Bioinformatics-Aided Identification, Characterization of Fungal Linalool Synthases and Applications in Linalool biosynthesis

Congqiang Zhang1,*, Xixian Chen1, Raphael Tze Chuen Lee2, Rehka T1, Sebastian Maurer-Stroh2,3#, Martin Rühl4#

1. Singapore Institute of Food and Biotechnology Innovation (SIFBI), Agency for Science, Technology and Research (A*STAR), Proteos level 4, 138673 Singapore
2. Bioinformatics Institute (BII), Agency for Science Technology and Research (A*STAR), 30 Biopolis Street, #07-01 Matrix, 138671, Singapore
3. Department of Biological Sciences (DBS), National University of Singapore (NUS), Singapore
4. Institute of Food Chemistry and Food Biotechnology, Justus Liebig University Giessen, Giessen, 35392 Giessen, Germany
*equal contribution

*To whom correspondence may be addressed. Email: zcqsimon@outlook.com or congqiang_zhang@sifbi.a-star.edu.sg

ABSTRACT: Enzymes empower chemical industries and are the keystone for metabolic engineering. For example, linalool synthases (LSs) are indispensable for the biosynthesis of linalool, an important fragrance used in 60-80% cosmetic and personal care products. However, plant LSs have low activities while expressed in microbes. Aided by bioinformatics analysis, four linalool/nerolidol synthases (LNSs) from various Agaricomycetes were accurately predicted and validated experimentally. Furthermore, we discovered a novel LS with exceptionally high levels of selectivity and activity from Agrocybe pediades, ideal for linalool bioproduction. It effectively converted glucose into enantiopure (R)-linalool in Escherichia coli, 44-fold and 287-fold more efficient than bacterial and plant counterparts, respectively. Phylogenetic analysis indicated the divergent evolution paths for plant, bacterial and fungal LSs. More critically, structural comparison provided catalytic insights in Ap.LS superior specificity and activity, and mutational experiments validated the key residues responsible for the specificity.

KEYWORDS: biocatalysts, biosynthesis, metabolic engineering, enzyme discovery, terpenoids, terpene synthases, monoterpenes, linalool synthase.
INTRODUCTION

Nature is the best inventor and breeds versatile enzymes. Among various enzymes, terpene synthases represent a unique class of biocatalysts with fascinating capabilities (e.g. introduction of carbon-carbon bonds, facilitation of cyclization and rearrangement of terpenes) 1. Terpene synthases are pivotal for the biosynthesis and diversity of terpenoids (>80000 different molecules), which constitute the largest group of natural products 1. Terpenoids have wide applications, including pharmaceuticals, nutraceuticals, flavorings, fragrances, and biofuels 2,3. However, the biosynthesis of most terpenoids has yet to achieve high titers and yields that are vital for commercial production. One major obstacle is that currently identified terpene synthases have low activities and/or low selectivities 2,4.

An example is linalool, a naturally occurring monoterpene alcohol (C10) found in several flowers such as lavender 5. With a pleasant floral smell, linalool is an important fragrance ingredient widely used in food, beverage and many personal care products (perfumes, body lotions, etc.). Natural linalool has two stereoisomers with different smells, (S)-linalool and (R)-linalool. (S)-linalool is floral, citrus and petitgrain-like (odor threshold 7.4 ppb) and (R)-linalool is woody and lavender-like (odor threshold 0.8 ppb) 6. Natural linalool has higher enantiopurity, thus, superior to synthetic linalool racemates in applications such as high-end perfumes and cosmetics. In 2018, the world consumption of linalool surpassed 11,000 metric tons and its global market is projected to reach 12.3 billion US$ in 2024 7. Despite great commercial interests, the biosynthesis of linalool has only achieved limited success (mg/L scale) 8,9. This contrasts with the rapidly growing demand for natural linalool. The plant linalool synthases (LSs, converting geranyl pyrophosphate (GPP) into linalool) are relatively abundant, yet proven to have low activities when expressed in microbial hosts (e.g. yeasts and Escherichia coli) 10. Recently, a bacterial bifunctional linalool/nerolidol synthase (LNS) has been identified and characterized 11. However, it produces more nerolidol (a sesquiterpene alcohol, the product of farnesyl pyrophosphate or FPP) than linalool when expressed in microbes. This is because, unlike plants that have special compartments (e.g. plastids) where GPP synthases (GPPSs) are localized 12, wildtype microbes have neither specialized organelles nor dedicated GPPSs. Rather, GPP is merely an intermediate compound of FPP synthases (e.g. ispA of E. coli, ERG20 of Saccharomyces cerevisiae) in microbes. Hence, FPP is more abundant than GPP in the cytosol 13. As such, a specific and more active LS is desired for microbial linalool production. Besides plants and bacteria, fungi such as the agaric mushroom Agrocybe aegerita (recently renamed into Cyclocebe aegerita), are also known to produce linalool and several other monoterprenes (α-pinene, p-cymene, limonene, and β-ocimene) 14. However, fungal LSs have not been identified until very recently. In our previous study 9, we described the bifunctional role of an LNS from A. aegerita that was used to produce linalool or nerolidol in E. coli but did not elaborate on the method of identification or its kinetic characterization.

Here guided by bioinformatics predictions and experimental validation, we identified four bifunctional LNSs and a specific LS (no sesquiterpene activity) from four fungal species. We characterized the kinetic parameters of the LNS from A. aegerita and...
the LS from *Agrocybe pediades*. Phylogenetic analysis indicated clear divergence among the 35 plant enzymes (4 nerolidol synthases, 9 LNs and 22 LNSs), 1 bacterial LNS (DSSL78) and 9 fungal enzymes (LNSs and LSs). Furthermore, *in vitro* and *in vivo* activities, structural and functional comparison were carried out for the plant, bacterial and fungal LSs. It was found that the fungal linalool synthase is highly active in *E. coli* and has an exceptional selectivity, thus ideal for linalool biosynthesis in metabolic engineering and other biocatalytic processes. On top of structural analysis, further mutation experiments identified the key residues responsible for the specificity of the fungal LS, as compared to fungal LNSs. The knowledge is essential for the design of artificial enzymes and interconversion between monoterpene synthases and sesquiterpene synthases.

**RESULTS AND DISCUSSION**

Plant metabolites are well recognized and explored. In contrast, the biosynthetic capability of fungi is relatively undervalued. Fungi (>3 million species) have enormous diversity and are rich in biocatalysts and secondary metabolites. In particular, fungi produce a large portfolio of volatile terpenoids, including sesquiterpenes (e.g. β-caryophyllene and cubenol) and monoterpenes (linalool, limonene and 1,8-cineole). To our knowledge, to date, there are only a couple of fungal monoterpene synthases reported in literature (Hyp3, a 1,8-cineole synthase and the very recently identified PpSTS25, a bifunctional myrcene/linalool synthase). Here, we addressed this gap and specifically explored fungal linalool synthases.

**Identification of linalool synthase in agaric mushroom.**

As *Agrocybe aegerita* is known to produce linalool and its genome has been recently sequenced, we tried to identify the LS in its genome. In our previous study, we identified 11 putative sesquiterpene synthases and nine of them are functional and produced various sesquiterpenes but not linalool. A re-evaluation of the raw genomic data led to an additional putative terpene synthase sequence (AAE3_109435). The gene exists in the Illumina sequencing data, whereas absent in the PacBio results. PCR amplification of the AAE3_109435 with subsequent sequencing confirmed the presence of the gene in the genome (Fig. S1).

The subsequent expression of AAE3_109435 in a GPP-accumulating *E. coli* that co-expressed the native enzymes DXS, IDI and ispA_S80F mutant (GPPS). The resulting strain (GPPS+9435) produced the acyclic monoterpene linalool as the main product and small amount of nerolidol (Fig. 1; mass spectra in Fig. S2). Geraniol was detected in GPPS+9435 as well as in its control strain (GPPS_ctrl), indicating that geraniol was not the product of AAE3_109435 but of a native *E. coli* enzyme (such as PhoA, a phosphatase, and NudB, a Nudix hydrolase). Furthermore, AAE3_109435 was expressed in the *E. coli* strain that accumulates FPP by overexpressing DXS and IDI (without ispA_S80F), and the strain was named ‘FPPS+9435’. Its main product (>96% regarding the peak area of all detected terpenes) was nerolidol and only traces of linalool could be detected (Fig. 1A). The control strain (FPPS_ctrl) with an empty vector produced neither linalool nor nerolidol. The results clearly proved that
the terpene synthase coded by AAE3_109435 is a bi-functional linalool/nerolidol synthase (LNS), which is able to convert FPP into nerolidol and GPP into linalool. Accordingly, it is named Aa.LNS. Furthermore, the linalool produced by Aa.LNS is mainly (R)-linalool (95% ee, Fig. 1B), the same stereoisomer as in lavender and sweet basil.

Bioinformatics prediction of other fungal LNSs.

Aa.LNS was used to probe other potential fungal LNSs. The first focus was laid on Agrocybe pediaes, another sequenced fungal species of the genus Agrocybe. Linalool was detected in the headspace of A. pediaes cultures in our lab. A blast search of Aa.LNS against the A. pediaes genome using the online tool of the Joint Genome Institute (https://genome.jgi.doe.gov/Agrped1/Agrped1.home.html) resulted in 11 terpene synthase homologues (Fig. 1C).

To improve the prediction confidence, we combined two strategies: 1) full sequence alignment (Fig. 1C) and 2) comparison of predicted active sites (Tables S1 and S2, the results were analyzed with 4LXW and 5NX5 (PDB ID) as templates). Four terpene synthase homologues (Agrped1_820868, Agrped1_694262, Agrped1_689671 and Agrped1_689675,) were found to be closely related to Aa.LNS (Fig. 1C). It was hypothesized that the two enzymes Agrped1_820868 and Agrped1_694262 are more similar to Agr8 (a γ-muurolene/β-cadinene synthase) as all three enzymes share almost identical active sites based on our algorithm (Table S1). Hence, we studied the other two enzymes. As predicted, the strain expressing Agrped1_689671 produced linalool and nerolidol (renamed Ap.LNS) similar to Aa.LNS (Fig. 1D). Surprisingly, linalool was the sole terpene product of the E. coli clone expressing Agrped1_689675, indicating it is a monofunctional LS (renamed Ap.LS). This attributes to the high specificity of Ap.LS, which is underpinned by the fact that FPP is more abundant in microbial cells than GPP. In addition, Ap.LS produced the enantiopure (R)-linalool (Fig. 1B).

Next, we asked if we could use Ap.LS to probe other fungal LSs. We carried out a UniProt BLAST search of Ap.LS and collected those hits with the highest alignment score (score >700, Fig. S3): three from Galerina marginata (Galma_223690, UniProt ID A0A067THX9; Galma_63556, A0A067T818; Galma_266794, A0A067T571); two from Hypholoma sublatentium (Hypsu1_148365, A0A0D2NH86; Hypsu1_148385, A0A0D2NA50) and one from Hebeloma cylindrosporum (M413_27416, A0A0C2YLE7).

Four of them (Galma_223690, Galma_63556, Hypsu1_148365 and Hypsu1_148385) clustered into a branch or subgroup with Ap.LS and Ap.LNS in both analyses using full sequence and active-site alignment (Fig. 2A and B). Overall, the six homologues share >90% identity of active site residues (Tables S1 and S2). Particularly, the active sites of Hypsu1_148385 and Galma_223690 show 95% (36/38) and 97-100% (37-38/38) identity with that of Ap.LS and Ap.LNS, respectively, indicating they are potential fungal LNSs. The other two amino acid sequences (Galma_266794 and M413_27416) were more closely related to Agr8, with as high as 95% active-site identity (Table S1). Thus, it was hypothesized that they were more likely to produce 1,10 cyclization products of FPP, e.g., muurolene, cadinene. This hypothesis was validated by the expression of Galma_223690,
Hypsu1_148385 or Galma_266794 in the FPP-accumulating *E. coli*. As predicted, Galma_223690 and Hypsu1_148385 were found to be bifunctional LNSs (Fig. 2C), thus renamed Gm.LNS and Hs.LNS, respectively. Moreover, like Agr8, the strain expressing Galma_266794 produced germacrene D (1,10 cyclization) as main product and a few minor products (γ-muurolene and (+)-δ-cadinene, Fig. S4), thus validating our hypothesis.

Here, the synergistic use of BLAST search, full-sequence alignment and active-site alignment was explored. As such, we achieved a relatively high predictability of hunting for biocatalysts of the same function (e.g. linalool synthases). BLAST search helps with identifying overall similar enzymes that form our initial screening candidates. Full sequence alignment and phylogenetic tree facilitate the classification of enzymes. Different classes often have distinct catalytic functions (e.g. different cyclization positions). Active-site prediction further supplements the prediction with two main roles. One is to filter out those enzyme candidates of incomplete binding pockets (e.g., Agr11 is missing the NSE triad, Tables S1 and S2). The other is to complement the enzyme classification of full sequence alignment. This is based on the hypothesis that enzymes of the same function may have overall low similarity but more conserved active sites. For example, the Galma_266794 shares comparable full-sequence similarities with Ap.LS (61%) and Agr8 (62%), however it has much higher active-site identity with Agr8 (95%) than with Ap.LS (71%, Table S1). Indeed, the products of Galma_266794 are similar to that of Agr8.

**Purification and characterization of fungal LS and LNSs.**

To date, a number of plant LSs and LNSs have been identified. However, only one bacterial LNS from *Streptomyces clavuligerus* was recently identified \(^\text{11}\), which only shares 15.2% identity with Aa.LS (Fig. S5). With the fungal enzymes studied in this work, LSs and LNSs in three kingdoms have been identified. Next, we sought to compare their catalytic activities and mechanisms by *in vitro*, *in vivo* assays, sequence alignments and 3D structural models.

Protein purification is the prerequisite to study the *in vitro* kinetics of the fungal enzymes. Though all the five bacterial strains expressing fungal enzymes produced linalool, their expression levels in *E. coli* were largely different. Aa.LNS had the highest expression level, followed by Gm.LNS, Ap.LNS and Ap.LS. The expression of Hs.LNS (Hypsu_148385) was so low that it was not detectable in a protein gel (Fig. S6A). As Aa.LNS had the highest expression level and Ap.LS is the only specific monoterpene synthase, they were chosen as the representatives of fungal LNS and LS for further studies. However, none of them was soluble based on solubility analysis with B-PER II reagent (Thermo Scientific™) (Fig. S6A). Many approaches were tested but failed to improve their solubility (such as abiotic condition optimization: lowering incubation temperature, tuning inducer dosages, media additives and protein fusion). Refolding of insoluble fraction could be another solution which we did not test because it is very time-consuming to optimize the best conditions. The N-terminal fusion of Aa.LNS with a maltose binding protein or thioredoxin did not help (Fig. S6B). Different chaperone systems (DnaK-dnaJ, GroES-GroEL) and trigger factor (TF) in *E. coli* were further tested. It was found that TF chaperone could slightly improve the solubility of the synthases. With the optimal condition
(3.3 mM arabinose to induce TF chaperone and 0.1 mM IPTG to induce Aa.LNS, Fig. S7) and further separation by size exclusion chromatography, we managed to purify enough soluble Aa.LNS for in vitro characterization. Yet its purity was quite low with ~16.3% (Fig. 3A). In contrast, relatively high purity of soluble Ap.LS (~71.2%) was obtained with the same experimental conditions (Fig. 3B). Consistent with the E. coli cultures producing the respective synthase, purified enzymes reconfirmed that Aa.LNS can use FPP and GPP to produce nerolidol and linalool, respectively. Whereas Ap.LS was only able to use GPP but not FPP (Fig. 3). The $K_m$ and $k_{cat}$ values of Aa.LNS for FPP were 9.0±2.3 μM and 3.3±0.3 min$^{-1}$, respectively; and slightly lower for GPP with 6.7±4.6 μM and 0.5±0.1 min$^{-1}$, respectively (Table 1). The $K_m$ value of Ap.LS for GPP with 3.8±0.7 μM was slightly lower than that for Aa.LNS, whereas $k_{cat}$ was much higher with 6±0.3 min$^{-1}$. To compare the catalytic efficiencies among the known LSs and LNSs, $k_{cat} / K_m$ value of Ap.LS was the highest, which is about 21-fold, 29-fold, threefold and fourfold higher than that of Aa.LNS, La.LS (Q2XSC5) from Lavandula angustifolia (Lavender) 5, Ma.LS (Q8H2B4) from Mentha aquatica 23, and of the bacterial Sc.LNS 11, respectively. As for Aa.LNS as a nerolidol synthase, although the $k_{cat}$ value of Aa.LNS for FPP was more than five times lower compared to the bacterial one, the $K_m$ value was similar to that of the bacterial Sc.LNS 11 and about half as that of Zm.LNS from Zea mays (Maize) (Table 1) 24.

**In vivo activity comparison of LSs and LNSs from three kingdoms.**

Due to potential issues such as poor expression and solubility when expressed in cells and the localization difference (cytosolic/membrane-bound in vivo versus a one-pot aqueous reaction), the advantages of in vitro enzyme kinetics (Table 1) may not be readily transferable into cellular applications, such as metabolic engineering, where in vivo activities are more critical than in vitro ones. To test the best candidate for microbial linalool production, our previously engineered E. coli strain was used to compare LNs from three kingdoms: Ap.LS, Sc.LNS and Cb.LS from Clarkia breweri (Q96376) as representatives for fungi, bacteria and plantae, respectively. They were separately cloned into pET-11a vector (Novagen). Together with a p15A vector carrying the whole mevalonate pathway genes 2, the bacterial strains grown in ZYM media produced linalool at 381.2, 8.7 and 1.3 mg/L for fungal, bacterial and plant linalool synthases, respectively (Fig. 3C). This means in vivo activity of Ap.LS (fungal) is about 44-fold and 287-fold as high as that of Sc.LNS (bacterial) and Cb.LS (plant), respectively. A previous study also supported that the bacterial Sc.LNS is better than plant LSs in terms of in vivo activity in TB media 10. In the same TB media, the linalool titre reached 601.2 mg/L for Ap.LS stain, about 65% higher than previously reported using Sc.LNS 10. Our study demonstrated that fungal Ap.LS is even superior to the bacterial one, in both activity and selectivity. Although Sc.LNS has a higher activity than plant Cb.LS, it prefers FPP (lower $K_m$ and higher $k_{cat}$) to GPP as the substrate 11. Therefore, Sc.LNS produced a larger amount of nerolidol than linalool in E. coli whose cytosol contained both FPP and GPP, in contrast, Ap.LS produced 100% linalool. The superior activity and selectivity of Ap.LS makes it more suitable for microbial production of linalool than its plant and bacterial counterparts. High activity contributes to high titers, rates and yields (TRYs) of linalool production and low mann-
factoring cost. High specificity would greatly simplify the downstream purification process and further reduces the overall produc-
tion cost. Thus, this study paves the way for commercialization of linalool bioproduction which is greener, safer, sustainable
and of exceptional enantiopurity ((R)-linalool), as compared to chemical synthesis.

Structural comparison of LSs and LNSs from three kingdoms (plants, fungi and bacteria).

Next, we generated a phylogenetic tree with 35 plant enzymes (including 4 nerolidol synthases, 9 LNs and 22 LNSs), 1
bacterial LNS (D5SL78) and 9 fungal enzymes (Table S3). The enzymes were clearly separated into two major clades (one is
plant, clade 1, and the other is microbial, clade 2, Fig. 4). The bacterial LNS was closer to fungal ones, in clade 2. Specifically,
the sequence identity among fungal, bacterial and plant LNSs or LSs are only 8%-15%, which includes those metal binding sites
(Fig. S5). Overall, plant LSs/LNSs are larger with 500-900 amino acids than microbial ones with 300-400 amino acids. One
particular enzyme D8RNZ9, which is a LNS isolated from the nonseed plant Selaginella moellendorfﬁi (Spikemoss) 25, is more
closely related to the bacterial LNS than plant LNSs. It was hypothesized that it could stem from horizontal gene transfer from
microbes to plants or that seed plants lost these LNS enzymes during evolution from nonseed plants 25. The phylogenetic tree
indicates the evolutionary divergence of fungal, bacterial and plant LSs and LNSs.

Subsequently, we compared the protein structures of the LSs from the fungus (A. pediades), the bacterium (S. clavuligerus)
and the plant (M. aquatica). The crystal structure of Sc.LNS is 5NX5 (PDB ID) 10. The homologue models of Ap.LS and Ma.LS
(Q8H2B4) were built based on the crystal structure of (+)-bornyl diphosphate synthase from Salvia ofﬁcinalis (1N1B/1N21) 26
and 1,8-cineole synthase from S. clavuligerus (5NX7) 10, respectively. As the sequence similarity between Ap.LS and 1,8-cineole
is only 20.3% (and active site identity is 49%), the Ap.LS model may have some deviations from its real structure. Nevertheless,
their active site regions are highly conserved (Fig. 5A). Fungal LS and bacterial LS are much more similar to each other than to
plant LS in both active site regions (Table S4) and overall structures (Fig. 5). As a typical plant monoterpene synthase, Ma.LS
has two domains (α and β domains) and, thus, is noticeably larger than the other two synthases with active site residing only in
the α domain (catalytic domain, Fig. S8). In contrast, microbial LSs, Sc.LNS and Ap.LS, are similar to typical class I terpene
cyclases with a single domain, despite with acyclic products (Fig. S9). Both GPP (its analogue, 2-fluorogeranyl diphosphate)
and FPP were docked into the three models. We mainly analyzed the interactions of the three LSs with GPP. With 9 hydrophobic
interactions with GPP, Ap.LS had the highest amount, as compared to 7 of Sc.LNS and 6 of Ma.LS (Fig. 5). Except for the
negatively charged pyrophosphate (Ppi) head, GPP is largely hydrophobic, thus, these hydrophobic interactions may contribute
to the high activity of Ap.LS. The number of hydrogen bonds identified for the three enzymes were similar. In addition, Ap.LS
had the highest binding affinity (-7.6 kcal/mol) to GPP, followed by Sc.LNS (-7 kcal/mol) and Ma.LS (-6.4 kcal/mol). The binding
affinity inversely correlated with the Km values of the three enzymes (Table 1), where higher binding affinity contributed to a
lower Km value. As compared for Ap.LS and Ap.LNS, Ap.LS has higher binding affinity to GPP than Ap.LNS (-7.3 kcal/mol), but
lower binding affinity (-8.5 kcal/mol) to FPP than Ap.LNS (-9.0 kcal/mol). The binding affinity data are nicely correlated with their difference in monoterpene and sesquiterpene activities.

Furthermore, we superimposed the 3D structures of the active sites of the three enzymes (Fig. 6A and Table S4). Residues in the binding pocket of plant LS showed the greatest divergence (green regions, Fig. 6A), although structure folding remains conserved. Overall, as shown in the grey regions (Fig. 6A) and highlighted in Table S4, there were 8 conserved residues among the three enzymes including the aspartate rich motif, D(D/E)XXD, responsible for Mg$^{2+}$ cofactor and substrate binding \(^1\) and NSE triad, (N/D)Dxx(S/T)xxxE, responsible for the substrate binding and coordination of the diphosphate and trinuclear Mg$^{2+}$ [PPi-(Mg$^{2+}$)] cluster \(^1\). All these structural analyses partially explain the activity difference among the three LSs from different kingdoms. Nevertheless, there are other factors which might also contribute to the high in vivo activity of Ap.LS in *E. coli*, such as, the non-active site residues, protein expression and solubility (Fig. S6) etc.

**Mechanism study on the Ap.LS selectivity**

Lastly, we attempted to understand the specificity of Ap.LS as compared to other fungal LNSs. Particularly, Ap.LNS and Ap.LS share the highest identity 77.9%, hence, we compared their difference of residues surrounding the substrate binding pocket. In total, five residues were found to be different (Fig. 6B and Table S5): 1) in the Ppi head region (A59:S58, E316:Q315, I153:V152, here, the former and latter residues refer to that of Ap.LS and Ap.LNS, respectively, especially that E316 is expected to have charged interactions with the Ppi group of GPP/FPP); 2) in the FPP tail region (L60:M59, which could affect the interaction with FPP); 3) in GPP tail region (G181:A180, which might affect flexibility of helix where the tail end of GPP resides). A series of single mutants were constructed for Ap.LS, A59S, L60M, G181A and E316Q. However, none of these single mutants had effect on the selectivity of Ap.LS (Fig. 6E). We speculated that the specificity might be the synergistic result of multiple residues. The region A59-L60 was particularly interesting, as A59 and L60 are in close proximity to both the Ppi head and hydrocarbon tail of FPP (Fig. 6D). Indeed, the combination of A59S and L60M mutations resulted in the production of a trace amount of nerolidol, ~2% of total amount of linalool and nerolidol produced (Fig. 6C and E), which indicated the two mutations are sufficient to convert Ap.LS from a monofunctional LS to a bi-functional LNS. Adjacent to A59-L60, another residue is also different between Ap.LS (V61) and Ap.LNS (I60). Although single mutation V61I had no effect on the selectivity of the wildtype Ap.LS, the introduction of V61I enhanced the nerolidol production by 12-fold (~40% of total linalool and nerolidol produced) and decreased linalool production by 45% on the basis of the double mutant A59S-L60M (Fig. 6C and E). It seems that the mutation L60M-A59S favors sesquiterpene activity (nerolidol formation) by stabilizing the ligand in a favorable position (A59S) and by promoting the easier leave of the Ppi group from the binding pocket (L60M, Fig. 6D). The third mutation V61I further enhances the effect by pushing M60 and S59 closer to FPP (Fig. 6D).
In addition, we also combined the mutations of two regions and obtained the quadruple mutant A59S-L60M-V61I-E316Q, however, it did not further increase nerolidol production. Furthermore, we observed that the wildtype Gm.LNS, whose corresponding residues are more similar to those 59-61 of Ap.LS (S59:A59, L60:L60, V61:V61, Fig. S9), produced the highest amount of linalool (15%) among all the wildtype fungal LNSs (Fig. 2C). As such, we concluded that the region residues (A59-V61) play an essential role for the highly specificity of Ap.LS.

MATERIALS AND METHODS

Sequence validation of AAE3_109435 in the genome.

AAE3_109435 sequence was analyzed from Illumina sequencing data with manual intron annotation. To verify the presence of the gene AAE3_109435, gDNA of A. aegerita was firstly extracted. Briefly, the vegetative mycelia of A. aegerita was ground with pestle and mortar using liquid nitrogen. Around 200 mg of ground mycelium was transferred in 1.5 mL E-cups and 500 µL of lysis buffer (400 mM Tris-HCl pH 8.0, 60 mM EDTA pH 8.0, 150 mM NaCl, 1% SDS) was added. The solution was vortexed and incubated at ambient temperature for 10 min before the addition of 150 µL potassium acetate (3 M, pH 4.8). The solution was then vortexed and centrifuged for 10 min at 3000 g. Supernatant containing gDNA was transferred into a new E-cup and the gDNA was obtained by ethanol precipitation method. The gDNA was used as the PCR template to amplify AAE3_109435 with primers AAE3_9435f (CCAAGATTGTCGTCAACGCC) and AAE3_9435r (CTGTGGGCCATTCTGTCCAT). The amplified region was subsequently sequenced and confirmed (Results in Fig. S1).

Bioinformatics prediction of fungal linalool synthases.

The fungal LS candidate genes were obtained by the combination of BLAST search in JGI fungal genomics and UniProt databases, full sequence alignment and predicted active site alignment. Those homologues with the highest similarity (usually >50% identity, or in the same cluster of the phylogenetic tree or sequence similarity network) to Aa.LNS (accession number MN954676) or Ap.LS in full sequences and in active sites were selected as the targets for experimental validation.

Prediction of active sites of fungal terpene synthases.

The full sequence alignment was generated by aligning the complete sequences of LS and LNS proteins from across the three kingdoms by Clustal Omega v1.2.2. In contrast, active-site alignment is proceeded by identifying amino acid residues surrounding the predicted active sites of each enzyme in 3D structures and used to extract the relevant residues from the full alignment. We used an in-house developed algorithm, BioTransformer v0.9, to predict and align the active sites. In brief, the algorithm first searches the PDB for appropriate templates. Next, the user gets to select the most appropriate PDB template or
templates with the most appropriate ligands (usually the ligands or combination of ligands that maximize the space within the active site). Using this approach, the PDB structures 4LXW (Epi-isozae synathase in complex with inorganic pyrophosphate (Ppi) and benzyl triethyl ammonium from *Streptomyces coelicolor* and 5NX5 (Linalool/Nerolidol synthase in complex with 2-fluorogeranyl diphosphate from *Streptomyces clavuligerus*) were chosen as templates for the prediction of the active sites. To maximize the number of residues found within the binding pockets of the PDB structures, residues found within 6.0 Angstrom from the substrate are considered as part of the active site, and the union set derived from both structural templates is used as the predicted active site.

**Phylogenetic analysis and sequence similarity network of terpene synthases.**

Full amino acid sequences were used to build a phylogenetic tree. Alignment was carried out by Clustal Omega program version 1.2.2 and the phylogenetic analysis was conducted with the Neighbor-Joining method or Maximum Likelihood method by MEGA version 7.0.26. For the predicted active sites of terpene synthases, the Enzyme Function Initiative-Enzyme Similarity Tool (EFI-EST, http://efi.igb.illinois.edu/efi-est/) was used to generate sequence similarity networks (SSNs) with the database-independent alignment score of 16. The resulting SSNs were visualized using the open source software Cytoscape version 3.5.1.

**Structural modelling and analysis.**

The homologue models of Ap.LS and Ma.LS (Q8H2B4) from *Mentha aquatica* were built based on the crystal structure of the most closely related terpene synthases: Epi-isozae synathase from *S. coelicolor* (PDB ID, 4LXW), 1,8-cineole synthase from *S. clavuligerus* (PDB ID, 5NX7) and (+)-bornyl diphosphate synthase from *Salvia officinalis* (PDB ID, 1N1B/1N21), using the Modeller software. The binding pockets, consisting of 21-22 residues within 6 angstroms from the substrate, were determined by PyMOL software v2.1.1. Structural alignment of proteins was conducted using MUSTANG implemented in YASARA. The web implementation of PLIP (Protein-Ligand Interaction Profiler) was used to identify interactions between ligand and the surrounding amino acid residues in the protein. Docking was performed with AutoDock Vina with an exhaustiveness of 200. AutoDock Tools was used to prepare the PDBQT files for the homology models for the enzymes Ap.LS, Ap.LNS, Sc.LNS and Ma.LS and also for the ligands GPP and FPP. A search space of 24x24x24 Å was used for GPP and a search space of 24x28x24 Å was used for FPP.

**Plasmid and strain construction.**

The fungal terpene synthases (TPSs), Sc.LNS from *Streptomyces clavuligerus* and Cb.LS from *Clarkia breweri* were codon optimized and cloned into pET-11a vector (Novagen) either alone or together with ispA_S80F (GPP synthase) or ispA (FPP synthase) from *E. coli*. For fungal TPS characterization, the chassis *E. coli* strain was used. The strain carried a plasmid p15A-
cam-T7-dxs-idi that overexpresses the enzymes DXS and IDI from *E. coli* to enhance the supply of terpene precursors \(^3\). For the monoterpene production study, the mevalonate pathway was overexpressed in p15A vectors under the T7 promoter variants \(^9\).

**Mutation study of Ap.LS.**

The targeted mutations were introduced to Ap.LS using in-house methods as described previously \(^2\). Single mutants (A59S, L60M, V61I, G181A and E316Q) and the double mutant (A59S-L60M) were directly introduced to the wildtype. Triple mutants (A59S-L60M-V61I, A59S-L60M-E316Q) were introduced on top of the double mutant A59S-L60M. E316Q was introduced to the triple mutant A59S-L60M-V61I to obtain the quadruple mutant A59S-L60M-V61I-E316Q. Primers used are as listed in Table S6.

**Terpenoid production in *E. coli*.**

*E. coli Bl21-Gold DE3* strain (Stratagene) was used for linalool production. The strains carrying mevalonate pathway and LSs from different species were grown in 1mL of ZYM medium or terrific broth (TB) supplemented with 0.4% glucose in 14 ml BD Falcon\textsuperscript{TM} tube at 28 °C/250 rpm for 3 days \(^2\). In addition, 200 μL isopropyl myristate was used for harvesting the linalool and/or nerolidol and antibiotics (34μg/ml chloramphenicol, 50 μg/ml spectinomycin and 100 μg/ml ampicillin) were used to maintain the plasmids. ZYM medium was prepared as previously described \(^2\) (1% tryptone, 0.5% yeast extract, 25 mM Na2HPO\(_4\), 50 mM KH2PO\(_4\), 50 mM NH4Cl, 5 mM Na2SO\(_4\), 2 mM MgSO\(_4\), 0.5% glycerol, 0.05% glucose, 15mM of α-lactose). In ZYM medium, the *E. coli* strains were automatically induced with the depletion of glucose and initiation of lactose consumption.

**Expression and purification of linalool synthases.**

*E. coli Bl21-Gold DE3* strains (Stratagene) carrying pET-11a-Aa.LNS or pET-11a-Ap.LS and the chaperone plasmid pTf16 (Takara Bio Inc., Japan), were grown in 4L of 2xZYM media with antibiotics (100 μg/ml ampicillin and 34μg/ml chloramphenicol) at 20 °C and 225 rpm. The cells were induced with 0.1 mM IPTG (to express LS) and 3 mM of arabinose (to express chaperone proteins) when OD600 reached 0.6-0.8. The cells were further cultured for another 20 hours and harvested. Cell were washed and re-suspended in 100 ml of the solution (50mM Tris of pH 8, 500mM NaCl, 0.1% TritonX100, 5% glycerol) with 2 tablet of protease cocktail inhibitor (Sigma-Aldrich, Singapore). The cells in the suspension were lysed by sonication at 4 °C (5s ON and 5s OFF, 40% amplitude). The proteins in the supernatant were then extracted by 1ml x HisTrap column. The column with bound proteins was washed by his-binding buffer (50mM Tris of pH 8, 500 mM NaCl and 20mM imidazole) and eluted by 20 ml of his-elution buffer (50mM Tris of pH 8, 500 mM NaCl and 500mM imidazole). The eluted protein was further concentrated by Vivaspin\textsuperscript{®} 10kDa cutoff spin column (Merck, Singapore). In the end, 2.7 mg/ml of Aa.LNS (16.3% purity) and 22.2 mg/ml of Ap.LS (71.2% purity) was obtained.
Kinetic analysis.

Steady-state kinetics of purified Aa.LNS and Ap.LS were determined by measuring PPI release via conversion to phosphate with inorganic pyrophosphatase using the EnzChek® Pyrophosphate Assay Kit (Thermo Fisher Scientific, Singapore). Substrate concentrations of GPP and (E,E)-FPP (Echelon bioscience, USA) were varied between 6.5-40 μM. Reactions were carried with 20 μg/ml (460 nM) of Ap.LNS or 10 μg/ml (250 nM) of Aa.LS at 37 °C for 1 hour. PPI concentrations were calculated by linear interpolation of the standard curve using the kit assay (0-60 μM). The scatter plots of initial rate versus substrate concentration were fitted to the equation $v = \frac{v_{\text{max}}[S]}{K_M + [S]}$ where $v_{\text{max}} = k_{\text{cat}}[E_0]$.

GC-MS detection and quantification of terpenes.

For characterization of fungal terpene synthases, the headspace compounds were sampled at 60 °C for 20 min by SPME with a DVB/CAR/PDMS (50/30 μm divinylbenzene/carboxen/polydimethylsiloxane) fiber (length 1 cm; Supelco, Steinheim, Germany). Subsequently, the compounds were desorbed for 1 min in the split inlet (250 °C; SPME liner, 0.75mm i.d.; Supelco) and analyzed by an Agilent 7980B gas chromatography equipped with an Agilent 5977B MSD. Samples were injected into Agilent DB5ms column with a split ratio of 40:1 at 240 °C. The oven program started at 80 °C for 1 min, was raised up to 210 °C at 10 °C/min, then to 310 °C at 60 °C/min and maintained at 310 °C for another 2 min. Mass spectrometer was operated in EI mode with full scan analysis (m/z 33–300, 9 scans/s). In addition to mass spectra, Kovats retention indices of detected compounds were calculated by calibrating the used GC-MS system with a C8–C30 alkane mix and compared with literature data in the National Institute of Standards and Technology database (Fig. S2).

For quantification of linalool and nerolidol of different strains, the organic layer with secreted terpenes was separated and diluted into ethyl acetate by 10-100 times. The samples were then analyzed with the same GC-MS program as the characterization method. The concentrations were calculated by interpolation using the standard curve of authentic linalool and nerolidol standards (Sigma-Aldrich, Singapore).

Chiral study of linalool produced by fungal LSs. The chirality of linalool produced by fungal LSs was analyzed by the GC chiral CycloSil-B column (30m, 0.25mm, 0.25u, Agilent, Singapore) in the same Agilent 7980B gas chromatography equipped with the 5977B MSD. The oven program started at 80 °C for 2 min, was raised up to 210 °C at 5 °C/min, then to 250 °C at 20 °C/min and maintained at 250 °C for another 2 min. Mass spectrometer was operated in EI mode with full scan analysis (m/z 33–300, 5.5 scans/s). The retention times and mass spectra of samples were compared with authentic standards of both (R)-linalool and a mixture of (R/S)-linalool.

SUPPLEMENTARY MATERIALS

Table S1. Predicted active sites of terpene synthases with 4LXW and 5NX5 as the templates.
Table S2. Comparison and combination of the predicted active sites of the two templates 4LXW and 5NX5.

Table S3. Summary of LSs and LNSs in this study.

Table S4. 3D alignment of active-site residues of Ap.LS, Sc.LNS and Ma.LS.

Table S5. Comparison of residues in the substrate-binding pockets of Ap.LS and Ap.LNS.

Table S6. Primers used for Ap.LS mutation.

Fig. S1. Nucleic acid sequence of AAE3_109435 in the genome.

Fig. S2. Mass spectra and retention indices for terpenes detected in this study.

Fig. S3. The alignment between Agrped_689675 and Galma_223690 and BLAST search results in UniProt database with Agrped_689675 (or Ap.LS).

Fig. S4. GCMS chromatograms and spectra for Galma_266794.

Fig. S5. Amino acid sequence alignment and identity table of LSs and LNSs from fungi, bacteria and plants.

Fig. S6. (A) Expression and solubility analysis of the fungal LNSs and LS.

Fig. S7. The optimization of Aa.LNS solubility with the co-expression of chaperone proteins.

Fig. S8. The sequence alignment and their secondary structures of plant terpene synthases.

Fig. S9. The sequence alignment and their secondary structures of microbial terpene synthases.

Acknowledgment: We thank Nic Lindley and Giulia Rancati for the support for the project. We thank Francisco Javier Ruiz-Dueñas for permission to use unpublished *Agrocybe pediades* genome data sequenced by the U.S. Department of Energy Joint Genome Institute, a DOE Office of Science User Facility, is supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231. We also thank Chin-Chin Lim and Sudha Shukal for assistance with some experiments, Shuke Wu for insightful comments. Funding: This work was funded by IAF-PP (HBMS Domain): H17/01/a0/006, and 2019 AME Young Individual Research Grants: A2084c0064, Agency for Science, Technology and Research (A*STAR), Singapore. Author Contributions: C.Z. conceived the project, designed the experiments, performed bioinformatics analysis, and wrote the manuscript. C.Z., R.T. and X.C. carried out the experiments and analyzed data. M.R. contributed to the initiation of the project, did the sequence annotation and validation of AAE3_109435. T.C.L and S.M.S. performed the active site prediction of fungal terpene synthases. All authors contributed to the discussion and approved the final manuscript. Competing interests: The authors declare that they have no competing interests. Data and materials availability: All data needed to evaluate the conclusions in the paper are present in the paper and/or the Supplementary Materials. Additional data related to this paper may
be requested from the authors. **Abbreviations:** GPP: geranyl diphosphate, FPP: farnesyldiphosphate, LS: linalool synthase, LNS: linalool/nerolidol synthase, TPS: terpene synthases, GPPS: GPP synthase, FPPS: FPP synthase, SSN: sequence similarity network, TRY: titre, rate and yield.

**REFERENCES AND NOTES**

1. Christianson, D.W. Structural and Chemical Biology of Terpenoid Cyclases. *Chemical Reviews* **117**, 11570-11648 (2017).
2. Shukal, S., Chen, X. & Zhang, C. Systematic engineering for high-yield production of viridiflorol and amorphanediene in auxotrophic *Escherichia coli*. *Metabolic Engineering* **55**, 170-178 (2019).
3. Zhang, C. et al. *Agrocybe aegerita* Serves As a Gateway for Identifying Sesquiterpene Biosynthetic Enzymes in Higher Fungi. *ACS Chemical Biology* **15**, 1268-1277 (2020).
4. Leonard, E. et al. Combining metabolic and protein engineering of a terpenoid biosynthetic pathway for overproduction and selectivity control. *Proceedings of the National Academy of Sciences of the United States of America* **107**, 13654-13659 (2010).
5. Landmann, C. et al. Cloning and functional characterization of three terpene synthases from lavender (*Lavandula angustifolia*). *Arch Biochem Biophys* **465**, 417-29 (2007).
6. Padrayuttawat, A., Yoshizawa, T., Tamura, H. & Tokunaga, T. Optical isomers and odor thresholds of volatile constituents in *Citrus sudachi*. *Food Science and Technology International, Tokyo* **3**, 402-408 (1997).
7. 360 Market Updates. Global Linalool Market 2019 by Manufacturers, Regions, Type and Application, Forecast Analysis to 2024. (2019).
8. Cao, X., Wei, L.-J., Lin, J.-Y. & Hua, Q. Enhancing linalool production by engineering oleaginous yeast *Yarrowia lipolytica*. *Bioresource Technology* **245**, 1641-1644 (2017).
9. Zhang, C., Seow, V.Y., Chen, X. & Too, H.P. Multidimensional heuristic process for high-yield production of astaxanthin and fragrance molecules in *Escherichia coli*. *Nat Commun* **9**, 1858 (2018).
10. Karuppiah, V. et al. Structural Basis of Catalysis in the Bacterial Monoterpene Synthases Linalool Synthase and 1,8-Cineole Synthase. *ACS Catalysis* **7**, 6268-6282 (2017).
11. Nakano, C., Kim, H.-K. & Ohnishi, Y. Identification and Characterization of the Linalool/Nerolidol Synthase from *Streptomyces clavuligerus*. *ChemBioChem* **12**, 2403-2407 (2011).
12. Gutensohn, M. et al. Cytosolic monoterpene biosynthesis is supported by plastid-generated geranyl diphosphate substrate in transgenic tomato fruits. *Plant J* **75**, 351-63 (2013).
13. Tashiro, M. et al. Bacterial Production of Pinene by a Laboratory-Evolved Pinene-Synthase. *ACS synthetic biology*, acssynbio.6b00140 (2016).
14. Rapior, S. et al. Volatile components of fresh *Agrocybe aegerita* and *Tricholoma sulphureum*. *Cryptogamie Mycologie* **19**, 15-23 (1998).
15. Skellam, E. Strategies for Engineering Natural Product Biosynthesis in Fungi. *Trends Biotechnol* **37**, 416-427 (2019).
16. Ohm, R.A. et al. Genome sequence of the model mushroom *Schizopyllum commune*. *Nat Biotechnol* **28**, 957-63 (2010).
17. Zhang, C. & Too, H.P. Revalorizing Lignocellulose for the Production of Natural Pharmaceuticals and Other High Value Bioproducts. *Curr Med Chem* **26**, 2475-2484 (2019).
18. Kramer, R. & Abraham, W.-R. Volatile sesquiterpenes from fungi: what are they good for? *Phytochemistry Reviews* 11, 15-37 (2012).

19. Shaw, J.J. et al. Identification of a fungal 1,8-cineole synthase from *Hypoxylon sp.* with specificity determinants in common with the plant synthases. *J Biol Chem* 290, 8511-26 (2015).

20. Ichinose, H. & Kitaoka, T. Insight into metabolic diversity of the brown-rot basidiomycete *Postia placenta* responsible for sesquiterpene biosynthesis: semi-comprehensive screening of cytochrome P450 monoxygenase involved in protoilludene metabolism. *Microbial biotechnology* 11, 952-965 (2018).

21. Gupta, D.K. et al. The genome sequence of the commercially cultivated mushroom *Agrocybe aegerita* reveals a conserved repertoire of fruiting-related genes and a versatile suite of biopolymer-degrading enzymes. *BMC Genomics* 19, 48 (2018).

22. Zada, B. et al. Metabolic engineering of *Escherichia coli* for production of mixed isoprenoid alcohols and their derivatives. *Biotechnology for Biofuels* 11, 210 (2018).

23. Crowell, A.L., Williams, D.C., Davis, E.M., Wildung, M.R. & Croteau, R. Molecular cloning and characterization of a new linalool synthase. *Arch Biochem Biophys* 405, 112-21 (2002).

24. Richter, A. et al. Characterization of Biosynthetic Pathways for the Production of the Volatile Homoterpenes DMNT and TMTT in *Zea mays*. *Plant Cell* 28, 2651-2665 (2016).

25. Li, G. et al. Nonseed plant *Selaginella moellendorffii* has both seed plant and microbial types of terpene synthases. *Proceedings of the National Academy of Sciences of the United States of America* 109, 14711-14715 (2012).

26. Whittington, D.A. et al. Bornyl diphosphate synthase: structure and strategy for carbocation manipulation by a terpenoid cyclase. *Proc Natl Acad Sci USA* 99, 15375-80 (2002).

27. Konagurthu, A.S., Whisstock, J.C., Stuckey, P.J. & Lesk, A.M. MUSTANG: A multiple structural alignment algorithm. *Proteins: Structure, Function, and Bioinformatics* 64, 559-574 (2006).

28. Land, H. & Humble, M.S. YASARA: A Tool to Obtain Structural Guidance in Biocatalytic Investigations. in *Protein Engineering: Methods and Protocols* (eds. Bornscheuer, U.T. & Höhne, M.) 43-67 (Springer New York, New York, NY, 2018).

29. Salentin, S., Schreiber, S., Haupt, V.J., Adasme, M.F. & Schroeder, M. PLIP: fully automated protein–ligand interaction profiler. *Nucleic Acids Research* 43, W443-W447 (2015).

30. Morris, G.M. et al. AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. *Journal of Computational Chemistry* 30, 2785-2791 (2009).
### Table 1. Comparison of kinetic parameters for LSs and LNSs from fungi, plant and bacteria.

|                              | Sesquiterpene synthase | Monoterpane synthase |
|------------------------------|------------------------|----------------------|
| **Enzymes**                  | Aa.LNS     | Zm.LNS    | Sc.LNS  | Ap.LS   | Aa.LNS | La.LS    | Ma.LS    | Sc.LNS  |
| **Accession no.**            | /          | Q29VN2    | D5SL78  | /       | /      | Q2XSC5   | Q8H2B4   | D5SL78  |
| **Product**                  | nerolidol  | nerolidol | nerolidol | (R)-linalool | (R)-linalool | (R)-linalool | (R)-linalool | (R)-linalool |
| **Organism**                 | Agrocybe aegerita | Zea mays | Streptomyces clavuligerus | Agrocybe pediades | Agrocybe aegerita | Lavandula angustifolia | Mentha aquatica | Streptomyces clavuligerus |
| **Kingdom**                  | fungi      | plantae   | bacteria | fungi    | fungi   | plantae  | plantae  | bacteria |
| **$K_m$ (μM)**               | 9.0±2.3    | 18±7.8    | 9.6±0.8  | 3.8±0.7  | 6.7±4.6 | 42.7     | 25±6     | 12.9±1.3 |
| **$k_{cat}$ (min$^{-1}$)**   | 3.3±0.3    | -         | 18.7±0.7 | 6±0.3    | 0.5±0.1 | 2.34     | 14.4±1.2 | 5.0±0.1 |
| **$k_{cat}/K_m$ (min$^{-1}$ μM$^{-1}$)** | 0.4       | -         | 1.9      | 1.6      | 0.1     | 0.05     | 0.6      | 0.4     |
| **Reference**                | This study  | 24        | 11     | This study | This study | 5        | 23       | 11     |

Reference numbers are indicated where available.
Fig. 1. Characterization of fungal linalool and nerolidol synthases (LNSs) and a fungal linalool synthase (LS) in *E. coli*. (A) GC-MS chromatograms of cultural supernatants of strains GPPS+9435, FPPS+9435 (co-expressing AAE3_109435 and GPP or FPP synthase) and the control strains GPPS_ctrl and FPPS_ctrl (expressing only GPP or FPP synthase). (B) Chiral separation of linalool produced by Aa.LNS and Ap.LS (Agrped1_689675) expressing strains. (C) Phylogenetic tree based on full-sequence alignment of Aa.LNS and terpene synthases (TPSs) in *A. pediades*. AAE3_109435 (a LNS) is marked with “●”, Agr8 (a γ-muurolene/β-cadinene synthase) with “★”. (D) GC-MS chromatograms of the supernatant of *E. coli* clones producing Agrped1_689671 and Agrped1_689675.

Fig. 2. Sequence analysis of fungal TPS homologues. (A) Phylogenetic tree based on full sequence alignment consisting of 12 TPS homologues from *A. aegerita* (Aa.LNS and Agr), 12 from *A. pediades* (Ap.LS, Ap.LNS and Agrped1), three from *Galerina marginata* (Galma), two from *Hypholoma sublateritium* (Hyps1) and one from *Hebeloma cylindrosporum* (M413_27416). (B) Sequence similarity network (SSN) built on the predicted active sites (Table S1). Agr10 is excluded as it shares limited similarity to the templates (PDB ID: 4LXM and 5NX5). Instead, D5SL78 from *Streptomyces clavuligerus* and D8RNZ9 from *Selaginella moellendorffii* (Spikemoss) are included. The rest of candidates are the same as those of full sequence alignment. LNS and muurolene/cadinene synthase group were highlighted in green and blue, respectively. (C) Experimental validation of predicted fungal TPS homologues. Ap.LS produced exclusively linalool; the other four produced both nerolidol (85-98%) and linalool (15-2%). Here, % refers to the peak area ratios of linalool or nerolidol to the sum of the peak areas of both substances present in each chromatogram.

Fig. 3. In vitro characterization of Aa.LNS and Ap.LS. (A) SDS PAGE gel image of purified Aa.LNS and its kinetic analysis. (B) Protein gel of purified Ap.LS and its kinetic analysis (No product was observed for FPP). (C) *In vivo* activity comparison of LSs from a fungus, a bacterium and a plant (Error bars, mean ± s.d., n=3). Ap.LS, Sc.LNS and Cb.LS from *Clarkia breweri* (Q96376) were selected as the representative of kingdoms fungi, bacteria and plantae, respectively.

Fig. 4. Phylogenetic tree of LNSs and LSs in plants, fungi and bacteria. Two major clades, plant and microbial, were clearly segmented. (R)-linalool synthases were marked with “▲”, (S)-linalool synthases with “△”, LNSs with “●” and nerolidol synthases with “■”. PpSTS25 is a myrcene/linalool synthase (MLS) from *Postia placenta*. For plant and bacterial enzymes, UniProt accession number were used. D8RNZ9, more similar to the bacterial LNS than plant ones, was a LNS isolated from the nonseed plant *Selaginella moellendorffii* (Spikemoss). Details of all the enzymes are in Table S3.
Fig. 5. Structural comparison of LSs in bacteria, fungi and plants. (A) The overlay of 3D structure (grey ribbon) and substrate-binding pocket of Ap.LS. Energy minimized homology model was used with 4LXW as template. GPP and FPP from 2ONG and 6A2C respectively were structurally aligned to 4LXW. Binding pocket surfaces for fungal (B), bacterial (C) and plant (D) LSs. GPP (2-fluorogeranyl diphosphate) and FPP ligands are shown as green and magenta spheres, respectively. The solvent-accessible surface of residues in the binding pocket is colored yellow (hydrophobic), cyan (polar), red (negative) or blue (positive). Best docking model from AutoDock Vina is used to submit to the Protein-Ligand Interaction Profiler (PLIP) server (Salentin et al. NAR 2015). Ap.LS, the most stable protein-ligand complex has the most hydrophobic interactions with the hydrophobic tail of GPP.

Fig. 6. Structure analysis for the understanding of activity and specificity. (A) The aligned 3D active site surfaces of Ap.LS, Sc.LNS and Ma.LS. The residues are listed in Table S4. Gray: conserved residues among the LSs of three kingdoms; green: not conserved in plant LS; blue: not conserved in bacterial LS; yellow: not conserved in fungal LS. (B) Highlight of the key residue difference between Ap.LS and Ap.LNS. Detailed comparison of binding-pocket residues in Table S5. (C) Mutation of the key residues responsible for Ap.LS specificity. (D) The interactions of the selected residues with FPP in Ap.LS wildtype (WT) and its mutants (distance by dash line, unit: Å). (E) Quantitative comparison of Ap.LS and its mutants.
**Agr8** (353 aa)

**Agrped1 820868** (515 aa)

**Agrped1 694262** (343 aa)

**AAE3 109435** (384 aa)

**Agrped1 689671** (343 aa)

**Agrped1 689675** (344 aa)

**Agrped1 665597** (358 aa)

**Agrped1 804989** (267 aa)

**Agrped1 693394** (368 aa)

**Agrped1 804996** (357 aa)

**Agrped1 740962** (345 aa)

**Agrped1 109003** (358 aa)

**Agrped1 640059** (344 aa)
Fig. 3

A  Purified Aa.LNS (purity ~16.3%)

B  Purified Ap.LS (purity ~71.2%)

C  287X

Substrate (μM)

V (nmol/min)

Substrate (μM)

Linalool/nerolidol (mg/L)

ZYM

TB

Fungal  44X

Bacterial  44X

Plant  44X

Fungal  601.2

Bacterial  168.6

Plant  168.6

Purified Aa.LNS (purity ~16.3%)

Purified Ap.LS (purity ~71.2%)

49kDa  1.5 ug  3.1 ug

38kDa

28kDa

62kDa

49kDa

38kDa

28kDa

100  200  300  400  500  600  700

Linalool/nerolidol (mg/L)

Fungal  981.2

Bacterial  81.4

Plant  13.3

Fungal  0

Bacterial  26.2

Plant  0

A  B  C
Fig. 4

1. Plant LSs, NSs and LNSs

2. Microbial LS and LNSs

Fungal MLS

Microbial type LNS in nonseed plants

Bacterial LNS
Fig. 5

A
Ap.LS
Fungi

B
Ap.LS
Fungi

C
Sc.LNS
Bacteria

D
Ma.LS
Plant

# hydrophobic int.: 9
# hydrogen bonds: 4
affinity (kcal/mol): -7.6

# hydrophobic int.: 7
# hydrogen bonds: 4
affinity (kcal/mol): -7.0

# hydrophobic int.: 6
# hydrogen bonds: 3
affinity (kcal/mol): -6.4

Ppi head

GPP (green ball)
FPP (magenta ball)
Supporting information
Bioinformatics-Aided Identification, Characterization of Fungal Linalool Synthases and Applications in Linalool biosynthesis

Congqiang Zhang*, Xixian Chen¹, Raphael Tze Chuen Lee², Rehka T¹, Sebastian Maurer-Stroh²,³#, Martin Rühl⁴#

1. Singapore Institute of Food and Biotechnology Innovation (SIFBI), Agency for Science, Technology and Research (A*STAR), Singapore
2. Bioinformatics Institute (BII), Agency for Science Technology and Research (A*STAR), 30 Biopolis Street, #07-01 Matrix, 138671, Singapore
3. Department of Biological Sciences (DBS), National University of Singapore (NUS), Singapore
4. Institute of Food Chemistry and Food Biotechnology, Justus Liebig University Giessen, Giessen, Germany

#equal contribution
*To whom correspondence may be addressed. Email: zcqsimon@outlook.com or congqiang_zhang@sifbi.a-star.edu.sg

Table of contents
Table S1. Predicted active sites of terpene synthases with 4LXW and 5NX5 as the templates.
Table S2. Comparison and combination of the predicted active sites of the two templates 4LXW and 5NX5.
Table S3. Summary of LSs and LNSs in this study.
Table S4. 3D alignment of active-site residues of Ap.LS, Sc.LNS and Ma.LS.
Table S5. Comparison of residues in the substrate-binding pockets of Ap.LS and Ap.LNS.
Table S6. Primers used for Ap.LS mutation.
Fig. S1. Nucleic acid sequence of AAE3_109435 in the genome.
Fig. S2. Mass spectra and retention indices for terpenes detected in this study.
Fig. S3. The alignment between Agrped_689675 and Galma_223690 and BLAST search results in UniProt database with Agrped_689675 (or Ap.LS).
Fig. S4. GCMS chromatograms and spectra for Galma_266794.
Fig. S5. Amino acid sequence alignment and identity table of LSs and LNSs from fungi, bacteria and plants.
Fig. S6. Expression and solubility analysis of the fungal LNSs and LS.
Fig. S7. The optimization of Aa.LNS solubility with the co-expression of chaperone proteins.
Fig. S8. The sequence alignment and their secondary structures of plant terpene synthases.
Fig. S9. The sequence alignment and their secondary structures of microbial terpene synthases.
Table S1. Predicted active sites of terpene synthases with 4LXW and 5NX5 as the templates.

| No. | Homologues | Predicted active sites by BioTransformer v0.9 or PyMOL v2.1.1 |
|-----|-------------|----------------------------------------------------------|
| 0   | consensus   | X L X M X F V D E T D V E Y R T X G X X N D S Y X E X X H N W X W S R Y |
| 1   | 5NX5       | V I G L T F L D D F D W K H R T I C L C L I N E S F K D - Q H N F Y W G R Y |
| 2   | 4LXW       | Y V M S F L V D D H D Y E Y R T F A H W A A N D S L K E - E H N W V F H R Y |
| 3   | DSSL78     | V I G L T F L D D F D W K H R T I C L C L I N E S F K D - Q H N F Y W G R Y |
| 4   | DBRN2b     | I Y A L V F L D D L E V O Y R T G F M V V I N D S F K E G Y H N V S W S R Y |
| 5   | Agr6       | F L A M Y F V D E T D V E Y R T I G A S I L C N D S Y V E R D H N W N W S R Y |
| 6   | Agr2       | F I T H W D F E T D G E Y R T S G A C I F N D S Y M E K A N W N W S R Y |
| 7   | Agr3       | S L T L L F H D N S D V O Y R T S G C S V T N D S Y V E R D H N W N W S R Y |
| 8   | Agr5       | F S N T L W D E T D V E Y R T S A V C V F N D S Y M E K S A N W N W S R Y |
| 9   | Agpded_1_109003 | A L T M L F H D N S D V O Y R T S G C C V T N D S Y V E R D H N W N W S R Y |
| 10  | Agpded_1_840059 | C L T L L F H D N S D V O Y R T S G C C V T N D S Y V E R D H N W N W S R Y |
| 11  | Agpded_1_665597 | F A T M F F V D E T D C O Y R N V G A T V I N D S Y R E T D H N W N W S R Y |
| 12  | Agpded_1_693584 | F L A M F F V D E T D C O Y R N I G A S I I N D S Y K E T D H N W N W N W S R Y |
| 13  | Agpded_1_694262 | F L A M Y F V D E T D V E Y R T S A G T I I N D S F R E C G H N W N W S R Y |
| 14  | Agpded_1_749682 | N L C V L F V D E S D V E F R N S A V C C V D N S Y M E K M N N W N W S R Y |
| 15  | Agpded_1_804996 | - F M F F V D E T D C O Y R N V G I S I I N D S Y R E T D H N W N W N W S R Y |
| 16  | Agr8       | L L A M Y F V D E T D V E Y R T C G A T I I N D S Y R E C H E N H W N W S R Y |
| 17  | Agpded_1_8208688 | L L A M Y F V D E T D V E Y R T C G A T I I N D S Y R E C G H N W N W S R Y |
| 18  | M413_27416 | L L A M Y F V D E T D V E Y R T C G A T I I N D S Y R E C A H N W N W S R Y |
| 19  | Galma_266794 | L L A M Y F V D E T D V E Y R T C G A T I I N D S Y R E C A H N W N W S R Y |
| 20  | Galma_63556 | - F M Y F A D E T D I E Y R T C G A S I A N D S Y L E R D H N C Y W S R Y |
| 21  | Hypsys1_148385 | L L A M Y F A D E T D V E Y R T C G A S I A N D S Y L E R D H N C Y W S R Y |
| 22  | Hypsys1_148385 | F L A M Y F A D E T D V E Y R T C G A S I A N D S Y L E R D H N C Y W S R Y |
| 23  | Agpded_1_698971 | F L A M Y F A D E T D V E Y R T C G A S I A N D S Y L E R D H N C Y W S R Y |
| 24  | Agpded_1_698975 | F L A M Y F A D E T D V E Y R T C G A S I A N D S Y L E R D H N C Y W S R Y |
| 25  | Galma_232690 | F L A M Y F A D E T D V E Y R T C G A S I A N D S Y L E R D H N C Y W S R Y |
| 26  | Agpded_1_804989 | - M F L V D K T D C O Y R N V G A A I T N D S Y K E T D H N W N W N W S R Y |
| 27  | Agr11      | T L A M Y F V D E T D V E Y R T S G A M T - - - - - - - R S A H N A E W A R Y |
| 28  | Agr1       | C L T L L F V D E T D V E Y R T S G A C V N D S Y M E K S A N W N W S R Y |
| 29  | Agr9       | F L A M Y F V D E T D V E Y R S S G C S I A N D S Y M E R E H N W N W T R Y |
| 30  | Agr4       | N L C V L F V D E T D V E Y R T S G A C V C N D S Y M E K S A N W N W S R Y |
| 31  | Agr7       | F L A M Y F V D E T D V E Y R T I G A S L C N D S Y V E R D H N W N W S R Y |
| 32  | AAE3_104935 | T L A M Y F V D E T D V E Y R T S G A C V C N D S Y M E K S A N W N W S R Y |
| 33  | Agr10      | T L A M Y F V D E T D V E Y R T S G A M T - - - - - - - T I G N D S Y M K - - - R W N W N T R Y |
| 34  | Agpded_1_705454 | P G H Y I V M D D T G L Q F Y T K G G L T N P G Y E R N A E W K W A M Y |
| 35  | Q8H2B4     | - P R W I T D D Y D E T I G A V Y V R D D F R K L - Y - - - - - - - - - - - - |

Murolene/cadinene synthase cluster and LS/LNS cluster predicted or characterized in this study were shown in blue and green, respectively (consistent with Fig. 2). The non-functional synthase Agr10 and Agr11 were highlighted in red.  
1 The two proteins were used as templates: DSSL78, Sc.LNS from Streptomyces clavuligerus, (PDB ID: 5NX5) and Q9K499, Epi-isozimaene synthase from Streptomyces coelicolor (PDB ID: 4LXW). For 4LXW, the ligands BTM, POP and MG were used for prediction of active sites. For 5NX5, the ligands 0FV and MG were used for prediction of active sites.  
2 Analyzed by PyMOL software version 2.1.1. The homologue model of Q8H2B4 was built on the structure of (+)-bornyl diphasphate synthase from Salvia officinalis (PDB entry ID, 1n1b/1n21) with Modeller software.
Table S2. Comparison and combination of the predicted active sites of the two templates 4LXW and 5NX5.

| Index Position | 4LXW_position | 5nx5_position | 4LXW_residue | 5nx5_residue |
|----------------|---------------|---------------|--------------|--------------|
| 52             | 72            | 49            | V            | V            |
| 55             | 73            | 52            | M            | I            |
| 56             | 92            | 75            | S            | L            |
| 76             | 95            | 76            | F            | T            |
| 79             | 96            | 77            | F            | F            |
| 80             | 99            | 80            | V            | D            |
| 81             | 100           | 83            | M            | L            |
| 83             | 102           | 84            | S            | D            |
| 84             | 103           | 86            | F            | D            |
| 86             | 172           | 88            | D            | D            |
| 87             | 175           | 90            | H            | D            |
| 153            | 190           | 91            | Y            | Y            |
| 156            | 194           | 92            | E            | R            |
| 171            | 197           | 93            | Y            | T            |
| 175            | 198           | 94            | R            | F            |
| 178            | 199           | 95            | F            | A            |
| 179            | 203           | 96            | D            | W            |
| 180            | 204           | 156           | H            | A            |
| 181            | 236           | 172           | A            | N            |
| 184            | 237           | 175           | D            | S            |
| 217            | 240           | 176           | L            | L            |
| 218            | 241           | 177           | K            | E            |
| 221            | 244           | 178           | E            | E            |
| 222            | 245           | 181           | H            | H            |
| 225            | 246           | 180           | W            | N            |
| 228            | 247           | 214           | V            | W            |
| 229            | 248           | 218           | F            | H            |
| 232            | 249           | 222           | H            | R            |
| 235            | 253           | 225           | Y            | Y            |
| 237            | 255           | 226           | T            | W            |
| 238            | 256           | 229           | A            | G            |
| 239            | 257           | 231           | N            | R            |
| 299            | 325           | 295           | F            | Y            |
| 300            | 329           | 299           | H            | W            |
| 303            | 332           | 302           | F            | G            |
| 306            | 333           | 307           | H            | R            |
| 307            | 338           | 312           | Y            | Y            |

1 Here, individual active sites of 4LXM and 5NX5 were predicted with BioTransformer v0.9, which are identical to PyMOL software prediction. Furthermore, to cover more broadly in our model, the active sites of the two templates were merged with each other to obtain 4LXW_m and 5NX5_m, respectively.
| No. | Accession no. | Entry/Gene name | Protein function | Organism | Length |
|-----|--------------|----------------|-----------------|----------|--------|
| 1   | Q8AW2V8      | ACS5_MAIZE     | Nerolidol synthase | Zea mays (Maize) | 590    |
| 2   | P0C6V26      | NES1_FRAAN     | Nerolidol synthase | Fragaria ananassa (Strawberry) (Fragaria chiloensis x Fragaria virginiana) | 519    |
| 3   | P0C6V95      | NES2_FRAAN     | Nerolidol synthase | Fragaria ananassa (Strawberry) (Fragaria chiloensis x Fragaria virginiana) | 578    |
| 4   | P0C6V96      | NES1_FRAVE     | Nerolidol synthase | Fragaria vesca (Woodland strawberry) (Potentilla vesca) | 580    |
| 5   | Q5SPN0       | LLO5_ARTAN     | R-linalool synthase | Artemisia annua (Sweet wormwood) | 567    |
| 6   | Q84UV0       | LNS_ARATH      | S-linalool synthase | Arabidopsis thaliana (Mouse-ear cress) | 569    |
| 7   | Q6ZH94       | LNS_ORYSJ      | S-linalool synthase | Oryza sativa subsp. japonica (Rice) | 595    |
| 8   | Q8H2B4       | LLO5_MENAO     | R-linalool synthase | Mentha aquatica (Water mint) | 606    |
| 9   | Q5SPN1       | LLOS_ARTAN     | R-linalool synthase | Medicago truncatula | 583    |
| 10  | Q5SPN3       | LLOS_OCIBA     | R-linalool synthase | Ocimum basilicum (Sweet basil) | 574    |
| 11  | Q96376       | LILS_CLABR     | S-linalool synthase | Clarkia breweri (Fairies) (Eucharidiaceae breweri) | 870    |
| 12  | Q2XSC5       | LALIN_LAVAN    | R-linalool synthase | Lavandula angustifolia (Lavender) | 564    |
| 13  | R46685       | R46685_MGNN    | S-linalool synthase | Cinnamomum osmophloeum | 585    |
| 14  | Q29VN2       | TPS2_MAIZE     | Nerolidol linalool synthase | Zea mays (Maize) | 581    |
| 15  | H6WBC5       | H6WBC5_VITVI   | Nerolidol linalool synthase | Vitis vinifera (Grape) | 577    |
| 16  | A0A06BBON9   | A0A06BBON9_9ROSA | Nerolidol linalool synthase | Prunus cerasoides var. campanulata | 618    |
| 17  | A0A06BB6B6   | A0A06BB6B6_9ROSA | Nerolidol linalool synthase | Prunus cerasoides var. campanulata | 561    |
| 18  | A0A072UZ75   | A0A072UZ75_9MEDTR | Nerolidol linalool synthase | Medicago truncatula (Barrel medic) (Medicago truncatuloides) | 570    |
| 19  | B1NAB4       | B1NAB4_9ANTMA  | Nerolidol linalool synthase | Antirhinum majus (Garden Snapdragon) | 596    |
| 20  | B1NAB5       | B1NAB5_9ANTMA  | Nerolidol linalool synthase | Antirhinum majus (Garden Snapdragon) | 566    |
| 21  | G5UB06       | G5UB06_9MEDTR  | Nerolidol linalool synthase | Medicago truncatula (Barrel medic) (Medicago truncatuloides) | 573    |
| 22  | G5CV39       | G5CV39_9SOLLCC | Nerolidol linalool synthase | Solanum lycopersicum (Tomato) (Lycopersicon esculentum) | 563    |
| 23  | G7KNU1       | G7N1U1_9MEDTR  | Nerolidol linalool synthase | Medicago truncatula (Barrel medic) (Medicago truncatuloides) | 519    |
| 24  | E5QAH7       | E5QAH7_VITVI   | Nerolidol linalool synthase | Vitis vinifera (Grape) | 545    |
| 25  | A0A07TG55   | A0A07TG55_GOSHI | Nerolidol linalool synthase | Gossypium hirsutum (Upland cotton) (Gossypium mexicanum) | 586    |
| 26  | A0A061G969   | A0A061G969_9THECC | Nerolidol linalool synthase | Theobroma cacao (Cacao) (Cocoa) | 583    |
| 27  | E5QAH4       | E5QAH4_VITVI   | Nerolidol linalool synthase | Vitis vinifera (Grape) | 545    |
| 28  | G7INZ1       | G7INZ1_9MEDTR  | Nerolidol linalool synthase | Medicago truncatula (Barrel medic) (Medicago truncatuloides) | 806    |
| 29  | E5QAH1       | E5QAH1_VITVI   | Nerolidol linalool synthase | Vitis vinifera (Grape) | 584    |
| 30  | E5QAH1       | E5QAH1_VITVI   | Nerolidol linalool synthase | Vitis vinifera (Grape) | 820    |
| 31  | G7INZ4       | G7INZ4_9MEDTR  | Nerolidol linalool synthase | Medicago truncatula (Barrel medic) (Medicago truncatuloides) | 819    |
| 32  | E5QAH0       | E5QAH0_VITVI   | Nerolidol linalool synthase | Vitis vinifera (Grape) | 840    |
| 33  | A0A061GEP1   | A0A061GEP1_9THECC | Nerolidol linalool synthase | Theobroma cacao (Cacao) (Cocoa) | 809    |
| 34  | E5QAH3       | E5QAH3_VITVI   | Nerolidol linalool synthase | Vitis vinifera (Grape) | 577    |
| 35  | D8RNS9       | MTS22_9SELML   | Nerolidol linalool synthase | Selaginella molendendorffii (Spikemoss) | 368    |
| 36  | D85L78       | SCLAV_p1185    | Nerolidol linalool synthase | Streptomyces clavuligerus | 333    |
| 37  | M1N16G304    | AA3_50204      | Not functional | Agrocybe aegerita (Cyclocybe aegerita) | 356    |
| 38  | MN954676     | AA3_109435     | Nerolidol/linalool synthase | Agrocybe aegerita (Cyclocybe aegerita) | 384    |

**Table S3.** Summary of LSs and LNSs in this study.
Table S4. 3D alignment of active-site residues of Ap.LS, Sc.LNS and Ma.LS. Conversed residues among the three enzymes are highlighted in green.

| Yellow | | Blue | | Green | |
|---|---|---|---|---|
| **Ap.LS_pos** | **Ap.LS_res** | **Sc.LNS_pos** | **Sc.LNS_res** | **Ma.LS_pos** | **Ma.LS_res** |
| 54 | F | | | | |
| 56 | L | 49 | V | | |
| 77 | M | 72 | L | 360 | T |
| 78 | N | 73 | G | 361 | A |
| 79 | F | 74 | W | 362 | L |
| 80 | Y | 75 | T | 363 | D |
| 81 | F | 76 | F | 364 | D |
| 82 | A | 77 | L | 365 | V |
| 83 | F | 78 | F | 366 | Y |
| 84 | D | 79 | D | 367 | D |
| 85 | E | 80 | D | 368 | I |
| | | 82 | F | | |
| 88 | D | 83 | D | | |
| 149 | Y | | | | |
| 153 | I | 150 | W | 438 | Y |
| 156 | E | 153 | K | 441 | E |
| 160 | R | 157 | R | 445 | Y |
| 171 | Y | | | | |
| 175 | R | 172 | R | 460 | A |
| 178 | T | 175 | T | 463 | T |
| 179 | C | 176 | I | 464 | I |
| | | 215 | I | 504 | R |
| 219 | N | 218 | N | 507 | D |
| 223 | S | 222 | S | 511 | T |
| 227 | E | 226 | D | 515 | E |
| 230 | R | | | 518 | R |
| 235 | H | | | | |
| 309 | Y | 309 | Y | 579 | A |
| 315 | L | 316 | G | 586 | G |
Table S5. Comparison of residues in the substrate-binding pockets of Ap.LS and Ap.LNS. The different residues of the same position are highlighted in green.

| LS_Ap_pos | LS_Ap_res | LNS_Ap_pos | LNS_Ap_res |
|-----------|-----------|------------|------------|
| 54        | F         | 53         | F          |
| 56        | L         | 55         | L          |
| 57        | L         | 56         | L          |
| 58        | G         | 57         | G          |
| 59        | A         | 58         | S          |
| 60        | L         | 59         | M          |
| 73        | S         | 72         | S          |
| 74        | C         | 73         | C          |
| 75        | D         | 74         | D          |
| 76        | L         | 75         | L          |
| 77        | M         | 76         | M          |
| 78        | N         | 77         | N          |
| 79        | F         | 78         | F          |
| 80        | Y         | 79         | Y          |
| 81        | F         | 80         | F          |
| 82        | A         | 81         | A          |
| 83        | F         | 82         | F          |
| 84        | D         | 83         | D          |
| 85        | E         | 84         | E          |
| 87        | T         | 86         | T          |
| 88        | D         | 87         | D          |
| 149       | Y         | 148        | Y          |
| 153       | I         | 152        | V          |
| 156       | E         | 155        | E          |
| 157       | A         | 156        | A          |
| 160       | R         | 159        | R          |
| 171       | Y         | 170        | Y          |
| 175       | R         | 174        | R          |
| 178       | T         | 177        | T          |
| 179       | C         | 178        | C          |
| 180       | G         | 179        | G          |
| 181       | G         | 180        | A          |
| 184       | S         | 183        | S          |

| LS_Ap_pos | LS_Ap_res | LNS_Ap_pos | LNS_Ap_res |
|-----------|-----------|------------|------------|
| 215       | I         | 214        | I          |
| 216       | A         | 215        | A          |
| 219       | N         | 218        | N          |
| 220       | D         | 219        | D          |
| 223       | S         | 222        | S          |
| 224       | Y         | 223        | Y          |
| 226       | L         | 225        | L          |
| 227       | E         | 226        | E          |
| 230       | R         | 229        | R          |
| 233       | D         | 232        | D          |
| 235       | H         | 234        | H          |
| 236       | N         | 235        | N          |
| 299       | Y         | 298        | Y          |
| 302       | W         | 301        | W          |
| 304       | Y         | 303        | Y          |
| 305       | E         | 304        | E          |
| 306       | T         | 305        | T          |
| 309       | Y         | 308        | Y          |
| 310       | Y         | 309        | Y          |
| 313       | N         | 312        | N          |
| 315       | L         | 314        | L          |
| 316       | E         | 315        | Q          |
| 332       | Y         | 331        | Y          |
Table S6. Primers used for Ap.LS mutation. The mutational regions are highlighted in red and caps.

| Mutations       | Forward primers            | Reverse primers            |
|-----------------|-----------------------------|----------------------------|
| A59S            | tgctgggtTCGctggttggtccgctgggtac | ccagCGAacccacgcagttgaatacatgc |
| L60M            | tgggtgcaATGtgtggtccgctgggtacaa | caacCATgcacccacgcagttgaatacatca |
| V61I            | tgcactgATTggtccgctgggtaccaaggaa | ggaccAACATcgtgcacccacgcagttgaata |
| G181A           | ctgtgggCgaacccacctgttctgttcttt | gtttcGcaccacaggtgtcagggc |
| E316Q           | aatggcctgCaaatccagaaaactgctcaga | tctggatttGcaggccatttttacctgaata |
| A59S-L60M       | ggtTcaAtggttggtccgtgggtgac | gaccaaccaTtgAaccacgcagttga |
| A59S-L60M-V61I  | ggtTcaAtgAtggttgccgtgggttacca | cggaccaATcaTtgAacccacgcagttga |
Fig. S1. Nucleic acid sequence of AAE3_109435 in the genome. The two primers (AAE3_9435f and AAE3_9435r) used for PCR amplification were highlighted here.
Fig. S2. Mass spectra and retention indices for terpenes detected in this study. Retention indices (RIs) were calculated by calibrating with GC-MS with a C8–C30 alkane mix and compared with literature data in National Institute of Standards and Technology database.

| Terpenes       | Calculated RI | Literature RI |
|----------------|---------------|---------------|
| linalool       | 1099          | 1086          |
| nerolidol      | 1567          | 1564          |
| geraniol       | 1259          | 1256          |
| γ-muurolene    | 1486          | 1472          |
| germacrene D   | 1498          | 1481          |
| (+)-δ-cadinene | 1527          | 1516          |
Fig. S3. The alignment between Agrped_689675 and Galma_223690 and BLAST search results in UniProt database with Agrped_689675 (or Ap.LS). The search was done in March-2018 with the top 6 proteins listed here. The results may change today as new proteins are being deposited into UniProt database.
Fig. S4. GCMS chromatograms and spectra for Galma_266794. Retention indices (RIs) were listed with the calculated values on the left and literature values on the right. Germacrene D and γ-muurolene are further verified by Cubeb essential oil. Cedrela woods oil was used as the authentic standard of (+)-δ-cadinene.

Germacrene D: RI 1498/1481
γ-muurolene: RI 1486/1472
(+)-δ-cadinene: RI 1527/1516
**Fig. S5.** Amino acid sequence alignment and identity table of LSs and LNSs from fungi, bacteria and plants. The two conserved regions, aspartate-rich motif and NSE triad were highlighted in blue boxes. The sequence identity table indicated the large difference of LSs among different kingdoms. The overall sequence identity is very low, and the fungal LS is slightly more similar to the bacterial LNS than to the plant LS and LNS. Alignment was done with Clustal Omega program v1.2.4.

|   | 1  | 2  | 3  | 4  |
|---|----|----|----|----|
| 1 | 15.2% | 13.4% | 10.0% |   |
| 2 | 15.2% | 7.8% | 10.4% |   |
| 3 | 13.4% | 7.8% | 24.0% |   |
| 4 | 10.0% | 10.4% | 24.0% |   |
Fig. S6. (A) Expression and solubility analysis of the fungal LNSs and LS. (B) Solubility analysis for the fusion Aa.LNS with maltose binding protein (MBP) or thioredoxin (TrxA). M: protein ladder, TP: total proteins, SF: soluble fraction, In: induced cells, No: not induced as control.

| No. | Enzymes          | kDa |
|-----|------------------|-----|
| 1   | AAE3_109435, Aa.LNS | 43.4|
| 2   | Agrped1_689675, Ap.LS    | 39.7|
| 3   | Agrped1_689671, Ap.LNS | 39.8|
| 4   | Galma_223690, Gm.LNS   | 39.4|
| 5   | Hysu_148385, Hs.LNS     | 39.2|

| No. | Enzymes          | kDa |
|-----|------------------|-----|
| I   | Aa.LNS           | 43.4|
| II  | MBP_Aa.LNS       | 85  |
| III | TrxA_Aa.LNS      | 57  |

M – marker
TP – total protein
SF – Soluble fraction
In – induced
No – not induced, as control
Fig. S7. The optimization of Aa.LNS solubility with the co-expression of chaperone proteins. Two chaperone systems were used here: pGro7 expressing groES-groEL proteins, pTf16 expressing the trigger factor protein (tig). IPTG and arabinose were used to induce the expression of Aa.LNS and chaperones, respectively. It was found arabinose concentration (or chaperone expression levels) has invisible effect on the solubility of Aa.LNS. Hence, 3.3mM arabinose and 0.1 mM IPTG were used for the large-scale expression and purification experiments. M: protein ladder, TP: total proteins, SF: soluble fraction, In: induced, No: not induced as control.
Fig. S8. The sequence alignment and their secondary structures of plant terpene synthases. The three regions, arginine effector motif, aspartate-rich motif and NSE triad, were highlighted in black boxes. The four proteins were (1) O81192, PDB ID 1n1b/1n24, (+)-bornyl diphosphate synthase from *Salvia officinalis*; (2) Q8H2B4, (R)-linalool synthase from *Mentha aquatica* (Water mint); (3) Q29VN2, linalool/nerolidol synthase from *Zea mays* (Maize); (4) Q96376, S-linalool synthase from *Clarkia breweri* (Fairy fans). The figure was prepared with the ESPript 3.0.
Fig. S9. The sequence alignment and their secondary structures of microbial terpene synthases. The three regions, arginine effector motif, aspartate-rich motif and NSE triad, were highlighted in black boxes. The eight proteins were (1) Q9K499 (PDB ID 4LXW), Epi-isozizaene synthase from *Streptomyces coelicolor*; (2) D5SL78, Sc.LNS from *S. clavuligerus*; (3) D8RNZ9, LNS from *Spikemoss*; (4) Ap.LS; (5) Ap.LNS; (6) Aa.LNS (7) Gm.LNS; (8) Hs.LNS. The figure was prepared with the ESPript 3.0 2.
Reference
1. Agger, S., Lopez-Gallego, F. & Schmidt-Dannert, C. Diversity of sesquiterpene synthases in the basidiomycete Coprinus cinereus. Molecular microbiology 72, 1181-1195 (2009).
2. Robert, X. & Gouet, P. Deciphering key features in protein structures with the new ENDscript server. Nucleic Acids Research 42, W320-W324 (2014).