Geranyllinalool synthases in solanaceae and other angiosperms constitute an ancient branch of diterpene synthases involved in the synthesis of defensive compounds

Falara, V.; Alba Cano, J.M.; Kant, M.; Schuurink, R.C.; Pichersky, E.

Published in:
Plant Physiology

DOI:
10.1104/pp.114.243246

Citation for published version (APA):

General rights
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: http://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.
Geranyllinalool Synthases in Solanaceae and Other Angiosperms Constitute an Ancient Branch of Diterpene Synthases Involved in the Synthesis of Defensive Compounds

Vasiliki Falara, Juan M. Alba, Merijn R. Kant, Robert C. Schuurink, and Eran Pichersky*

Department of Molecular, Cellular, and Developmental Biology, University of Michigan, Ann Arbor, Michigan 48109 (V.F., E.P.); and Department of Population Biology, Institute for Biodiversity and Ecosystem Dynamics (J.M.A., M.R.K.), and Department of Plant Physiology (R.C.S.), University of Amsterdam, 1098 XH Amsterdam, The Netherlands

Many angiosperm plants, including basal dicots, eudicots, and monocots, emit (E,E)-4,8,12-trimethyltrideca-1,3,7,11-tetraene, which is derived from geranyllinalool, in response to biotic challenge. An Arabidopsis (Arabidopsis thaliana) geranyllinalool synthase (GLS) belonging to the e/f clade of the terpene synthase (TPS) family and two Fabaceae GLSs that belong to the TPS-g clade have been reported, making it unclear which is the main route to geranyllinalool in plants. We characterized a tomato (Solanum lycopersicum) TPS-e/f gene, TPS46, encoding GLS (SlGLS) and its homolog (NaGLS) from Nicotiana attenuata. The Km value of SIGLS for geranylgeranyl diphosphate was 18.7 μM, with a turnover rate value of 6.85 s⁻¹. In leaves and flowers of N. attenuata, which constitutively synthesize 17-hydroxygeranyllinalool glycosides, NaGLS is expressed constitutively, but the gene can be induced in leaves with methyl jasmonate. In tomato, SIGLS is not expressed in any tissue under normal growth but is induced in leaves by alamethicin and methyl jasmonate treatments. SIGLS, NaGLS, AtGLSs, and several other GLSs characterized only in vitro come from four different eudicot families and constitute a separate branch of the TPS-e/f clade that diverged from kaurene synthases, also in the TPS-e/f clade, before the gymnosperm-angiosperm split. The early divergence of this branch and the GLS activity of genes in this branch in diverse eudicot families suggest that GLS activity encoded by these genes predates the angiosperm-gymnosperm split. However, although a TPS sequence belonging to this GLS lineage was recently reported from a basal dicot, no representative sequences have yet been found in monocot or nonangiospermous plants.

Geranyllinalool is an acyclic diterpene alcohol with a wide distribution in the plant kingdom; it has been identified as component of essential oils of distantly related plant species such as Jasmin grandiflorum, Michelia champaca, and Homamellis virginiana (Sandee, 2009). Geranyllinalool is the precursor of 4,8,12-trimethyltrideca-1,3,7,11-tetraene (TMTT), a volatile C₁₆-homoterpene emitted from the foliage of many angiosperm species including Arabidopsis (Arabidopsis thaliana), tomato (Solanum lycopersicum), maize (Zea mays), fava bean (Vicia faba), lima bean (Phaseolus lunatus), alfalfa (Medicago sativa), and Eucalyptus spp. (Van Poecke et al., 2001; Ament et al., 2004; Williams et al., 2005; Hopke et al., 1994; Leitner et al., 2010; Webster et al., 2010). In addition, various hydroxygeranyllinalool glycosides have been isolated from many Solanaceous species such as Capsicum annum, Lycium chinense, and at least 26 Nicotiana species (Yahara et al., 1993; Iorizzi et al., 2001; Snook et al., 1997). The biosynthetic pathway leading to geranyllinalool, as for all other terpenoids, begins with the condensation of isopentenyl diphosphate and its allylic isomer, dimethylallyl diphosphate. Sequential condensation of one isopentenyl diphosphate molecule with three dimethylallyl diphosphate molecules produces geranylgeranyl diphosphate (GGPP), the C₂₀ intermediate of the diterpenoid pathway. Next, a terpene synthase (TPS) catalyzes a two-step reaction in which carbocation formation of the C₂₀ precursor is followed by an allylic rearrangement that results in the production of the tertiary alcohol geranyllinalool (Herde et al., 2008).

Although geranyllinalool and its derivatives, TMTT and geranyllinalool glycosides, have been reported in a wide variety of plant species, a geranyllinalool synthase (GLS) involved in TMTT biosynthesis was only recently identified in Arabidopsis (Herde et al., 2008). ATPS04 belongs to the TPS-e/f subfamily along with the previously identified Clarkia spp. linalool synthases (Chen et al., 2011). More recently, two TPSs from Vitis...
and Phaseolus lunata, closely related Fabaceae species. On the other hand, two TPS-g subfamily proteins from the endoplasmic reticulum (Herde et al., 2008). On the localization was obtained in Arabidopsis by observing transit peptide, and direct evidence for nonplastidial inhibition GLS activity in vitro (Martin et al., 2010; Zerbe et al., 2013). However, no in planta information has been presented for these, nor any evidence showing their involvement in TMTT biosynthesis.

The common characteristic of the TPS-e/f GLSs so far identified is that they lack a predicted plastidial transit peptide, and direct evidence for nonplastidial localization was obtained in Arabidopsis by observing the AtTPS04-GUS fusion protein in the cytosol and endoplasmic reticulum (Herde et al., 2008). On the other hand, two TPS-g subfamily proteins from the closely related Fabaceae species Medicago truncatula and Phaseolus lunata (MtTPS03 and PitTPS2, respectively) were shown to be plastidic and to catalyze the formation of geranyllinalool in vitro when GGPP was provided as a substrate and also when expressed in a heterologous plant species (Arimura et al., 2008; Brillada et al., 2013). However, the same enzymes also produced linalool and nerolidol when supplied with geranyl diphosphate (GGP) and farnesyl diphosphate (FPP), respectively (Arimura et al., 2008; Brillada et al., 2013). Given the present paucity of in vivo and in vitro studies of geranyllinalool biosynthesis in plants, it is not clear whether geranyllinalool in plants is typically produced via TPS-g or TPS-e/f type TPSs, or both.

The role of geranyllinalool itself in plant tissues is not well established. Often geranyllinalool coexists in floral or vegetative tissues with its homoterpene derivative TMTT. The contribution of TMTT to the floral scent of insect-pollinated species suggests a putative role in attraction of pollinators (Tholl et al., 2011). On the other hand, in many angiosperm species, including tomato, TMTT is a component of volatile blends released from vegetative tissues upon herbivore attack, sometimes in parallel with its constitutive emission from floral tissues (Hopke et al., 1994; Ament et al., 2004; de Boer et al., 2004; Kant et al., 2004; Williams et al., 2005, Herde et al., 2008). The latter case suggests that TMTT might play a defensive role in both vegetative and floral tissues. TMTT production from insect-infested plants is considered as an indirect defense mechanism because TMTT attracts insect predators of the insect herbivores (Brillada et al., 2013). Interestingly, production of TMTT, and the homoterpene (E)-4,8-dimethyl-1,3,7-nonatriene, from herbivore-attacked lima bean plants has been found to correlate with enhanced expression of defense genes in neighboring nonaffected control plants (Arimura et al., 2000). In these cases, homoterpenses are believed to act as stress-responsive signals that enable intraspecies plant-to-plant communication.

A plant defense role has also been suggested for 17-hydroxygeranyllinalool diterpene glycosides (HGL-DTGs) present in leaves and flowers of Nicotiana species, with higher concentrations measured in buds (Heiling et al., 2010; Jassbi et al., 2010). Several studies have found negative correlation between total leaf HGL-DTG content and the mass of the larvae that feed on them (Jassbi et al., 2008; Dinh et al., 2013). Eleven HGL-DTGs that differ in sugar moieties and number of malonyl esters have been isolated from Nicotiana attenuata. The sugar groups of these compounds are Glc and rhamnose and are conjugated to the hydroxygeranyllinalool skeleton via bonds at C3 and C17 hydroxylated carbons. Additional sugars may be added to these sugars on their hydroxyl groups at C2, C4, and C6, and manolyl esters are typically formed at the C6 hydroxyl group of the glucose. The concentrations of these HGL-DTGs are higher in young and reproductive tissues. While their total levels appear to be constant, the concentration of individual compounds change upon herbivore attack, with a proportionally greater increase in malonylated compounds. Unlike many other defense-related specialized metabolites, the N. attenuata HGL-DTGs are not found on the leaf surface or the trichomes, but, instead, they accumulate inside the leaves (Heiling et al., 2010).

Here, we show that in the Solanaceae species cultivated tomato and N. attenuata, geranyllinalool is synthesized by TPSs that belong to the TPS-e/f subfamily and that the corresponding genes are related to Arabidopsis TPS04. The tomato and N. attenuata enzymes were biochemically characterized, and the kinetic parameters were determined. We also describe a detailed quantitative expression of these genes in different parts of the plant. In addition, we establish that the expression of the geranyllinalool synthase genes correlates well with the induced emission of TMTT in tomato leaves after alamethicin and methyl jasmonate (MeJA) treatments and with the total concentrations of HGL-DTGs in N. attenuata leaves and floral organs.

RESULTS

Tomato TPS46 Encodes GLS

In a previous study, we identified 44 TPS genes in the genome of tomato, of which 29 were shown or predicted to encode functional enzymes (Falara et al., 2011). Although tomato plants were previously shown to make geranyllinalool (Ament et al., 2006), a GLS could not be identified among these 29 genes based on pattern of expression or in vitro enzymatic activity (Falara et al., 2011). However, a search of an updated version of the tomato genome sequence from 2011 (http://solgenomics.net/organism/Solanum_lycopersicum/genome) identified an additional TPS gene on chromosome 3, which we named TPS46 (TPS45 is a gene found in Solanum habrochaites but not in tomato; Matsuba et al., 2013). TPS46 encodes a protein of 818 amino acids that belongs to the TPS-e/f clade and is 39% identical to Arabidopsis GLS, a higher identity level than its level of identity to any other tomato TPS protein (Fig. 1; Supplemental Fig. S1). The sequence of the N terminus of the tomato TPS46 protein (SiTPS46) is similar to that of AtGLS (Fig. 1), suggesting that it too does not have a transit peptide, and no transit peptide is predicted by SignalP. However, SiTPS46 has an internal deletion of...
Figure 1. Alignments of TPS-e/f proteins with experimentally demonstrated GLS activity and several additional TPS sequences from P. inflata, N. benthamiana, N. attenuata, and potato with high sequence identity to SlGLS. The deletions in the tomato and tobacco sequences in the γ-domain are outlined. Accession numbers of these proteins are as follows: At (Arabidopsis thaliana), Q93YV0; Sl (Solanum lycopersicum), KJ755870; St (Solanum tuberosum), KJ755872; Pi (Petunia inflata), KJ755871; Na (Nicotiana attenuata), KJ755868; Nb (Nicotiana benthamiana), KJ755869; Gh (Grindelia hirsutula), AGN70888; Vv1 (Vitis vinifera1), NP_001268201; and Vv2 (Vitis vinifera2), NP_001268004. [See online article for color version of this figure.]
42 amino acids relative to AtGLS (Fig. 1). This deletion is at the γ domain (Köksal et al., 2011) in the N-terminal half of the protein (amino acid positions 205–246 in the Arabidopsis TPS04 protein).

It was previously shown that treating detached Arabidopsis leaves with the fungal peptide alamethicin from the fungus *Trichoderma viride* led to induction of AtGLS and TMTT emission (Herde et al., 2008). To test whether alamethicin also induces TMTT emission in tomato, we placed the petioles of detached tomato leaves in water solution containing alamethicin for 24 h and observed that TMTT was emitted, whereas no TMTT emission was observed in detached leaves that were placed in water only for 24 h or in leaves that were just detached from the stem (Fig. 2). This experiment also confirmed that the major terpene volatiles from the leaf are 2-carene, β-phellandrene, germacrene, β-caryophyllene, and α-humulene, which are known to be produced mostly in the trichomes by TPS20, TPS9, and TPS12 (Schilmiller et al., 2010; Falara et al., 2011). We next measured the levels of transcripts of each of the 30 active tomato TPS genes in leaves treated with alamethicin by reverse transcription (RT)-PCR with pairs of specific primers and compared the results to the levels of the respective transcripts in leaves treated with water only for 24 h or in leaves from which RNA was isolated immediately after detachment. Alamethicin induced the expression of several tomato TPS genes to some degree, although some of them already showed expression in the two control samples (Fig. 3). However, all the TPS genes showing any level of induction by alamethicin, with the exception of TPS46, encode proteins that have been previously characterized, with none of them active with GGPP (Falara et al., 2011). TPS5, TPS7, and TPS8 encode monoterpene synthases of the TPS-b clade, TPS17 is a sesquiterpene synthase of the TPS-a clade, and TPS20 and TPS21 are, respectively, monoterpene and diterpene synthases of the TPS-e/f clade, and neither one uses GGPP as a substrate or produces geranyl-linalool (Falara et al., 2011; Matsuba et al., 2013). We also observed the induction of TPS37, a member of the TPS-g clade, but its protein has also been previously characterized and shown to catalyze the formation of linalool from GPP and nerolidol from FPP and to have no activity with GGPP (Falara et al., 2011).
The activity of SLGLS were determined with GGPP and FPP (Table I). In contrast, incubation with GPP did not give any product. The kinetic parameters for the production of nerolidol (Fig. 4, B and C). Incubation with GGPP and FPP, and GPP. GC-MS analysis of the reaction products indicated that NaGLS had activity only with GGPP (Fig. 4, D and E). The $K_m$ value of NaGLS for GGPP was 31.5 ± 4.1 $\mu$M, with a $K_{cat}$ of 7.53 ± 0.9 s⁻¹ (Table I).

**Figure 3.** RT-PCR of transcripts of all functional tomato TPS genes upon alamethicin treatment. The analysis includes the newly identified TPS46 gene in addition to all functional TPS genes described in Falara et al. (2011). The samples used were control (at 0 h) leaves from intact plants, mock control of detached leaves treated for 24 h, and leaves treated with alamethicin solution for 24 h as described in the Figure 2 legend. RT-PCR of actin transcripts was also carried out for control.

Conversely, TPS46 showed clear induction by alamethicin treatment and not by water treatment alone. We therefore synthesized the complete open reading frame of SITPS46 recoded for optimal expression in *Escherichia coli* and inserted it into the expression vector PEXP5-N-terminal His-tag. His-tagged SITPS46 was purified from *E. coli* cells by affinity purification (Fig. 4A) and used for in vitro enzyme assays with various substrates, after which products were extracted and analyzed by gas chromatography (GC)-mass spectrometry (MS). Incubation with GGPP resulted in the production of one major terpenoid product that comigrated with an authentic geranyllinalool standard and had an identical mass spectrum to the standard. The internal deletion found in SlTPS46 (Fig. 1) led to a similar pattern of increases in transcript levels and TMTT emission, peaking at 24 h after the start of the treatment (Fig. 6C). No induction of SITGLS transcripts was observed after methyl salicylate (MeSA) treatment (Fig. 6D).

To determine if TMTT emission came primarily from trichomes or the rest of the leaf tissue, we treated detached leaves with alamethicin for 24 h, removed the trichomes or the rest of the leaf tissue, and measured TMTT emission and SITGLS transcript levels from leaves with and without trichomes (Fig. 7). The levels of SITGLS transcripts in untreated whole leaves (i.e. leaves with trichomes) and in trichome derived from such leaves, relative to the housekeeping gene *ELONGATION FACTOR1A* (*SlEF1A*), was quite low (Fig. 7A), consistent with no emission of TMTT being observed from such leaves (Fig. 6). Untreated leaves from which trichomes were removed by brushing (which causes some bruising on the leaves) showed a slight induction of SITGLS. By contrast, whole leaves treated with alamethicin for 24 h showed a 30-fold increase in SITGLS transcript levels. Brushed leaves a protein with high similarity to SITPS46 (Fig. 1), which we designated as NbGLS. Because a detailed analysis of the presence and concentration of glycosylated geranyllinalool derivatives was performed on *N. attenuata* (Heiling et al., 2010), we use the sequence of NbGLS to isolate a full-length complementary DNA (cDNA) of the corresponding gene from this species, which we designated as NaGLS (Fig. 1; Supplemental Fig. S1). The NbGLS and NaGLS encode proteins that are 87% identical to each other, and they fall into the TPS-e/f clade, the same clade as SITPS46, to which they show 66.7% and 67.5% identity, respectively (Figs. 1 and 5). The two *Nicotiana* spp. GLS proteins do not contain the internal deletion found in SITPS46 (Fig. 1). The NaGLS protein was produced in *E. coli* with an N-terminal His-tag, purified (Fig. 4D), and tested for activity with GGPP, FPP, and GPP. GC-MS analysis of the reaction products indicated that NaGLS had activity only with GGPP (Fig. 4, D and E). The $K_m$ value of NaGLS for GGPP was 31.5 ± 4.1 $\mu$M, with a $K_{cat}$ of 7.53 ± 0.9 s⁻¹ (Table I).

**SIGLS Transcript Levels under Normal Conditions and after Treatments with Alamethicin, MeJA, and MeSA**

Although tomato leaf GLS enzymatic activity was previously shown to be inducible by MeJA treatment (Ament et al., 2006), transcripts of SIGLS were not previously reported in EST databases, suggesting a very low level of expression under our regular growth conditions. Analysis of SIGLS transcript levels in tomato plants grown in the growth room without insect challenge found extremely low levels of such transcripts in all tissues (Fig. 6A). Alamethicin treatment of detached leaves for 48 h showed peak transcript levels after 24 h, corresponding with peak TMTT emission (Fig. 6B). Treatment of intact leaves with MeJA led to a similar pattern of increases in transcript levels and TMTT emission, peaking at 24 h after the start of the treatment (Fig. 6C). No induction of SIGLS transcripts was observed after methyl salicylate (MeSA) treatment (Fig. 6D).

Nicotiana benthamiana and N. attenuata Have TPS Genes Encoding Proteins with High Similarity to SITPS46

We examined the recently released genome sequence of *N. benthamiana* and identified a single gene encoding
without trichome showed a slightly higher level of induction, consistent with an additive effect of the alamethicin and physical damage elicitations. Isolated trichomes from treated leaves showed about 20-fold increase (Fig. 7A).

Measurements of TMTT emission from treated whole leaves and from treated leaves after their trichomes were removed showed that TMTT emission levels were only slightly lower after the trichomes had been removed (Fig. 7B). Overall, these results indicate that SlGLS is expressed and induced in both trichomes and nontrichome leaf cells and that TMTT is produced in both cell populations. However, because trichome cells in tomato constitute only a small portion of the total mass of the leaf, it must be concluded that the majority of TMTT emission comes from nontrichome leaf cells.

To further explore the response of SiGLS to herbivory attack, intact tomato plants were challenged with two different spider mite lines, Tetranycus urticae Santpoort-2, a benchmark defense-inducing line (Kant et al., 2008), which induces TMTT emission in tomato (Ament et al., 2004), and Tetranycus evansi Vicosa-1, a benchmark defense-suppressing line, which does not induce TMTT emission in tomato (Sarmento et al., 2011). SlGLS transcript levels were measured at 1, 4, and 7 days postchallenge with each mite line.

**Table 1. Kinetic parameters of SITPS46 and NaGLS**

<table>
<thead>
<tr>
<th>Substrate</th>
<th>$K_m$ (μM)</th>
<th>$V_{cat}$ (pmol/min)</th>
<th>$K_{cat}/K_m$ (s⁻¹)</th>
</tr>
</thead>
<tbody>
<tr>
<td>SITPS46</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GGPP</td>
<td>18.7 ± 2.3</td>
<td>6.85 ± 0.71</td>
<td>0.37</td>
</tr>
<tr>
<td>FPP</td>
<td>63.7 ± 8.6</td>
<td>0.43 ± 0.26</td>
<td>0.0067</td>
</tr>
<tr>
<td>NaGLS</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GGPP</td>
<td>31.5 ± 4.1</td>
<td>7.53 ± 0.9</td>
<td>0.24</td>
</tr>
</tbody>
</table>
7 d after mite infestation. SI\( \text{GLS} \) transcript levels were already substantially and significantly increased after only 1 d with the defense-inducing spider mite line, reaching a 59-fold induction 7 d after the initiation of the infestation. Conversely, defense-suppressing mites had no effect on SI\( \text{GLS} \) transcript levels after 1 d and only minor effects after 4 and 7 d (Fig. 8A). To investigate whether the higher and lower GLS transcript levels correlated with TMTT production, we also collected the headspace of spider mite-infested and control tomato plants. Infestation with the defense-inducing spider mite line resulted in increased TMTT production, whereas the defense-suppressing mites resulted in increased TMTT production, whereas the defense-suppressing mites had no effect on TMTT levels (Fig. 8B).
**NaGLS Transcript Levels Are Constitutively High in *N. attenuata* Flowers and Are Elevated in Leaves by MeJA Treatment**

Previous investigations showed that levels of HGL-DTGs are constitutively present in buds and flowers of *N. attenuata* and, to a lower extent, in leaves found on the stem and that this species does not emit TMTT (Heiling et al., 2010). We examined the transcript levels of NaGLS in different parts of the plant (Fig. 9A) and found that they were highest in reproductive organs, which show the highest levels of HGL-DTGs, and next highest in stem leaves, which have the second-highest levels of HGL-DTGs (Heiling et al., 2010), thus showing a positive correlation, although not exactly proportional, with HGL-DTGs biosynthesis. We also examined the changes in transcript levels upon MeJA and MeSA treatments of rosette and stem leaves. Rosette leaves, which have very low levels of NaGLS transcript under unchallenged conditions, exhibited approximately 100-fold increase after 24 h of treatment with MeJA, whereas stem leaves, which have 30-fold higher levels of NaGLS transcript levels compared with the levels of NaGLS transcripts in rosette leaves in the unchallenged conditions, exhibited just about a 2-fold increase after 24 h of MeJA treatment (Fig. 9B). Treatment with MeSA resulted in no increase in NaGLS transcript levels in both types of leaves (Fig. 9C).

**Figure 6.** Expression of *SIGLS* and TMTT emission in different parts of the plant and during alamethicin, MeJA, and MeSA treatments. A, Transcript levels of *SIGLS* in different parts of the plant (trichomes, young and old leaves, stems, flowers, green and red fruit, and root tissues) relative to the transcript levels of the reference gene EF1A were analyzed in plants grown under normal conditions. No TMTT was detected from the plant under normal conditions. B, Relative increase in level of *SIGLS* transcripts (relative to 0-h time point of control) during a 48-h treatment with alamethicin. The petioles of detached leaves were submerged in alamethicin water solution. The mock control consisted of submerging the petioles of detached leaves in water. C, Relative increase in level of *SIGLS* transcripts (relative to 0-h time point) during a 48-h treatment with MeJA vapors in an airtight container. D, Relative increase in level of *SIGLS* transcripts (relative to 0-h time point) during a 48-h treatment with MeSA vapors in an airtight container. Transcript levels were determined by real-time qPCR with at least three biological replicates and three technical replicates for each biological sample. The error bars represent the ± sd. In B, C, and D, all values are compared to the level in control leaves, which is set at 1. DW, Dry weight.
DISCUSSION

**SlTPS46 Encodes GLS**

We previously demonstrated that a total of 44 TPS genes could be found in the version of the tomato genome sequence released in 2010, of which 29 were shown or predicted to encode functional enzymes (Falara et al., 2011). Only three of the tomato TPS enzymes characterized to date were found to be diterpene synthases, including the two TPS enzymes involved in GA biosynthesis, copalyl diphosphate synthase (CPS) and kaurene synthase (KS; Falara et al., 2011), and TPS21, which uses neryl diphosphate to make a cyclic diterpene (Matsuba et al., 2013). However, the presence of the specialized linear diterpene metabolite geranyllinalool (which could be enzymatically converted to TMTT) had been reported in tomato, together with an enzymatic activity that uses GGPP to catalyze the formation of geranyllinalool (Ament et al., 2006). Thus, the identity of the TPS gene encoding GLS was unresolved.

A recently updated tomato genome sequence (http://solgenomics.net/organism/Solanum_lycopersicum/genome) contains a previously unavailable TPS sequence on chromosome 3 that encodes a protein with similarity to Arabidopsis GLS (Figs. 1 and 5). Because alamethicin is a known inducer of Arabidopsis GLS and of TMTT emission and it also induces TMTT emission in tomato (Fig. 2), we used it to determine which of the tomato TPS genes encode GLS. Levels of transcript accumulation of all functional tomato TPS genes, including the newly identified SlTPS46 (a TPS45 gene is found in *S. habrochaites* but not in tomato), were measured in leaf tissue after 24 h of alamethicin treatment and compared to levels in nontreated leaves (Fig. 3). Some genes showed increase transcript accumulation following alamethicin treatment, but none of these genes, with the exception of SlTPS46, encoded proteins that can use GGPP as a substrate (Falara et al., 2011). On the other hand, SlTPS46, a gene highly similar to AtTPS04, showed strong induction with alamethicin (Figs. 4 and 7), and the purified recombinant TPS46 exhibited GLS activity when given GGPP as a substrate. Its $K_m$ value for GGPP was 18.7 $\mu$M, and its $K_{cat}$ value was 6.85 s$^{-1}$ (Table I). SIGLS also exhibits some activity with FPP, catalyzing the formation of nerolidol, but its efficiency with GGPP was 55-fold higher than with FPP (Table I). It had no activity with GPP. To the best of our knowledge, no kinetic parameters have been reported for previously identified GLSs. Thus, our data identified SlTPS46 as encoding GLS; however, in the absence of a mutant line defective in SlTPS46, we cannot rule out the possibility that the tomato genome has additional gene(s) encoding GLS activity.

---

**Figure 7.** Induced *SIGLS* transcript levels and TMTT emission in whole leaves and leaves without trichomes. Quantitative RT-PCR analysis of *SIGLS* transcript levels (A) and measurements of TMTT emission (B) for whole leaves treated with alamethicin and from treated leaves after trichomes have been removed by brushing. RT-PCR was also done on *SIGLS* transcripts in trichomes isolated from treated leaves. Samples were analyzed 24 h after detached leaves were placed in 50 $\mu$g ml$^{-1}$ alamethicin solution. The error bars represent the $SD$ of three biological replicates. Bars annotated with the same letter are not significantly different according to Fisher’s LSD post hoc test ($P < 0.05$).
Interestingly, SIGLS has a deletion of 42 amino acids in its γ domain in the N terminus. This domain is usually not active in Type I diterpene synthases that use GGPP to directly make a diterpene olefin without a copalyl disphosphate intermediate, although it is often conserved in these enzymes for unknown reasons (Hillwig et al., 2011). A very closely related sequence from potato (Solanum tuberosum), a species in the same genus as tomato, also has a similar (but slightly shorter) deletion at this position, but N. attenuata GLS and related sequences from N. benthamiana and Petunia inflata (all in family Solanaceae) do not. The N termini of all of these proteins, including SIGLS, coincide with the N terminus of AtGLS, which was shown to have no transit peptide and to be localized in the cytosol. It is therefore likely that tomato GLS is located in the cytosol.

**SIGLS and NaGLS Have Very Different Expression Characteristics**

Our analyses of the expression of GLS in tomato and N. attenuata showed that whereas SIGLS is strictly inducible in response to cues of biotic challenge (Figs. 7–9), NaGLS is constitutively expressed in flowers and, to a lower degree, in leaves (Fig. 9), consistent with previous reports showing the presence of HGL-DTGs in these organs (Heiling et al., 2010). However, some induction in N. attenuata leaves was observed upon MeJA treatment, with a higher fold induction in rosette leaves, whose constitutive level of expression was much lower than in the stem leaves. We previously showed that attack by the generalist spider mite T. urticae Santpoort-2 on tomato induces jasmonate responses and causes emission of TMTT (Kant et al., 2008), thereby augmenting the plant’s attractiveness to predatory mites, the spider mite’s natural enemy (Ament et al., 2004). Here, we show that during such an interaction between T. urticae Santpoort-2 and tomato, the levels of steady-state SIGLS transcripts as well as TMTT emission were greatly increased, suggesting that the increase in TMTT emission is partially determined by pretranslational regulation (Fig. 8, A and B). When the tomato specialist T. evansi Viçosa-1 was feeding on the plants, the increase in steady-state level of SIGLS transcripts was approximately 8-fold lower compared with T. urticae Santpoort-2 infestation, and no increase in TMTT emission was observed (Fig. 8, A and B), consistent with previous results showing that T. evansi Viçosa-1 suppressed jasmonate responses in tomato.

![Figure 8. Transcript levels of GLS in and TMTT emission from spider mite-infested tomato leaflets. Control plants are free of spider mites; Santpoort-2 is a T. urticae, and Viçosa-1 is a T. evansi line. The bars in A represent the mean GLS transcript levels plus the standard error of the mean in four plants, normalized for EF1a, with the lowest value set to 1. The bars in B represent the amount of TMTT emitted by three plants during 24 h from the fourth to the fifth day post infestation. Numbers above or in the bar indicate the mean value represented by the bar. Bars annotated with the same letter are not significantly different according to Fisher's LSD post hoc test (P > 0.05).](image-url)
and that their infestation of tomato did not lead to TMTT production (Sarmento et al., 2011). We note that in the experiments measuring TMTT emission (Fig. 8B), one of the control plants emitted some TMTT, but the other two had undetectable levels of TMTT emission. While the outlier control plant raised the average of TMTT emission for the control group, analysis showed that there was no statistically significant difference in emission between the control plants and plants infested with *T. evansi* Viçosa-1.

Overall, the function of GLS in both species appears to be related to defense. However, the end products, TMTT in tomato and HGL-DTGs in *N. attenuata*, appear to have different functions. HGL-DTGs are non-volatile compounds that are toxic to the native insect herbivores, including the larvae of the specialist tobacco hornworm (*Manduca sexta*; Heiling et al., 2010). TMTT, on the other hand, is a volatile that has been shown in tomato and other plant species to be emitted under insect attack and thereby attract insect predators of the herbivorous insects (Kant et al., 2004; Brillada et al., 2013). Thus, while the gene encoding the enzyme GLS appears to have been present in their common ancestor, the South American species tomato and the western North American species *N. attenuata* evolved the ability to make different end products from geranyllinalool, likely the result of selection exerted by local herbivores. Our results underscore the role that contingency and chance play in the evolution of new biochemical pathways.

Cytosolic GLSs Constitute a Deep Branch of the TPS-e/f Clade of the TPS Family

TMTT emission has so far been reported from basal dicot families, eudicots, and monocots but not from any plant species outside the angiosperms (Tholl et al., 2011). To date, all experimentally demonstrated GLS enzymes have been characterized only from eudicot angiosperms, and five of these seven proteins, found in four different families, belong to the TPS-e/f clade, while the others, two closely related sequences from *M. truncatula* and *P. lunatus* (both in Fabaceae), belong to the TPS-g clade. The TPS-g clade originated within the angiosperms, and it contains mostly mono- and sesquiterpene synthases (Chen et al., 2011), including *SlTPS37*, which is inducible by alamethicin (Fig. 3) and encodes a protein that catalyzes the formation in vitro of linalool from GPP and nerolidol from FPP but is inactive with GGPP (Falara et al., 2011). It is therefore likely that the reported *M. truncatula* and *P. lunatus* GLSs evolved their ability to make geranyllinalool from GGPP only recently and perhaps their main role lies elsewhere, because they are more efficient in catalyzing the formation of nerolidol from FPP and linalool from GPP (Arimura et al., 2008; Brillada et al., 2013).

The GLSs of the TPS-e/f clade, on the other hand, form a deep branch in this clade that diverged from the KS lineage before the split between the gymnosperms and eudicots. The GLSs of this clade have been observed in eudicots, but they are not present in any monocots or in any angiosperm outside the apocynaceous clade (Chen et al., 2011).

Figure 9. Expression of NaGLS in different parts of the plant and during MeJA and MeSA treatments. A, Transcript levels of NaGLS in different parts of the plant (rosette and stem leaves, stems, roots, flower, and bud and their individual parts, sepals, petals, ovary, anthers, stamens, stigma, and style) relative to the transcript levels of the reference gene EF1A were analyzed in plants grown under normal conditions. No TMTT was detected from the plant under normal conditions. B, Relative increase in levels of NaGLS transcripts (relative to 0-h time point) during a 48-h treatment with MeJA vapors in an airtight container. C, Relative increase in level of SlGLS transcripts (relative to 0-h time point) during a 48-h treatment with MeSA vapors in an airtight container. Transcript levels were determined by real-time qPCR with at least three biological replicates and three technical replicates for each biological sample. The error bars represent the so.
and the angiosperms. TPSs in this lineage in four disparate angiosperm families act as GLS suggests that the founding gene of this lineage also encoded GLS. Interestingly, while many additional sequences from eudicots, as well as one sequence from the basal dicot Amborella trichopoda, belonging to this clade can be found in the database, no such sequences are found in monocot species or from any nonangiosperm species (Supplemental Fig. S2). While sequence information for gymnosperms is still incomplete, the absence of GLS-like sequences in the monocots, for which several full genome sequences are available, suggests that these sequences must have been lost in the monocot lineage. It thus appears that monocot species must have evolved a different mechanism to make TMTT.

When terpenes evolved their various ecological functions is not clear. It was recently found that Selaginella moellendorffii, classified as a fern ally, contains a set of genes related to microbial TPSs (called microbial TPS-like enzymes) that specify the production of monoterpenes and sesquiterpenes, but microbial TPS-like enzymes have not yet been identified in genomes of nonbryophyte plants. Physcomitrella patens, a bryophyte, has only one functional TPS gene encoding a bifunctional enzyme with both CPS and KS activities. This gene is believed to be ancestral to the entire plant TPS family, giving rise initially to a TPS-c clade that contained CPS genes and a TPS-e/f clade that contained KS genes (Chen et al., 2011). While CPS and KS work in tandem to convert GGPP to which then serves as a precursor to GAs, it has also been shown that ent-kaurene itself is emitted from some plants (Otsuka et al., 2004). Our phylogenetic analysis suggests that the ancestor of GLS and GLS-like genes in angiosperm species began diverging from the KS gene lineage at a time prior to the split of the angiosperm and gymnosperm lineages. Thus, geranylvinalool is likely to be one of the first specialized metabolites that were produced by the action of plant enzymes belonging to the plant TPS family.

**MATERIALS AND METHODS**

**Plant Growth and Treatment Conditions**

Seeds of tomato (Solanum lycopersicum) were obtained from the Tomato Genetic Resource Center (http://tgc.ucdavis.edu). When not specifically indicated, the tomato plants used were of cv MVP. Seedlings were grown in Jiffy peat pots (Hummert International) in a controlled-growth room maintained for 16 h in the light at 28°C and 8 h in the dark at 20°C. Nicotiana attenuata seeds were kindly provided by Dr. Ian T. Baldwin, and seed germination was performed with the BD SMART RACE (Invitrogen). PCR amplification was performed with the KOD DNA polymerase (Novagen) and the following PCR conditions: 2 min at 95°C and then 20 s at 95°C, 10 s at 58°C, and 30 s at 68°C for 30 cycles. For SlTPS46, the gene was obtained by RT-PCR amplification with gene-specific primers (Supplemental Table S1) based on genomic DNA sequence information. The RT-PCR product was excised from the agarose gel and purified and quantified by measuring absorbance at 260 nm using a Nanodrop spectrophotometer. The gene was amplified using primers T46F/T46R and transferred to pEXP5-NT/TOPO to be fused to an N-terminal His-tag coding region. Expression in *E. coli* was induced by adding isopropyl-D-galactopyranoside at 0.1 mM to the growth media 4 h before harvest and then induced with 0.4 mM isopropyl-D-galactopyranoside at 18°C for 16 h. Cell pellets were kept at −80°C until used for affinity purification with His-Select HF Nickel Affinity Gel (Sigma). Because levels of expression in *E. coli* were relatively low, a 200-ml culture was induced with isopropyl 1-thio-D-galactopyranoside as described above.

**Biotic Stresses**

Spider mite *Tetranychus urticae* Santpoort-2 (Cytochrome C oxidase subunit I-sequence KF447575 in GenBank) was described previously by Kant et al. (2008), where it was referred to as line KMB and has been propagated on detached leaves of Phaseolus vulgaris. *Speedy* in a climate room at 25°C and a 16-h/8-h light regime with 300 μmol m⁻² s⁻¹ and 60% relative humidity. Spider mite *Tetranychus evansi* Viçosa-1 (Cytochrome C oxidase subunit I-sequence KF447575 in GenBank) was described previously by Sarmento et al. (2011) and has been propagated on detached leaves of tomato 'Castlemart' under the same conditions. Twenty-one-day-old tomato plants were infested with 45 adult female mites as described in Kant et al. (2004). Total RNA was isolated with a phenol-D-LiCl-based method and treated with Turbo-DNAsafe (Ambion). Subsequently, 2 μg of the DNA-free samples was used for cDNA synthesis using M-MuLV RT (Fermentas) as a template for quantitative PCR (qPCR) using the Platinum SYBR Green qPCR SuperMix-UDG Kit (Invitrogen) in the ABI 7500 Real-Time PCR System (Applied Biosystems). The cycling program was set to 45 cycles of 15 s at 95°C and 1 min at 60°C, followed by a melting curve analysis. The normalized expression (NE) data were calculated as follows: \( NE = \left( \frac{1}{PE_{target}} \right) / \left( \frac{1}{PE_{reference}} \right) \), where PE refers to primer efficiency and Ct refers to cycle threshold. The primer efficiencies were determined by fitting a linear regression line on the cycle threshold values of a standard cDNA dilution series. NE of each target gene was compared per time point independently using a generalized linear model (γ error distribution and logarithmic as link function). The model included spider mite strain as factor, and the technical replicate (for which each reaction was nested to biological replicate and included as nested factor. Means of each group were contrasted by Fisher’s LSD post hoc test using PASW Statistics 17.0.

The volatiles in the headspace of control and spider mite-infested plants were collected and analyzed according to Ament et al. (2010) with the following modifications: volatiles were trapped on 5-mm-wide glass tubes filled with 100 mg of Porapak Q/100 mesh (Supelco); compounds were separated on a capillary HP-5MS column (5%-Phenyl)-methylpolysiloxane; 30 m × 250 μm × 0.25 μm; Agilent 5973 N GC columns, http://www.chrom.agilent.com with the temperature of the GC oven set to 40°C for 3 min, after which the temperature was raised to 250°C at 10°C per min and the column flow was 1.2 mL of He per min. Mass spectra of the eluting compounds were collected on an Agilent 7820 Q-TOF MS, with an acquisition rate of 20 spectra s⁻¹. TMTT was identified and quantified on the basis of pure TMTT after normalization to the internal standard (benzyl acetate). Differences in volatile production between treatments and control were tested with a general linear model (log transformed) with replicates as a random factor, followed by LSD post hoc test.

**Characterization of SITPS46 and NaGLS**

Full-length cDNA for SITPS46 were obtained by RT-PCR amplification with gene-specific primers (Supplemental Table S1) based on genomic DNA sequence information. The RT-PCR product was excised from the agarose gel and subjected to electrophoresis in a 1.5% agarose gel containing 0.5 ng/ml ethidium bromide in TAE buffer. The cDNA was cloned into pEXP5-NT and sequenced. The KOI gene was amplified from *Phaseolus vulgaris* cDNA by PCR using primers 5′-CTG GGA TGA GCA GCT GGT ATG-3′ and 5′-CTT CAC TGG TGT CGG ACT AGA-3′. The PCR product was cloned into pET32a (Novagen) and transferred to E. coli BL21-plysS cells (Invitrogen) containing the plasmids pEXP5-NT. The fusion protein was purified using a Ni-affinity column and used for mass spectrometric analysis using Agilent 6520 Q-TOF LC/MS. The model included spider mite strain as factor, and the technical replicate (for which each reaction was nested to biological replicate and included as nested factor. Means of each group were contrasted by Fisher’s LSD post hoc test using PASW Statistics 17.0.

**SITPS46 and NaGLS Expression in *E. coli* and Affinity Purification**

*E. coli* BL21-lysS cells (Invitrogen) containing the plasmids pEXP5-NT-TOPO were grown in Luria-Bertani medium containing the appropriate antibiotics until optical density of the culture at 600 nm reached 0.5 to 0.7 and then induced with 0.4 mM isopropyl 1-thio-D-galactopyranoside at 18°C for 16 h. Cell pellets were kept at −80°C until used for affinity purification with His-Select HF Nickel Affinity Gel (Sigma). Because levels of expression in *E. coli* were relatively low, a 200-ml culture was induced with isopropyl 1-thio-D-galactopyranoside as described above.
SITPS46 and NaGLS Enzymatic Assays

Elution buffer containing the purified His-tagged SITPS46 or NaGLS protein was exchanged with assay buffer containing 50 mM HEPES, 7.5 mM MgCl₂, 100 mM KCl, 5 mM dithiothreitol, and 10% (v/v) glycerol, pH 7.0, using Ultrafiltr (Amicon Ultra). The enzyme reactions contained 40 μM of partially purified protein in assay buffer and 60 μM prenyl diphosphate substrates: GPP, FPP, and GGPP (Echelon Biosciences).

GC-MS Analyses

Volatiles were collected by solid phase micro extraction (SPME) or by hexane extraction as described in Falara et al. (2011). Samples were injected into an EC-WAX column (Grace Davison; 30-m length, 0.25-μm film thickness, and 0.32-mm i.d.) on a GC17-A (Shimadzu) coupled to a QP-5000 GC-MS system. Injector temperature was 220°C and working on splitless mode. In the case where SPME fiber was used for extraction, the tissue was exposed to the fiber in a 2-ml glass vial for 15 min at 42°C.

Gene Expression Analyses

Expression Analysis of Tomato TPS Genes Performed by RT-PCR

Total RNA was isolated with the RNeasy kit (Qiagen), treated with the DNA-free kit (Ambion) to remove genomic DNA contamination, and used for first-strand cDNA synthesis using SuperScript II reverse transcriptase and oligo(dT)₁₂–₁₈ primer (Invitrogen) according to the manufacturer’s protocol. PCR amplification was performed with GoTaq Green Master Mix (Promega) using an initial denaturation step for 2 min at 94°C, followed by denaturation for 30 s at 94°C, annealing for 30 s at 68°C, extension for 30 s at 72°C, and a final extension of 5 min for 30 cycles. Tomato TPS gene-specific primers and actin primers used can be found in Falara et al. (2011). Reactions were done with RNA isolated from three identical experiments.

Expression Analysis of SITPS46 and NaGLS Performed by Quantitative RT-PCR

Total RNA extraction and first-strand cDNA synthesis were performed as mentioned above. The resulting cDNA was diluted 10-fold, and 1 μl was used as a template for PCR amplification in a 30-μl reaction using Power SYBR Green PCR master mix (Applied Biosystems) and gene-specific primers (Supplemental Table S1). Reactions were performed with the StepOnePlus Real-Time PCR System (Applied Biosystems) with the following cycles: 95°C for 30 s at 94°C, followed by 40 cycles of 95°C for 15 s and 60°C for 1 min. A final extension of 5 min for 30 cycles. Tomato TPS gene-specific primers and actin primers used can be found in Falara et al. (2011). Reactions were done with RNA isolated from three independent experiments.

Sequence data from this article can be found in the GenBank accession nos. X14449. Measurements were done by triplicates on RNA isolated from each of three independent experiments.

Supplemental Data

The following materials are available in the online version of this article.

Supplemental Figure S1. Sequences of SITPS36 and NaGLS.

Supplemental Figure S2. A phylogenetic tree of GLS protein sequences and biochemically characterized and uncharacterized sequences that fall inside the GLS clade.

Supplemental Table S1. A list of the oligonucleotide primers used in this work.

ACKNOWLEDGMENTS

We thank Dr. Ian T. Baldwin for providing seeds of Nicotiana attenuata and Drs. Yuki Matsuba and Reza Sohrabi for logistics help with some of the experiments.

Received June 1, 2014; accepted July 21, 2014; published July 22, 2014.

LITERATURE CITED


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


Iorizzi M, Lanzotti V, De Marino S, Zollo F, Blanco-Molina M, Macho A, Muñoz E (2001) New gla-1 up to 275°C, and hold for 1 min. Solvent extractions were done with methyl tert-butyl ether containing tetradecane as internal standard. In the case where SPME fiber was used for extraction, the tissue was exposed to the fiber in a 2-ml glass vial for 15 min at 42°C.

Drs. Yuki Matsuba and Reza Sohrabi for logistics help with some of the experiments.
Jassbi AR, Zamanizadehnaajari S, Baldwin IT (2010) 17-Hydroxygeranylinalool glycosides are major resistance traits of Nicotiana obtusifolia against attack from tobacco hornworm larvae. Phytochemistry 71: 1115–1121


Solanum lycopersicum

TPS46, cDNA (SlTPS46, Accession number KJ755870)

ATGGATTCCTCACCATTGCTTCCTATGTCAATCATCGTGAACAAAAGAATGTCGCAATTTGCAAAATCATAGACCCATTGCAAGAGAAGCTATGGTGAATGCTTCCATGCT

SlTPS46, optimized

ATGGACTCCTCCCCGCTGTCCTCAATCGAAAGCCTGGTTCACGACATCAAAAAAGACCTGTTCAGTAATCAAGAA

1
>SlTFS46, protein
MDSSLSSIESLVDIKDLFSNQENFNTLFTKTFAYDIWAMIPFDNQKEANGPMFVSCSLWILNNQEQGFW

GESENGQNSPTINTLPTILAMCVMKKNVQERIHKGLFVHANMYTLKENSEQCFCPCNMTIVFPSILQAKAGG
LEIIVNSKLSSLDNISDNEAFAQFSTGNGKSSDNKLCQYLMISYQPCGNSVPFSDVEKMLC
MVDNITQRMGKHEKIEQILQDYVNYQINNKLNNREYLSSEESKTSDDLDPLKKDLSSFVRLRLLRQGYKINQG
SFCWFLQPHLKDIEKINKEFQFTYVMNYRYATLDMFNGESQMEEARFARNFLNNNDYNLSLFSLQKYIEHELNVP
WFARLEHDHLDRWILIENQPSLGSKSSFYLWSLYQNEKLMLQAVQNYEQSYIRKEELDRLKNSKEKGLVD1GF
REKTYSYPASSAASASSFPELSFLRFLRIVAKCSIITIVADDFYDEAEASLHELHSEAIQWRDANKLGDPSKIFEEA
LDDLVCVDVAKMYHLQHIDITPQLRHMWEFDMESLMSRTSNPLSRDKLYQVMISVGAIHLLHAASLGC
LPNQVFPGINGQYENITKLLLMTTRRDQSYQEREVGKMNYYLHNLNPRQIESIDFVKYLGDKKEFLE
NVIMDGNFDMRTKNLHLSLCLNVFHFNSCNFLDTSKSALEDIMRAIYFILPGQEHTKVAIQEKKIKKIEIVK
VNACVNYEINMKQLQVQVGSYVQKRNPKNSFLSTKFGSHNLRCPM

>Nicotiana attenuata putative GLS, cDNA (NaGLS, Accession number KJ755868)
ATGGATTCCTCCTCCATTGTGTCTTCCATTGAATCTCTAGTCAATGAGATCAAGAAAGAATATGTTCTCAAACCAAGA
ATTTAACACTTTTTGTACCTACCAATATCCTGGCATAGCAGTCTGGTTGAGTTGGAATGAGGAAATGGATGTGACAGC
AGGGCATGATCAACCTCTTTTGGCATATCTAGAGAGCTTGCCAAATTATTTGTTGATCAAAAAGACCAAATCCTAAAACATTTAAGTGAAGATGGTTCTTTGTTTCAATCTCCCTCTGCTACA
GCACAAGCATTCATGTCACAGGGAACCAAAAATGTCTGAGCAATGAGGACAACTTATTACTACTTCCTTC
ACTTCAAAAAGTGTAGTTGAGCATGAGCTACATCTCTTCTTGGCATATTGAGCTGGCAAATTTTGAATGAGGAATGGTTGAATGTTCCATGGGTTGCTCGATTAGAACATCTTGATCACAGATTATGGGA
TTGAATTAAGCAAACGTTCTTCTTGGCTGAAATGAGCTTTTGAACCTTTGAACTTGGAACAGATCTGTGAGAATG
TTTGGAACATTTGGACAGTCTGAGTCTATGAAAGAAGAATATGAGAGCTATTTACATTCCTCTTCAAGAACAAACCTCAATTCCATCATTAAAAACTTTC

2
>NaGLS, optimized
GAATTCGACGACGACGATCTGCTGCTACGCTCTATCGACGAAACGCTGGTCAACGAAATCTAAGACACATGGTTTACCA
ATCAAGAATTTAACACTTTTGTGACCGCTGTATTCGTGATGATATACATCGAGTGGCCATGAGAAGTATGCTCCCTGCT
TTGCAATGGGAAAACTTCAAACATCAAGTTGGCATTGGAATAAGTTTTGTGTGGAAGTAAAGCTCCTAAGAATCATC
CTTTTGGTTGGAATACAAAAGGATTTGCTTCACATAAGCTAAGTTCATGTTTCATGTAA
>NaGLS, protein
MDSSSIVSSIESLVNEIKKDMFSNQEFNTFVTPISAYDTAWLAMISYNNQEKATNDHSFSGPMFESCLNWILNNQNE
QGFWGSLSGENLPTIYALPTTLACMIALKKWDVGETNIKKGLKFIHANTEILLKENSQQFPRWFTIVFPAMVQLAKS
VGLEIIFVNGSERRLSVDVVKRVRILESEKLMVDTANGHQPLLAPLESIPNYFDQDKILKHLSEDGSLFSQPSAT
AQAFMSTGNCKCELYMSIHSMVSNEMNTTATCGTCTTTCGCGGCTGATGCATATTATCGGCCATGAGAAGTACTGCTG
GCGCGAGTTCTCCTATGTTTGCCTGGTCGAATGCTGTCGTCGGCAGAATGCTCACTTGCCCGCTGCAACTCGGTTACG
GCAGCAATGATTCGAGAAGGACACTCTAAGTCTTGATGTTTCTGAGTTTTTGAAATGATGCTACGGCAGCTGTCGACC
GCTGAGTATGAAAATATCACCAAACTGCTGGACCCGCGTGTCAAAGGCTACTCAGGGATGTGATGCTGCTGCTGCTG
TCCCAATTTGGAATACAAAGAGAAGAGAAAGAGGAGAAAGAGTGGAGAATATTCCAACTAAAGTTTCTGCTT
GTCTCAATGGGAAAAACTTCAAACATCAAGTTGGCATTGGAATAAGTTTTGTGTGGAAGTAAAGCTCCTAAGAATCATC
CTTTTGGTTGGAATACAAAAGGATTTGCTTCACATAAGCTAAGTTCATGTTTCATGTAA
>NaGLS, protein
MDSSSIVSSIESLVNEIKKDMFSNQEFNTFVTPISAYDTAWLAMISYNNQEKATNDHSFSGPMFESCLNWILNNQNE
QGFWGSLSGENLPTIYALPTTLACMIALKKWDVGETNIKKGLKFIHANTEILLKENSQQFPRWFTIVFPAMVQLAKS
VGLEIIFVNGSERRLSVDVVKRVRILESEKLMVDTANGHQPLLAPLESIPNYFDQDKILKHLSEDGSLFSQPSAT
AQAFMSTGNCKCELYMSIHSMVSNEMNTTATCGTCTTTCGCGGCTGATGCATATTATCGGCCATGAGAAGTACTGCTG
GCGCGAGTTCTCCTATGTTTGCCTGGTCGAATGCTGTCGTCGGCAGAATGCTCACTTGCCCGCTGCAACTCGGTTACG
GCAGCAATGATTCGAGAAGGACACTCTAAGTCTTGATGTTTCTGAGTTTTTGAAATGATGCTACGGCAGCTGTCGACC
GCTGAGTATGAAAATATCACCAAACTGCTGGACCCGCGTGTCAAAGGCTACTCAGGGATGTGATGCTGCTGCTGCTG
TCCCAATTTGGAATACAAAGAGAAGAGAAAGAGGAGAAAGAGTGGAGAATATTCCAACTAAAGTTTCTGCTT
GTCTCAATGGGAAAAACTTCAAACATCAAGTTGGCATTGGAATAAGTTTTGTGTGGAAGTAAAGCTCCTAAGAATCATC
CTTTTGGTTGGAATACAAAAGGATTTGCTTCACATAAGCTAAGTTCATGTTTCATGTAA
>NaGLS, protein
MDSSSIVSSIESLVNEIKKDMFSNQEFNTFVTPISAYDTAWLAMISYNNQEKATNDHSFSGPMFESCLNWILNNQNE
QGFWGSLSGENLPTIYALPTTLACMIALKKWDVGETNIKKGLKFIHANTEILLKENSQQFPRWFTIVFPAMVQLAKS
VGLEIIFVNGSERRLSVDVVKRVRILESEKLMVDTANGHQPLLAPLESIPNYFDQDKILKHLSEDGSLFSQPSAT
AQAFMSTGNCKCELYMSIHSMVSNEMNTTATCGTCTTTCGCGGCTGATGCATATTATCGGCCATGAGAAGTACTGCTG
GCGCGAGTTCTCCTATGTTTGCCTGGTCGAATGCTGTCGTCGGCAGAATGCTCACTTGCCCGCTGCAACTCGGTTACG
GCAGCAATGATTCGAGAAGGACACTCTAAGTCTTGATGTTTCTGAGTTTTTGAAATGATGCTACGGCAGCTGTCGACC
GCTGAGTATGAAAATATCACCAAACTGCTGGACCCGCGTGTCAAAGGCTACTCAGGGATGTGATGCTGCTGCTGCTG
TCCCAATTTGGAATACAAAGAGAAGAGAAAGAGGAGAAAGAGTGGAGAATATTCCAACTAAAGTTTCTGCTT
GTCTCAATGGGAAAAACTTCAAACATCAAGTTGGCATTGGAATAAGTTTTGTGTGGAAGTAAAGCTCCTAAGAATCATC
CTTTTGGTTGGAATACAAAAGGATTTGCTTCACATAAGCTAAGTTCATGTTTCATGTAA
>NaGLS, protein
MDSSSIVSSIESLVNEIKKDMFSNQEFNTFVTPISAYDTAWLAMISYNNQEKATNDHSFSGPMFESCLNWILNNQNE
QGFWGSLSGENLPTIYALPTTLACMIALKKWDVGETNIKKGLKFIHANTEILLKENSQQFPRWFTIVFPAMVQLAKS
VGLEIIFVNGSERRLSVDVVKRVRILESEKLMVDTANGHQPLLAPLESIPNYFDQDKILKHLSEDGSLFSQPSAT
AQAFMSTGNCKCELYMSIHSMVSNEMNTTATCGTCTTTCGCGGCTGATGCATATTATCGGCCATGAGAAGTACTGCTG
GCGCGAGTTCTCCTATGTTTGCCTGGTCGAATGCTGTCGTCGGCAGAATGCTCACTTGCCCGCTGCAACTCGGTTACG
GCAGCAATGATTCGAGAAGGACACTCTAAGTCTTGATGTTTCTGAGTTTTTGAAATGATGCTACGGCAGCTGTCGACC
GCTGAGTATGAAAATATCACCAAACTGCTGGACCCGCGTGTCAAAGGCTACTCAGGGATGTGATGCTGCTGCTGCTG
TCCCAATTTGGAATACAAAGAGAAGAGAAAGAGGAGAAAGAGTGGAGAATATTCCAACTAAAGTTTCTGCTT
GTCTCAATGGGAAAAACTTCAAACATCAAGTTGGCATTGGAATAAGTTTTGTGTGGAAGTAAAGCTCCTAAGAATCATC
CTTTTGGTTGGAATACAAAAGGATTTGCTTCACATAAGCTAAGTTCATGTTTCATGTAA
Supplemental Figure 2. A phylogenetic tree of GLS protein sequences and biochemically characterized and uncharacterized sequences that fall inside the GLS clade. These sequences come from species in diverse eudicot families and a basal dicot, Amborella trichopoda, but none have been found in monocots or non-angiospermous plants. For each sequence, the name of the species it comes from and the accession number of the corresponding protein are provided, as well as the enzymatic function when characterized (GLS: geranyllinalool synthase; LIN: linalool synthase). Phylogenetic analysis was conducted using the MAFFT algorithm (Katoh et al., 2013). Kaurene synthase (KS) protein sequences from S. lycopersicum, Oryza sativa, and Picea glauca serve as the outgroup.
<table>
<thead>
<tr>
<th>Primer sequence (5′-3′)</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Full length cloning primers for cDNA amplification</strong></td>
<td></td>
</tr>
<tr>
<td>FL_SlTPS46-FOR</td>
<td>ATGGATTCCTCACATTCTCTTATGAATC</td>
</tr>
<tr>
<td>FL_SlTPS46-REV</td>
<td>TTACATGAAAATACATATGATAGTGAAGC</td>
</tr>
<tr>
<td>FL_NaGLS-FOR</td>
<td>ATGGATTCCTCCATATTGTGAAGC</td>
</tr>
<tr>
<td>FL_NaGLS-REV</td>
<td>TTACATGAAAATGACTTTAGTTATAGT</td>
</tr>
<tr>
<td><strong>Degenerate primers for N. attenuata GLS amplification and gene-specific primers for 3′ and 5′ RACE</strong></td>
<td></td>
</tr>
<tr>
<td>degTPS46-FOR</td>
<td>GCNTAYGAYAYNGCNTGGYT</td>
</tr>
<tr>
<td>degTPS46-REV</td>
<td>TYAYATRAAACCATGATGGATTTGTGAAGCAAA</td>
</tr>
<tr>
<td>Na_46_1F</td>
<td>TGGTTGGCCACCTTGCTCATTT</td>
</tr>
<tr>
<td>Na_46_2F</td>
<td>CTTGATGGAATGGAAAG</td>
</tr>
<tr>
<td>Na_46_1R</td>
<td>CCACAAATCCTTACCTTTT</td>
</tr>
<tr>
<td>Na_46_2R</td>
<td>GAAGCAGCATGAAGAACAA</td>
</tr>
<tr>
<td><strong>Full length cloning primers for E. coli expression</strong></td>
<td></td>
</tr>
<tr>
<td>FL_optTPS46-FOR</td>
<td>ATGGACTCCTCCCCGCTGTC</td>
</tr>
<tr>
<td>FL_optTPS46-REV</td>
<td>TCACATGAAAGCAGACAGCAAGCATTATG</td>
</tr>
<tr>
<td><strong>RT-PCR primers for gene expression</strong></td>
<td></td>
</tr>
<tr>
<td>RT_SlACT-FOR</td>
<td>ATTCCCTGACTGTTTGTAGT</td>
</tr>
<tr>
<td>RT_SlACT-REV</td>
<td>TCCAACAAATACGGGTGTG</td>
</tr>
<tr>
<td>RT_SlTPS46-FOR</td>
<td>GTCTTCAAGATTTCCGAGTGAAGA</td>
</tr>
<tr>
<td>RT_SlTPS46-REV</td>
<td>CAATCCATAATCTGTGATCAAGATGTTCTAATCG</td>
</tr>
<tr>
<td><strong>qRT-PCR primers for gene expression</strong></td>
<td></td>
</tr>
<tr>
<td>qPCR_SlEF1A-FOR</td>
<td>AGCTTCACTGCCCAGGGCTATCATC</td>
</tr>
<tr>
<td>qPCR_SlEF1A-REV</td>
<td>ACCATACCAGCATACCAGGCTTTCA</td>
</tr>
<tr>
<td>qPCR_SlTPS46-FOR</td>
<td>AAGTGGGATGCAAAAGAAATC</td>
</tr>
<tr>
<td>qPCR_SlTPS46-REV</td>
<td>CATTTTCGGCACAATCACAC</td>
</tr>
<tr>
<td>qPCR_NaEF1A-FOR</td>
<td>TGTTGGCCATGATTTCTTA</td>
</tr>
<tr>
<td>qPCR_NaEF1A-REV</td>
<td>CCCAAATCCTTGGCTATTTT</td>
</tr>
<tr>
<td>qPCR_NaGLS-FOR</td>
<td>CCTCTTGGAGGCTCTTGACC</td>
</tr>
<tr>
<td>qPCR_NaGLS-REV</td>
<td>GACACAGTTCCTCCACAGC</td>
</tr>
</tbody>
</table>