Biochemistry and Metabolism

Running title: Variation in cis-prenyltransferase activity

For Correspondence:

Cornelius S. Barry,
Department of Horticulture,
Michigan State University,
East Lansing,
MI 48824,
USA,
Phone: (1) 517 884 6959
Fax: (1) 517 353 0890
Email: barrycs@msu.edu
Determination of residues responsible for substrate and product specificity of *Solanum habrochaites* short-chain *cis*-prenyltransferases

Jin-Ho Kang†, Eliana Gonzales-Vigil‡, Yuki Matsuba, Eran Pichersky and Cornelius S. Barry*

Department of Horticulture (J.-H.K., E.G.-V., C.S.B.), Michigan State University, East Lansing, MI 48824, USA and Department of Molecular, Cellular and Developmental Biology (Y.M., E.P.), University of Michigan, Ann Arbor, MI 48109, USA

* For Correspondence: Cornelius Barry, Department of Horticulture, Michigan State University, East Lansing, MI 48824, USA.
Phone: (1) 517 884 6959
Fax: (1) 517 353 0890
Email: barrycs@msu.edu

**One Sentence Summary:** The relative positions of aromatic amino acids and adjacent residues within domain II of short-chain *cis*-prenyltransferases contributes to the evolution of volatile terpene biosynthesis in *Solanum* trichomes.
Financial Source: This research was supported by NSF IOS-1025636 to C.S.B. and E.P. and a Strategic Partnership Grant from the Michigan State University Foundation to C.S.B..

† These authors contributed equally to this research.
‡ Present Address: Department of Wood Science, The University of British Columbia, Vancouver, British Columbia, Canada.
*To whom correspondence should be addressed. Email: barrycs@msu.edu
ABSTRACT

Isoprenoids are diverse compounds that have their biosynthetic origin in the initial condensation of isopentenyl diphosphate (IPP) and dimethylallyl diphosphate (DMAPP) to form C10 prenyl diphosphates that can be elongated by the addition of subsequent IPP units. These reactions are catalyzed by either *cis*-prenyltransferases (CPTs) or *trans*-prenyltransferases (TPTs). The synthesis of volatile terpenes in plants typically proceeds through either geranyl diphosphate (C10) or *trans*-farnesyl diphosphate (C15), to yield monoterpenes and sesquiterpenes, respectively. However, terpene biosynthesis in glandular trichomes of tomato (*Solanum lycopersicum*) and related wild relatives also occurs via the *cisoid* substrates neryl diphosphate (NPP) and *2Z,6Z*-farnesyl diphosphate (*Z,Z*-FPP). NPP and *Z,Z*-FPP are synthesized by neryl diphosphate synthase 1 (NDPS1) and *Z,Z*-farnesyl diphosphate synthase (zFPS), which are encoded by the orthologous *CPT1* locus in tomato and *Solanum habrochaites*, respectively. In this study, comparative sequence analysis of NDPS1 and zFPS enzymes from *S. habrochaites* accessions that synthesize either monoterpenes or sesquiterpenes was performed to identify amino acid residues that correlate with the ability to synthesize NPP or *Z,Z*-FPP. Subsequent structural modeling, coupled with site-directed mutagenesis highlighted the importance of four amino acids located within conserved domain II of CPT enzymes that form part of the second alpha helix, for determining substrate and product specificity of these enzymes. In particular, the relative positioning of aromatic amino acid residues at positions 100 and 107 determines the ability of these enzymes to synthesize NPP or *Z,Z*-FPP. This study provides insight into the biochemical evolution of terpene biosynthesis in the glandular trichomes of *Solanum* species.
INTRODUCTION

Plant isoprenoids are structurally heterogeneous, giving rise to pigments, hormones, quinones, and sterols, together with a variety of specialized metabolites that are often restricted to specific genera or families (Kirby and Keasling, 2009). Among these specialized metabolites are volatile short-chain terpenoids with chain lengths of C10, C15, and C20 defined as mono-, sesqui-, and diterpenes, respectively. Subsequently, these hydrocarbons can be modified by cleavage, oxidation, acylation, or glycosylation to yield diverse volatile and non-volatile compounds with varied activities (Richman et al., 2005; Heiling et al., 2010; Lee et al., 2010; Nguyen et al., 2010; Lange and Turner, 2013). Terpenoids possess multiple roles in plants, contributing to floral scents that attract pollinators and aroma compounds that facilitate seed dispersal (Aharoni et al., 2004; Gang, 2005; Dudareva et al., 2013). In addition, terpenoids can be toxic, or act as repellants or signaling molecules that deter or limit herbivory (Schnee et al., 2006; Degenhardt et al., 2009; Bleeker et al., 2011; Schmelz et al., 2011). The chemical properties of terpenoids have led to their exploitation by humans as flavorings, fragrances, medicines and biofuels, with considerable interest focused in engineering their biosynthesis to improve production efficiencies (Wu et al., 2006; Kirby and Keasling, 2009; Niehaus et al., 2011; Zhang et al., 2011; Westfall et al., 2012; Lange and Ahkami, 2013; Yeo et al., 2013).

The C5 compounds isopentenyl diphosphate (IPP) and dimethylallyl diphosphate (DMAPP) serve as the precursors of isoprenoids and can be formed through either the plastid localized 2-C-methyl-D-erythritol 4-phosphate pathway (MEP) or the cytosolic mevalonate pathway (Rodriguez-Concepcion and Boronat, 2002; Eisenreich et al., 2004; Chen et al., 2011). Condensation of IPP and DMAPP to form either geranyl diphosphate (GPP) in the trans configuration or neryl diphosphate (NPP) in the cis configuration represents the initial step of isoprenoid biosynthesis and these reactions are catalyzed by trans-prenyltransferases (TPTs) or cis-prenyltransferases (CPTs), respectively (Burke et al., 1999; Oh et al., 2000; Schilmiller et al., 2009; Akhtar et al., 2013). Successive head to tail additions of IPP units in either cis or trans configurations...
determine the final chain length of isoprenoids which typically ranges between C15 and C120, but can be greater than C10,000 in the case of natural rubber (Takahashi and Koyama, 2006).

TPTs and CPTs are distinct and operate through different catalytic mechanisms (Takahashi and Koyama, 2006; Liang, 2009). Several plant TPTs have been characterized including those involved in the synthesis of GPP, 2\(E,6E\)-farnesyl diphosphate \((E,E\text{-FPP})\), geranylgeranyl diphosphate (C20) and solanesyl diphosphate (C45) that serve as the precursors of terpenes, sterols, hormones and carotenoids (Burke et al., 1999; Hirooka et al., 2003; Lange and Ghassemian, 2003; Ducluzeau et al., 2012). Similarly, plant CPTs involved in polyisoprenoid biosynthesis have been identified, including those involved in the synthesis of rubber (Cunillera et al., 2000; Asawatreratanakul et al., 2003; Schmidt et al., 2010; Surmacz and Swiezewska, 2011; Kera et al., 2012; Post et al., 2012). Recently, characterization of the seven member CPT family of tomato \((Solanum lycopersicum)\) led to the identification of four SlCPT enzymes involved in the synthesis of polyisoprenoids, including SlCPT3 that rescues a dolichol deficient mutant of yeast (Akhtar et al., 2013). A further three enzymes, SlCPT1 (neryl diphosphate synthase 1 (NDPS1)), SlCPT2, and SlCPT6, synthesize the short chain \(cis\)-prenyl diphosphates, neryl diphosphate (NPP), nerylnereryl diphosphate and \(2Z,6Z\)-farnesyl diphosphate \((Z,Z\text{-FPP})\), respectively and each gene is associated with a terpene synthase gene cluster (Schilmiller et al., 2009; Falara et al., 2011; Akhtar et al., 2013; Matsuba et al., 2013).

Plants typically synthesize volatile terpenes from either GPP or \(E,E\text{-FPP}\) giving rise to diverse mono- and sesquiterpenes, respectively (Chen et al., 2011). However, type VI glandular trichomes of tomato and closely related species synthesize mono- and sesquiterpenes from the \(cisoid\) substrates NPP and \(Z,Z\text{-FPP}\) that are formed in the plastid through the action of NDPS1 and \(Z,Z\text{-farnesyl diphosphate synthase (zFPS), respectively (Sallaud et al., 2009; Schilmiller et al., 2009; Fig. 1). NDPS1 is highly expressed in type VI glandular trichomes of \(S. lycopersicum\) where it acts together with the terpene synthase \(\beta\text{-phellandrene synthase 1 (PHS1) to form }\beta\text{-phellandrene and
several additional monoterpenes (Schilmiller et al., 2009). The loci encoding these enzymes reside within a gene cluster on chromosome 8 and an orthologous cluster in S. pennellii synthesizes a mixture of terpenes in which α-phellandrene predominates (Falara et al., 2011). In contrast, in S. habrochaites accession LA1777, zFPS catalyzes the formation of Z,Z-FPP, which is subsequently cyclized by santalene / bergamotene synthase (SBS) to form several sesquiterpenes (Sallaud et al., 2009). These enzymes are also encoded by genes that reside within the chromosome 8 TPS cluster and while PHS1 and SBS are encoded by separate but tightly linked loci (TPS20 and TPS45, respectively), NDPS1 and zFPS are orthologous and are both encoded by the CPT1 locus within this cluster, suggesting that their different activities arose through sequence divergence alone (van der Hoeven et al., 2000; Sallaud et al., 2009; Matsuba et al., 2013).

Metabolite profiling of 79 S. habrochaites accessions revealed considerable intra-specific qualitative and quantitative variation in terpene synthesis leading to the subsequent isolation of a second sesquiterpene synthase, zingiberene synthase (ShZIS), which is an allele of TPS45 locus and encodes an enzyme that also utilizes Z,Z-FPP as a substrate (Gonzales-Vigil et al., 2012; Matsuba et al., 2013). In addition, three sequences that encode monoterpane synthases (ShPHS1, ShPIS and ShLIS) that utilize NPP to synthesize β-phellandrene, α-pinene, and limonene, respectively were also identified and shown to be alleles of the TPS20 locus (Gonzales-Vigil et al., 2012, Matsuba et al., 2013). The presence of monoterpane synthases that use NPP as their substrate suggested the existence in S. habrochaites accessions of an NDPS1 enzyme rather than zFPS. In the present study, cDNAs corresponding to the CPT1 locus were isolated from the trichomes of chemically diverse S. habrochaites accessions and comparative sequence analysis, together with homology modeling were employed to identify specific amino acid residues that correlate with either NDPS1 or zFPS activity. The role of these residues in determining substrate and product specificity of short-chain CPTs was confirmed through site-directed mutagenesis, revealing an essential role for the relative positions of aromatic amino acids within a hydrophobic cleft between
helices II and III. These data provide insight into the biochemical evolution of trichome-derived specialized metabolites that function as insect deterrents.

RESULTS

Identification and characterization of NDPS1 from S. habrochaites

To identify sequences encoding NDPS1 enzymes from S. habrochaites and to begin to address the evolutionary relationship of NDPS1 and zFPS, cDNAs from the CPT1 locus were amplified and sequenced from multiple accessions of S. habrochaites that predominantly synthesize either monoterpenes or sesquiterpenes from the cisoid pathway. In total, 25 cDNAs were recovered from 17 accessions that share between 97 and 100% identity at the nucleotide level. Phylogenetic analysis revealed that these sequences group into two major clades (Fig. 2). One clade is defined by the presence of the previously characterized zFPS gene from S. habrochaites accession LA1777 (Sallaud et al., 2009) and the second, containing fewer sequences, is more closely related to NDPS1 from S. lycopersicum (Schilmiller et al., 2009).

The phylogeny of the S. habrochaites CPT1 sequences suggest that several may encode enzymes with NDPS1 activity and this hypothesis was tested by expressing a codon optimized synthetic version of CPT1 from S. habrochaites accession LA2409, a monoterpenes producing accession (Gonzales-Vigil et al., 2012), in E. coli and assaying the activity of the purified recombinant enzyme. Upon incubation with DMAPP and IPP at equimolar concentrations, the recombinant enzyme synthesized the C10 isoprenoid NPP that was detected as nerol following dephosphorylation, indicating that the enzyme possesses NDPS1 activity (Fig. 3A). Furthermore, as the ratio of IPP to DMAPP was increased, the product specificity of the enzyme remained unchanged. In contrast, purified recombinant enzyme derived from a codon optimized version of CPT1 from S. habrochaites accession LA1393, a sesquiterpene producing accession that is identical to the previously characterized zFPS enzyme from S. habrochaites accession LA1777 (Sallaud et al., 2009; Gonzales-Vigil et al., 2012), preferentially synthesized the C15 product, Z,Z-FPP (Fig. 3A). The activities of CPT1 from LA2409 and LA1393 were also determined utilizing NPP and IPP as substrates. As previously documented, zFPS is able to synthesize Z,Z-FPP from NPP and IPP (Sallaud et al., 2009) although CPT1
from LA2409 is unable to utilize this substrate combination (Fig. 3B). These data indicate that \textit{CPT1} from \textit{S. habrochaites} accession LA2409 encodes an enzyme with NDPS1 activity.

\textit{Identification of amino acid residues important for the activity of NDPS1 and zFPS}  
The predicted protein sequences of NDPS1 encoded by \textit{CPT1} from accession LA2409 and zFPS encoded by \textit{CPT1} from accession LA1393 are highly similar to one another, sharing 94% identity and 97% similarity over the entire length of their sequence, which corresponds to 18 amino acid differences. A comparative sequence approach was utilized to further refine the identity of the amino acids responsible for specifying either NDPS1 and zFPS activity. Amino acid alignments of putative NDPS1 and zFPS proteins of \textit{S. habrochaites} were investigated for conserved residues that correlate with both the ability of the parent accession to synthesize either monoterpenes or sesquiterpenes (Fig. 2), and in the case of putative \textit{S. habrochaites} NDPS1 proteins, also possess an identical residue at the corresponding position in \textit{S. lycopersicum} NDPS1. In total, 11 residues that reside within the predicted mature protein of NDPS1 and zFPS matched these criteria and these were generally spaced throughout the protein sequence but several were located within conserved domains II, III, and V of CPTs (Table I; Supplemental Fig. S1).

This analysis was followed by changing 10 residues that correlate with changes in substrate in NDPS1 from accession LA2409 to those found in zFPS from accession LA1393 (Pro/Ser45, which corresponds to the second residue in the predicted mature protein, was excluded), and we made the reciprocal changes in zFPS (LA1393). The two resulting recombinant proteins were named NDPS1-M10 and zFPS-M10. The activities of the two M10 variants were determined and compared to the activities of the parent enzymes. An IPP:DMAPP ratio of 5:1 was utilized for these experiments as it corresponds to the typical ratio synthesized by 4-hydroxy-3-methyl-2-(E)-butenyl-diphosphate reductase (Rohdich et al., 2002). The NDPS1-M10 variant was able to utilize DMAPP and IPP to synthesize \(Z,Z\)-FPP at levels that are similar to that of the zFPS enzyme from LA1393 (Fig. 4A). The ability of NDPS1-M10 to synthesize \(Z,Z\)-FPP
coincided with a greatly reduced ability to synthesize NPP. Similarly, the zFPS-M10 recombinant protein lost the ability to synthesize Z,Z-FPP but gained NDPS1 activity and synthesized NPP at levels similar to NDPS1 from accession LA2409 (Fig. 4A). Furthermore, substitution of the M10 residues also changed the substrate specificity of NDPS1 and zFPS with respect to their ability to utilize NPP. For example, the NDPS1-M10 variant gained the ability to synthesize Z,Z-FPP from NPP and IPP, whereas this activity was lost in the zFPS-M10 variant (Fig. 4B). Together, these data illustrate that substitution of the 10 conserved amino acids influenced the substrate and product specificities of trichome expressed short chain CPTs and were sufficient to interchange NDPS1 and zFPS activities.

Homology modeling of NDPS1 and zFPS
To attempt to refine which subset of the M10 amino acids determine the substrate and product specificities of NDPS1 and zFPS, homology models were generated to visualize the relative positions of these amino acids within three-dimensional space (Fig. 5). Currently, there are no available crystal structures of plant CPTs. However, BLAST searches of the protein data bank database revealed that NDPS1 and zFPS, without the chloroplast transit peptide, each share approximately 40% amino acid identity with an *E. coli* undecaprenyl pyrophosphate synthase (PDB: 1X07A), which catalyzes the formation of a C55 undecaprenyl pyrophosphate involved in peptidoglycan synthesis in the bacterial cell wall (Guo et al., 2005). Homology models of NDPS1 and zFPS were obtained using the SWISS-MODEL Workspace with QMEAN z-score values of -2.41 and -2.20 for NDPS1 and zFPS, respectively, which represent acceptable models.

In general, the models of NDPS1 and zFPS aligned to the template with reasonable overlap of the predicted helices observed (Fig. 5A and B). However, the third alpha-helix of both NDPS1 and zFPS is considerably shorter than that observed in the *E. coli* undecaprenyl pyrophosphate synthase and instead of an extended helix structure, the sequence of both NDPS1 and zFPS form disordered loops. The dimension of the hydrophobic cleft formed between helices II and III is known to influence the product chain length of CPTs (Kharel et al., 2006; Noike et al., 2008). Recently, six additional
tomato CPTs (SlCPT2 through SlCPT7) were identified and characterized that synthesize prenyl diphosphates ranging between C15 and C65 in length (Akhtar et al., 2013). The enzymes that synthesize the shorter chain length products lack amino acid residues within a variable domain downstream of CPT conserved domain III. Homology models with acceptable QMEAN z-score values ranging between -1.52 and -3.44 were generated between five additional tomato CPT enzymes (SlCPT2, 4, 5, 6 and 7) and various CPT templates (Supplemental Fig. S2). A reliable model was not obtained for SlCPT3. Structure homology models of SlCPT4, and SlCPT5, which synthesize C55 and C60 products (Akhtar et al., 2013), each possess a complete third alpha-helix that resembles that of the *E. coli* undecaprenyl pyrophosphate synthase whereas SlCPT2, SlCPT6 and SlCPT7, which synthesize short chain prenyl diphosphates of between C15 and C35 in length, lack a complete third alpha-helix (Supplemental Fig. S2). The incomplete alpha-helix in NDPS1, SlCPT2, SlCPT6 and SlCPT7 is the result of a shorter intervening sequence downstream of CPT conserved domain III (Akhtar et al., 2013; Fig. 5 and Supplemental Fig. S2). Together, these data suggest a correlation between the length of the third alpha-helix in the tomato CPTs, the size of the resulting hydrophobic cleft between the second and third helices, and the product chain length.

The structural models of NDPS1 and zFPS are similar to each other with no difference in the length of the third alpha helix predicted (Fig. 5A and B), suggesting that the different catalytic activities of these enzymes is unlikely to be caused by major structural changes within this region. However, several of the divergent amino acids between NDPS1 and zFPS are located within CPT conserved domain II and III, which lie within or adjacent to either the second or third alpha helix (Fig. 5; Table I). In particular, the relative positions of the aromatic amino acids corresponding to Tyr100 in NDPS1 and Phe107 in zFPS, which are located within the second alpha helix within CPT conserved domain II, differ substantially between the two proteins (Fig. 5; Table I). Notably, the presence of bulky amino acids within the hydrophobic cleft between helices II and III is known to influence product chain length of CPTs (Kharel et al., 2006; Noike et al., 2008), suggesting that the relative positions of NDPS1 Tyr100 and zFPS Phe107 may impact the substrate and product specificities of these enzymes.
The relative positions of residues within region II determine the substrate and product specificities of NDPS1 and zFPS

As the domain II residues, and particularly the relative positions of Tyr100 and Phe107, constitute the major differences between the NDPS1 and zFPS structural models (Fig. 5), we converted the four conserved amino acids within domain II, Glu98, Tyr100, Ile106 and Ile107 of NDPS1 to those present in zFPS, generating NDPS1-M4 (Fig. 6A). We also converted these four domain II amino acids in zFPS-M4 to the four residues from NDPS1. Incubation of zFPS-M4 with DMAPP and IPP led to the synthesis of NPP at a similar amount to that observed for NDPS1 from LA2409 (Fig. 6B). Furthermore, recombinant zFPS-M4 protein was unable to utilize NPP as a substrate (Fig. 6C). Conversely, the NDPS1-M4 recombinant enzyme displays catalytic properties that are similar to zFPS and is able to utilize both DMAPP + IPP and NPP + IPP to synthesize Z,Z-FPP. Together, these data indicate that the residues within conserved domain II of NDPS1 and zFPS, which lie within the second alpha-helix, are important for determining substrate and product specificity of these short chain CPTs.

To test the hypothesis that the relative positions of Tyr100 in NDPS1 and Phe107 in zFPS are responsible for determining their activities, site-directed mutagenesis was performed to generate mutant versions of each enzyme in which the positions of these aromatic amino acids were switched and replaced with those present in the other enzyme (Fig. 6A). The resultant NDPS1 Tyr100Ser, Ile107Phe and zFPS Ser100Tyr, Phe107Ile constructs were designated NDPS1-M2 and zFPS-M2, respectively. The activity of each of the M2 constructs was determined using both DMAPP and IPP and NPP and IPP as substrates. Whereas the NDPS1-M4 construct gained the activity typically associated with zFPS, the M2 construct synthesized a mixture of C10 and C15 prenyl diphosphates when using DMAPP and IPP as substrates (Fig. 6B). In addition, while the NDPS1-M2 enzyme was able to utilize NPP as a substrate to synthesize Z,Z-FPP, this reaction was less efficient than was observed with the NDPS1-M4 and zFPS enzymes (Fig. 6C). In contrast, the activity of the zFPS-M2 enzyme was similar to that of zFPS-M4 and mirrored that of the wild type NDPS1 enzyme (Fig. 6). However, the
zFPS-M2 enzyme also retained a residual level of activity when supplied with NPP and IPP, although this was minor when compared to the activity of zFPS from LA1393 (Fig. 6C). In combination, these data indicate that while the relative positioning of the bulky aromatic amino acids between the predicted second and third helices of these short chain CPTs are sufficient to convert the substrate and product specificity of zFPS to that of NDPS1, they are insufficient to confer the complete reciprocal change to NDPS1.

The dimensions and structure of the hydrophobic cleft of CPTs may be particularly important for specifying enzyme activity when multiple rounds of chain elongation occur, as is the case when Z,Z-FPP is synthesized from DMAPP and IPP. As such, the amino acids immediately flanking the aromatic residues within conserved domain II could be important for determining the CPT activity through modifying their spatial orientation. To examine the role of the additional residues within the hydrophobic cleft in more detail, two additional mutants of NDPS1; NDPS1-M3V1 (Tyr100Ser, Ile106Leu, Ile107Phe) and NDPS1-M3V2 (Glu98Asp, Tyr100Ser, Ile107Phe) were constructed (Fig. 6A). In particular, we were interested to determine whether the Ile106Leu substitution, which lies immediately adjacent to Phe107 in the NDPS1-M4 and NDPS1-M2 constructs, influences the activity of these recombinant enzymes.

The activity of the resulting NDPS1-M3V1 recombinant enzyme more closely resembled zFPS and produced predominantly Z,Z-FPP, but still produced NPP as approximately 10% of the total product whereas the NDPS1-M3V2 enzyme produces NPP and Z,Z-FPP in approximately equal proportions (Fig. 6B). Similarly, the NDPS1-M3V1 and NDPS1-M3V2 enzymes utilized NPP and IPP as substrates at similar relative proportions to that observed when using DMAPP and IPP as substrates (Fig. 6C). Furthermore, as anticipated from the activities of the zFPS-M4 and M2 enzymes, both zFPS-M3V1 (Asp98Glu Ser100Tyr, Phe107Ile) and zFPS-M3V2 (Ser100Tyr, Leu106Ile, Phe107Ile) enzymes synthesize NPP when supplied with DMAPP and IPP as substrates and essentially lost the ability to efficiently utilize NPP as a substrate (Fig. 6B and C). Overall, these data indicate that substitution of the aromatic amino acids at positions 100 and 107 are sufficient to convert zFPS into NDPS1 but the complete
reciprocal change in activity from NDPS1 to zFPS requires additional amino acid substitutions at positions 98 and 106.

**Kinetic analysis of mutant variants of NDPS1 and zFPS**

The activities of mutant variants of NDPS1 and zFPS revealed qualitative changes in the synthesis of prenyl diphosphates that are consistent with a role of amino acid residues within CPT conserved region II as determinants of substrate and product specificity (Figs. 4 and 6). Furthermore, the majority of the mutant enzymes synthesize prenyl diphosphates at abundances similar to those of the corresponding parent enzyme. For example, NDPS1 and zFPS-M10 synthesize approximately the same amounts of NPP when supplied with DMAPP and IPP as substrates. However, it is possible that the kinetic properties of the mutant enzyme variants differ from the wild type enzymes and these differences may not be apparent if substrate concentrations are limiting. Therefore, kinetic analysis was performed on wild type and select mutant variants of both NDPS1 and zFPS.

Kinetic analysis of NDPS1 from LA2409 revealed a $K_m$ for DMAPP and IPP of 112 and 140 µM, respectively (Table II), which are similar to the values of 177 and 152 µM previously reported for *S. lycopersicum* NDPS1 (Schilmiller et al., 2009). The $K_m$ of the M10 and M2 variants of zFPS using DMAPP as a substrate, are similar to that obtained for NDPS1 although the $K_{cat}$ and subsequent catalytic efficiency of the zFPS-M2 enzyme is slightly lower than that observed for NDPS1. The $K_m$ for IPP of zFPS-M2 is lower than those measured for the NDPS1 and zFPS-M10 enzymes, although the catalytic efficiencies of each enzyme are similar (Table II). The $K_m$ of zFPS from LA1393 varied slightly when compared to previously published data (Sallaud et al., 2009). The $K_m$ for DMAPP and NPP was 76 ± 12 µM and 51 ± 9 µM (Table II) compared with 35 ± 5 µM and 10 ± 2 µM from the previous study. In addition, $K_m$ for IPP, when measured using NPP as a substrate, was 147 ± 59 µM. The kinetic properties of the NDPS1-M10 and NDPS1-M4 enzymes were also determined and compared to those of the wild type zFPS enzyme. The $K_m$ for each substrate was comparable for all three enzymes and although the catalytic efficiencies of the NDPS1-M10 enzyme are slightly lower than
those calculated for the wild type zFPS enzyme from LA1393, the values for NDPS1-M4 enzyme are very similar to those of the wild type zFPS enzyme (Table II).

DISCUSSION
Tomato and closely related wild species exhibit both qualitative and quantitative variation in trichome-derived volatile terpenes, several of which are synthesized by enzymes encoded within a rapidly evolving gene cluster on the short arm of chromosome 8 that is characterized by gene duplications, deletions, rearrangements and gene conversion (Matsuba et al., 2013). The TPS genes within the chromosome 8 terpene cluster are atypical in that they encode enzymes that use cis-prenyl diphosphates of either C10, C15 or C20 chain lengths that are synthesized by CPTs encoded by genes also present within this cluster (Sallaud et al., 2009; Schilmiller et al., 2009; Gonzales-Vigil et al., 2012; Matsuba et al., 2013).

Within Solanum, the capacity to synthesize monoterpenes from NPP appears to be present in multiple species and accessions, including some S. habrochaites accessions, and is associated with the CPT1 locus that encodes NDPS1 and the TPS19 or TPS20 loci that encode monoterpane synthases that use NPP (Schilmiller et al., 2009; Falara et al., 2011; Gonzales-Vigil et al., 2012; Matsuba et al., 2013). However, the ability to synthesize sesquiterpenes from Z,Z-FPP is thus far restricted to specific accessions of S. habrochaites and involves both the evolution of the TPS45 locus (whose alleles encode either ShSBS or ShZIS), and the evolution of an allele at the CPT1 locus that encodes zFPS activity (Sallaud et al., 2009; Gonzales-Vigil et al., 2012; Matsuba et al., 2013). The presence of TPS45 only in S. habrochaites suggests that synthesis of sesquiterpenes from Z,Z-FPP is of recent origin. The trichome-derived sesquiterpenes and sesquiterpene acids of select S. habrochaites accessions are potent herbivore deterrents, and the ability to synthesize these compounds likely offer an advantage to the plants in their native habitat (Frelichowski and Juvik, 2001; Bleeker et al., 2011).

Comparative sequence analysis identified a set of 11 amino acid residues within NDPS1 and zFPS that correlate with the ability of their parent accessions to synthesize
either mono- or sesquiterpenes (Table I; Supplemental Fig. S1). Reciprocal amino acid substitutions involving 10 of these residues confirmed that they were sufficient to switch the activity of NDPS1 to an enzyme that shares characteristics of zFPS and to convert zFPS to an enzyme that possesses NDPS1 activity (Fig. 4). Subsequent refinement using a combination of homology modeling, site directed mutagenesis and enzyme activity assays, supplemented by kinetic analysis of recombinant enzymes led to the discovery that the relative positioning of aromatic amino acids and immediately adjacent residues within the conserved CPT domain II are important determinants of substrate and product specificity of NDPS1 and zFPS (Figs. 5 and 6; Table II). Notably, the amino acid changes within the CPT domain II that specify either NDPS1 or zFPS activity require only one or two nucleotide changes (Fig. 6; Table I). These data are congruent with previous studies of terpene biosynthesis that indicate that single nucleotide changes, which may occur relatively freely during evolution, can lead to qualitative changes in product profiles (Kollner et al., 2004; Kampranis et al., 2007; Xu et al., 2007; Keeling et al., 2008; Kollner et al., 2009; Green et al., 2011).

The chain length of isoprenoids varies dramatically from C10 in the case of monoterpenes to >C10 000 for natural rubber, and this wide range of diverse products is catalyzed by prenyltransferases with different substrate and product specificities (Takahashi and Koyama, 2006). In the case of the trans-prenyltransferases, which are structurally and mechanistically distinct from the cis-prenyltransferases, bulky aromatic residues at the interface between two alpha helices are known to determine product chain length and a combination of crystallography, structural modeling, bioinformatics, and biochemical analyses has led to the development and validation of predictive models of their functional properties (Ohnuma et al., 1996; Tarshis et al., 1996; Ohnuma et al., 1997; Ohnuma et al., 1998; Wallrapp et al., 2013).

Although less well defined, determinants of substrate and productivity of cis-prenyltransferases have also been reported. These studies, in which mutagenesis of specific residues and the incorporation of additional residues at the flexible loop adjacent to domain III allowed products of longer chain length to be synthesized,
highlighted the importance of the number of amino acid residues within the third alpha helix and residues within conserved domain III for the determination of product chain length (Kharel et al., 2006; Noike et al., 2008). For example, alanine substitution of bulky amino acids such as Leu84 in the Z,E-farnesyl diphosphate synthase of *Mycobacterium tuberculosis* allows an increase in product chain length over the wild type enzyme (Noike et al., 2008). In the structural models of CPTs (Fig. 5), domain III residues form the third alpha helix and the preceding disordered loop (Fig. 5C), which together with the second alpha-helix, forms the hydrophobic cleft close to the substrate binding sites where chain elongation occurs (Guo et al., 2005; Noike et al., 2008). The *Solanum* short chain CPTs encoded by *CPT1*, *CPT2*, and *CPT6* lack amino acid residues downstream of CPT domain III (Akhtar et al., 2013), which shortens the length of the third alpha helix in these proteins compared to those of *CPT4*, *CPT5*, and *CPT7* that catalyze the formation of longer chain cis-prenyl diphosphates (Akhtar et al., 2013; Supplemental Fig. S2). Therefore, a correlation exists between the length of the third alpha helix and the product chain length. However, this correlation does not explain the difference in product chain length of NDPS1 and zFPS and our data indicate a hereto unknown role for the relative positioning of aromatic residues within conserved domain II, which is part of the second alpha helix (Fig. 5), as important for specifying the relative activity of these short-chain CPTs. Further experiments are required to determine whether this region has broad utility across diverse CPTs.

In general, the function and evolutionary relationships of plant CPTs, together with the structural features that determine their substrate and product specificity, remain poorly understood. By utilizing an integrated approach involving comparative sequence analysis of chemically distinct germplasm, homology modeling, and site-directed mutagenesis, coupled with enzyme activity assays and kinetics, we have identified and characterized allelic diversity at the *CPT1* locus within a natural population of *S. habrochaites*. This allelic diversity causes variation between NDPS1 and zFPS through altering the relative position of aromatic amino acids within a structurally conserved hydrophobic cleft that is important for product chain length determination of CPTs. This variation led to the formation of Z,Z-FPP in the chloroplasts of type VI glandular
trichomes, which represents an essential step in the evolution sesquiterpene biosynthesis.

MATERIALS AND METHODS

Plant Material and Growth Conditions

Seeds of *Solanum habrochaites* accessions were obtained from the C.M. Rick Tomato Genetics Resource Center (TGRC) ([http://tgrc.ucdavis.edu/](http://tgrc.ucdavis.edu/)). Seeds were germinated on filter paper in the dark. Following germination, 2-3 plants per accession were grown on Jiffy-7 Peat Pellets ([http://www.hummert.com/](http://www.hummert.com/)) for 3 weeks as previously described (Schilmiller *et al.*, 2009). Plants were transplanted into peat-based compost supplemented with fertilizer in greenhouses equipped with environmental controls and supplemental lighting at Michigan State University, East Lansing, MI.

Trichome Isolation and Gene Cloning

Trichomes were isolated from the stems and petioles from 2-3 fully grown plants of the same accession and RNA extraction, cDNA synthesis, and gene cloning was performed as previously described (Gonzales-Vigil *et al.*, 2012) using primers conserved for *NDPS1* and *zFPS* (Sallaud *et al.*, 2009; Schilmiller *et al.*, 2009). Nucleotide sequences are deposited in GenBank under the following accession numbers KF494850-KF494874.

DNA Sequence Analysis, Multiple Sequence Alignments, Phylogenetic Analyses, and Homology Modeling

DNA sequences were edited and assembled using Sequencher™ software version 4.8 ([http://genecodes.com/](http://genecodes.com/)). Sequence analysis was performed with MEGA version 5 (Tamura *et al.*, 2011) using MUSCLE (Edgar, 2004) for constructing the sequence alignments and the Maximum Likelihood method based on the Tamura-Nei model (Tamura and Nei, 1993) for phylogenetic analysis. A bootstrap test was used to assess the reliability of the tree. Percent identity between nucleotide and predicted amino acid sequences of TPS enzymes was calculated using the alignment made by MUSCLE to calculate distance. Homology models were constructed within the SWISS-MODEL...
workspace using the automatic alignment algorithm (Arnold et al., 2006) and models were visualized and annotated using PyMOL (www.pymol.org). Models of NDPS1 and zFPS were generated using the crystal structure of *E.coli* undecaprenyl pyrophosphate synthase chain A (PDB: 1X07A) as a template (Guo et al., 2005). Models of SICPT4, SICPT5 and SICPT7 were generated using the crystal structure of *E.coli* undecaprenyl pyrophosphate synthase chain A (PDB: 1X09A) (Guo et al., 2005). Models of SICPT2 and SICPT6 were generated using the crystal structure of *Staphylococcus aureus* undecaprenyl diphosphate synthase chain A (PDB: 4H8E) as a template (Zhu et al., 2013).

**Synthesis of Codon Optimized Genes and Site-Directed Mutagenesis**

Codon optimized versions of NDPS1 and zFPS, lacking the chloroplast targeting sequences, together with M10 versions of each sequence (Supplemental Fig. S3), were synthesized by Genscript Corporation (http://www.genscript.com/). Each synthetic gene contained *Bam*HI and *Sal*I restriction enzyme recognition sequences at the 5′ and 3′ ends, respectively. The *CPT1* inserts were released from the pUC57 cloning vector by digestion with *Bam*HI and *Sal*I and ligated into the pHIS8 vector, previously linearized with the same enzymes. Recombinant clones were transformed into *E.coli* BL21 cells. Site-directed mutagenesis was performed using codon optimized versions of NDPS1 and zFPS in the pUC57 vector as templates and utilizing standard methods (Sambrook and Russell, 2001). Primers utilized for site directed mutagenesis are listed in Supplemental Table S1 and the sequences of the codon-optimized and mutant variants of NDPS1 and zFPS are provided in Supplemental Dataset S1.

**Recombinant Protein Expression and Purification**

A single *E. coli* colony was inoculated into 3 ml Terrific Broth (TB) medium containing 50 µg ml⁻¹ kanamycin, and the culture grown overnight at 37°C. A 500 µL aliquot of this culture was used to inoculate 100 ml TB medium containing 50 µg ml⁻¹ kanamycin. The culture was grown to log phase (*A*₆₀₀ nm = 0.3 – 0.7) at 37°C, at which point isopropyl-thio-β-D-galactopyranoside (www.sigmaaldrich.com) was added to final concentration of 0.1 mM. The culture was grown for an additional 16 h at 20°C. Cells were harvested by
centrifugation at 5,000 \( \times \) g for 20 min at 4°C. One third of the cell pellet was resuspended in 3 ml of lysis buffer (50 mM HEPES pH 8, 5% glycerol, 100 mM KCl, 7.5 mM MgCl\(_2\), 10 mM imidazole, 1 mg ml\(^{-1}\) lysozyme) containing 1X Protease Inhibitor Cocktail Tablets (www.roche-applied-science.com) and incubated on ice for 30 min prior to sonication. Soluble proteins were recovered from the supernatant following centrifugation at 10,000 \( g \) for 20 min at 4°C. 0.5 ml of Ni-NTA slurry (www.qiagen.com) was added to the cleared supernatant and the mixture was incubated for 1 h at 4°C on a rotary shaker prior to passing through a Poly-Prep chromatography column. The column was washed with 10 ml of washing buffer (50 mM HEPES pH 8, 5% glycerol, 100 mM KCl, 7.5 mM MgCl\(_2\), and 20 mM imidazole). His-tagged proteins were eluted from the column with lysis buffer containing 250 mM imidazole but lacking lysozyme and protease inhibitor. Fractions containing purified proteins were pooled and buffer-exchanged to 20 mM HEPES pH 8, 20% glycerol, 4 mM DTT with an Amicon Ultra-15 10-kD filter (www.millipore.com). Purified proteins were quantified with Pierce® BCA Protein Assay Kit (www.thermoscientific.com) according to the manufacturer’s instructions. Purified recombinant proteins were approximately 90% pure as determined by SDS-PAGE.

**Enzyme Activity Assays**

Enzyme assays were modified from a previously published protocol (Thulasiram and Poulter, 2006). The standard enzyme assays were performed in 200 µl of reaction buffer (35 mM HEPES pH 7.6, 10 mM MgCl\(_2\), 5% glycerol, 4 mM DTT) containing 1 µg of purified protein and substrates [10 µM each (1:1 ratio) of either DMAPP and IPP, NPP and IPP, and 2Z,6Z-FPP and IPP (http://www.echelon-inc.com)]. Additional reactions with increased IPP ratios of 2:1, 5:1 and 10:1 were also prepared. Reactions were incubated at 30°C for 1 h following which the enzyme was heat inactivated for 10 min at 65°C and the mix allowed to cool to room temperature. The reaction products were dephosphorylated by incubation for 1 h at 37°C by the addition of 80 µl of a buffer containing 250 mM Tris-HCl pH 7.9, 500 mM NaCl, and 50 mM MgCl\(_2\), together with 20 units of Calf Intestinal Phosphatase (www.neb.com). Following dephosphorylation, 0.1 g of NaCl was added and the reaction mix was extracted three times with 200 µl Methyl
tert-butyl ether (MTBE) containing 1 ng/µl of tetradecane. Following each addition of MTBE, vials were vortexed and centrifuged for 2 min at 2500 rpm. The supernatants were pooled and evaporated to a volume of 50 µl under N₂ gas and transferred to vials for gas-chromatography mass-spectrometry (GC-MS).

GC-MS analysis was performed using a 6890N network GC system with 5975B inert XL MSD detector (http://www.agilent.com). Separation was achieved by injection of 1 µl of extract onto a 127-501N DB-5 column (10 m length, 0.1 mm i.d, 0.34 µm thick stationary phase) at an initial temperature of 80°C for 1 min coupled with the following temperature profile: 5°C min⁻¹ to 110°C; 40°C min⁻¹ to 165°C; 5°C min⁻¹ to 190°C; 40°C min⁻¹ to 210°C; 5°C min⁻¹ to 270°C; 40°C min⁻¹ to 320°C for 2 min. Metabolites were normalized to the tetradecane internal standard. Nerol and E,E-farnesol (www.sigmaaldrich.com) were used as standards to quantify metabolites.

**Kinetic Analysis**
Kinetic analysis was performed using radio-labeled [¹⁴C] IPP (60 mCi/mmol) (http://www.perkinelmer.com/) as a substrate with between 1 and 4 µg of purified recombinant enzyme in each reaction. Estimates of the $K_m$ for DMAPP, IPP and NPP were performed over a range of substrate concentrations using a fixed concentration of the co-substrate as follows: NDPS1-LA2409, 30 - 80 µM [¹⁴C] IPP at 350 µM DMAPP and 20 - 100 µM DMAPP at 50 µM [¹⁴C] IPP; NDPS1-M10, 10 - 100 µM DMAPP at 50 µM [¹⁴C] IPP, 20 - 80 µM [¹⁴C] IPP at 150 µM NPP and 10 - 100 µM NPP at 50 µM [¹⁴C] IPP; NDPS1-M4, 10 - 100 µM DMAPP at 50 µM [¹⁴C] IPP, 20 - 80 µM [¹⁴C] IPP at 150 µM NPP and 10 - 100 µM NPP at 50 µM [¹⁴C] IPP; zFPS-LA1393, 10 - 100 µM DMAPP at 50 µM [¹⁴C] IPP, 20 - 80 µM [¹⁴C] IPP at 150 µM NPP and 10 - 100 µM NPP at 50 µM [¹⁴C] IPP; zFPS-M10, 20 - 80 µM [¹⁴C] IPP at 150 µM DMAPP, 10 - 100 µM DMAPP at 50 µM [¹⁴C] IPP; zFPS-M2, 20 - 80 µM [¹⁴C] IPP at 150 µM DMAPP and 10 - 100 µM DMAPP at 50 µM [¹⁴C] IPP. All reactions were performed in a 50 µl volume containing 50 mM HEPES, 100 mM KCl, 7.5 mM MgCl₂, 5% (v/v) glycerol, and 5 mM DTT (pH 8.0) and were incubated at 30°C for 10 to 15 min. The reaction was stopped by addition of 50 µl of 1N HCl and hydrolysis of the reaction products allowed to proceed at 37°C for
30 min. The radio-labeled products were extracted with 150 µl of ethyl acetate and the radio-label was quantified in 100 µl of the extract by scintillation counting using an LS6500 multipurpose scintillation counter (www.beckmancoulter.com). Blank reactions, without added enzyme were used to subtract the background radiolabel associated with [14C] IPP. Apparent \( V_{\text{max}} \) and \( K_{\text{m}} \) values were determined using nonlinear regression of the Michaelis-Menten equation and \( k_{\text{cat}} \) values were calculated using the \( V_{\text{max}} \) value, the concentration of the enzyme in the assay, and the molecular weight of the monomeric form of each recombinant protein.

Supplemental Data

Supplemental Figure S1. Amino acid alignment of CPT1 proteins from \textit{S. habrochaites}.

Supplemental Figure S2. Homology models of tomato CPT enzymes.

Supplemental Figure S3. Amino acid alignment of mutant CPT1 proteins utilized in enzyme assays.

Supplemental Table S1. Oligonucleotide primers utilized in this study.

Supplemental Dataset S1. Codon-optimized nucleotide and amino acid sequences of mutant variants of NDPS1 and zFPS utilized in this study.

AKNOWLEDGEMENTS

We thank Dr. Joe Noel (Salk Institute) for providing the pHIS8 expression vector.

LITERATURE CITED

Aharoni A, Giri AP, Verstappen FWA, Bertea CM, Sevenier R, Sun ZK, Jongsma MA, Schwab W, Bouwmeester HJ (2004) Gain and loss of fruit flavor compounds produced by wild and cultivated strawberry species. Plant Cell \textbf{16}: 3110-3131

Akhtar TA, Matsuba Y, Schauvinhold I, Yu G, Lees HA, Klein SE, Pichersky E (2013) The tomato \textit{cis}–prenyltransferase gene family. Plant J \textbf{73}: 640-652

Arnold K, Bordoli L, Kopp J, Schwede T (2006) The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. Bioinformatics \textbf{22}: 195-201
Asawatreratanakul K, Zhang YW, Wititsuwannakul D, Wititsuwannakul R, Takahashi S, Rattanapittayaporn A, Koyama T (2003) Molecular cloning, expression and characterization of cDNA encoding cis-prenyltransferases from Hevea brasiliensis - A key factor participating in natural rubber biosynthesis. Eur J Biochem. 270: 4671-4680

Bleeker PM, Diergaarde PJ, Ament K, Schutz S, Johne B, Dijkink J, Hiemstra H, de Gelder R, de Both MTJ, Sabelis MW, Haring MA, Schuurink RC (2011) Tomato produced 7-epizingiberene and R-curcumene act as repellents to whiteflies. Phytochemistry 72: 68-73

Burke CC, Wildung MR, Croteau R (1999) Geranyl diphosphate synthase: Cloning, expression, and characterization of this prenyltransferase as a heterodimer. Proc Natl Acad Sci USA 96: 13062-13067

Chen F, Tholl D, Bohlmann J, Pichersky E (2011) The family of terpene synthases in plants: a mid-size family of genes for specialized metabolism that is highly diversified throughout the kingdom. Plant J 66: 212-229

Cunillera N, Arro M, Fores O, Manzano D, Ferrer A (2000) Characterization of dehydrodolichyl diphosphate synthase of Arabidopsis thaliana, a key enzyme in dolichol biosynthesis. FEBS Lett 477: 170-174

Degenhardt J, Hiltpold I, Kollner TG, Frey M, Gierl A, Gershenzon J, Hibbard BE, Ellersieck MR, Turlings TC (2009) Restoring a maize root signal that attracts insect-killing nematodes to control a major pest. Proc Natl Acad Sci USA 106: 13213-13218

Ducluzeau AL, Wamboldt Y, Elowsky CG, Mackenzie SA, Schuurink RC, Basset GJC (2012) Gene network reconstruction identifies the authentic trans-prenyl diphosphate synthase that makes the solanesyl moiety of ubiquinone-9 in Arabidopsis. Plant J 69: 366-375

Dudareva N, Klempien A, Muhlemann JK, Kaplan I (2013) Biosynthesis, function and metabolic engineering of plant volatile organic compounds. New Phytol 198: 16-32

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

Eisenreich W, Bacher A, Arigoni D, Rohdich F (2004) Biosynthesis of isoprenoids via the non-mevalonate pathway. Cell Mol Life Sci 61: 1401-1426
Falara V, Akhtar T, Nguyen TT, Spyropoulou EA, Bleeker PM, Schauvinhold I, Matsuba Y, Bonini ME, Schilmiller AL, Last RL, Schuurink RC, Pichersky E (2011) The tomato (Solanum lycopersicum) terpene synthase gene family. Plant Physiol 157: 770-789

Frelichowski JE, Juvik JA (2001) Sesquiterpene carboxylic acids from a wild tomato species affect larval feeding behavior and survival of Helicoverpa zea and Spodoptera exigua (Lepidoptera : Noctuidae). J Econ Entomol 94: 1249-1259

Gang DR (2005) Evolution of flavors and scents. Ann Rev Plant Biol 56: 301-325

Gonzales-Vigil E, Hufnagel DE, Kim J, Last RL, Barry CS (2012) Evolution of TPS20-related terpene synthases influences chemical diversity in the glandular trichomes of the wild tomato relative Solanum habrochaites. Plant J 71: 921-935

Green S, Baker EN, Laing W (2011) A non-synonymous nucleotide substitution can account for one evolutionary route to sesquiterpene synthase activity in the TPS-b subgroup. FEBS Lett 585: 1841-1846

Guo RT, Ko TP, Chen APC, Kuo CJ, Wang AHJ, Liang PH (2005) Crystal structures of undecaprenyl pyrophosphate synthase in complex with magnesium, isopentenyl pyrophosphate, and farnesyl thiopyrophosphate - Roles of the metal ion and conserved residues in catalysis. J Biol Chem 280: 20762-20774

Heiling S, Schuman MC, Schoettner M, Mukerjee P, Berger B, Schneider B, Jassbi AR, Baldwin IT (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

Hirooka K, Bamba T, Fukusaki El, Kobayashi A (2003) Cloning and kinetic characterization of Arabidopsis thaliana solanesyl diphosphate synthase. Biochem J 370: 679-686

Kampranis SC, Ioannidis D, Purvis A, Mahrez W, Ninga E, Katerelos NA, Anssour S, Dunwell JM, Degenhardt J, Makris AM, Goodenough PW, Johnson CB (2007) Rational conversion of substrate and product specificity in a Salvia monoterpenic synthase: Structural insights into the evolution of terpene synthase function. Plant Cell 19: 1994-2005
Keeling CI, Weisshaar S, Lin RPC, Bohlmann J (2008) Functional plasticity of paralogous diterpene synthases involved in conifer defense. Proc Natl Acad Sci USA 105: 1085-1090

Kera K, Takahashi S, Sutoh T, Koyama T, Nakayama T (2012) Identification and characterization of a cis,trans-mixed heptaprenyl diphosphate synthase from Arabidopsis thaliana. FEBS J 279: 3813-3827

Kharel Y, Koyama T (2003) Molecular analysis of cis-prenyl chain elongating enzymes. Nat Prod Rep 20: 111-118

Kharel Y, Takahashi S, Yamashita S, Koyama T (2006) Manipulation of prenyl chain length determination mechanism of cis-prenyltransferases. FEBS J 273: 647-657

Kirby J, Keasling JD (2009) Biosynthesis of plant isoprenoids: Perspectives for microbial engineering. Ann Rev Plant Biol 60: 335-355

Kollner TG, Gershenzon J, Degenhardt J (2009) Molecular and biochemical evolution of maize terpene synthase 10, an enzyme of indirect defense. Phytochemistry 70: 1139-1145

Kollner TG, Schnee C, Gershenzon J, Degenhardt J (2004) The variability of sesquiterpenes emitted from two Zea mays cultivars is controlled by allelic variation of two terpene synthase genes encoding stereoselective multiple product enzymes. Plant Cell 16: 1115-1131

Lange BM, Ahkami A (2013) Metabolic engineering of plant monoterpenes, sesquiterpenes and diterpenes-current status and future opportunities. Plant Biotechnol J 11: 169-196

Lange BM, Ghassemian M (2003) Genome organization in Arabidopsis thaliana: a survey for genes involved in isoprenoid and chlorophyll metabolism. Plant Mol Biol 51: 925-948

Lange BM, Turner GW (2013) Terpenoid biosynthesis in trichomes-current status and future opportunities. Plant Biotechnol J 11: 2-22

Lee S, Badieyan S, Bevan DR, Herde M, Gatz C, 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
Liang PH (2009) Reaction kinetics, catalytic mechanisms, conformational changes, and inhibitor design for prenyltransferases. Biochemistry 48: 6562-6570

Matsuba Y, Nguyen TTH, Wiegert K, Falara V, Gonzales-Vigil E, Leong B, Schafer P, Kudrna D, Wing RA, Bolger AM, Usadel B, Tissier A, Fernie AR, Barry CS, Pichersky E (2013) Evolution of a complex locus for terpene biosynthesis in Solanum. Plant Cell 25: 2022 - 2036

Nguyen DT, Gopfert JC, Ikezawa N, MacNevin G, Kathiresan M, Conrad J, Spring O, Ro DK (2010) Biochemical conservation and evolution of germacrene A oxidase in Asteraceae. J Biol Chem 285: 16588-16598

Niehaus TD, Okada S, Devarenne TP, Watt DS, Sviripa V, Chappell J (2011) Identification of unique mechanisms for triterpene biosynthesis in Botryococcus braunii. Proc Natl Acad Sci USA 108: 12260-12265

Noike M, Arnbo T, Kikuch S, Suzuki T, Yamashita S, Takahashi S, Kurokawa H, Mahapatra S, Crick DC, Koyama T (2008) Product chain-length determination mechanism of Z,E-farnesyl diphosphate synthase. Biochem Bioph Res Com 377: 17-22

Oh SK, Han KH, Ryu SB, Kang H (2000) Molecular cloning, expression, and functional analysis of a cis-prenyltransferase from Arabidopsis thaliana - Implications in rubber biosynthesis. J Biol Chem 275: 18482-18488

Ohnuma S, Hirooka K, Ohto C, Nishino T (1997) Conversion from archaeal geranylgeranyl diphosphate synthase to farnesyl diphosphate synthase. Two amino acids before the first aspartate-rich motif solely determine eukaryotic farnesyl diphosphate synthase activity. J Biol Chem 272: 5192-5198

Ohnuma S, Hirooka K, Tsuruoka N, Yano M, Ohto C, Nakane H, Nishino T (1998) A pathway where polyprenyl diphosphate elongates in prenyltransferase. Insight into a common mechanism of chain length determination of prenyltransferases. J Biol Chem 273: 26705-26713

Ohnuma S, Narita K, Nakazawa T, Ishida C, Takeuchi Y, Ohto C, Nishino T (1996) A role of the amino acid residue located on the fifth position before the first aspartate-rich motif of farnesyl diphosphate synthase on determination of the final product. J Biol Chem 271: 30748-30754
Post J, van Deenen N, Fricke J, Kowalski N, Wurbs D, Schaller H, Eisenreich W, Huber C, Twyman RM, Prufer D, Gronover CS (2012) Laticifer-specific cis-prenyltransferase silencing affects the rubber, triterpene, and inulin content of *Taraxacum brevicorniculatum*. Plant Physiol **158**: 1406-1417

Richman A, Swanson A, Humphrey T, Chapman R, McGarvey B, Pocs R, Brandle J (2005) Functional genomics uncovers three glucosyltransferases involved in the synthesis of the major sweet glucosides of *Stevia rebaudiana*. Plant J **41**: 56-67

Rodriguez-Concepcion M, Boronat A (2002) Elucidation of the methylerythritol phosphate pathway for isoprenoid biosynthesis in bacteria and plastids. A metabolic milestone achieved through genomics. Plant Physiol **130**: 1079-1089

Rohdich F, Hecht S, Gartner K, Adam P, Krieger C, Amslinger S, Arigoni D, Bacher A, Eisenreich W (2002) Studies on the nonmevalonate terpene biosynthetic pathway: Metabolic role of IspH (LytB) protein. Proc Natl Acad Sci USA **99**: 1158-1163

Sallaud C, Rontein D, Onillon S, Jabes F, Duffe P, Giacalone C, Thoraval S, Escoffier C, Herbette G, Leonhardt N, Causse M, Tissier A (2009) A novel pathway for sesquiterpene biosynthesis from \( \text{Z,}\text{Z}-\text{farnesyl pyrophosphate} \) in the wild tomato *Solanum habrochaites*. Plant Cell **21**: 301-317

Sambrook J, Russell DW (2001) Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.

Schilmiller AL, Schauvinhold I, Larson M, Xu R, Charbonneau AL, Schmidt A, Wilkerson C, Last RL, Pichersky E (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

Schmelz EA, Kaplan F, Huffaker A, Dafoe NJ, Vaughan MM, Ni XZ, Rocca JR, Alborn HT, Teal PE (2011) Identity, regulation, and activity of inducible diterpenoid phytoalexins in maize. Proc Natl Acad Sci USA **108**: 5455-5460

Schmidt T, Lenders M, Hillebrand A, van Deenen N, Munt O, Reichelt R, Eisenreich W, Fischer R, Prufer D, Gronover CS (2010) Characterization of rubber particles and rubber chain elongation in *Taraxacum koksaghyz*. BMC Biochemistry **11**: 11
Schnee C, Kollner TG, Held M, Turlings TCJ, Gershenzon J, 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

Surmacz L, Swiezewska E (2011) Polyisoprenoids - Secondary metabolites or physiologically important superlipids? Biochem Bioph Res Com 407: 627-632

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

Tamura K, Nei M (1993) Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. Mol Biol Evol 10: 512-526

Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S (2011) MEGA5: Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol Biol Evol 28: 2731-2739

Tarshis LC, Proteau PJ, Kellogg BA, Sacchettini JC, Poulter CD (1996) Regulation of product chain length by isoprenyl diphosphate synthases. Proc Natl Acad Sci USA 93: 15018-15023

Thulasiram HV, Poulter CD (2006) Farnesyl diphosphate synthase: The art of compromise between substrate selectivity and stereoselectivity. J Am Chem Soc 128: 15819-15823

van der Hoeven RS, Monforte AJ, Breeden D, Tanksley SD, Steffens JC (2000) Genetic control and evolution of sesquiterpene biosynthesis in Lycopersicon esculentum and L. hirsutum. Plant Cell 12: 2283-2294

Wallrapp FH, Pan J-J, Ramamoorthy G, Almonacid DE, Hillerich BS, Seidel R, Patskovsky Y, Babbitt PC, Almo SC, Jacobson MP, Poulter CD (2013) Prediction of function for the polyprenyl transferase subgroup in the isoprenoid synthase superfamily. Proc Natl Acad Sci USA 110: 1196-1202

Westfall PJ, Pitera DJ, Lenihan JR, Eng D, Woolard FX, Regentin R, Horning T, Tsuruta H, Melis DJ, Owens A, Fickes S, Diola D, Benjamin KR, Keasling JD, Leavell MD, McPhee DJ, Renninger NS, Newman JD, Paddon CJ (2012) Production
of amorphadiene in yeast, and its conversion to dihydroartemisinic acid, precursor to the antimalarial agent artemisinin. Proc Natl Acad Sci USA 109: 111-118

Wu SQ, Schalk M, Clark A, Miles RB, Coates R, Chappell J (2006) Redirection of cytosolic or plastidic isoprenoid precursors elevates terpene production in plants. Nat Biotech 24: 1441-1447

Xu MM, Wilderman PR, Peters RJ (2007) Following evolution's lead to a single residue switch for diterpene synthase product outcome. Proc Natl Acad Sci USA 104: 7397-7401

Yeo YS, Nybo SE, Chittiboyina AG, Weerasooriya AD, Wang YH, Gongora-Castillo E, Vaillancourt B, Buell CR, DellaPenna D, Celiz MD, Jones AD, Wurtele ES, Ransom N, Dudareva N, Shaaban KA, Tibrewal N, Chandra S, Smillie T, Khan IA, Coates RM, Watt DS, Chappell J (2013) Functional identification of valerena-1,10-diene synthase, a terpene synthase catalyzing a unique chemical cascade in the biosynthesis of biologically active sesquiterpenes in Valeriana officinalis. J Biol Chem 288: 3163-3173

Zhang FZ, Rodriguez S, Keasling JD (2011) Metabolic engineering of microbial pathways for advanced biofuels production. Curr Opin Biotech 22: 775-783

Zhu W, Zhang Y, Sinko W, Hensler ME, Olson J, Molohon KJ, Lindert S, Cao R, Li K, Wang K, Wang Y, Liu Y-L, Sankovsky A, de Oliveira CsAF, Mitchell DA, Nizet V, McCammon JA, Oldfield E (2013) Antibacterial drug leads targeting isoprenoid biosynthesis. Proc Natl Acad Sci USA 110: 123-128
Figure Legends

**Figure 1.** The pathway for synthesis of short-chain cis-prenyl diphosphates in the trichomes of *Solanum*. The synthesis of neryl diphosphate (NPP) occurs through the single head to tail condensation of dimethylallyl diphosphate (DMAPP) and isopentenyl diphosphate (IPP) and is catalyzed by neryl diphosphate synthase 1 (NDPS1). The synthesis of 2Z,6Z-farnesyl diphosphate (2Z,6Z-FPP) is catalyzed by Z,Z-farnesyl diphosphate synthase (zFPS) and can proceed through two routes; either directly from DMAPP through successive head to tail condensations with two molecules of IPP or through the single condensation of NPP and IPP.

**Figure 2.** Phylogenetic relationship of CPT1 from *S. habrochaites*. An unrooted phylogenetic tree was constructed using the Maximum Likelihood method from a codon alignment of 25 cDNA sequences from 17 *S. habrochaites* accessions. Previously characterized NDPS1 from *S. lycopersicum* (SINDPS1) and zFPS from *S. habrochaites* accession LA1777 were included for reference. Codon alignments were generated by MUSCLE and phylogenetic analysis was performed using MEGA5 (Tamura et al., 2011). M and S refer to *S. habrochaites* accessions that synthesize either monoterpenes or sesquiterpenes from the cisoid substrates NPP and Z,Z-FPP, respectively. Bootstrap values greater than 50 are displayed.

**Figure 3.** *In vitro* enzymatic assay of CPT1 from *S. habrochaites* accessions LA2409 and LA1393. A) The synthesis of nerol (C10) and farnesol (C15) using various ratios of IPP and DMAPP. B) The synthesis of farnesol (C15) using various ratios of IPP and NPP. Experimental details are provided in the Materials and Methods. Each data point represents the mean ± SD of two or three replicates.

**Figure 4.** *In vitro* enzymatic assay of M10 mutant variants of NDPS1 and zFPS from *S. habrochaites*. A) The synthesis of nerol (C10) and farnesol (C15) by wild type (WT) and M10 mutant variants of NDPS1 and zFPS at a 5:1 ratio of IPP to DMAPP. B) The
synthesis of farnesol (C15) by wild type (WT) and M10 mutant variants of NDPS1 and zFPS at a 5:1 ratio of IPP to NPP. Experimental details are provided in the Materials and Methods. Each data point represents the mean ± SD of two or three replicates.

**Figure 5.** Homology models of NDPS1 and zFPS. A) A model of the structure of NDPS1 (green) superimposed on to the crystal structure of *E. coli* undecaprenyl pyrophosphate synthase A subchain (PDB: 1X07A) (gray). B) A model of zFPS (blue) superimposed on to PDB: 1X07A (gray). Residues highlighted in yellow correspond to the M10 residues that are different between NDPS1 and zFPS (see text for details). Note the relative positions of Tyr100 and Phe107 (red arrows) that are located in conserved domain II of CPTs, within the second alpha helix, which together with the third alpha helix, form the characteristic hydrophobic cleft of CPTs that is important for product chain length determination. Also note the truncated length of the third alpha helix in NDPS1 and zFPS compared to undecaprenyl pyrophosphate synthase (white arrows). C) A model of the structure of NDPS1. The five conserved domains (I – V) of CPTs (Kharel and Koyama, 2003) are highlighted in different colors. Conserved domain II forms part of the second alpha helix and conserved domain III forms part of the third alpha helix. All models were generated using the Swiss Model Workspace and visualized using PyMOL.

**Figure 6.** *In vitro* enzymatic assay of M4, M3 and M2 variants of NDPS1 and zFPS of *S. habrochaites*. A) Amino acid residues targeted for functional analysis in the M4, M3 and M2 constructs. The residues depicted in bold represent the residues mutated in the corresponding construct. B) The synthesis of nerol (C10) and farnesol (C15) by wild type (WT), M4, M3 and M2 mutant variants of NDPS1 and zFPS at a 5:1 ratio of IPP to DMAPP. C) The synthesis of farnesol (C15) by wild type (WT) M4, M3 and M2 mutant variants of NDPS1 and zFPS at a 5:1 ratio of IPP to NPP. Experimental details are provided in the Materials and Methods. Each data point represents the mean ± SD of two or three replicates.
Table I. Conserved amino acid residues of NDPS1 and zFPS that correlate with terpene profile

| Residue: NDPS1 / zFPS | Change in amino acid property | Position: full length protein | Position: mature protein | Codon NDPS1<sup>a</sup> | Codon zFPS | CPT domain | Role of residue tested in construct |
|-----------------------|------------------------------|------------------------------|--------------------------|------------------------|----------------|------------|-----------------------------------|
| Pro / Ser             | nonpolar → polar             | 45                           | 2                        | CCT                    | TCT            | N-terminus | -                                 |
| Glu / Asp             | none                         | 98                           | 55                       | GAA                    | GAÇ            | II         | M10, M4, M3V2                     |
| Tyr / Ser             | aromatic → polar             | 100                          | 57                       | TAT                    | TCC            | II         | M10, M4, M3V1, M3V2, M2           |
| Ile / Leu             | none                         | 106                          | 63                       | ATT                    | CTÇ            | II         | M10, M4, M3V1, M3V2, M2           |
| Ile / Phe             | nonpolar → aromatic          | 107                          | 64                       | ATT                    | TTT            | II         | M10, M4, M3V1, M3V2, M2           |
| Ile / Val             | none                         | 124                          | 81                       | ATT                    | GTT            | III        | M10                                |
| Ser / Ala             | polar → nonpolar             | 138                          | 95                       | CTC / TCT              | GCC            | III        | M10                                |
| Glu / Gly             | negative → nonpolar          | 140                          | 97                       | GAG                    | GGÇ            | III        | M10                                |
| Glu / Asp             | none                         | 253                          | 210                      | GAA                    | GAÇ            | V          | M10                                |
| Asn / Lys             | polar → basic                | 273                          | 230                      | AAC                    | AAA            | V          | M10                                |
| Met/ Ile              | none                         | 289                          | 246                      | ATG                    | ATA            | C-terminus | M10                                |

<sup>a</sup>Nucleotide substitutions that confer codon changes between NDPS1 and zFPS are underlined.
| enzyme       | substrate | $K_m$ (µM) | $k_{cat}$ (S⁻¹) | $k_{cat} / K_m$ (µM⁻¹ S⁻¹) |
|--------------|-----------|------------|-----------------|-----------------------------|
| NDPS1 (LA2409) | IPP (DMAPP 350 µM) | 140 ± 11 | (4.7 ± 0.5) x 10⁻² | (3.3 ± 0.1) x 10⁻⁴ |
|              | DMAPP (IPP 50 µM)   | 112 ± 4  | (2.6 ± 0.3) x 10⁻² | (2.3 ± 0.1) x 10⁻⁴ |
| zFPS-M10     | IPP (DMAPP 150 µM) | 228 ± 43 | (6.6 ± 1.2) x 10⁻³ | (3.0 ± 0.7) x 10⁻⁴ |
|              | DMAPP (IPP 50 µM)   | 160 ± 17 | (2.2 ± 0.2) x 10⁻³ | (1.4 ± 0.1) x 10⁻⁴ |
| zFPS-M2      | IPP (DMAPP 150 µM) | 37 ± 5   | (1.5 ± 0.6) x 10⁻² | (3.1 ± 0.6) x 10⁻⁴ |
|              | DMAPP (IPP 50 µM)   | 108 ± 25 | (1.0 ± 0.2) x 10⁻² | (1.0 ± 0.1) x 10⁻⁴ |
| zFPS (LA1393)| DMAPP (IPP 50 µM)  | 76 ± 12  | (5.5 ± 1.5) x 10⁻³ | (9.2 ± 2.4) x 10⁻⁴ |
|              | IPP (NPP 150 µM)    | 147 ± 59 | (1.9 ± 1.1) x 10⁻¹ | (1.3 ± 0.6) x 10⁻³ |
|              | NPP (IPP 50 µM)     | 51 ± 9   | (6.9 ± 1.6) x 10⁻² | (1.3 ± 0.1) x 10⁻³ |
| NDPS1-M10    | DMAPP (IPP 50 µM)  | 47 ± 9   | (6.4 ± 0.7) x 10⁻³ | (1.5 ± 0.1) x 10⁻⁴ |
|              | IPP (NPP 150 µM)    | 157 ± 16 | (1.1 ± 0.4) x 10⁻² | (7.3 ± 0.5) x 10⁻⁵ |
|              | NPP (IPP 50 µM)     | 14 ± 2   | (3.7 ± 0.8) x 10⁻³ | (2.8 ± 0.9) x 10⁻⁴ |
| NDPS1-M4     | DMAPP (IPP 50 µM)  | 54 ± 3   | (8.8 ± 3.0) x 10⁻³ | (1.6 ± 0.6) x 10⁻³ |
|              | IPP (NPP 150 µM)    | 111 ± 50 | (1.6 ± 0.9) x 10⁻¹ | (1.8 ± 1.3) x 10⁻³ |
|              | NPP (IPP 50 µM)     | 97 ± 9   | (1.8 ± 0.7) x 10⁻¹ | (1.8 ± 0.1) x 10⁻³ |
Figure 1. The pathway for synthesis of short-chain cis-prenyl diphenosphates in the trichomes of *Solanum*. The synthesis of neryl diphosphate (NPP) occurs through the single head to tail condensation of dimethylallyl diphosphate (DMAPP) and isopentenyl diphosphate (IPP) and is catalyzed by neryl diphosphate synthase 1 (NDPS1). The synthesis of 2Z,6Z-farnesyl diphosphate (2Z,6Z-FPP) is catalyzed by Z,Z-farnesyl diphosphate synthase (zFPS) and can proceed through two routes; either directly from DMAPP through successive head to tail condensations with two molecules of IPP or through the single condensation of NPP and IPP.
Figure 2. Phylogenetic relationship of CPT1 from S. habrochaites. An unrooted phylogenetic tree was constructed using the Maximum Likelihood method from a codon alignment of 25 cDNA sequences from 17 S. habrochaites accessions. Previously characterized NDPS1 from S. lycopersicum (SINDPS1) and zFPS from S. habrochaites accession LA1777 were included for reference. Codon alignments were generated by MUSCLE and phylogenetic analysis was performed using MEGA5 (Tamura et al., 2011). M and S refer to S. habrochaites accessions that synthesize either monoterpene or sesquiterpenes from the cisoid substrates NPP and Z,Z-FPP, respectively. Bootstrap values greater than 50 are displayed.
Figure 3. *In vitro* enzymatic assay of CPT1 from *S. habrochaite* accessions LA2409 and LA1393. A) The synthesis of nerol (C10) and farnesol (C15) using various ratios of IPP and DMAPP. B) The synthesis of farnesol (C15) using various ratios of IPP and NPP. Experimental details are provided in the Materials and Methods. Each data point represents the mean ± SD of two or three replicates.
Figure 4. In vitro enzymatic assay of M10 mutant variants of NDPS1 and zFPS from S. habrochaites. A) The synthesis of nerol (C10) and farnesol (C15) by wild type (WT) and M10 mutant variants of NDPS1 and zFPS at a 5:1 ratio of IPP to DMAPP. B) The synthesis of farnesol (C15) by wild type (WT) and M10 mutant variants of NDPS1 and zFPS at a 5:1 ratio of IPP to NPP. Experimental details are provided in the Materials and Methods. Each data point represents the mean ± SD of two or three replicates.
Figure 5. Homology models of NDPS1 and zFPS. A) A model of the structure of NDPS1 (green) superimposed on to the crystal structure of E. coli undecaprenyl pyrophosphate synthase A subchain (PDB: 1X07A) (gray). B) A model of zFPS (blue) superimposed on to PDB: 1X07A (gray). Residues highlighted in yellow correspond to the M10 residues that are different between NDPS1 and zFPS (see text for details). Note the relative positions of Y100 and F107 (red arrows) that are located in conserved domain II of CPTs, within the second alpha helix, which together with the third alpha helix, form the characteristic hydrophobic cleft of CPTs that is important for product chain length determination. Also note the truncated hydrophobic length of the third alpha helix in NDPS1 and zFPS compared to undecaprenyl pyrophosphate synthase (white arrows). C) A model of the structure of NDPS1. The five conserved domains (I – V) of CPTs (Kharel and Koyama, 2003) are highlighted in different colors. Conserved domain II forms part of the second alpha helix and conserved domain III forms part of the third alpha helix. All models were generated using the Swiss Model Workspace and visualized using PyMOL.
A) Amino acid residues targeted for functional analysis in the M4, M3 and M2 constructs. The residues depicted in bold represent the residues mutated in the corresponding construct. B) The synthesis of nerol (C10) and farnesol (C15) by wild type (WT), M4, M3 and M2 mutant variants of NDPS1 and zFPS at a 5:1 ratio of IPP to DMAPP. C) The synthesis of farnesol (C15) by wild type (WT) M4, M3 and M2 mutant variants of NDPS1 and zFPS at a 5:1 ratio of IPP to NPP. Experimental details are provided in the Materials and Methods. Each data point represents the mean ± SD of two or three replicates.

**Figure 6. In vitro enzymatic assay of M4, M3 and M2 variants of NDPS1 and zFPS of S. habrochaites.**