. USA*, **89**, 11088–11092.

Falara, V., Pichersky, E. and Kanellis, A.K. (2010) A copal-8-ol diphosphate synthase from the angiosperm *Cistus creticus* subsp. creticus is a putative key enzyme for the formation of pharmacologically active, oxygen-containing labdane-type diterpenes. *Plant Physiol.* **154**, 301–310.

Fäldt, J., Martin, D., Miller, B., Rawat, S. and Bohlmann, J. (2003a) Traumatic resin defense in Norway spruce (*Picea abies*): methyl jasmonate-induced terpene synthase gene expression, and cDNA cloning and functional characterization of (+)-3-carene synthase. *Plant Mol. Biol.* **51**, 119–133.

Fäldt, J., Arimura, G., Gershenzon, J., Takabayashi, J. and Bohlmann, J. (2003b) Functional identification of AtTPS03 as (*E*)-beta-ocimene synthase: a monoterpene synthase catalyzing jasmonate- and wound-induced volatile formation in *Arabidopsis thaliana*. *Planta*, **216**, 745–751.

Fraenkel, G.F. (1959) The raison d'être of secondary plant substances. *Science*, **129**, 1466–1470.

Frost, C.J., Mescher, M.C., Dervinis, C., Davis, J.M., Carlson, J.E. and De Moraes, C.M. (2008) Priming defense genes and metabolites in hybrid poplar by the green leaf volatile *cis*-3-hexenyl acetate. *New Phytol.* **180**, 22–34.

Gennadios, H.A., Gonzalez, V., Di Costanzo, L., Li, A., Yu, F., Miller, D.J., Allemann, R.K. and Christianson, D.W. (2009) Crystal structure of (+)-delta-cadinene synthase from *Gossypium arboreum* and evolutionary divergence of metal binding motifs for catalysis. *Biochemistry*, **48**, 6175–6183.

Gershenzon, J. and Dudareva, N. (2007) The function of terpene natural products in the natural world. *Nat. Chem. Biol.* **3**, 408–414.

Goff, S.A., Ricke, D., Lan, T.H. *et al.* (2002) A draft sequence of the rice genome (*Oryza sativa* L. ssp. japonica). *Science*, **296**, 92–100.

Greenhagen, B.T., O'Maille, P.E., Noel, J.P. and Chappell, J. (2006) Identifying and manipulating structural determinates linking catalytic specificities in terpene synthases. *Proc. Natl Acad. Sci. USA*, **103**, 9826–9831.

Hall, D.E., Robert, J.A., Keeling, C.I., Domanski, D., Qesada, A.L., Jancsik, S., Kuzyk, M., Hamberger, B., Borchers, C.H. and Bohlmann, J. (2011) An integrated genomic, proteomic, and biochemical analysis of (+)-3-carene biosynthesis in Sitka spruce (*Picea sitchensis*) genotypes which are resistant or susceptible to white pine weevil. *Plant J.* **65**, 936–948.

Hartmann, T. (2007) From waste products to ecochemicals: fifty years research of plant secondary metabolism. *Phytochemistry*, **68**, 2831–2846.

Hayashi, K., Kawaide, H., Notomi, M., Sakigi, Y., Matsuo, A. and Nozaki, H. (2006) Identification and functional analysis of bifunctional ent-kaurene synthase from the moss *Physcomitrella patens*. *FEBS Lett.* **580**, 6175–6181.

Hayashi, K., Horie, K., Hiwatashi, Y. *et al.* (2010) Endogenous diterpenes derived from ent-kaurene, a common gibberellin precursor, regulate protonema differentiation of the moss *Physcomitrella patens*. *Plant Physiol.* **153**, 1085–1097.

Heiling, S., Schuman, M.C., Schoettner, M., Mukerjee, P., Berger, B., Schneider, B., Jassbi, A.R. and Baldwin, I.T. (2010) Jasmonate and ppHsystemin regulate key malonylation steps in the biosynthesis of 17-hydroxygeranyllinalool diterpene glycosides, an abundant and effective direct defense against herbivores in *Nicotiana attenuata*. *Plant Cell*, **22**, 273–292.

Herde, M., Gärtner, K., Köllner, T.G., Fode, B., Boland, W., Gershenzon, J., Gatz, C. and Tholl, D. (2008) Identification and regulation of TPS04/GES, an *Arabidopsis* geranyllinalool synthase catalyzing the first step in the formation of the insect-induced volatile C16-homoterpene TMTT. *Plant Cell*, **20**, 1152–1168.

Huang, M., Abel, C., Sohrabi, R., Petri, J., Haupt, I., Cosimano, J., Gershenzon, J. and Tholl, D. (2010) Variation of herbivore-induced volatile terpenes among *Arabidopsis* ecotypes depends on allelic differences and subcellular targeting of two terpene synthases, TPS02 and TPS03. *Plant Physiol.* **153**, 1293–1310.

Hyatt, D.C., Youn, B., Zhao, Y., Santhamma, B., Coates, R.M., Croteau, R. and Kang, C. (2007) Structure of limonene synthase, a simple model for terpenoid cyclase catalysis. *Proc. Natl Acad. Sci. USA*, **104**, 5360–5365.

Jaillon, O., Aury, J.M., Noel, B. *et al.* (2007) The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. *Nature*, **449**, 463–467.

Kampranis, S.C., Ioannidis, D., Purvis, A. *et al.* (2007) Rational conversion of substrate and product specificity in a *Salvia* monoterpene synthase: structural insights into the evolution of terpene synthase function. *Plant Cell*, **19**, 1994–2005.

Kanno, Y., Otomo, K., Kenmoku, H. *et al.* (2006) Characterization of a rice gene family encoding type-A diterpene cyclases. *Biosci. Biotechnol. Biochem.* **70**, 1702–1710.

Kato, H., Kodama, O. and Akatsuka, T. (1994) Oryzalexin F, a diterpene phytoalexin from UV-irradiated rice leaves. *Phytochemistry*, **36**, 299–301.

Kato-Noguchi, H. and Ino, T. (2003) Rice seedlings release momilactone B into the environment. *Phytochemistry*, **63**, 551–554.

Keeling, C.I. and Bohlmann, J. (2006a) Genes, enzymes and chemicals of terpenoid diversity in the constitutive and induced defence of conifers against insects and pathogens. *New Phytol.* **170**, 657–675.

Keeling, C.I. and Bohlmann, J. (2006b) Diterpene resin acids in conifers. *Phytochemistry*, **67**, 2415–2423.

Keeling, C.I., Weisshaar, S., Lin, R.P. and Bohlmann, J. (2008) Functional plasticity of paralogous diterpene synthases involved in conifer defense. *Proc. Natl Acad. Sci. USA*, **105**, 1085–1090.

Keeling, C.I., Dullat, H.K., Yuen, M., Ralph, S.G., Jancsik, S. and 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 Physiol.* **152**, 1197–1208.

Keeling, C.I., Weißhaar, S., Ralph, S.G., Jancsik, S., Hamberger, B., Dullat, H.K. and Bohlmann, J. (2011) Transcriptome mining, functional characterization, and phylogeny of a large terpene synthase gene family in spruce (*Picea* spp.). *BMC Plant Biol.*, in press.

Kirby, J. and Keasling, J.D. (2009) Biosynthesis of plant isoprenoids: perspectives for microbial engineering. *Annu. Rev. Plant Biol.* **60**, 335–355.

Köksal, M., Zimmer, I., Schnitzler, J.P. and Christianson, D.W. (2010a) Structure of isoprene synthase illuminates the chemical mechanism of teragram atmospheric carbon emission. *J. Mol. Biol.* **402**, 363–373.

Köksal, M., Jin, Y., Coates, R.M., Croteau, R. and Christianson, D.W. (2010b) Taxadiene synthase structure and evolution of modular architecture in terpene biosynthesis. *Nature*, **469**, 116–120.

Köllner, T.G., Held, M., Lenk, C., Hiltpold, I., Turlings, T.C., Gershenzon, J. and Degenhardt, J. (2008) A maize (*E*)-beta-caryophyllene synthase implicated in indirect defense responses against herbivores is not expressed in most American maize varieties. *Plant Cell*, **20**, 482–494.

Köllner, T.G., Gershenzon, J. and Degenhardt, J. (2009) Molecular and biochemical evolution of maize terpene synthase 10, an enzyme of indirect defense. *Phytochemistry*, **70**, 1139–1145.

Lee, S. and Chappell, J. (2008) Biochemical and genomic characterization of terpene synthases in *Magnolia grandiflora*. *Plant Physiol.* **147**, 1017–1033.

Lee, S., Badieyan, S., Bevan, D.R., Herde, M., Gatz, C. and Tholl, D. (2010) Herbivore-induced and floral homoterpene volatiles are biosynthesized by a single P450 enzyme (CYP82G1) in *Arabidopsis*. *Proc. Natl Acad. Sci. USA*, **107**, 21205–21210.

Lewinsohn, E., Schalechet, F., Wilkinson, J. *et al.* (2001) Enhanced levels of the aroma and flavor compound S-linalool by metabolic engineering of the terpenoid pathway in tomato fruits. *Plant Physiol.* **127**, 1256–1265.

Lichtenthaler, H.K. (1999) The 1-deoxy-D-xylulose 5-phosphate pathway of isoprenoid biosynthesis in plants. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **50**, 47–65.

Loivamäki, M., Mumm, R., Dicke, M. and Schnitzler, J.P. (2008) Isoprene interferes with the attraction of bodyguards by herbaceous plants. *Proc. Natl Acad. Sci. USA*, **105**, 17430–17435.

van Loon, J.J.A., de Boer, J.G. and Dicke, M. (2000) Parasitoid–plant mutualism: parasitoid attack of herbivore increases plant reproduction. *Entomol. Exp. Appl.* **97**, 219–227.

Lücker, J., Bouwmeester, H.J., Schwab, W., Blaas, J., van der Plas, L.H. and Verhoeven, H.A. (2001) Expression of Clarkia S-linalool synthase in transgenic petunia plants results in the accumulation of S-linalyl-β-D-glucopyranoside. *Plant J.* **27**, 315–324.

Lücker, J., Bowen, P. and Bohlmann, J. (2004) *Vitis vinifera* terpenoid cyclases: functional identification of two sesquiterpene synthase cDNAs encoding (+)-valencene synthase and (–)-germacrene D synthase and

expression of mono- and sesquiterpene synthases in grapevine flowers and berries. *Phytochemistry*, **65**, 2649–2659.

Lund, S.T. and Bohlmann, J. (2006) The molecular basis for wine grape quality – a volatile subject. *Science*, **311**, 804–805.

Martin, D.M. and Bohlmann, J. (2004) Identification of *Vitis vinifera* (–)-alpha-terpineol synthase by *in silico* screening of full-length cDNA ESTs and functional characterization of recombinant terpene synthase. *Phytochemistry*, **65**, 1223–1229.

Martin, D.M., Fäldt, J. and Bohlmann, J. (2004) Functional characterization of nine Norway spruce *TPS* genes and evolution of gymnosperm terpene synthases of the TPS-d subfamily. *Plant Physiol.* **135**, 1908–1927.

Martin, D.M., Toub, O., Chiang, A., Lo, B.C., Ohse, S., Lund, S.T. and Bohlmann, J. (2009) The bouquet of grapevine (*Vitis vinifera* L. cv. Cabernet Sauvignon) flowers arises from the biosynthesis of sesquiterpene volatiles in pollen grains. *Proc. Natl Acad. Sci. USA*, **106**, 7245–7250.

Martin, D.M., Aubourg, S., Schouwey, M.B., Daviet, L., Schalk, M., Toub, O., Lund, S.T. and Bohlmann, J. (2010) Functional annotation, genome organization and phylogeny of the grapevine (*Vitis vinifera*) terpene synthase gene family based on genome assembly, FLcDNA cloning, and enzyme assays. *BMC Plant Biol.* **10**, 226.

Mau, C.J. and West, C.A. (1994) Cloning of casbene synthase cDNA: evidence for conserved structural features among terpenoid cyclases in plants. *Proc. Natl Acad. Sci. USA*, **91**, 8497–8501.

McCaskill, D.J. and Croteau, R. (1997) Prospects for bioengineering of isoprenoid biosynthesis. In *Advances in Biochemical Engineering/Biotechnology* (Berger, R.G., de Bont, J.A.M., Cheetham, P.S.J. and Croteau, R., eds). Berlin: Springer-Verlag, pp. 107–146.

McGarvey, D.J. and Croteau, R. (1995) Terpenoid metabolism. *Plant Cell*, **7**, 1015–1026.

Miller, B., Oschinski, C. and Zimmer, W. (2001) First isolation of an isoprene synthase gene from poplar and successful expression of the gene in *Escherichia coli*. *Planta*, **213**, 483–487.

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

Morrone, D., Jin, Y.H., Xu, M.M., Choi, S.Y., Coates, R.M. and Peters, R.J. (2006) An unexpected diterpene cyclase from rice: functional identification of a stemodene synthase. *Arch. Biochem. Biophys.* **448**, 133–140.

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

Mumm, R. and Hilker, M. (2006) Direct and indirect chemical defence of pine against folivorous insects. *Trends Plant Sci.* **11**, 351–358.

Nagegowda, D.A., Gutensohn, M., Wilkerson, C.G. and Dudareva, N. (2008) Two nearly identical terpene synthases catalyze the formation of nerolidol and linalool in snapdragon flowers. *Plant J.* **55**, 224–239.

Nemoto, T., Cho, E.M., Okada, A. *et al.* (2004) Stemar-13-ene synthase, a diterpene cyclase involved in the biosynthesis of the phytoalexin oryzalexin S in rice. *FEBS Lett.* **571**, 182–186.

O'Maille, P.E., Malone, A., Dellas, N. *et al.* (2008) Quantitative exploration of the catalytic landscape separating divergent plant sesquiterpene synthases. *Nat. Chem. Biol.* **4**, 617–623.

Otomo, K., Kenmoku, H., Oikawa, H. *et al.* (2004a) Biological functions of ent- and syn-copalyl diphosphate synthases in rice: key enzymes for the branch point of gibberellin and phytoalexin biosynthesis. *Plant J.* **39**, 886–893.

Otomo, K., Kanno, Y., Motegi, A. *et al.* (2004b) Diterpene cyclases responsible for the biosynthesis of phytoalexins, momilactones A, B, and oryzalexins A–F in rice. *Biosci. Biotechnol. Biochem.* **68**, 2001–2006.

Paterson, A.H., Bowers, J.E., Bruggmann, R. *et al.* (2009) The *Sorghum bicolor* genome and the diversification of grasses. *Nature*, **457**, 551–556.

Peters, R.J. (2006) Uncovering the complex metabolic network underlying diterpenoid phytoalexin biosynthesis in rice and other cereal crop plants. *Phytochemistry*, **67**, 2307–2317.

Peters, R.J., Flory, J.E., Jetter, R., Ravn, M.M., Lee, H.J., Coates, R.M. and Croteau, R. (2000) Abietadiene synthase from grand fir (*Abies grandis*): characterization and mechanism of action of the 'pseudomature' recombinant enzyme. *Biochemistry*, **39**, 15592–15602.

Peters, R.J., Ravn, M.M., Coates, R.M. and Croteau, R. (2001) Bifunctional abietadiene synthase: free diffusive transfer of the (+)-copalyl diphosphate intermediate between two distinct active sites. *J. Am. Chem. Soc.* **123**, 8974–8978.

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

Pichersky, E. and Gang, D.R. (2000) Genetics and biochemistry of secondary metabolites in plants: an evolutionary perspective. *Trends Plant Sci.* **5**, 439–445.

Pichersky, E. and Gershenzon, J. (2002) The formation and function of plant volatiles: perfumes for pollinator attraction and defense. *Curr. Opin. Plant Biol.* **5**, 237–243.

van Poecke, R.M., Posthumus, M.A. and Dicke, M. (2001) Herbivore-induced volatile production by *Arabidopsis thaliana* leads to attraction of the parasitoid *Cotesia rubecula:* chemical, behavioral, and gene-expression analysis. *J. Chem. Ecol.* **27**, 1911–1928.

Prisic, S., Xu, M.M., Wilderman, P.R. and Peters, R.J. (2004) Rice contains two disparate ent-copalyl diphosphate synthases with distinct metabolic functions. *Plant Physiol.* **136**, 4228–4236.

Raguso, R.A. (2009) Floral scent in a whole-plant context: moving beyond pollinator attraction. *Funct. Ecol.* **23**, 837–840.

Ralph, S.G., Chun, H.J.E., Kolosova, N. *et al.* (2008) A conifer genomics resource of 200,000 spruce (*Picea* spp.) ESTs and 6,464 high-quality, sequence-finished full-length cDNAs for Sitka spruce (*Picea sitchensis*). *BMC Genomics*, **9**, 484.

Rensing, S.A., Lang, D., Zimmer, A.D. *et al.* (2008) The *Physcomitrella* genome reveals evolutionary insights into the conquest of land by plants. *Science*, **319**, 64–69.

Ro, D.K., Arimura, G., Lau, S.Y., Piers, E. and Bohlmann, J. (2005) Loblolly pine abietadienol/abietadienal oxidase PtAO (CYP720B1) is a multifunctional, multisubstrate cytochrome P450 monooxygenase. *Proc. Natl Acad. Sci. USA*, **102**, 8060–8065.

Ro, D.K., Ehlting, J., Keeling, C.I., Lin, R., Mattheus, N. and Bohlmann, J. (2006) Microarray expression profiling and functional characterization of *AtTPS* genes: Duplicated *Arabidopsis thaliana* sesquiterpene synthase genes At4g13280 and At4g13300 encode root-specific and wound-inducible (Z)-γ-bisabolene synthases. *Arch. Biochem. Biophys.* **448**, 104–116.

Rohloff, J. and Bones, A.M. (2005) Volatile profiling of *Arabidopsis thaliana* – putative olfactory compounds in plant communication. *Phytochemistry*, **66**, 1941–1955.

Sakamoto, T., Miura, K., Itoh, H. *et al.* (2004) An overview of gibberellin metabolism enzyme genes and their related mutants in rice. *Plant Physiol.* **134**, 1642–1653.

Sallaud, C., Rontein, D., Onillon, S. *et al.* (2009) A novel pathway for sesquiterpene biosynthesis from Z,Z-farnesyl pyrophosphate in the wild tomato *Solanum habrochaites*. *Plant Cell*, **21**, 301–317.

Sapir-Mir, M., Mett, A., Belausov, E., Tal-Meshulam, S., Frydman, A., Gidoni, D. and Eyal, Y. (2008) Peroxisomal localization of *Arabidopsis* isopentenyl diphosphate isomerases suggests that part of the plant isoprenoid mevalonic acid pathway is compartmentalized to peroxisomes. *Plant Physiol.* **148**, 1219–1228.

Sasaki, K., Ohara, K. and Yazaki, K. (2005) Gene expression and characterization of isoprene synthase from *Populus alba*. *FEBS Lett.* **579**, 2514–2518.

Schilmiller, A.L., Schauvinhold, I., Larson, M. *et al.* (2009) Monoterpenes in the glandular trichomes of tomato are synthesized from a neryl diphosphate precursor rather than geranyl diphosphate. *Proc. Natl Acad. Sci. USA*, **106**, 10865–10870.

Schnee, C., Köllner, T.G., Gershenzon, J. and Degenhardt, J. (2002) The maize gene terpene synthase 1 encodes a sesquiterpene synthase catalyzing the formation of (*E*)-beta-farnesene, (*E*)-nerolidol, and (*E,E*)-farnesol after herbivore damage. *Plant Physiol.* **130**, 2049–2060.

Schnee, C., Köllner, T.G., Held, M., Turlings, T.C., Gershenzon, J. and Degenhardt, J. (2006) The products of a single maize sesquiterpene synthase form a volatile defense signal that attracts natural enemies of maize herbivores. *Proc. Natl Acad. Sci. USA*, **103**, 1129–1134.

von Schwartzenberg, K., Schultze, W. and Kassner, H. (2004) The moss *Physcomitrella patens* releases a tetracyclic diterpene. *Plant Cell Rep.* **22**, 780–786.

Sharkey, T.D. and Yeh, S. (2001) Isoprene emission from plants. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **52**, 407–436.

Snoeren, T.A., Kappers, I.F., Broekgaarden, C., Mumm, R., Dicke, M. and Bouwmeester, H.J. (2010) Natural variation in herbivore-induced volatiles in *Arabidopsis thaliana. J. Exp. Bot.* **61**, 3041–3056.

Starks, C.M., Back, K., Chappell, J. and Noel, J.P. (1997) Structural basis for cyclic terpene biosynthesis by tobacco 5-epi-aristolochene synthase. *Science*, **277**, 1815–1820.

Steele, C.L., Crock, J., Bohlmann, J. and Croteau, R. (1998) Sesquiterpene synthases from grand fir (*Abies grandis*). Comparison of constitutive and wound-induced activities, and cDNA isolation, characterization, and bacterial expression of delta-selinene synthase and gamma-humulene synthase. *J. Biol. Chem.* **273**, 2078–2089.

Sun, T.P. and Kamiya, Y. (1994) The *Arabidopsis* GA1 locus encodes the cyclase ent-kaurene synthetase A of gibberellin biosynthesis. *Plant Cell*, **6**, 1509–1518.

Swiegers, J.H., Bartowsky, E.J., Henschke, P.A. and Pretorius, I.S. (2005) Yeast and bacterial modulation of wine aroma and flavour. *Aust. J. Grape Wine Res.* **11**, 139–173.

Takahashi, S. and Koyama, T. (2006) Structure and function of cis-prenyl chain elongating enzymes. *Chem. Rec.* **6**, 194–205.

Tantillo, D.J. (2010) The carbocation continuum in terpene biosynthesis – where are the secondary cations? *Chem. Soc. Rev.* **39**, 2847–2854.

Tholl, D. (2006) Terpene synthases and the regulation, diversity and biological roles of terpene metabolism. *Curr. Opin. Plant Biol.* **9**, 297–304.

Tholl, D., Chen, F., Petri, J., Gershenzon, J. and Pichersky, E. (2005) Two sesquiterpene synthases are responsible for the complex mixture of sesquiterpenes emitted from *Arabidopsis* flowers. *Plant J.* **42**, 757–771.

Toyomasu, T. (2008) Recent advances regarding diterpene cyclase genes in higher plants and fungi. *Biosci. Biotechnol. Biochem.* **72**, 1168–1175.

Toyomasu, T., Kawaide, H., Ishizaki, A., Shinoda, S., Otsuka, M., Mitsuhashi, W. and Sassa, T. (2000) Cloning of a full-length cDNA encoding ent-kaurene synthase from *Gibberella fujikuroi:* functional analysis of a bifunctional diterpene cyclase. *Biosci. Biotechnol. Biochem.* **64**, 660–664.

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

Tuskan, G.A., Difazio, S., Jansson, S. *et al.* (2006) The genome of black cottonwood, *Populus trichocarpa* (Torr. & Gray). *Science*, **313**, 1596–1604.

Unsicker, S.B., Kunert, G. and Gershenzon, J. (2009) Protective perfumes: the role of vegetative volatiles in plant defense against herbivores. *Curr. Opin. Plant Biol.* **12**, 479–485.

Vogel, B.S., Wildung, M.R., Vogel, G. and Croteau, R.B. (1996) Abietadiene synthase from grand fir (*Abies grandis*). cDNA isolation, characterization,

and bacterial expression of a bifunctional diterpene cyclase involved in resin acid biosynthesis. *J. Biol. Chem.* **271**, 23262–23268.

Wang, K.C. and Ohnuma, S.I. (2000) Isoprenyl diphosphate synthases. *Biochim. Biophys. Acta*, **1529**, 33–48.

Whittington, D.A., Wise, M.L., Urbansky, M., Coates, R.M., Croteau, R. and Christianson, D.W. (2002) Bornyl diphosphate synthase: structure and strategy for carbocation manipulation by a terpenoid cyclase. *Proc. Natl Acad. Sci. USA*, **99**, 15375–15380.

Wilderman, P.R., Xu, M.M., Jin, Y.H., Coates, R.M. and 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 Physiol.* **135**, 2098–2105.

Wildung, M.R. and Croteau, R. (1996) A cDNA clone for taxadiene synthase, the diterpene cyclase that catalyzes the committed step of taxol biosynthesis. *J. Biol. Chem.* **271**, 9201–9204.

Xu, M.M., Hillwig, M.L., Prisic, S., Coates, R.M. and 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 J.* **39**, 309–318.

Xu, M.M., Wilderman, P.R., Morrone, D. *et al.* (2007) Functional characterization of the rice kaurene synthase-like gene family. *Phytochemistry*, **68**, 312–326.

Yamaguchi, S. (2008) Gibberellin metabolism and its regulation. *Annu. Rev. Plant Biol.* **59**, 225–251.

Yamaguchi, S., Sun, T., Kawaide, H. and Kamiya, Y. (1998) The GA2 locus of *Arabidopsis thaliana* encodes ent-kaurene synthase of gibberellin biosynthesis. *Plant Physiol.* **116**, 1271–1278.

Yoshikuni, Y., Ferrin, T.E. and Keasling, J.D. (2006) Designed divergent evolution of enzyme function. *Nature*, **440**, 1078–1082.

Yuan, J.S., Köllner, T.G., Wiggins, G., Grant, J., Degenhardt, J. and Chen, F. (2008) Molecular and genomic basis of volatile-mediated indirect defense against insects in rice. *Plant J.* **55**, 491–503.

Zulak, K.G., Lippert, D.N., Kuzyk, M.A., Domanski, D., Chou, T., Borchers, C.H. and Bohlmann, J. (2009) Targeted proteomics using selected reaction monitoring reveals the induction of specific terpene synthases in a multilevel study of methyl jasmonate-treated Norway spruce (*Picea abies*). *Plant J.* **60**, 1015–1030.

Zulak, K.G., Dullat, H.K., Keeling, C.I., Lippert, D. and Bohlmann, J. (2010) Immunofluorescence localization of levopimaradiene/abietadiene synthase in methyl jasmonate treated stems of Sitka spruce (*Picea sitchensis*) shows activation of diterpenoid biosynthesis in cortical and developing traumatic resin ducts. *Phytochemistry*, **71**, 1695–1699.