

**Plant Biology:**

**Bifunctional *cis*-Abienol Synthase from  
*Abies balsamea* Discovered by  
Transcriptome Sequencing and Its  
Implications for Diterpenoid Fragrance  
Production**

Philipp Zerbe, Angela Chiang, Macaire Yuen,  
Björn Hamberger, Britta Hamberger, Jason A.  
Draper, Robert Britton and Jörg Bohlmann  
*J. Biol. Chem.* 2012, 287:12121-12131.

doi: 10.1074/jbc.M111.317669 originally published online February 15, 2012



Access the most updated version of this article at doi: [10.1074/jbc.M111.317669](https://doi.org/10.1074/jbc.M111.317669)

Find articles, minireviews, Reflections and Classics on similar topics on the [JBC Affinity Sites](https://www.jbc.org/).

Alerts:

- [When this article is cited](#)
- [When a correction for this article is posted](#)

[Click here](#) to choose from all of JBC's e-mail alerts

Supplemental material:

<http://www.jbc.org/content/suppl/2012/02/15/M111.317669.DC1.html>

This article cites 56 references, 15 of which can be accessed free at  
<http://www.jbc.org/content/287/15/12121.full.html#ref-list-1>

# Bifunctional *cis*-Abienol Synthase from *Abies balsamea* Discovered by Transcriptome Sequencing and Its Implications for Diterpenoid Fragrance Production<sup>\*[S]</sup>

Received for publication, October 26, 2011, and in revised form, February 14, 2012. Published, JBC Papers in Press, February 15, 2012, DOI 10.1074/jbc.M111.317669

Philipp Zerbe<sup>‡</sup>, Angela Chiang<sup>‡</sup>, Macaire Yuen<sup>‡</sup>, Björn Hamberger<sup>‡1</sup>, Britta Hamberger<sup>‡1</sup>, Jason A. Draper<sup>§</sup>, Robert Britton<sup>§</sup>, and Jörg Bohlmann<sup>‡1,2</sup>

From the <sup>‡</sup>Michael Smith Laboratories, University of British Columbia, 301-2185 East Mall, Vancouver, British Columbia V6T 1Z4, Canada and the <sup>§</sup>Department of Chemistry, Simon Fraser University, Burnaby, British Columbia V5A 1S6, Canada

**Background:** Balsam fir produces *cis*-abienol, a natural product of value to the fragrance industry.

**Results:** We describe the genomics-based discovery of balsam fir *cis*-abienol synthase.

**Conclusion:** *cis*-Abienol synthase is a bifunctional diterpene synthase that produces a bicyclic diterpenol in the class II active site.

**Significance:** *cis*-Abienol synthase is a new enzyme for metabolic engineering of plants or microorganisms to produce high value fragrance compounds.

The labdanoid diterpene alcohol *cis*-abienol is a major component of the aromatic oleoresin of balsam fir (*Abies balsamea*) and serves as a valuable bioproduct material for the fragrance industry. Using high-throughput 454 transcriptome sequencing and metabolite profiling of balsam fir bark tissue, we identified candidate diterpene synthase sequences for full-length cDNA cloning and functional characterization. We discovered a bifunctional class I/II *cis*-abienol synthase (*AbCAS*), along with the paralogous levopimaradiene/abietadiene synthase and isopimaradiene synthase, all of which are members of the gymnosperm-specific TPS-d subfamily. The *AbCAS*-catalyzed formation of *cis*-abienol proceeds via cyclization and hydroxylation at carbon C-8 of a postulated carbocation intermediate in the class II active site, followed by cleavage of the diphosphate group and termination of the reaction sequence without further cyclization in the class I active site. This reaction mechanism is distinct from that of synthases of the isopimaradiene- or levopimaradiene/abietadiene synthase type, which employ deprotonation reactions in the class II active site and secondary cyclizations in the class I active site, leading to tricyclic diterpenes. Comparative homology modeling suggested the active site residues Asp-348, Leu-617, Phe-696, and Gly-723 as potentially important for the specificity of *AbCAS*. As a class I/II bifunc-

tional enzyme, *AbCAS* is a promising target for metabolic engineering of *cis*-abienol production.

Conifers produce a diverse array of diterpenoids as major oleoresin components that play a key role in the chemical defense against herbivores and pathogens, such as bark beetles and their associated fungi (1–3). Oleoresin diterpenoids are also used as large volume, renewable raw material for the production of a suite of industrial resins and coatings and other bioproducts (4, 5). Although tricyclic diterpene resin acids are ubiquitously abundant in the pine family (*Pinaceae*), the oleoresin of balsam fir (*Abies balsamea* (L.) Mill.) contains *cis*-abienol 4 (see Fig. 1), a bicyclic, tertiary diterpene alcohol, as the major diterpenoid (6). Among other applications, *cis*-abienol 4 and other oxygen-containing diterpenoids of plant origin (e.g. sclareol and manool) are important in the fragrance industry to produce Ambrox®, which serves as a sustainable replacement for the use of ambergris in high end perfume formulations (7). Whereas Ambrox® is produced from plant terpenoids, ambergris is an animal product secreted from the intestines of sperm whales, which are listed as an endangered species.

In addition to balsam fir, a gymnosperm tree, only a few angiosperm plant sources, such as tobacco (*Nicotiana tabacum*; family *Solanaceae*) trichomes (8, 9) or the tuberous roots of Bolivian sunroot (*Polymnia sonchifolia*; family *Asteraceae*) (10) are known to produce *cis*-abienol 4 in amounts that are relevant for industrial scale extraction. The importance of *cis*-abienol 4 as a plant-derived precursor for the fragrance and bioproducts industry has spawned an interest in the discovery of the relevant biosynthetic genes and enzymes and their future application in metabolically engineered microbial or plant production systems.

Based on knowledge of diterpenoid biosynthesis in other conifer species (11–18), we explored balsam fir as a target for a genomics-enabled discovery strategy of *cis*-abienol 4 biosynthesis. In general, biosynthesis of oleoresin diterpenoids in

\* This research was supported by Genome Alberta, Genome British Columbia, Genome Quebec, and the Government of Canada through Genome Canada in support of the PhytoMetaSyn Project and by grants from the Natural Sciences and Engineering Research Council of Canada (to J. B.).

The nucleotide sequences reported in this paper have been submitted to the short read archive at NCBI with the accession number SRX039631; cDNA sequences described in this paper have been submitted to the GenBank™/EBI Data Bank with accession numbers JN254805 (*AbdiTPS1*), JN254806 (*AbdiTPS2*), JN254807 (*AbdiTPS3*), and JN254808 (*AbdiTPS4*).

[S] This article contains supplemental Tables 1–4 and Figs. 1 and 2.

<sup>1</sup> Present address: Dept. of Plant Biology and Biotechnology, University of Copenhagen, Thorvaldsensvej 40, 1871 Frederiksberg C, Copenhagen, Denmark.

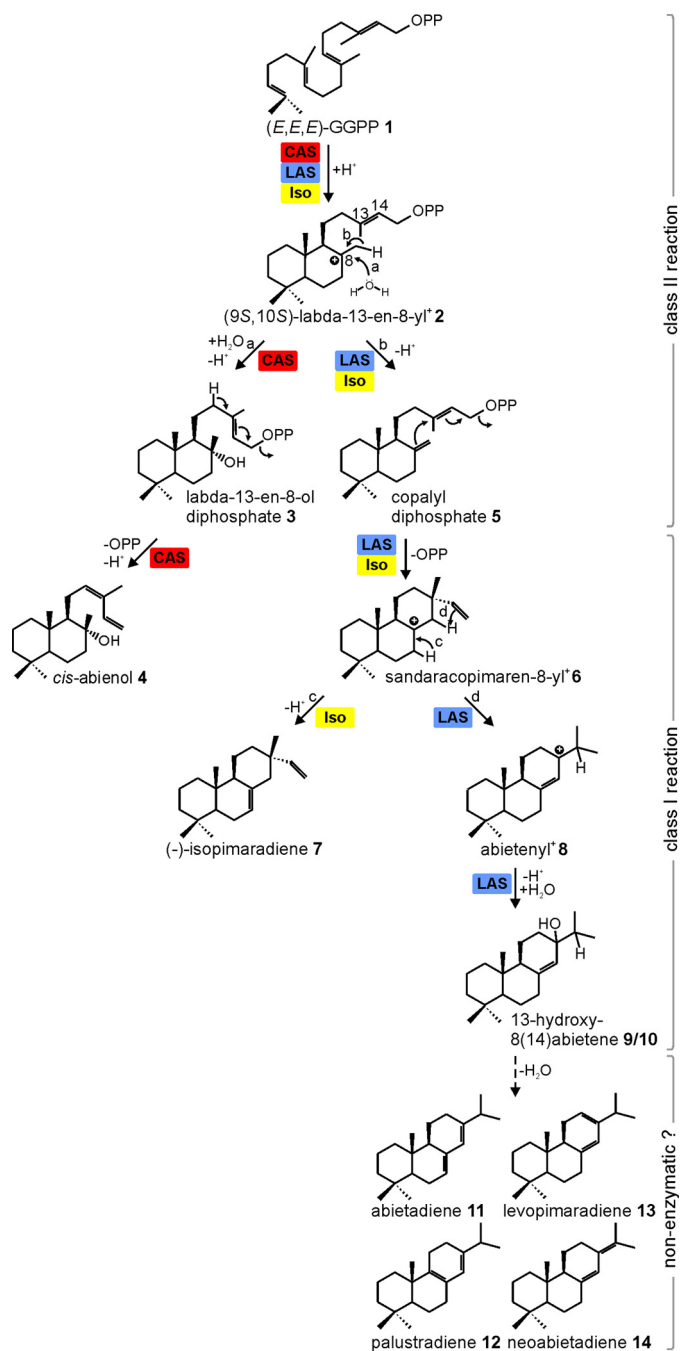
<sup>2</sup> Supported in part by the University of British Columbia Distinguished University Scholars program. To whom correspondence should be addressed. Tel.: 604-822-0282; Fax: 604-822-2114; E-mail: bohlmann@msl.ubc.ca.

## Bifunctional *cis*-Abienol Synthase

conifers involves the sequential cycloisomerization of (*E,E,E*)-geranylgeranyl diphosphate (GGPP)<sup>3</sup> **1**, catalyzed by bifunctional class I/II diterpene synthases (diTPSs) of the TPS-d subfamily (Fig. 1) (19). Oxygen functionality of conifer oleoresin diterpenes can be introduced by diTPSs (18) and/or by separate activity of cytochrome P450-dependent monooxygenases of the CYP720B subfamily (20–22). Previously characterized conifer diTPSs (11–18) produce either isopimaradiene **7** and minor amounts of sandaracopimaradiene (Iso-type diTPSs) or epimeric forms of 13-hydroxy-8(14)-abietene **9/10** (LAS-type diTPSs) (Fig. 1). *In vitro* and perhaps *in vivo*, 13-hydroxy-8(14)-abietene **9/10** is readily dehydrated, resulting in a mixture of abietadiene **11**, palustradiene **12**, levopimaradiene **13**, and neoabietadiene **14** (18). The reaction mechanism of the Iso- and LAS-type diTPSs involves the initial protonation-initiated bicyclization of GGPP **1** at the class II active site, resulting in copalyl diphosphate (CPP) **5** of (9*S*,10*S*)-configuration (*i.e.* normal or (+)-configuration). (9*S*,10*S*)-CPP **5** then translocates to the class I active site, where it undergoes a secondary ionization-dependent cyclization and enzyme-specific rearrangement of intermediate carbocations. In Iso-type diTPSs, direct deprotonation of the sandaracopimarenyl cation **6** results in the formation of isopimaradiene **7**, whereas LAS-type diTPSs catalyze additional rearrangement and water capture at C-13, resulting in 13-hydroxy-8(14)-abietene **9/10** as the initial product (18).

In contrast to the well characterized biosynthesis of tricyclic diterpenes by conifer diTPSs of the Iso- and LAS-type, a conifer gene or enzyme for the biosynthesis of *cis*-abienol **4** has not yet been reported. A plausible reaction sequence of *cis*-abienol **4** formation, catalyzed by a bifunctional conifer class I/II diTPS could proceed via water capture of a carbocation intermediate at carbon C-8 and subsequent ionization of the allylic diphosphate group without further cycloisomerization (Fig. 1). Recently, a monofunctional angiosperm diTPS catalyzing the formation of a bicyclic oxygen-containing diterpenoid, copal-8-ol diphosphate synthase from *Cistus creticus* (*Cistaceae*), has been reported (23). This enzyme represents a class II diTPS, which catalyzes the protonation-initiated cyclization of GGPP **1** to form the hydroxylated CPP compound. In addition, Mafu *et al.* (24) described a bifunctional class I/II diTPS, *SmCPSKSL1*, from the lycophyte *Sellaginella moellendorffii* for the formation of labda-7,13-dien-15-ol, where the primary hydroxyl group is introduced during the reaction of the class I active site.

We describe here the development of a 454 transcriptome resource for balsam fir, which, in conjunction with terpenoid metabolite profiling, was used for the identification and functional characterization of a class I/II gymnosperm *cis*-abienol synthase (CAS). The same sequence resource also revealed two additional bifunctional class I/II diTPSs, representing the balsam fir Iso and LAS enzymes. Comparative homology modeling



**FIGURE 1. Schematic of the proposed biosynthesis of diterpene olefins and alcohols by class I/II bifunctional diTPSs in conifers.** The known activity of LAS- and Iso-type diTPSs in the formation of tricyclic diterpenes involves the stepwise cyclization of GGPP **1** via (9*S*,10*S*)-CPP **5** (*i.e.* CPP of normal or (+)-stereochemistry). Protonation-initiated cyclization of GGPP **1** to CPP **5** is catalyzed by the class II active site of LAS- and Iso-type diTPSs. At the class I active site of Iso enzymes, ionization-dependent secondary cyclization of CPP **5** and deprotonation of the resulting sandaracopimaren-8-yl cation **6** lead to the formation of isopimaradiene **7**. Alternatively, in the class I active site of LAS enzymes, additional proton transfer and methyl migration afford the tertiary abietenyl cation **8**, and further deprotonation and hydroxylation (*i.e.* water capture) result in the formation of the unstable 13-hydroxy-8(14)-abietene **9/10** product of the LAS activity. Dehydration of this tertiary diterpene alcohol yields several different diterpene olefins, including abietadiene **11**, palustradiene **12**, levopimaradiene **13**, and neoabietadiene **14**. A different reaction sequence is proposed for the formation of the bicyclic diterpene alcohol *cis*-abienol **4** by the class I/II bifunctional AbCAS enzyme. The class II activity of AbCAS converts GGPP **1** to labda-13-en-8-ol diphosphate **3** via water capture at the C-8 carbon of the labda-13-en-8-yl<sup>+</sup> cation **2**. Subsequent ionization of the allylic diphosphate at the class I active site, without additional cyclization, yields *cis*-abienol **4**.

<sup>3</sup> The abbreviations used are: GGPP, geranylgeranyl diphosphate; CAS, *cis*-abienol synthase; CPP, copalyl diphosphate; diTPS, diterpene synthase; LAS, levopimaradiene/abietadiene synthase; Iso, isopimaradiene synthase; TPS, terpene synthase; FL, full-length; RACE, rapid amplification of cDNA ends; APCI, atmospheric pressure chemical ionization.



and site-directed mutagenesis revealed unique features of *cis*-abienol **4** biosynthesis and suggest functional divergence in the evolution of CAS relative to the paralogous Iso- and LAS-type diTPSs associated with substitution of key residues in the active sites. A novel feature of CAS is the formation of a tertiary diterpenol at the class II active site of a bifunctional class I/II diTPS.

## EXPERIMENTAL PROCEDURES

**Plant Material**—Two-year-old saplings of *Abies balsamea* var. *phanerolepsis* (L.) Mill. were purchased from Arbutus Grove Nursery Ltd. (North Saanich, Canada) and maintained in a greenhouse as described previously (25). Needles, xylem/wood, and phloem/bark were harvested from the upper interwhorls and used fresh or flash frozen in liquid N<sub>2</sub> for metabolite and RNA extraction, respectively.

**Diterpene Standards**—Authentic diterpene resin acid standards were purchased from Orchid Celmark (New Westminster, Canada). The corresponding diterpene olefins were synthesized from the acids at Best West Laboratories Inc. (Salt Lake City, UT) as described previously (20).

**Metabolite Analysis**—Terpenoids were extracted from samples of 2 g of pulverized tissue with 1.5 ml of diethyl ether for 16 h at room temperature. Water was removed by the addition of anhydrous Na<sub>2</sub>SO<sub>4</sub>, and extracts were passed through 0.22- $\mu$ m GHP membrane filters (PALL Corp., De Miniac, Canada). GC-MS analysis was performed on an Agilent 6890N GC (Agilent Technologies Inc., Mississauga, Canada), 7683B series autosampler and a 5975 Inert XL MS Detector at 70 eV and 1 ml min<sup>-1</sup> helium as carrier gas using a SGE Solgel-Wax column (polyethylene glycol, 30 m, 250  $\mu$ m inner diameter, 0.25- $\mu$ m film) with the following GC temperature program: 40 °C for 5 min, 3 °C min<sup>-1</sup> to 80 °C, 8 °C min<sup>-1</sup> to 250 °C, 10 °C min<sup>-1</sup> to 270 °C, hold 5 min, pulsed splitless injector held at 250 °C. Diterpene resin acids were extracted from samples of ~150 mg of tissue with 1.5 ml of *tert*-butylmethylether following the method described by Lewinsohn *et al.* (26) and derivatized with 2 M trimethylsilyl diazomethane (Sigma) prior to GC-MS analysis using an Alltech AT-1000 column (polyethylene glycol acid-modified, 30 m, 250- $\mu$ m inner diameter, 0.25- $\mu$ m film) and GC specifications as follows: 150 °C initial temperature, 1.5 °C min<sup>-1</sup> to 220 °C, 20 °C min<sup>-1</sup> to 240 °C. Compound identification was achieved by comparison of mass spectra with those of authentic standards and reference mass spectral databases of the National Institute of Standards and Technology MS library searches (Wiley W9N08). Quantifications were based on three independent biological replicates.

**RNA Isolation and cDNA Synthesis**—Total RNA was isolated from samples of 150 mg of bark tissue according to the method of Kolosova *et al.* (27), and mRNA was purified on Dynabeads (Invitrogen). RNA integrity and amounts were determined on a Bioanalyzer 2100 using an RNA Pico Chip (Agilent). Synthesis of cDNA as a template for PCR amplification of candidate genes was carried out with random hexamer oligonucleotides using the SMARTer cDNA rapid amplification of cDNA ends (RACE) amplification kit (Clontech, Mountain View, CA).

**cDNA Library Construction and 454 Transcriptome Sequencing**—Construction of a non-normalized bark tissue cDNA library and subsequent 454 transcriptome sequencing

was conducted at the McGill University and Génome Québec Innovations Centre (Montreal, Canada). The cDNA library was constructed from 200 ng of fragmented mRNA using the cDNA Rapid Library Preparation kit, GS FLX Titanium series (Roche Applied Science) following the manufacturer's protocols. Yield of cDNA and the size range of cDNA fragments were assessed using a Bioanalyzer 2100 Pico Chip (Agilent), and 200 ng of the cDNA library were subjected to a half-plate reaction of 454 pyrosequencing using the Roche GS FLX Titanium technology.

**De Novo Sequence Assembly**—The 454 transcriptome sequencing generated 797,060 raw reads. After trimming of 5' adapter sequences and exclusion of reads smaller than 45 bp, *de novo* assembly of the bark transcriptome was performed using the GS *De Novo* Assembler version 2.5p1 (Roche Applied Science). BLASTx search (28) of the resulting isotigs against a comprehensive in-house curated sequence database of plant diTPS sequences was used to identify a subset of candidate isotigs. Reads from these isotigs were reassembled in PHRAP (29) for closer investigation.

**Isolation of Full-length cDNAs**—A full-length (FL) sequence of *AbdiTPS4* was retrieved from the assembled transcriptome sequences. Completion of the 3'-sequences of *AbdiTPS1* and -2 was achieved by 3'-RACE using the SMARTer cDNA RACE cDNA amplification kit (Clontech) and Phusion DNA polymerase (New England Biolabs, Pickering, Canada) with the following PCR program: initial denaturation at 98 °C for 30 s, five cycles of denaturation at 98 °C for 20 s and extension at 72 °C for 90 s, five cycles of denaturation at 98 °C for 20 s, annealing at 70 °C for 20 s and extension at 72 °C for 90 s, 27 cycles of denaturation at 98 °C for 20 s, annealing at 65 °C for 20 s and extension at 72 °C for 90 s, and a final extension at 72 °C for 3 min. The obtained amplicons were gel-purified and ligated into pJET using the CloneJET kit (Clontech). Pseudomature versions of the FL cDNAs (starting at (K/N)REFPP) were amplified (initial denaturation at 98 °C for 30 s followed by 30 cycles of denaturation at 98 °C for 10 s, annealing at 65–68 °C for 30 s, and extension at 72 °C for 80 s and final extension at 72 °C for 10 min), cloned into pJET, and subcloned into the *NheI*/*SalI* and *NotI* restriction sites of pET28b(+) (EMD Biosciences, San Diego, CA). All cDNA constructs were sequence-verified at the Nucleic Acid Protein Service Unit (University of British Columbia, Vancouver, Canada) prior to expression in *Escherichia coli*. Oligonucleotides used in PCR procedures are listed in supplemental Table 1.

**Functional Characterization of Diterpene Synthases**—Recombinant proteins were expressed in *E. coli* BL21DE3-C41, Ni<sup>2+</sup> affinity-purified as described elsewhere (16), and desalted into 20 mM HEPES (pH 7.2), 150 mM NaCl, 10% glycerol, 5 mM DTT using PD MiniTrap G-25 columns (GE Healthcare). Enzyme assays were carried out as described before (16) in 50 mM HEPES (pH 7.2), 10  $\mu$ M MgCl<sub>2</sub>, 5% glycerol, 5 mM DTT, using 100  $\mu$ g of purified protein (100  $\mu$ g each for coupled assays) and 15  $\mu$ M (*E,E,E*)-GGPP **1** (Sigma) with incubation for 1 h at 30 °C. After extraction of reaction products with 500  $\mu$ l of pentane, GC-MS analysis was conducted on an Agilent 7890A GC, 7683B series autosampler and a 7000A Triple Quad MS Detector at 70 eV and 1.2 ml min<sup>-1</sup> helium flow using an HP5ms column (5% phenyl methyl siloxane, 30 m, 250  $\mu$ m

inner diameter, 0.25- $\mu$ m film) and the following GC temperature program: 40 °C for 2 min, 20 °C min<sup>-1</sup> to 300 °C, hold 2 min, pulsed splitless injection (injector held at 250 °C). For the detection of diphosphate intermediates, reaction products were dephosphorylated prior to extraction by incubation with 10 units of calf intestinal alkaline phosphatase (Invitrogen) for 16 h at 37 °C and analyzed on a Solgel-Wax column as described above (see "Metabolite Analysis") with modified GC parameters: 40 °C for 2 min, 25 °C min<sup>-1</sup> to 250 °C, hold 5 min, pulsed splitless injector held at 250 °C. Analysis of reaction products via LC-MS was performed on an Agilent 1100 series LC/MSD Trap XCT Plus MS with atmospheric pressure chemical ionization (APCI) in positive mode on an Agilent Zorbax RX-Sil silica column (4.6 mm inner diameter  $\times$  150 mm  $\times$  5  $\mu$ m) as reported previously (18).

**Nuclear Magnetic Resonance (NMR) Analysis—*cis*-Abienol 4** was prepared from a pool of 10 individual enzyme assays as described above using 20  $\mu$ M GGPP and an incubation time of 2 h to maximize product formation. To enhance product purity, pentane was purified on alumina prior to use, and assays were performed in buffer pre-extracted with alumina-purified pentane. Nuclear magnetic resonance (NMR) spectra were recorded on a Bruker Avance 600 spectrometer equipped with a QNP or TCI cryoprobe (600 MHz) using deuteriochloroform (CDCl<sub>3</sub>) as the solvent, which was neutralized by distillation and filtration through potassium carbonate prior to preparation of the sample. Signal positions ( $\delta$ ) were calculated in parts per million (ppm) as compared with tetramethylsilane ( $\delta$  0) and were measured relative to the signal of the solvent (CDCl<sub>3</sub>:  $\delta$  7.26, <sup>1</sup>H NMR;  $\delta$  77.0, <sup>13</sup>C NMR).

**Generation of Site-directed Protein Variants of AbdiTPS4—**Alanine substitutions of Asp-402, Asp-404, or Asp-621 of AbdiTPS4 were generated following the QuikChange site-directed mutagenesis protocol (Stratagene, Mississauga, Canada) with the cDNA in pET28b(+) as template and the following PCR program: initial denaturation at 98 °C for 90 s followed by 29 cycles of denaturation at 98 °C for 50 s, annealing at 60 °C for 30 s and extension at 72 °C for 4 min, and final extension at 72 °C for 10 min. Oligonucleotides used for mutagenesis are listed in supplemental Table 1.

**Phylogenetic Analysis—**Multiple protein sequence alignments were performed using the CLC bio Main Workbench 5.7.1 (CLC bio, Århus, Denmark). Phylogenetic analyses were conducted on the basis of the maximum likelihood algorithm using PhyML 3.0 (30) with four rate substitution categories, LG substitution model, BIONJ starting tree and 100 bootstrap repetitions, and displayed as phylogram using treeview32 1.6.6.

**Computational Structure Analysis—**Homology models of AbLAS, AbIso, and AbCAS were built using the CPHmodels 3.0 server (31) based on the tertiary structure of TbTXS (32) (PDB code 3p5pA, chain A) and certified as high quality exceeding 91% residues assigned to most favored regions in Ramachandran plot statistics using PROCHECK (33). Lack of structural errors in the models was validated using the ProSA-web server (34). Pairwise comparison of these modeled structures with the DaliLite server (35) demonstrated a high structural similarity of AbLAS, AbIso, AbCAS, and TbTXS with root mean square deviations of  $\leq$  1 Å for the  $\beta\gamma$  domain and the  $\alpha$  domain, respec-

tively, thus allowing for a structural comparison of the active sites. Using Molegro Virtual Docker 2010.4.0.0 (36), a semiautomated docking approach was applied to place labda-13-en-8-ol diphosphate 3 in the class I and class II active site of AbCAS (Fig. 6). For this purpose, active site cavities were predicted using a probe radius of 1.0 Å and a grid resolution of 0.6 Å with further manual optimization. Substrate docking was then restricted to the predicted cavities, inclusive of the required Mg<sup>2+</sup> complex. Energy-minimized Protein Data Bank coordinates of the ligands were generated using the PRODRG server (37).

## RESULTS

**Terpenoid Metabolite Profiling Confirms *cis*-Abienol as Major Diterpene in Bark Tissue of Balsam Fir—**To identify a suitable tissue source for transcriptome mining of genes of *cis*-abienol 4 biosynthesis, we established diterpenoid metabolite profiles for wood and xylem, bark and phloem, and needles of balsam fir sapling trees (supplemental Table 2). As supporting information, we also measured mono- and sesquiterpenoids. Although only trace amounts of these terpenoids were detected in wood tissue (data not shown), bark/phloem and needle samples had similar total amounts of terpenoids, consisting mostly of diterpenoids, with lesser amounts of monoterpenoids and relatively minor amounts of sesquiterpenoids (supplemental Table 2). Both the bark/phloem and the needle samples contained the diterpene resin acids, abietic acid, isopimaric acid, dehydroabietic acid, and palustric acid, as major diterpenoids. These are all commonly abundant in conifers. *cis*-Abienol 4 was found almost exclusively in the bark/phloem tissue, where it was the most abundant terpenoid metabolite, accounting for more than 25% of the total diterpenoid content of balsam fir stem tissue (supplemental Table 2).

**Discovery of diTPS Genes via Transcriptome Sequencing and de Novo Sequence Assembly—**Based on the high abundance of *cis*-abienol 4 in bark/phloem of balsam fir sapling stems, we used this tissue source for the preparation of a non-normalized cDNA library and subsequent transcriptome sequencing. A half-plate reaction of Roche 454 sequencing generated a total of 797,060 sequence reads with an average GC content of 46% and an average read length of 359 bp (supplemental Table 3). After adapter trimming, the remaining high quality reads were subjected to a *de novo* assembly in GS De Novo Assembler 2.5p1 with a size exclusion of 45 bp. The assembly contained 85% of all input reads, yielding 14,699 isogroups from 17,122 isotigs with an average size of 1,114 nucleotides. For the discovery of TPS genes, these isotigs were searched against an in-house curated database of 146 known plant TPSs, resulting in a subset of isotigs that represented candidate genes for nine different putative mono- and sesqui-TPSs and five putative diTPSs (supplemental Table 4). Of the diTPSs, four candidate isotigs (AbdiTPS1, AbdiTPS2, AbdiTPS3, and AbdiTPS4) resembled bifunctional class I/II gymnosperm diTPSs of the TPS-d group (14, 17, 19) containing the characteristic DXDD, DDXXD, and NSE/DTE motifs (38–40). To optimize and validate the sequence assemblies, reads corresponding to these four isotigs were reassembled in PHRAP (29), from which three unique partial diTPS cDNA sequences (AbdiTPS1, AbdiTPS2, and

AbdiTPS3) and one FL cDNA sequence (AbdiTPS4) were obtained. The encoded proteins of AbdiTPS1, AbdiTPS2, and AbdiTPS3 showed highest similarity with grand fir (*Abies grandis*) abietadiene synthase (AgAS) (11, 12), with more than 90% protein sequence identity. In contrast, the deduced protein sequence AbdiTPS4 showed highest similarity with Norway spruce (*Picea abies*) isopimaradiene synthase (PaIso) (14), although at a substantially lower level of sequence identity of only 75%, highlighting AbdiTPS4 as a unique candidate diTPS sequence (supplemental Fig. 1).

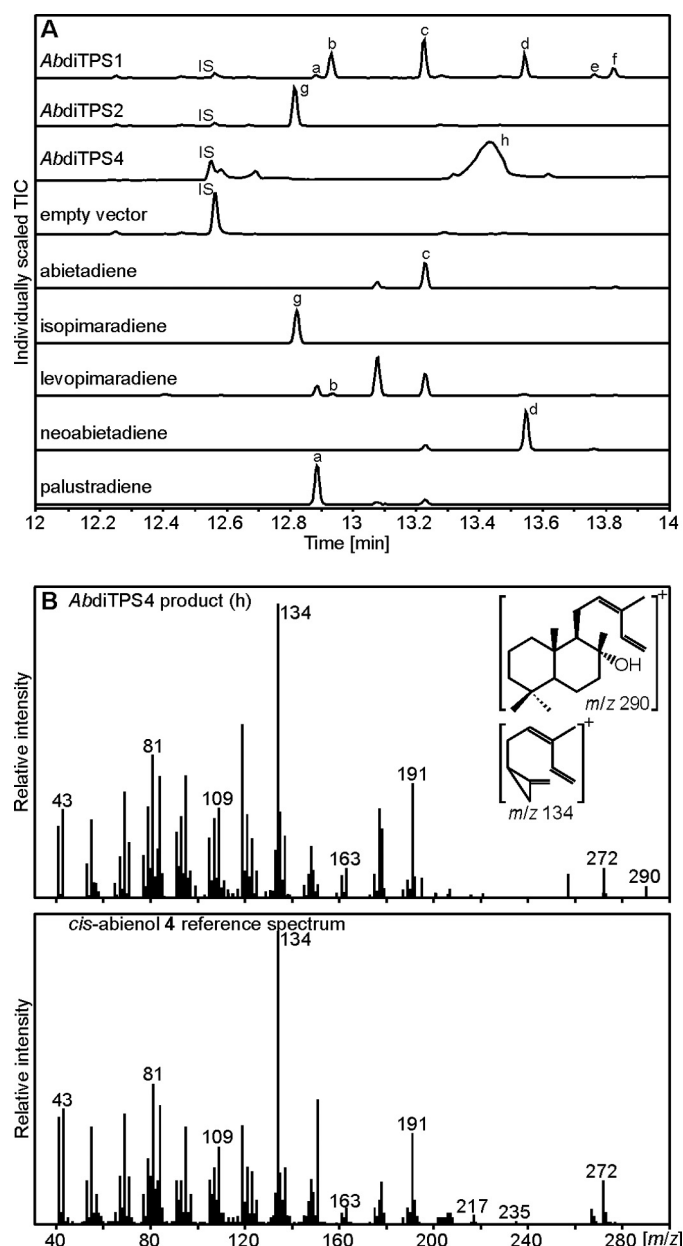
**Functional Characterization of Balsam Fir diTPSs and Discovery of Novel *cis*-Abienol Synthase**—The FL cDNA clone for AbdiTPS4 was amplified based on the *in silico* assembled FL sequence. FL or nearly FL cDNAs for AbdiTPS1 and AbdiTPS2, respectively, were recovered by 3'-RACE, with AbdiTPS2 lacking ~40–50 bp of the plastidial transit peptide. Despite using a number of different primer combinations and PCR conditions, we could not amplify a cDNA clone corresponding to the AbdiTPS3 sequence assembly. For functional characterization of AbdiTPS1, AbdiTPS2, and AbdiTPS4, we cloned pseudomature versions of these diTPSs starting at the conserved (N/K)RX<sub>6</sub>W motif and lacking the putative plastidial transit peptide (16). Recombinant proteins were expressed in *E. coli* and Ni<sup>2+</sup> affinity-purified, resulting in soluble proteins of the expected molecular mass of 90–91 kDa (supplemental Fig. 2). Enzyme assays with GGPP **1** as substrate followed by GC-MS or LC-MS analysis of reaction products against proper controls and, where available, authentic standards identified functions for AbdiTPS1, AbdiTPS2, and AbdiTPS4 as described below.

The product profile of AbdiTPS1 was identified by GC-MS to consist of four peaks corresponding to abietadiene **11**, palustradiene **12**, levopimaradiene **13**, and neoabietadiene **14** in a ratio of ~4:3:2:1 (Fig. 2). This profile closely matches that previously identified for grand fir AgAS (12), a diTPS that is 99% identical with AbdiTPS1 on the amino acid level. According to these patterns, AbdiTPS1 can be classified as an LAS-type bifunctional class I/II diTPS, which appears to be orthologous with AgAS. Following the recent discovery of epimers of a tertiary alcohol, 13-hydroxy-8(14)-abietene **9/10**, as the initial enzyme products of Norway spruce PaLAS (18), we also investigated the product of AbdiTPS1 by LC-MS, which confirmed the formation of 13-hydroxy-8(14)-abietene **9/10** (Fig. 3). The detected mass fragment of *m/z* 273 in the LC-MS analysis (Fig. 3) corresponds to dehydration product of 13-hydroxy-8(14)-abietene **9/10** as detailed by Keeling *et al.* (18).

AbdiTPS2 formed exclusively isopimaradiene **7** (Fig. 2), similar to the previously characterized isopimaradiene synthases from Norway spruce (PaIso) and Sitka spruce (PsIso) (14, 16, 17). AbdiTPS2 was thereby identified as an Iso-type bifunctional class I/II diTPS.

AbdiTPS4 showed a unique single-peak product profile (Figs. 2 and 3), which corresponded to *cis*-abienol **4** according to retention time consistent with the compound extracted from plant tissue and shared characteristic mass fragments (e.g. *m/z* 290 [*M*<sup>+</sup>], *m/z* 272 [*M*<sup>+</sup>–H<sub>2</sub>O], and *m/z* 134) in comparison with reference mass spectra (41).

When the product of AbdiTPS4 was analyzed by GC-MS, under a number of different conditions, we observed poor res-



**FIGURE 2. Activity of balsam fir diTPSs AbdiTPS1 (AbLAS), AbdiTPS2 (AbIso), and AbdiTPS4 (AbCAS).** A, total ion chromatograms (TIC) of reaction products from *in vitro* assays with purified recombinant enzymes using GGPP **1** as a substrate. GC-MS analysis was performed on an Agilent HP5ms column with electronic ionization at 70 eV. Enzymatic activity assays were confirmed with three independent experiments. Peak IS, internal standard 1.6  $\mu$ M 1-eicosene; peak a, palustradiene **12**; peak b, levopimaradiene **13**; peak c, abietadiene **11**; peak d, neoabietadiene **14**; peaks e and f, epimers of 13-hydroxy-8(14)-abietene **9/10**; peak g, isopimaradiene **7**; peak h, *cis*-abienol **4**; the authentic standard of abietadiene **11** contained an unknown contamination; the authentic standard of levopimaradiene **13** contains other diterpene isomers, which could not be removed. B, mass spectrum of *cis*-abienol **4**, produced by recombinant AbdiTPS4 with GGPP **1** as substrate and comparison with the reference mass spectrum of *cis*-abienol **4** as described by Vlad *et al.* (41) and obtained from the National Institute of Standards and Technology MS library searches (Wiley W9N08). Authentic *cis*-abienol **4** standard is not commercially available.

olution of the shape of the *cis*-abienol **4** peak (Fig. 2), relative to the peak shape of diterpene olefins, and the occurrence of trace amounts of two additional compounds. This profile is probably due to degradation of *cis*-abienol **4** during GC-MS as reported



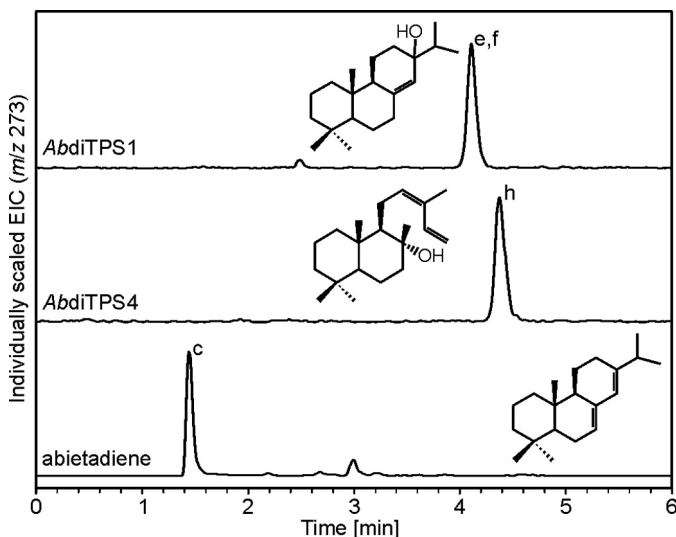


FIGURE 3. LC-MS analysis of diterpenol products of *AbdiTPS1* (*AbLAS*) and *AbdiTPS4* (*AbCAS*). Assay products were analyzed by normal phase LC-APCI-MS in positive mode and are shown as extracted ion chromatograms (EIC) of the base peak  $m/z$  273, which is indicative of dehydration of the unstable diterpene alcohol compounds in the APCI interface. Comparison with abietadiene **11** standard showed separation of olefin compounds, such as abietadiene **11** (peak *c*) and the more polar hydroxylated 13-hydroxy-8(14)-abietene epimers **9/10** (peaks *e* and *f*) and *cis*-abienol **4** (peak *h*).

previously (42–44). However, LC-MS analysis confirmed *cis*-abienol **4** as a single product of *AbdiTPS4*, with  $m/z$  273 in the LC-MS analysis corresponding to the predicted dehydration product of the diterpene alcohol (Fig. 3). Additional structural and stereochemical analysis by proton and carbon NMR and comparison with previously reported analyses (45, 46) confirmed the identity of the *AbdiTPS4* product as *cis*-abienol **4**.

Based on functional characterization, *AbdiTPS1*, *AbdiTPS2*, and *AbdiTPS4* will be referred to as *AbLAS*, *AbIso*, and *AbCAS*, respectively. *AbCAS* represents a new type of conifer diTPS, producing a bicyclic, tertiary diterpene alcohol, as opposed to tricyclic products of the LAS- and Iso-type diTPSs. With the present study, LAS- or Iso-type diTPSs have now been identified in three different genera of the pine family, namely in true firs (*Abies*), spruce (*Picea*), and pine (*Pinus*), which allows for analysis of gene orthology within this family (Fig. 4). It appears that gene duplications and neofunctionalization leading to paralogous pairs of LAS and Iso genes occurred independently in *Abies* and *Picea*, after the separation of these genera. Within the spruce genus, the dichotomy of LAS and Iso genes happened apparently prior to the speciation of Norway spruce and Sitka spruce.

The bifunctional class I/II *AbCAS* described here has only been cloned from balsam fir, matching the major diterpenoid produced by this species (supplemental Table 2). Within the group of conifer class I/II diTPSs, the *AbCAS* gene is separate from and appears basal to a clade of LAS and Iso genes from firs, spruces, and pine (Fig. 4).

**Site-directed Mutagenesis of DXDD (Class II Active Site) and DDXXD Motifs (Class I Active Site) of *AbCAS* Demonstrate That Formation of the Tertiary Alcohol and Cyclization Occur during the Class II Reaction**—A plausible mechanism for the formation of *cis*-abienol **4** requires hydroxylation via water quenching of a labda-13-en-8-yl carbocation **2** at C-8 (Fig. 1). In

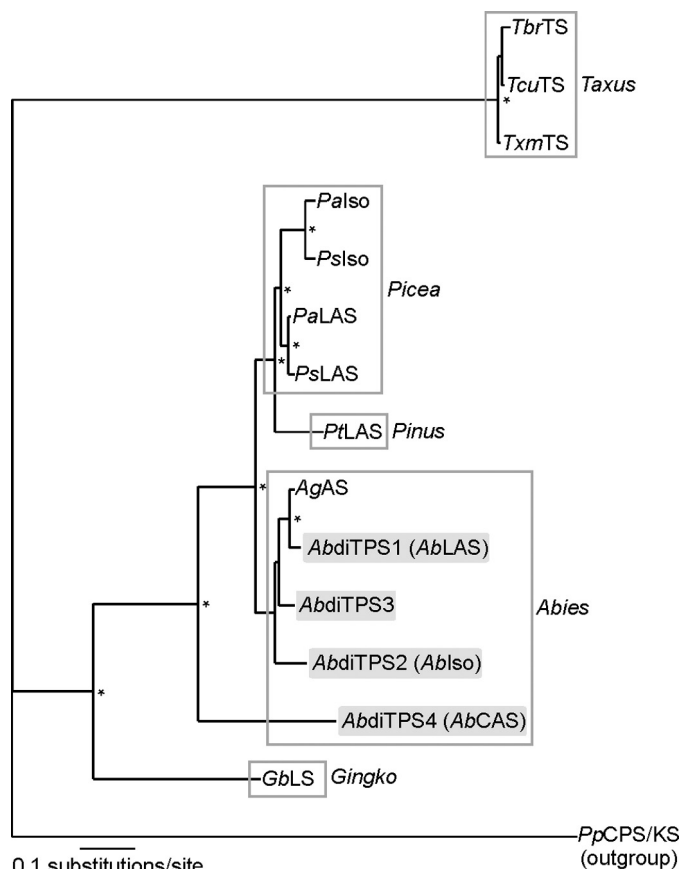
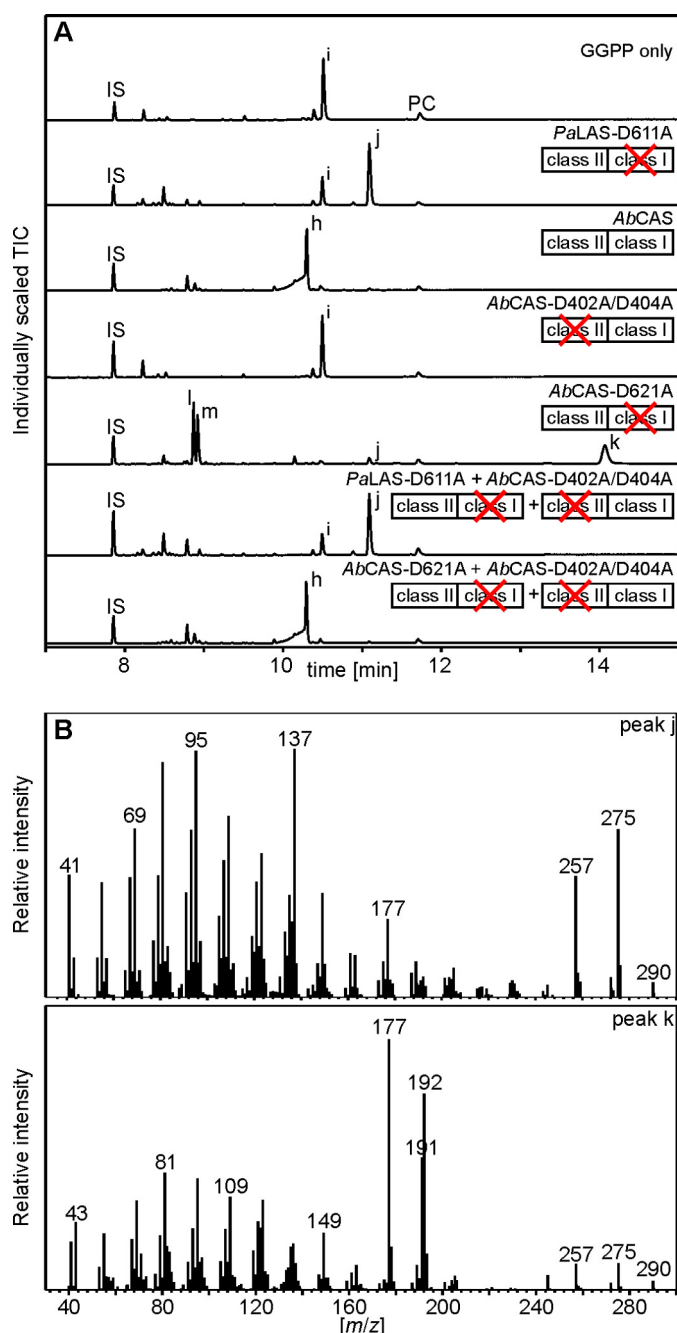


FIGURE 4. Phylogeny of balsam fir diTPSs of the LAS-, Iso-, and CAS-type in the context of the TPS-d3 subfamily of conifer diTPSs. The phylogenetic tree was generated by sequence alignment in CLC bio, phylogenetic analysis with PhyML (four rate substitution categories, LG substitution model, BIONJ starting tree, 100 bootstrap repetitions), and visualization in treeview by rooting with the outgroup copalyl diphosphate synthase/kaurene synthase from the moss *P. patens* (*PpCPS/KS*). *TbrTS*, *Taxus brevifolia* taxadiene synthase (NCBI accession number AAC49310); *TcuTS*, *Taxus cuspidata* taxadiene synthase (NCBI accession number ABW82997); *TxmTS*, *Taxus x media* taxadiene synthase (NCBI accession number AAS18603); *Palso*, *P. abies* isopimaradiene synthase (NCBI accession number AAS47690); *PsIso*, *Picea sitchensis* isopimaradiene synthase (NCBI accession number ADZ45512); *PaLAS*, *P. abies* (NCBI accession number AAS47691); *PsLAS*, *P. sitchensis* levopimaradiene/abietadiene synthase (NCBI accession number ADZ45517); *PtLAS*, *P. taeda* levopimaradiene synthase (NCBI accession number AY779541); *AgAS*, *A. grandis* abietadiene synthase (NCBI accession number AAK83563); *GbLS*, *G. biloba* levopimaradiene synthase (NCBI accession number AAL09965); *PpCPS/KS*, *P. patens* ent-copalyl diphosphate/ent-kaurene synthase (NCBI accession number BAF61135). Asterisks indicate nodes supported by >90% bootstrap values.

principle, hydroxylation can occur during reactions at the class II active site or at the class I active site. To delineate which of the two active sites of bifunctional *AbCAS* catalyzed the formation of the tertiary alcohol, we generated a set of alanine substitutions of the DXDD (class II active site) and DDXXD (class I active site) motifs to obtain monofunctional *AbCAS* variants that contain either a non-functional class II (*AbCAS*:D402A/D404A) or a non-functional class I (*AbCAS*:D621A) active site.

*AbCAS*:D621A converted GGPP **1** (peak *i* in Fig. 5) into trace amounts of CPP **5** (peak *j*) and a product with a significantly longer retention time, indicative of a hydroxylated diphosphate (peak *k*). This compound was identified, upon cleavage of the diphosphate group, as labda-13-en-8,15-diol based on characteristic mass fragments ( $m/z$  290 [ $M^+$ ],  $m/z$  275,  $m/z$  257,  $m/z$



**FIGURE 5. Characterization of *AbCAS* protein variants.** Activity analysis of protein variants of *AbCAS* was conducted with 20  $\mu$ M GGPP **1** as substrate and dephosphorylation of the reaction products prior to GC-MS analysis (*AbCAS*: D402A/D404A) or in the form of coupled assays, combining the *AbCAS*:D621A variant with a point mutant of *PaLAS* (*PaLAS*:D611A), containing a non-functional class I active site (C. I. Keeling and J. Bohlmann, unpublished results), and direct GC-MS analysis of the reaction products. A, total ion chromatograms (TIC) of reactions products analyzed on an Agilent Solgel-Wax column with electronic ionization at 70 eV. Peak IS, internal standard 1.6  $\mu$ M 1-eicosene; peak PC, plasticizer contamination; peak h, *cis*-abienol **4**; peak i, geranylgeraniol; peak j, (9*S*,10*S*)-copalol; peak l, *epi*-manoyl oxide; peak m, manoyl oxide; peak k, labda-13-en-8,15-diol. B, characteristic mass spectra of (9*S*,10*S*)-copalol (top) and labda-13-en-8,15-diol (bottom), produced by *PaLAS*:D611A and *AbCAS*:D621A, respectively.

192, and *m/z* 177) as compared with the mass spectrum of the authentic compound (23). *epi*-Manoyl oxide and manoyl oxide (peaks l and m in Fig. 5) were apparently not products of *AbCAS* but rather the result from ether formation of the dephos-

phorylated labda-13-en-8-ol diphosphate **3** under GC-MS conditions because they were not seen in the wild-type enzyme product profile and did not occur under different GC conditions. Results from enzyme assays with *AbCAS*:D621A suggested that the class II active site of *AbCAS* catalyzes protonation-initiated formation of bicyclic labda-13-en-8-ol diphosphate **3** via water quenching of the parental carbocation.

Alanine substitution of Asp-402 and Asp-404 (*AbCAS*:D402A/D404A) in the class II active site resulted in complete loss of enzymatic activity with GGPP **1** as substrate. Activity of *AbCAS*:D402A/D404A could not be restored by combining this mutant with a protein variant of *PaLAS* (*PaLAS*:D611A), which contains an inactive class I active site and accumulates (9*S*,10*S*)-CPP **5** (peak j in Fig. 5). This result suggested that, unlike the LAS and Iso-type enzymes, the class I active site of *AbCAS* is not active with (9*S*,10*S*)-CPP **5** as an intermediate substrate.

Formation of *cis*-abienol **4** (peak h in Fig. 5) from GGPP **1** was restored in coupled assays of the two monofunctional *AbCAS* class I and class II protein variants, *AbCAS*:D402A/D404A and *AbCAS*:D621A, confirming that the non-mutated active sites remained functional in the two mutated proteins.

Together, the mutational analysis of class I and class II active sites of *AbCAS* suggested a reaction sequence of the bifunctional class I/II enzyme in which both the bicyclization and water capture occur in the class II active site, resulting in an intermediary labda-13-en-8-ol diphosphate **3**, which undergoes cleavage of the diphosphate group and final deprotonation at the class I active site (Fig. 1).

## DISCUSSION

Variations on a common theme of diTPS-catalyzed cycloisomerization of GGPP **1** contribute substantially to the chemical diversity of diterpene metabolites found in the oleoresin of conifers and in nature in general. Introduction of hydroxy functions further increases the diversity of naturally occurring diTPS products. Hydroxylation of conifer diterpenes can result from activity of P450s acting on products of diTPSs (20, 22) or can result from capture of water by carbocation intermediates during the diTPS reaction (18). For the fragrance industry, bicyclic hydroxylated diterpenes, such as *cis*-abienol **4** and sclareol, are of particular value as plant-derived precursors for the sustainable production of Ambrox®, which replaces the controversial use of animal-derived ambergris in perfume formulations.

Surprisingly, relatively little is known about diTPSs catalyzing cyclohydration reactions. Reported examples are the biosynthesis of copal-8-ol diphosphate by a monofunctional class II diTPS from *Cistus creticus* (23), the formation of *ent*-16 $\alpha$ -hydroxy-kaurene as a product of the bifunctional class I/II *ent*-copalyl diphosphate/*ent*-kaurene synthase from the non-vascular plants *Physcomitrella patens* and *Jungermannia subulata* (47, 48), and the biosynthesis of labda-7,13-dien-15-ol catalyzed by an *ent*-copalyl diphosphate/*ent*-kaurene synthase-like diTPS from the lycophyte *Sellaginella moellendorffii* (24). In addition, a recent study on the product specificity of Norway spruce *PaLAS* (18) demonstrated a tricyclic, tertiary diterpene alcohol as the initial, but highly unstable, product of this diTPS. In the case of *PaLAS*, the allylic diterpenol product dehydrates



to a set of diterpene olefins. Against this background of prior knowledge, the identification of bifunctional class I/II *AbCAS* adds a unique new catalyst to the known portfolio of diTPSs, with possible applications in the bioproducts industry. The genomics-based discovery of *AbCAS*, alongside *AbLAS* and *AbIso*, also substantially contributes to our understanding of the molecular, biochemical, and evolutionary underpinnings of conifer diterpenoid diversity.

The present study highlights the powers of combined metabolite profiling, tissue-specific deep transcriptome sequencing, and functional (*i.e.* biochemical) genomics for the successful discovery and characterization of new enzymes of natural product biosynthesis. The general framework for such an approach has also recently been applied for the discovery of a suite of new TPS genes in, for example, tomato (49) and *Euphorbia fischeriana* (50). Of fundamental importance for the success of a genomics- or transcriptomics-based approach to natural product enzyme discovery is the traditional or new knowledge of the occurrence of specialized metabolites associated with particular plant species or particular tissues. Balsam fir was ideally suited for the discovery of CAS because this species accumulates large amounts of *cis*-abienol **4** in the oleoresin of bark/phloem tissue (supplemental Table 2). Xylem/wood or needles of balsam fir do not contain *cis*-abienol **4** in relevant amounts and therefore were not included in the transcriptome sequencing for CAS discovery.

Beyond the selection of species and tissues via metabolite profiling prior to 454 sequencing, another key element in the successful discovery strategy of *AbCAS* was the application of a directed BLASTx search of the *de novo* transcriptome assembly. Specifically, we directed the BLASTx search of the *de novo* transcriptome assembly against a comprehensive sequence database of characterized TPSs. This strategy allowed for the curator-optimized annotation of TPSs genes, which substantially streamlined the efforts of functional gene characterization toward four candidate *AbdiTPS*s. Of these four candidates, the corresponding enzymes of *AbdiTPS1* (*AbLAS*), *AbdiTPS2* (*AbIso*), and *AbdiTPS4* (*AbCAS*) were successfully characterized.

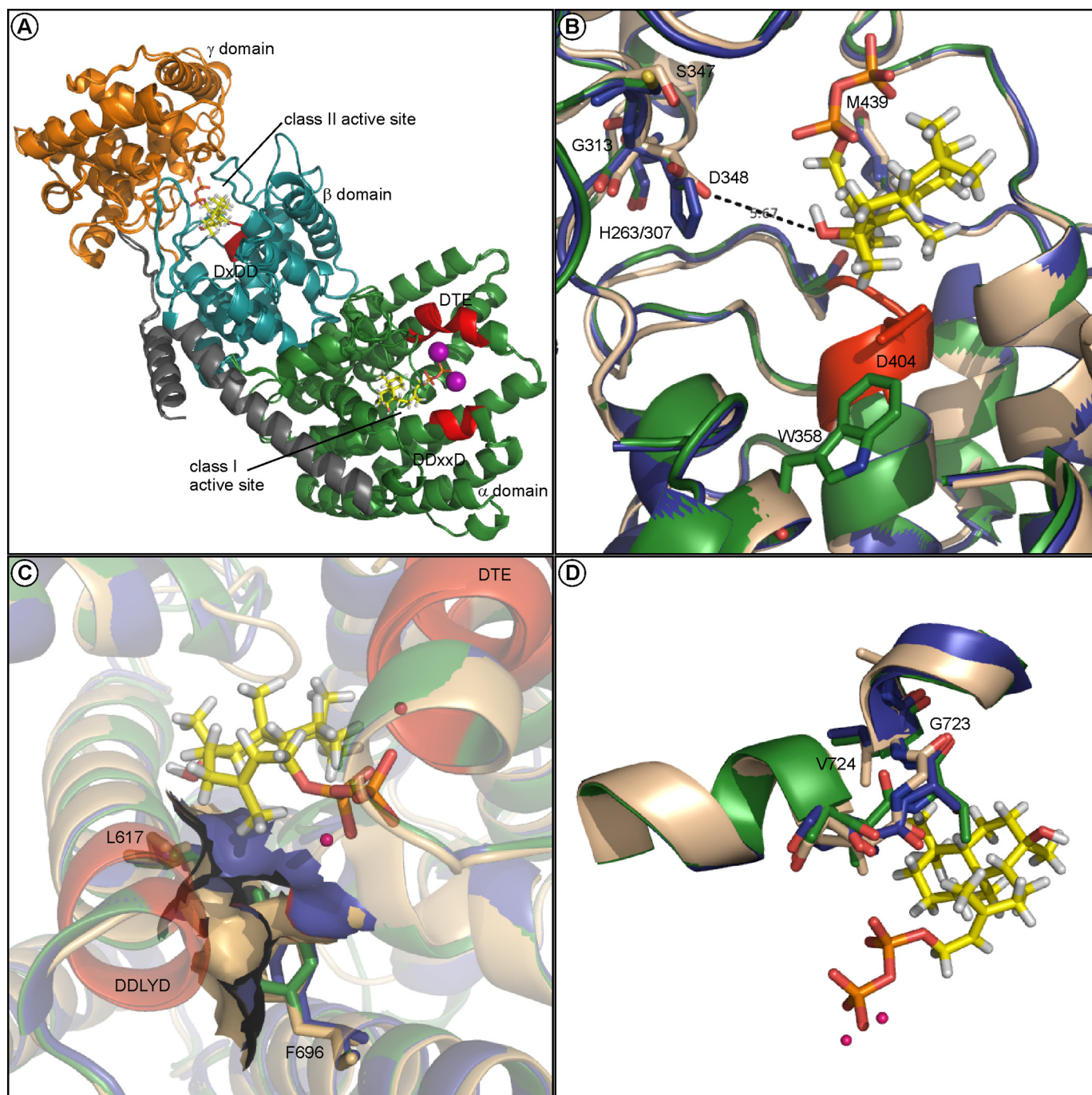
Based on the relatively high protein sequence identity of more than 90% with *AgAS*, a similar catalytic function was predicted for *AbdiTPS1* and *AbdiTPS2*, whereas the lower sequence identity of *AbdiTPS4* (75%) was indicative of a distinct function. This general prediction was confirmed by functional characterization, which remains a necessary part of TPS gene annotation. The functional characterization of *AbdiTPS1* (*AbLAS*) with 13-hydroxy-8(14)-abietene **9/10** as the initial reaction product and multiple olefin compounds occurring upon dehydration of the alcohol was important to substantiate a recent discovery of the formation of a tricyclic tertiary C-13 alcohol as the primary product of Norway spruce *PaLAS* (18). Likewise, the discovery of *AbdiTPS2* (*AbIso*) was important because it was the first Iso-type diTPS to be characterized outside of the spruce genus (*Picea*), with phylogenetic patterns suggesting that functional divergence of LAS- and Iso-type diTPS occurred independently in the spruce and fir (*Abies*) lineages (Fig. 4). In agreement with predictions based on sequence divergence, *AbdiTPS4* (*AbCAS*) represents a novel diTPS func-

tion. A bifunctional *cis*-abienol synthase has not been previously reported for any plant species. Together, the set of three types of bifunctional diTPSs, *AbIso*, *AbLAS*, and *AbCAS*, account for the majority of diterpene structures found in the specialized diterpene metabolism of balsam fir bark tissue.

In the absence of pure *cis*- or *trans*-abienol standards, stereochemistry of the product of *AbdiTPS4* was assigned based on comparison of retention times, under several different GC and LC conditions, with those of the *cis*-abienol metabolite in the resin mixture of balsam fir. Previous studies demonstrated that abienol from balsam fir exhibits the *cis*-configuration (51). Our proton and carbon NMR analysis further corroborated the stereochemistry of the reaction product of *AbdiTPS4* as *cis*-abienol.

Despite extensive efforts of TPS gene discovery in species of spruce (14, 17) and in grand fir (52), there is no known gene in any plant species that is closely related (>70% protein identity) to *AbCAS*. It is therefore possible that a CAS-type bifunctional class I/II diTPS is unique to balsam fir, which is the first conifer species for which three types of functionally distinct diTPSs are now known. A phylogenetic position of *AbCAS* between *Ginkgo biloba* *GbLS* and other known conifer diTPSs of specialized metabolism in spruce, fir, and pine (Fig. 4) may suggest an evolutionary origin of *AbCAS* by gene duplication prior to speciation within the pine family with its possible loss in some or all of the other species of this comparison.

Bifunctional class I/II diTPSs of conifer specialized metabolism are members of the TPS-d group (14, 19). Within this group, new genes and enzyme functions appear to have evolved from a common ancestor by repeated gene duplication. Neo- or sub-functionalization of members of this gene family involved mutation of specific active site residues (16, 53–55). The mutational analysis of *AbCAS* and comparative structural analysis of *AbCAS* with *AbIso* and *AbLAS* identified distinct catalytic mechanisms and revealed unique residues in the *AbCAS* class I and class II active sites that may determine product specificity. The initial bicyclization of GGPP **1**, resulting in a proposed labda-13-en-8-yl<sup>+</sup> intermediate **2** of (9*S*,10*S*) stereochemistry, is probably a common feature of the class II reactions of CAS-, LAS-, and Iso-type diTPSs (Fig. 1). Previous work showed that naturally occurring *cis*-abienol **4** is optically pure and represents the (9*S*,10*S*) configuration (51). In a unique sequence of the class II reaction, *AbCAS* may then promote water capture at the C-8 carbon of the carbocation to form labda-13-en-8-ol diphosphate **3**, whereas the LAS- and Iso-type diTPSs form (9*S*,10*S*)-CPP **5** through deprotonation at the C-8 position (Fig. 1). Structural modeling and molecular docking of labda-13-en-8-ol diphosphate **3** in the class I and class II active sites of *AbCAS* (Fig. 6) revealed only one unique amino acid in the class II active site, namely Asp-348, in proximity of the docked intermediate that may control this particular hydroxylation reaction. Located at the posterior of the active site opposite of the DIDD motif and Trp-358, which have previously been reported to contribute to the *AgAS*-catalyzed class II reaction (39), the negatively charged side chain of Asp-348 may stabilize the positive charge at C-8 for water quenching to occur in the formation of labda-13-en-8-ol diphosphate **3** (Fig. 6*B*). This local negative charge is not present in any LAS- or Iso-type enzymes,



**FIGURE 6. Unique active site residues of *AbcCAS* relative to *AblAS* and *AbISO* and implication for the formation of *cis*-abienol.** Structural models were built based on the tertiary structure of *T. brevifolia* taxadiene synthase (32) (Protein Data Bank code 3p5p, chain A). A, superimposed homology models of *AbcCAS*, *AblAS*, and *AbISO* resembling the common  $\alpha$ -helical folding pattern, comprised of the N-terminal  $\beta$  domain (cyan) and  $\gamma$  domain (orange) and C-terminal  $\alpha$  domain (green). The position of labda-13-en-8-ol diphosphate **3** (yellow) was estimated in the active sites by molecular docking using Molegro Virtual Docker 2010.4.0.0 (36), with the diphosphate group bound to the  $Mg^{2+}$  cluster (pink), which in return is coordinated by the DDXXD and NDXXTXXE motifs (red). B, superimposition of putative catalytic residues in the class II active site of *AblAS* (blue), *AbISO* (green), and *AbcCAS* (light brown). Asp-348 is located at the posterior of the class II active site, opposite the DIDD motif, with its side chain protruding toward C-8 of labda-13-en-8-ol diphosphate **3**. C, Leu-617 and Phe-696 are located in the class I active site cavity upstream of the DDLVD motif, creating an expansion of the hydrophobic pocket relative to *AblAS* and *AbISO*. D, Gly-723 and Val-724 contribute to a change in the hinge region between helix G1 and G2, which has previously been shown to be critical for the plasticity of conifer diTPSs (16, 51, 54).

which contain a conserved histidine in this position (supplemental Fig. 1).

Mutational analysis and complementation assays suggest that the class I active site of *AbcCAS* is selective for labda-13-en-8-ol diphosphate **3** as a substrate but does not convert (9*S*,10*S*)-

CPP **5**, which is the class I active site substrate of LAS-, and Iso-type diTPSs (Fig. 5). Unlike LAS- and Iso-enzymes, the class I active site of *AbcCAS* catalyzes the ionization of the diphosphate group without cyclization of a C-ring (Fig. 1). Several unique residues were found in a radius of 7 Å around the



hydroxy group of labda-13-en-8-ol diphosphate **3** docked within the class I active site of *AbCAS* that may control the *AbCAS* reaction (Fig. 6, *C* and *D*). Among these, Leu-617, Phe-696, and Gly-723 are of particular interest because they contribute to the contour of the active site cavity. Leu-617 and Phe-696 are located in the class I active site cavity upstream of the DDXXD motif on helix D, creating an expansion of the hydrophobic pocket relative to *AbLAS* and *AbIso*. Gly-723 and Val-724 account for a change in the hinge region between helix G1 and G2. These residues may contribute to the release of a bicyclic product rather than facilitating a secondary cyclization. Interestingly, residues corresponding to Leu-617, Phe-696, and Gly-723 have previously been shown to be critical for the catalytic plasticity of conifer diTPSs (16, 53–56) and thus appear to represent key positions for the functional evolution of these enzymes.

Plant TPSs are useful enzymes for the metabolic engineering of bioproducts and biofuels in yeast and *E. coli* (4, 57). Metabolic engineering of microbial hosts or plants to produce oxygenated diterpenoids, such as *cis*-abienol **4** or sclareol, may provide a sustainable production platform for these compounds for the fragrance industry and other applications. The patent literature (58, 59) suggests that genes for *cis*-abienol **4** and sclareol biosynthesis have been cloned from angiosperms. However, unlike the bifunctional conifer class I/II diTPSs, the known angiosperm diTPSs are exclusively monofunctional enzymes (19). Thus, using the angiosperm enzymes for pathway engineering would require the dual expression optimization of separate class I and class II enzymes. In contrast, the use of a bifunctional class I/II diTPS, such as *AbCAS*, would require only the expression of a single gene. Optimization of protein expression is also likely to be less complicated for the bifunctional class I/II diTPS (*i.e.* *AbCAS*), because the two active sites required for the conversion of GGPP **1** to *cis*-abienol **4** will be in equal amounts as part of the same protein and will be in close physical proximity of each other. In essence, for applications of metabolic engineering, the bifunctional class I/II *AbCAS* offers a scaffolded arrangement of the two essential active sites. This scaffold evolved in nature for the high volume production of *cis*-abienol **4** in the oleoresin of balsam fir. Because *AbCAS* is only very distantly related to angiosperm diTPSs (less than 30% protein sequence identity), this enzyme would also be a preferred choice for metabolic engineering of *cis*-abienol **4** production in industrial crops, such as tobacco, because co-suppression effects on endogenous diTPSs are unlikely.

**Acknowledgments**—We thank Karen Reid for excellent laboratory and project management support, Lina Madilao for support with GC-MS and LC-MS analysis, and Christopher I. Keeling for helpful discussions. Transcriptome sequencing was performed at the McGill University/Genome Quebec Innovation Centre with database assistance for handling of sequence raw data at the University of Calgary.

## REFERENCES

- Trapp, S., and Croteau, R. (2001) Defensive Resin Biosynthesis in Conifers. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **52**, 689–724
- Keeling, C. I., and Bohlmann, J. (2006) Diterpene resin acids in conifers. *Phytochemistry* **67**, 2415–2423
- Keeling, C. I., and Bohlmann, J. (2006) Genes, enzymes, and chemicals of terpenoid diversity in the constitutive and induced defence of conifers against insects and pathogens. *New Phytol.* **170**, 657–675
- Bohlmann, J., and Keeling, C. I. (2008) Terpenoid biomaterials. *Plant J.* **54**, 656–669
- Hillwig, M. L., Mann, F. M., and Peters, R. J. (2011) Diterpenoid biopolymers. New directions for renewable materials engineering. *Biopolymers* **95**, 71–76
- Gray, P. S., and Mills, J. S. (1964) The isolation of abienol from Canada Balsam, the oleoresin of *Abies balsamea* (L.) *J. Chem. Soc.* **1964**, 5822–5825
- Barrero, A. F., Alvarez-Manzaneda, E. J., Altarejos, J., Salido, S., and Ramos, J. M. (1993) Synthesis of Ambrox® from (–)-sclareol and (+)-*cis*-abienol. *Tetrahedron* **49**, 10405–10412
- Guo, Z., Severson R. F., and Wagner, G. J. (1994) Biosynthesis of the diterpene *cis*-abienol in cell-free extracts of tobacco trichomes. *Arch. Biochem. Biophys.* **308**, 103–108
- Guo, Z., and Wagner, G. J. (1995) Biosynthesis of labdenediol and sclareol in cell-free extracts from trichomes of *Nicotiana glutinosa*. *Planta* **197**, 627–632
- Miyazawa, M., and Tamura, N. (2008) *J. Essent. Oil Res.* **20**, 12–14
- Vogel, B. S., Wildung, M. R., Vogel, G., and Croteau, R. (1996) Abietadiene synthase from grand fir (*Abies grandis*). cDNA isolation, characterization, and bacterial expression of a bifunctional diterpene cyclase involved in resin acid biosynthesis. *J. Biol. Chem.* **271**, 23262–23268
- Peters, R. J., Flory, J. E., Jetter, R., Ravn, M. M., Lee, H. J., Coates, R. M., and Croteau, R. B. (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. B. (2001) Bifunctional abietadiene synthase: free diffusive transfer of the (+)-copalyl diphosphate intermediate between two distinct active sites. *J. Am. Chem. Soc.* **123**, 8974–8978
- 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
- Ro, D. K., and Bohlmann, J. (2006) Diterpene resin acid biosynthesis in loblolly pine (*Pinus taeda*). Functional characterization of abietadiene/levopimaradiene synthase (PtTPS-LAS) cDNA and subcellular targeting of PtTPS-LAS and abietadienol/abietadienol oxidase (PtAO, CYP720B1). *Phytochemistry* **67**, 1572–1578
- 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. U.S.A.* **105**, 1085–1090
- 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.* **11**, 43
- Keeling, C. I., Madilao, L. L., Zerbe, P., Dullat, H. K., and Bohlmann, J. (2011) The primary diterpene synthase products of *Picea abies* levopimaradiene/abietadiene synthase (PaLAS) are epimers of a thermally unstable diterpenol. *J. Biol. Chem.* **286**, 21145–21153
- Chen, F., Tholl, D., Bohlmann, J., and Pichersky, E. (2011) The family of terpene synthases in plants. A mid-size family of genes for specialized metabolism that is highly diversified throughout the kingdom. *Plant J.* **66**, 212–229
- Ro, D. K., Arimura, G., Lau, S. Y., Piers, E., and Bohlmann, J. (2005) Loblolly pine abietadienol/abietadienol oxidase PtAO (CYP720B1) is a multifunctional, multisubstrate cytochrome P450 monooxygenase. *Proc. Natl. Acad. Sci. U.S.A.* **102**, 8060–8065
- Hamberger, B., and Bohlmann, J. (2006) Cytochrome P450 mono-oxygenases in conifer genomes. Discovery of members of the terpenoid oxygenase superfamily in spruce and pine. *Biochem. Soc. Trans.* **34**, 1209–1214
- Hamberger, B., Ohnishi T., Hamberger, B., Séguin, A., and Bohlmann, J. (2011) Evolution of diterpene metabolism. Sitka spruce CYP720B4 catalyzes multiple oxidations in resin acid biosynthesis of conifer defense against insects. *Plant Physiol.* **157**, 1677–1695



23. 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
24. Mafu, S., Hillwig, M. L., and Peters, R. J. (2011) A novel labda-7,13e-dien-15-ol-producing bifunctional diterpene synthase from *Selaginella moellendorffii*. *Chembiochem.* **12**, 1984–1987
25. 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 putative octadecanoid pathway transcripts in Sitka spruce. *Plant Physiol.* **137**, 369–382
26. Lewinsohn, E., Gijzen, M., Muzika, R. M., Barton, K., and Croteau, R. (1993) Oleoresinosis in grand fir (*Abies grandis*) saplings and mature trees (modulation of this wound response by light and water stresses). *Plant Physiol.* **101**, 1021–1028
27. Kolosova, N., Miller, B., Ralph, S., Ellis, B.E., Douglas, C., Ritland, K., and Bohlmann, J. (2004) Isolation of high quality RNA from gymnosperm and angiosperm trees. *BioTechniques* **36**, 821–824
28. Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990) Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410
29. De la Bastide, M., and McCombie, W. R. (2007) Assembling genomic DNA sequences with PHRAP. *Curr. Protoc. Bioinformatics* Chapter 11, Unit 11.4
30. Guindon, S., Dufayard, J.F., Lefort, V., Anisimova, M., Hordijk, W., and Gascuel, O. (2010) New algorithms and methods to estimate maximum-likelihood phylogenies. Assessing the performance of PhyML 3.0. *Syst. Biol.* **59**, 307–321
31. Nielsen, M., Lundegaard, C., Lund, O., and Petersen, T. N. (2010) CPH-models-3.0. Remote homology modeling using structure-guided sequence profiles. *Nucleic Acids Res.* **38**, W576–W581
32. Köksal, M., Jin, Y., Coates, R. M., Croteau, R., and Christianson, D. W. (2011) Taxadiene synthase structure and evolution of modular architecture in terpene biosynthesis. *Nature* **469**, 116–120
33. Laskowski, R. A., MacArthur, M. W., Moss, D. S., and Thornton, J. M. (1993) PROCHECK: A program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.* **26**, 283–291
34. Wiederstein, M., and Sippl, M.J. (2007) ProSA-web. Interactive web service for the recognition of errors in three-dimensional structures of proteins. *Nucleic Acids Res.* **35**, W407–W410
35. Holm, L., and Park, J. (2000) DaliLite workbench for protein structure comparison. *Bioinformatics* **16**, 566–567
36. Thomsen, R., and Christensen, M.H. (2006) MolDock. A new technique for high accuracy molecular docking. *J. Med. Chem.* **49**, 3315–3321
37. Schüttelkopf, A. W., and van Aalten, D. M. (2004) PRODRG. A tool for high-throughput crystallography of protein-ligand complexes. *Acta Crystallogr. D Biol. Crystallogr.* **60**, 1355–1363
38. Peters, R. J., Carter, O. A., Zhang, Y., Matthews, B. W., and Croteau, R. B. (2003) Bifunctional abietadiene synthase. Mutual structural dependence of the active sites for protonation-initiated and ionization-initiated cyclizations. *Biochemistry* **42**, 2700–2707
39. Peters, R. J., and Croteau, R. B. (2002) Abietadiene synthase catalysis. Conserved residues involved in protonation-initiated cyclization of geranylgeranyl diphosphate to (+)-copalyl diphosphate. *Biochemistry* **41**, 1836–1842
40. Zhou, K., and Peters, R.J. (2009) Investigating the conservation pattern of a putative second terpene synthase divalent metal binding motif in plants. *Phytochemistry* **70**, 366–369
41. Vlad, P. F., Khariton, K. H. S. H., Koltsa, M. N., and Bordakh, O. D. (1974) Mass-spectrometric investigation of the abienols. Diterpene alcohols. *Khimiya Prirodnykh Soedinenii* **1**, 30–35
42. Carman, R., and Duffield, A. (1993) The biosynthesis of labdanoids. The optical purity of naturally occurring manool and abienol. *Aust. J. Chem.* **46**, 1105–1114
43. Severson, R. F., Arrendale, R. F., Chortyk, O. T., Johnson, A. W., Jackson, D. M., Gwynn, G. R., Chaplin, J. F., and Stephenson, M. G. (1984) Quantitation of the major cuticular components from green leaf of different tobacco types. *J. Agric. Food Chem.* **32**, 566–570
44. Ding, L., Xie, F., Zhao, M., Xie, J., and Xu, G. (2007) Quantification of *cis*-abienol in oriental tobacco leaves by LC. *Chromatographia* **66**, 529–532
45. Mills, J. S. (1967) The conversion of *cis*- into *trans*-abienol. Some reactions with mercuric acetate. *J. Chem. Soc.* **23**, 2514–2520
46. Yang, X. W., Feng, L., Li, S. M., Liu, X. H., Li, Y. L., Wu, L., Shen, Y. H., Tian, J. M., Zhang, X., Liu, X. R., Wang, N., Liu, Y., and Zhang, W. D. (2010) Isolation, structure, and bioactivities of abiesadines A-Y, 25 new diterpenes from *Abies georgei* Orr. *Bioorg. Med. Chem.* **18**, 744–754
47. Hayashi, K., Kawaide, H., Notomi, M., Sakigi, Y., Matsuo, A., and Nozaki, H. (2006) Identification and functional analysis of bifunctional entkaurene synthase from the moss *Physcomitrella patens*. *FEBS Lett.* **580**, 6175–6181
48. Kawaide, H., Hayashi, K., Kawanabe, R., Sakigi, Y., Matsuo, A., Natsume, M., and Nozaki, H. (2011) Identification of the single amino acid involved in quenching the ent-kaurenylation by a water molecule in ent-kaurene synthase of *Physcomitrella patens*. *FEBS J.* **278**, 123–133
49. Bleeker, P. M., Spyropoulou, E. A., Diergaarde, P. J., Volpin, H., De Both, M. T., Zerbe, P., Bohlmann, J., Falara, V., Matsuba, Y., Pichersky, E., Haring, M. A., and Schuurink, R. C. (2011) RNA-seq discovery, functional characterization, and comparison of sesquiterpene synthases from *Solanum lycopersicum* and *Solanum habrochaites* trichomes. *Plant Mol. Biol.* **77**, 323–336
50. Barrero, R. A., Chapman, B., Yang, Y., Moolhuijzen, P., Keeble-Gagnère, G., Zhang, N., Tang, Q., Bellgard, M. L., and Qiu, D. (2011) *De novo* assembly of *Euphorbia fischeriana* root transcriptome identifies prostratin pathway-related genes. *BMC Genomics* **12**, 600
51. Carman, R. M. (1966) The stereochemistry and absolute configuration of abienol. *Aust. J. Chem.* **19**, 1535–1537
52. 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
53. Wilderman, P. R., and Peters, R. J. (2007) A single residue switch converts abietadiene synthase into a pimaradiene-specific cyclase. *J. Am. Chem. Soc.* **129**, 15736–15737
54. Peters, R.J., and Croteau, R.B. (2002) Abietadiene synthase catalysis. Mutational analysis of a prenyl diphosphate ionization-initiated cyclization and rearrangement. *Proc. Natl. Acad. Sci. U.S.A.* **99**, 580–584
55. Leonard, E., Ajikumar, P. K., Thayer, K., Xiao, W. H., Mo, J. D., Tidor, B., Stephanopoulos, G., and Prather, K. L. (2010) Combining metabolic and protein engineering of a terpenoid biosynthetic pathway for overproduction and selectivity control. *Proc. Natl. Acad. Sci. U.S.A.* **107**, 13654–13659
56. Zerbe, P., Chiang, A., and Bohlmann, J. (2012) Mutational analysis of white spruce (*Picea glauca*) ent-kaurene synthase (PgKS) reveals common and distinct mechanisms of conifer diterpene synthases of general and specialized metabolism. *Phytochemistry* **74**, 30–39
57. Peralta-Yahya, P. P., Ouellet, M., Chan, R., Mukhopadhyay, A., Keasling, J. D., and Lee, T. S. (2011) Identification and microbial production of a terpene-based advanced biofuel. *Nat. Commun.* **2**, 483
58. Sallaud, C., Rontein, D., and Tissier, A. (January 1, 2008) International Patent WO/2008/007031
59. Schalk, M. (August 20, 2009) International Patent WO 2009/101126 A1