

Spirocyclo Terpene Compounds and Their Biosynthesis Thereof

Technical Field

- 5 The present invention relates, in general terms, to spirocyclo terpene compounds and their biosynthesis thereof.

Background

- 10 The fruiting bodies of milk cap mushrooms (genus *Lactarius*) produce a milky latex, which is exuded upon tissue injury, containing sesquiterpene compounds that are responsible for their characteristic color or flavor (Figure 1).

- Examples are the blue azulene pigment in *L. indigo* (Figure 1A) and orange
15 dihydroazulene pigment in *L. deliciosus* (Figure 1B). In many cases, the terpene compounds are fatty acyl esters, and tissue damage leads to the cleavage of the fatty acyl group and conversion of the terpene moiety into a variety of products. In *L. deliciosus*, this results in the production of a variety of substituted azulene derivatives, causing the latex to take on its characteristic
20 hue. In other milk cap mushrooms (e.g., *L. fuliginosus* and *L. chrysorrheus*), this leads to the development of compounds imparting a bitter or acrid flavor. Additionally, volatile sesquiterpene-derived dialdehydes isovelleral and piperdial (Figure 1C and 1D, respectively) are released upon injury of the fruiting bodies of *L. rufus* and *L. piperatus*, respectively, and contribute to their pungent flavour
25 and provide chemical defense against microbes and parasites.

- These terpene compounds are expected to display a broad range of physiological properties, including antibiotic, antitumor, antiviral, cytotoxic, immunosuppressive, phytotoxic, antifungal, insect antifeedant, and hormonal
30 activities. These terpene compounds may also be used as a starting material to obtain functionalized terpene derivatives for such applications.

Despite the large number of terpenoids that have been isolated and identified in *Lactarius* mushrooms (including caryophyllene, drimane and isolactarorufin), their biosynthetic pathways are not yet understood. The first step in the biosynthetic pathway of sesquiterpene natural products is generation of the terpene scaffold, catalyzed by sesquiterpene synthases (STSs). Fungal STSs, particularly those in mushrooms (basidiomycetes), are much less well studied compared to plant STSs, and the limited homology between fungal and plant STSs makes it difficult to predict their functions based on studies of the plant enzymes.

10

The mushroom STS has remained a fascinating topic for studies over the past few years, and several mushroom STSs have been identified, though their full catalytic repertoire remains unexplored. A more complete understanding of mushroom STSs would enable the identification of new terpene-derived natural products and provide new terpene scaffolds for biotechnological applications.

15

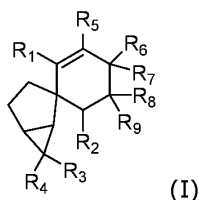
It would be desirable to overcome or ameliorate at least one of the above-described problems.

20 **Summary**

Several candidate sesquiterpene synthases were identified from the genome of the saffron milk cap mushroom *L. deliciosus* and expressed in a sesquiterpene overproducing *Escherichia coli* strain. In addition to enzymes that produce several known terpenes, we identified an enzyme belonging to a previously unknown clade of sesquiterpene synthases that produces a terpene with a unique spiro-tricyclic scaffold. These findings add to the rich diversity of terpene scaffolds and mushroom terpene synthases and are valuable for biotechnological applications in producing these terpenoids.

30

The present disclosure provides a compound of Formula (I) or a salt, solvate or stereoisomer thereof:



wherein

R_1 , R_2 , R_3 and R_4 are independently selected from H, oxo, optionally substituted alkyl, optionally substituted alkenyl; and

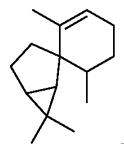
- 5 R_5 , R_6 , R_7 , R_8 and R_9 are independently selected from H, oxo, optionally substituted alkyl, optionally substituted alkenyl.

In some embodiments, R_1 , R_2 , R_3 and R_4 are independently optionally substituted C_1 - C_4 alkyl.

10

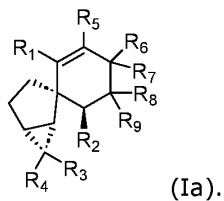
In some embodiments, R_5 , R_6 , R_7 , R_8 and R_9 are independently selected from H, or optionally substituted C_1 - C_4 alkyl.

In some embodiments, the compound of Formula (I) is



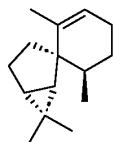
15

In some embodiments, the compound of Formula (I) is a compound of Formula (Ia):

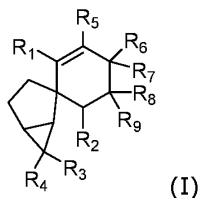


20

In some embodiments, the compound of Formula (I) is



The present disclosure also provides a biosynthetic method of preparing a compound of Formula (I) or a salt, solvate or stereoisomer thereof:



5

wherein

R₁, R₂, R₃ and R₄ are independently selected from H, oxo, optionally substituted alkyl, optionally substituted alkenyl; and

10 R₅, R₆, R₇, R₈ and R₉ are independently selected from H, oxo, optionally substituted alkyl, optionally substituted alkenyl;

the method comprising:

a) fermenting a cellular organism comprising a sesquiterpene synthetase (STS) enzyme, wherein the STS enzyme is characterised by SEQ 7 and/or SEQ 11.

15

In some embodiments, the method further comprises transfecting or transforming a clonal construct comprising the STS enzyme into the cellular organism.

20 In some embodiments, the clonal construct comprises a vector backbone with a pET28a and/or p15A origin.

In some embodiments, the cellular organism is a single celled organism.

25 In some embodiments, the cellular organism is an engineered single celled

organism modified to overexpress at least one mevalonate pathway gene.

In some embodiments, the cellular organism is *E. coli* BL21.

- 5 In some embodiments, the compound of Formula (I) is characterised by a yield of more than 60 mg/L.

In some embodiments, the compound of Formula (I) is characterised by a yield of more than 80 mg/L.

10

In some embodiments, the method is characterised by an impurity selected from 9-aristolene and 1-aristolene.

- 15 In some embodiments, the method is characterised by a ratio of the compound of Formula (I) to the impurity of about 2:1 to about 6:1.

Brief description of the drawings

- 20 Embodiments of the present invention will now be described, by way of non-limiting example, with reference to the drawings in which:

Figure 1 shows examples of terpene-derived compounds in *Lactarius* mushrooms. (A) The blue azulene pigment of *L. indigo*, (B) orange dihydroazulene pigment of *L. deliciosus* ($R = \text{CH}_3(\text{CH}_2)_{16}\text{CO}_2$), (C) isovelleral, and (D) piperdial.

Figure 2 shows maximum likelihood phylogenetic tree of previously characterized basidiomycete STS, together with predicted *L. deliciosus* STS.

Figure 3 shows GC-MS detection of the terpene products of the clade 5 enzymes LdSTS7 and LdSTS11. (A) Total ion chromatograms of the fermentation products of *E. coli* overexpressing LdSTS7 and LdSTS11, showing a major product 1 and two other minor products (2 & 3). (B) GC-MS trace of the main compound 1 produced by LdSTS7.

Figure 4 shows (A) Numbering nomenclature for compound (1), 9-aristolene (2), and 1-aristolene (3) used in NMR analysis. (B) Structure elucidation process for compound 1. (C) Observed key HMBC correlations for 1. (D) Observed key NOE correlations for 1.

- 5 **Figure 5** shows proposed bifurcated mechanistic pathway for conversion of farnesyl pyrophosphate to 1, 2, and 3.

Detailed description

"Alkyl" refers to monovalent alkyl groups which may be straight chained or
10 branched and preferably have from 1 to 10 carbon atoms or more preferably 1 to 6 carbon atoms. Examples of such alkyl groups include methyl, ethyl, n-propyl, iso-propyl, n-butyl, iso-butyl, n-hexyl, and the like.

"Alkenyl" refers to a monovalent alkenyl group which may be straight chained
15 or branched and preferably have from 2 to 10 carbon atoms and more preferably 2 to 6 carbon atoms and have at least 1 and preferably from 1-2, carbon to carbon, double bonds. Examples include ethenyl ($-\text{CH}=\text{CH}_2$), n-propenyl ($-\text{CH}_2\text{CH}=\text{CH}_2$), iso-propenyl ($-\text{C}(\text{CH}_3)=\text{CH}_2$), but-2-enyl ($-\text{CH}_2\text{CH}=\text{CHCH}_3$), and the like.

20 "Alkoxy" refers to the group alkyl-O- where the alkyl group is as described above. Examples include, methoxy, ethoxy, n-propoxy, iso-propoxy, n-butoxy, tert-butoxy, sec-butoxy, n-pentoxy, n-hexoxy, 1,2-dimethylbutoxy, and the like.

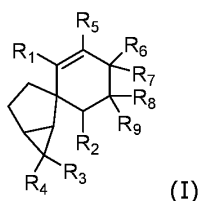
25 "Halo" or "halogen" refers to fluoro, chloro, bromo and iodo.

"Oxo/hydroxy" refers to groups $=\text{O}$, $\text{HO}-$.

30 "Amino" refers to the group $-\text{NR}^1\text{R}^2$ where each R¹ is independently hydrogen, alkyl, cycloalkyl, aryl, heteroaryl, and heterocyclyl and where each of alkyl, cycloalkyl, aryl, heteroaryl and heterocyclyl is as described herein.

In this specification "optionally substituted" is taken to mean that a group may or may not be further substituted or fused (so as to form a condensed polycyclic group) with one or more groups selected from hydroxyl, acyl, alkyl, alkoxy, alkenyl, alkenyloxy, alkynyl, alkynyloxy, amino, aminoacyl, thio, arylalkyl, arylalkoxy, aryl, aryloxy, carboxyl, acylamino, cyano, halogen, nitro, phosphono, sulfo, phosphorylamino, phosphinyl, heteroaryl, heteroarylalkyl, heteroaryloxy, heterocyclyl, heterocyclylalkyl, heterocyclyoxy, oxyacyl, oxime, oxime ether, hydrazone, oxyacylamino, oxysulfonylamino, aminoacyloxy, trihalomethyl, trialkylsilyl, pentafluoroethyl, trifluoromethoxy, difluoromethoxy, trifluoromethanethio, trifluoroethenyl, mono- and di-alkylamino, mono- and di-(substituted alkyl)amino, mono- and di-arylamino, mono- and di-heteroarylamino, mono- and di-heterocyclyl amino, and unsymmetric di-substituted amines having different substituents selected from alkyl, aryl, heteroaryl and heterocyclyl, and the like, and may also include a bond to a solid support material, (for example, substituted onto a polymer resin). For instance, an "optionally substituted amino" group may include amino acid and peptide residues.

The present disclosure provides a compound of Formula (I) or a salt, solvate or stereoisomer thereof:



wherein

R₁, R₂, R₃ and R₄ are independently selected from H, oxo, optionally substituted alkyl, optionally substituted alkenyl; and

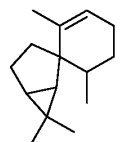
R₅, R₆, R₇, R₈ and R₉ are independently selected from H, oxo, optionally substituted alkyl, optionally substituted alkenyl.

In some embodiments, R_1 , R_2 , R_3 and R_4 are independently optionally substituted alkyl. In some embodiments, R_1 , R_2 , R_3 and R_4 are independently optionally substituted C_1 - C_4 alkyl. The alkyl may be methyl, ethyl, propyl iso-propyl, n-butyl, t-butyl, sec-butyl or iso-butyl. The optional substituent may be
5 halo, oxo, alkoxy, or amino.

In some embodiments, R_5 , R_6 , R_7 , R_8 and R_9 are independently selected from H, or optionally substituted alkyl. In some embodiments, R_5 , R_6 , R_7 , R_8 and R_9 are independently selected from H, or optionally substituted C_1 - C_4 alkyl. In some
10 embodiments, R_5 , R_6 , R_7 , R_8 and R_9 are independently selected from optionally substituted C_1 - C_4 alkyl. The alkyl may be methyl, ethyl, propyl iso-propyl, n-butyl, t-butyl, sec-butyl or iso-butyl. The optional substituent may be halo, oxo, alkoxy, or amino. In some embodiments, R_5 , R_6 , R_7 , R_8 and R_9 are independently H.

15

In some embodiments, the compound of Formula (I) is

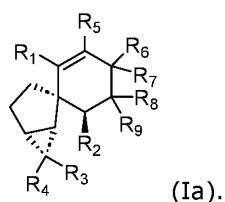


“Isomer” includes especially optical isomers or stereoisomers (for example
20 essentially pure enantiomers, essentially pure diastereomers, and mixtures thereof) as well as conformation isomers (i.e. isomers that differ only in their angles of at least one chemical bond), position isomers (particularly tautomers), and geometric isomers (e.g. cis-trans isomers).

25 Compounds described herein can comprise one or more asymmetric centers, and thus can exist in various isomeric forms, e.g., enantiomers and/or diastereomers. For example, the compounds described herein can be in the form of an individual enantiomer, diastereomer or geometric isomer, or can be in the form of a mixture of stereoisomers, including racemic mixtures and mixtures

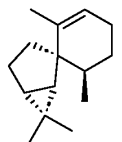
enriched in one or more stereoisomer. Isomers can be isolated from mixtures by methods known to those skilled in the art, including chiral high pressure liquid chromatography (HPLC) and the formation and crystallization of chiral salts; or preferred isomers can be prepared by asymmetric syntheses. The invention additionally encompasses compounds described herein as individual isomers substantially free of other isomers, and alternatively, as mixtures of various isomers. "Optically-enriched," as used herein, means that the compound is made up of a significantly greater proportion of one enantiomer. In certain embodiments the compound of the present invention is made up of at least about 90% by weight of a preferred enantiomer. In other embodiments the compound is made up of at least about 95%, 98%, or 99% by weight of a preferred enantiomer. Preferred enantiomers may be isolated from racemic mixtures by any method known to those skilled in the art, including chiral high pressure liquid chromatography (HPLC) and the formation and crystallization of chiral salts or prepared by asymmetric syntheses. See, for example, Jacques et al., *Enantiomers, Racemates and Resolutions* (Wiley Interscience, New York, 1981); Wilen et al., *Tetrahedron* 33:2725 (1977); Eliel, *Stereochemistry of Carbon Compounds* (McGraw-Hill, NY, 1962); and Wilen, *Tables of Resolving Agents and Optical Resolutions* p. 268 (E.L. Eliel, Ed., Univ. of Notre Dame Press, Notre Dame, IN 1972).

In some embodiments, the compound of Formula (I) or a salt, solvate or stereoisomer thereof is a compound of Formula (Ia):



25

In some embodiments, the compound of Formula (I) is



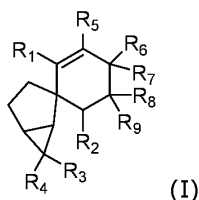
The compound of the invention can be salt thereof. Suitable salts include, but are not limited to salts of inorganic acids such as hydrochloric, sulphuric, phosphoric, nitric, carbonic, boric, sulfamic, and hydrobromic acids, or salts of pharmaceutically acceptable organic acids such as acetic, propionic, butyric, tartaric, maleic, hydroxymaleic, fumaric, maleic, citric, lactic, mucic, gluconic, benzoic, succinic, oxalic, phenylacetic, methanesulphonic, toluenesulphonic, benzenesulphonic, salicylic sulphonic, aspartic, glutamic, edetic, stearic, palmitic, oleic, lauric, pantothenic, tannic, ascorbic and valeric acids.

Base salts include, but are not limited to, those formed with pharmaceutically acceptable cations, such as sodium, potassium, lithium, calcium, magnesium, ammonium and alkylammonium. In particular, the present invention includes within its scope cationic salts eg sodium or potassium salts, or alkyl esters (eg methyl, ethyl) of the phosphate group.

Basic nitrogen-containing groups may be quarternised with such agents as lower alkyl halide, such as methyl, ethyl, propyl, and butyl chlorides, bromides and iodides; dialkyl sulfates like dimethyl and diethyl sulfate; and others.

The compound of the invention may be in crystalline form either as the free compound or as a solvate (e.g. hydrate) and it is intended that both forms are within the scope of the present invention. Methods of solvation are generally known within the art.

The present disclosure also provides a biosynthetic method of preparing a compound of Formula (I) or a salt, solvate or stereoisomer thereof:



wherein

R₁, R₂, R₃ and R₄ are independently selected from H, oxo, optionally substituted alkyl, optionally substituted alkenyl; and

- 5 R₅, R₆, R₇, R₈ and R₉ are independently selected from H, oxo, optionally substituted alkyl, optionally substituted alkenyl;

the method comprising:

- a) fermenting a cellular organism comprising a sesquiterpene synthetase (STS) enzyme, wherein the STS enzyme is characterised by SEQ 7 and/or SEQ

10 11 .

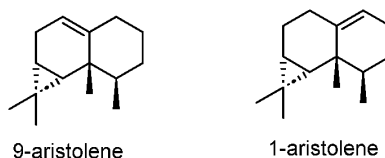
In some embodiments, the method further comprises transfecting or transforming a clonal construct comprising the STS enzyme into the cellular organism. Transfection or transforming is the process of deliberately introducing
 15 naked or purified nucleic acids into cells, the difference being transfection usually refers to eukaryotic cells while transformation refers to bacteria or non-animal eukaryotic cells.

In some embodiments, the clonal construct comprises a vector backbone with
 20 a pET28a and/or p15A origin. In some embodiments, the vector expresses a gene ispA (encoding for E. coli FPP synthase).

In some embodiments, the cellular organism is a single celled organism. In other embodiments, the cellular organism is a bacterial cell. In some embodiments,
 25 the cellular organism is an engineered single celled organism modified to overexpress at least one mevalonate pathway gene. In some embodiments, the cellular organism is E. coli BL21.

In some embodiments, the compound of Formula (I) is characterised by a yield of more than 60 mg/L. In some embodiments, the compound of Formula (I) is characterised by a yield of more than 80 mg/L.

- 5 In some embodiments, the method is characterised by an impurity selected from 9-aristolene and 1-aristolene. In some embodiments, the method is characterised by a 9-aristolene impurity. In some embodiments, the method is characterised by a 1-aristolene impurity.



10

In some embodiments, the method is characterised by a ratio of the compound of Formula (I) to the impurity of about 2:1 to about 6:1. In other embodiments, the ratio is about 2:1 to about 4:1.

- 15 In some embodiments, the method is characterised by a ratio of the compound of Formula (I) to 9-aristolene of about 2:1 to about 6:1. In other embodiments, the ratio is about 2:1 to about 4:1.

Examples

20 Identification and Analysis of *L. deliciosus* STS Sequences.

The genome of *L. deliciosus* was previously sequenced and assembled by Li et al. Sci. Rep. 2018, 8 (1), 9982, the reference of which is incorporated herein. Protein coding genes were predicted using AUGUSTUS18 (version 3.2.2, default parameters with *Coprinus cinereus* as training set). Candidate STS sequences
25 were identified by a HMMER search, with a profile created from an alignment of 393 putative basidiomycete STSs. This yielded 14 complete sequences (Table 1; SEQ 1-14). The amino acid sequences revealed conservation of the metal-binding region (NSE motif and D(D/E/N)xx(D/E) motif).

Table 1. Predicted *L. deliciosus* Sesquiterpene Synthases, Their Cyclization Clades, And Products When Expressed in a Terpene-Overproducing *E. coli* Strain

gene	enzyme	clade	products ^a
g3408	LdSTS6	I	α -muurolene, γ -gurjunene, α -selinene, etc.
g833	LdSTS1	II	1,10-di- <i>epi</i> -cubenol
g1979	LdSTS3	II	myrcene, <i>trans</i> - β -ocimene, etc.
g2294	LdSTS4	III	NI
g2295	LdSTS5	III	NI
g4554	LdSTS8	III	ND
g4592	LdSTS9	III	NI
g5296	LdSTS10	III	ND
g7402	LdSTS13	III	NI
g1333	LdSTS2	V	ND
g4057	LdSTS7	V	new
g5987	LdSTS11	V	new
g6779	LdSTS12	V	ND
g9127	LdSTS14	V	ND

^aNI = not investigated; ND = no product detected. Product identities are determined by comparison of their GC-MS traces against the NIST spectra library.

5

The classification of basidiomycete STSs into several cyclization clades was previously reported. To classify the newly identified STS, a multiple sequence alignment including the new STS and previously characterized basidiomycete STSs was performed with Clustal Omega, and a phylogenetic analysis was

10 conducted with PhyML (Figure 2, Table 1). Of the predicted STSs, one belongs to clade I (α -muurolene, 1,10 cyclase), two belong to clade II (δ -cadinene, 1,10 cyclases), six belong to clade III (putative 1,11 cyclases), and the remaining five belong to an unknown clade V.

15 Determination of STS Terpene Products by GC-MS.

Since *L. deliciosus* is known to produce guaiene-based terpenes, which are products of 1,10 cyclization (Figure 2), we first focused our experiments on the STSs from clades I and II and subsequently on the unknown clade V. The genes for selected STSs were synthesized, cloned into pET28a vectors, and
5 transformed into an engineered *E. coli* BL21 strain overproducing the sesquiterpene precursor farnesyl pyrophosphate (FPP).

In an attempt to identify a *L. deliciosus* guaiene synthase, we first investigated the *L. deliciosus* STS belonging to the δ -cadinene (LdSTS1, LdSTS3) and α -
10 muurolene (LdSTS6) clades (Table 1), which catalyze 1,10-cyclization as expected for guaiene STSs. LdSTS1 produced a single major product (1,10-di-epi-cubenol). LdSTS3 is a highly promiscuous terpene synthase with ~40 terpene products including monoterpenes (myrcene, trans- β -ocimene, etc.) and sesquiterpenes (copaene, β -cubebene, γ -cadinene, etc.). LdSTS6 is also a
15 promiscuous terpene synthase with several sesquiterpene products (α -muurolene, γ -gurjunene, α -selinene, etc.). Contrary to our expectation, none of the three 1,10 cyclases produced a guaiene terpene as its major product under our experimental conditions. We also investigated two of the STSs belonging to the Δ^6 -protoilludene clade (LdSTS8 and LdSTS10), but no products were
20 detected.

Next, we turned our attention to the *L. deliciosus* STSs in the unknown clade V (LdSTS2, LdSTS7, LdSTS11, LdSTS12, LdSTS14, Table 1). No products were detected for LdSTS2, LdSTS12, and LdSTS14. Intriguingly, LdSTS7 produced
25 one major and one minor product, both with identical m/z of 204.2 according to GC-MS analysis (Figure 3). Preliminary analysis suggested that the terpene product could be the guaiene terpene that we were after. LdSTS11 produced a lower yield of the same compounds corresponding to the major product of LdSTS7, based on their identical retention times and mass spectra. Further
30 experiments were conducted to determine the structures of the terpene products of LdSTS7.

Purification of the Unknown LdSTS7 Terpene Product.

To enhance the yield of LdSTS7 terpene products, the vector backbone was changed from pET28a to a p15A origin of replication-based vector that expressed the gene *ispA* (encoding for *E. coli* FPP synthase). This LdSTS7
5 plasmid was transformed into an engineered *E. coli* BL21 strain containing the SPS01 plasmid that overexpresses all the mevalonate pathway genes (Zhang, C. et al.; Nat. Commun. 2018, 9 (1), 1858, the reference of which is incorporated herein) for the large-scale fermentation. At the end of the production process, the extracted terpenoids were analyzed using GC-MS to
10 determine the level of the terpenoid products along with 10 mg/L of caryophyllene standard. The calculated yield of the product mixture was estimated at ~80 mg/L.

The isolated terpene sample contained primarily a mixture of the major 1 and
15 minor product 2 (ca. 2:1 peak area based on GC-TIC, Figure 3A) with trace amounts of product 3. Due to the volatile nature of the compounds, further attempts to purify the components by flash column chromatography or preparative thin-layer chromatography (TLC) were unsuccessful. The GC-MS spectra of both products had $m/z = 204$, which was corroborated by LC-HRMS,
20 suggesting that the products could be structural isomers with very similar physical characteristics. The IR spectrum did not show the presence of any hydroxy groups ($2500\text{--}3600\text{ cm}^{-1}$). Combining elemental analysis (absence of nitrogen group) with the chemical formula ($\text{C}_{15}\text{H}_{24}$), a sesquiterpene hydrocarbon with 4 degrees of unsaturation is proposed.

25

NMR Analysis of the LdSTS7 Terpene Products.

Initial 1D and 2D NMR experiments revealed significant peak overlaps as well as low signal intensity due to limited sample quantity. Thus, selective and semiselective 1D and 2D NMR experiments were employed for *de novo* structure
30 elucidation to resolve and rigorously identify individual signals from each component in the volatile colorless oil mixture while concurrently obtaining accurate J-couplings and ^1H multiplicities. Initial attempts at separating the

compounds from LdSTS7 by flash column chromatography were met with very limited success, and only managed to improve the ratio of compounds 1/2 slightly from about 2:1 (by GC-MS) to about 4:1 (by NMR). The postchromatography outcome (with 4:1 ratio) was used subsequently for NMR structure elucidation.

Briefly, 1D selective TOCSY experiments identified two well-defined spin systems for the major compound 1, leaving two CH₃ singlets and two quaternary carbons based on the chemical formula (Figure 4B). The identities of the spin systems were confirmed by 2D HSQC experiments. The connectivities of the fragments were obtained by 2D semiselective HMBC experiments, revealing the main spiro[4.5]dec-6-ene skeleton bearing a key spiro center, as well as a gem-dimethylcyclopropyl motif (Figure 4C). 1D selective NOESY experiments then identified protons that were close in space (Figure 4D), thus confirming the relative stereochemistry about the 4 stereogenic centers and affording the final structure of compound 1.

To further analyze if this scaffold has been previously reported, a search of the first spin system on Reaxys revealed 74 structurally similar hits such as premnaspirodiene and hinesene that bear the spiro[4.5]dec-6-ene skeleton. However, none of them fulfilled the subsequent structural conditions of (a) possessing the elucidated second spin system and (b) containing no further olefinic groups.

The remaining minor compounds 2 and 3 (Figure 4A) were identified to be 9-aristolene (2) and 1-aristolene (3) based on similar de novo structural elucidation from (semiselective) 1D and 2D NMR experiments, followed by subsequent verification after a structure search (Reaxys) and comparison of ¹H and ¹³C NMR chemical shifts from literature data (see SI, sections S4.3 and S4.4).

Proposed Mechanism for Formation of 1, 2, and 3.

Our phylogenetic analysis suggests that clade V STSs (which include LdSTS7) are closely related to clade III STSs (which produce Δ^6 -protoilludene, a product of 1,11 cyclization) and to the enzymes Agr8 and Agr9 (which produce several sesquiterpenes including γ -muurolene and β -cadinene, products of 1,10 cyclization). To account for the formation of major product 1 and minor products 2 and 3, we hypothesize the mechanistic pathway shown in Figure 5, which draws on vetispiradiene and 5-epi-aristolochene. Cyclization of farnesyl diphosphate (FPP, 4) to form the cyclopropyl moiety could proceed through either 1,10 cyclization to form a germacrenyl carbocation 7, 1,11 cyclization to form a humulyl carbocation 5, or direct formation of the nonclassical carbocation intermediate/transition state 6, involving a threecenter two-electron bond between C1, C10, and C11, as proposed for viridiflorol synthase. Deprotonation of 5/7 gives bicyclogermacrene.

Subsequently, the neutral 8 is reprotonated, followed by Markovnikov-directed 1,6 cyclization to give maaliane carbocation 9. Finally, a 1,2-hydride shift from 9 would lead to common intermediate 10. Conversion of common intermediate 10 to the observed products then follows a bifurcated pathway: 10 can undergo either (a) a Wagner–Meerwein type rearrangement with ring contraction to yield 1 (pathway a) or (b) a 1,2-methyl group migration that results in 2 and 3 upon deprotonation (pathway b). These pathways are consistent with those involved in the conversion of germacrene A to hinesene and selina-4,11-diene. Based on the current results, the presence of the aristolane isomers coupled with the syn-protons observed on the dimethylcyclopropane products (cf. lepidozene) suggests that they might have originated from a common biosynthetic source activated by a bicyclogermacrene synthase similar to those responsible for other structurally related sesquiterpenes.

CONCLUSIONS

In conclusion, our studies of STSs in *L. deliciosus* have led to the identification of a new clade of STSs producing compound 1 with a structurally unique spirobicyclo[3.1.0]terpene sesquiterpene scaffold. In contrast to other known

structurally similar terpene systems, such as bicyclo[3.1.0] in sabinene and spiro[4,5]decane in acoradiene and epi-isozizaene, we believe compound 1 has a distinct mechanistic origin. The involvement of this STS in the biosynthesis of terpene natural products in *L. deliciosus* is to-date unknown, but its discovery highlights the latent biosynthetic capabilities and potential for discovery of new enzymatic activity in basidiomycete fungi. Spirocycles possess a combination of structural complexity and rigidity and are privileged motifs in medicinal chemistry. Recent reports showed terpene synthases may be paired with P450 hydroxylases for combinatorial biosynthesis of functionalized terpene derivatives. Similarly, we envision this new enzyme may add to the rich repertoire of terpene synthases whereupon subsequent oxidation or postmodification reactions on the spirocyclic carbon skeleton and rigid dimethylcyclopropane scaffold could provide a combinatorial biosynthesis platform to assess previously unexplored terpenoid chemical space.

15

MATERIALS AND METHODS

Strain and Plasmid Construction. The genes for selected STSs were synthesized, cloned into pET28a vectors, and transformed into an engineered *E. coli* BL21 strain overproducing the sesquiterpene precursor farnesyl pyrophosphate (FPP) used in a previous study.

20

Heterologous Expression of Terpene Synthase Genes in *E. coli*. The cells were grown in 4 mL of autoinduction media, 1× ZYM (1% tryptone, 0.5% yeast extract, 25 mM Na₂HPO₄, 25 mM KH₂PO₄, 50 mM NH₄Cl, 5 mM Na₂SO₄, 2 mM MgSO₄, 0.5% glycerol, 0.05% glucose, and 15 mM lactose, pH of 7.0), prepared based on Studier's recipe (Studier, F. W. Protein Expr. Purif. 2005, 41 (1), 207–234, the reference of which is incorporated herein). It was supplemented with 100 µg/mL of spectinomycin and 50 µg/mL of kanamycin to maintain the plasmids. The cells were grown at 28 °C for 1 day in SPME vials before the direct analysis of metabolites by GC-MS.

30

Gas ChromatographyMass Spectroscopy (GC-MS) Analysis of Terpenoids.

GC-MS was used for the detection and quantification of terpenes: For the initial screening of LdSTSs, the Agilent GC 7890B was equipped with an Agilent 5977B MSD and DB-5 ms Ultra Inert GC column (30 m × 250 μm × 0.25 μm). The cells were cultivated at 28 °C for 1 day in solid phase microextraction (SPME) 4 mL tubes, which were directly used for GC-MS analysis. The headspace compounds were sampled at 60 °C for 15 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 5 min in the split inlet (250 °C; SPME liner, 0.75 mm i.d.; Supelco). Samples were injected into an Agilent DB5 ms column with a split ratio of 40:1 at 250 °C. The oven program started at 80 °C for 1 min, raised to 210 °C at 10 °C/min, then to 310 °C at 80 °C/min, and maintained at 310 °C for another 1 min. The mass spectrometer was operated in EI mode with scanning ranging from 33 to 300 m/z. The detected peaks were predicted through literature data in the National Institute of Standards and Technology (NIST) database.

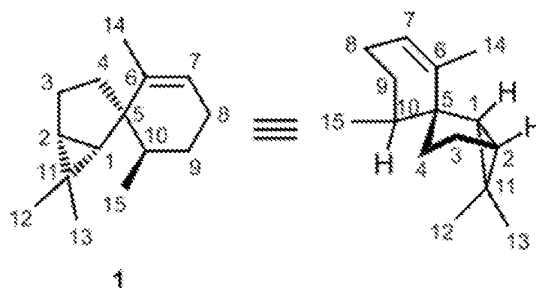
For the large-scale production (400 mL) and product purification, the products of LdSTS7 were determined with an Agilent Intuvo 9000 equipped with an Agilent 5977B MSD. For each extracted fraction, 10 μL of sample was diluted 20 times with 190 μL hexane, then 1 μL was injected onto the Agilent DB-WAX Ultra Inert GC column (30 m × 250 μm × 0.25 μm). The split ratio of 20:1 was set at 250 °C. The oven temperature was initially held at 100 °C for 1 min and increased to 150 °C at a rate of 50 °C/min and further increased to 220 °C at a rate of 15 °C/min. Finally, the temperature was increased to 240 °C at a rate of 50 °C/min and held for 2 min. Mass spectrometer was operated in EI mode with scanning after a 3 min solvent delay (ranging from 50 to 350 m/z). To calculate the relative yield of the terpene products, caryophyllene (10 mg/L) was used as an internal standard. The detected peaks were predicted through literature data in the NIST database.

Nuclear Magnetic Resonance (NMR) Spectroscopy. All spectra were acquired on a Bruker AV Neo 400 MHz spectrometer equipped with a RT 5 mm z-gradient BBFO probe or on a Bruker AVIII 400 MHz spectrometer equipped with a 5 mm z-gradient $^{13}\text{C}/^1\text{H}/^2\text{H}$ DCH observe cryoprobe. All NMR experiments were performed without sample rotation at 298 K in CDCl_3 . The following selective refocusing pulses were used to select desired frequencies through carrier-shift: Gaus1_180r.1000 (80 or 120 ms) for 1D TOCSY, CLIP-COSY and NOESY, and Q3_surbop.1 (815 μs) for 2D band-selective HSQC and HMBC. For clarity, all 1D TOCSY, CLIP-COSY, and NOESY experiments in this paper refer to their selective variants that employ selective refocusing (excitation sculpting via spfge).

Data acquisition was performed using the software Bruker TopSpin 4.0.9 using pulse sequences obtained therein. Initial data processing was performed using Topspin 4.0.9, then further data processing, spectral analysis, and spectra generation were performed using Mestrenova 12.0.2 unless otherwise stated. Baseline correction for 1D and 2D spectra was not used unless otherwise stated. Forward linear prediction on F1 dimension was used for all 2D spectra (except JRES). JRES data was processed with 45° tilt and J-symmetrization with sensitivity enhancement (Mestrenova).

Compound 1

rel-(1S,2S,5R,6'R)-2',6,6,6'-Tetramethylspiro[bicyclo[3.1.0]hexane-2,1'-cyclohexan]-2'-ene



^1H NMR (400.1 MHz, CDCl_3) δ 5.209 (1H, ddquint, $J = 5.1, 2.5, 1.3$ Hz, C7H), 2.040 (1H, dddquint, $J = 17.8, 12.5, 6.0, 2.3$ Hz, C8H_AH_B), 2.012 (1H, dddd, $J = 13.4, 11.1, 8.8, 6.1$ Hz, C3H_AH_B), 1.866 (1H, ddd, $J = 14.1, 10.8, 1.6$ Hz, C4H_AH_B), 1.854 (1H, qt, $J = 7.0, 3.5$ Hz, C10H), 1.809 (1H, dtquint, $J = 17.8, 5.8, 1.5$ Hz, C8H_AH_B), 1.734 (3H, dt, $J = 2.8, 1.4$ Hz, C14H₃), 1.727 (1H, ddd, $J = 13.5, 9.9, 2.0$ Hz, C3H_AH_B), 1.594 (1H, tdd, $J = 12.5, 6.0, 3.4$ Hz, C9H_AH_B), 1.484 (1H, dt, $J = 13.8, 9.3$ Hz, C4H_AH_B), 1.448 (1H, dddt, $J = 12.7, 6.1, 3.4, 1.4$ Hz, C9H_AH_B), 1.192 (1H, t, $J = 6.4$ Hz, C2H), 1.165 (3H, s, C12H₃), 1.018 (3H, s, C13H₃), 1.001 (3H, d, $J = 7.1$ Hz, C15H₃), 0.861 (1H, d, $J = 6.6$ Hz, C1H).

$^{13}\text{C}\{^1\text{H}\}$ NMR (100.6 MHz, CDCl_3) δ 141.35 (C6), 118.35 (C7), 51.43 (C5), 43.40 (C4H₂), 39.08 (C1H), 36.69 (C10H), 32.36 (C2H), 29.01 (C13H₃), 27.02 (C3H₂), 26.88 (C9H₂), 20.81 (C8H₂), 20.66 (C14H₃), 20.15 (C11), 17.19 (C12H₃), 16.46 (C15H₃).

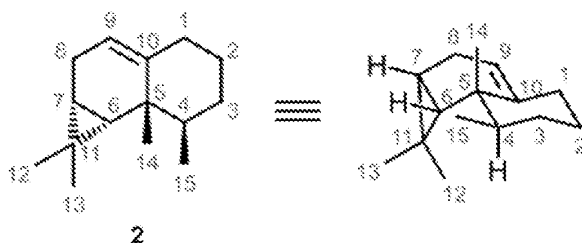
FT-IR (mixture, neat) 3000, 2963, 2867, 1456, 1374, 1203, 1124, 1030, 993, 954, 891, 813, 791, 621 cm^{-1} .

HRMS (mixture, ESI^+) calcd. for $\text{C}_{15}\text{H}_{25}$ m/z (M^+H^+): 205.19508, found: 205.1970.

9-Aristolene (2)

rel-(1aR,7R,7aR,7bS)-1,1,7,7a-Tetramethyl-1a,2,3,5,6,7,7a,7b-octahydro-1H-

cyclopropa[a]-naphthalene



¹H NMR (400.1 MHz, CDCl₃) δ 5.113 (1H, dq, J = 4.4, 2.5 Hz, C9H), 2.328 (1H, dddd, J = 18.8, 7.2, 4.3, 2.9 Hz, C8H_AH_B), 2.202 (1H, dddd, J = 15.7, 13.1, 4.3, 2.0 Hz, C1H_AH_B), 2.064 (1H, ddd, J = 18.8, 4.4, 2.1 Hz, C8H_AH_B), 1.996 (1H, ddt, J = 13.8, 4.2, 2.2 Hz, C1H_AH_B), 1.688 (1H, ddq, J = 11.9, 4.9, 2.5 Hz, C2H_AH_B), 1.627 (1H, dd, J = 11.6, 7.3 Hz, C4H), 1.532 – 1.432 (1H, m, C3H_AH_B), 1.360 (1H, qd, J = 12.6, 3.1 Hz, C3H_AH_B), 1.263 (1H, qt, J = 12.9, 4.0 Hz, C2H_AH_B), 1.105 (3H, s, C12H₃), 1.073 (3H, s, C14H₃), 1.049 (3H, s, C13H₃), 0.963 (3H, d, J = 6.8 Hz, C15H₃), 0.781 (1H, dd, J = 9.4, 7.1 Hz, C7H), 0.629 (1H, d, J = 9.4 Hz, C6H).

10

¹³C{¹H} NMR (100.6 MHz, CDCl₃) δ 141.64 (C10), 118.30 (C9H), 37.79 (C4H), 36.75 (C5), 33.01 (C1H₂), 32.05 (C6H), 31.38 (C3H₂), 29.91 (C12H₃), 27.21 (C2H₂), 21.76 (C8H₂), 21.43 (C14H₃), 19.21 (C7H), 17.93 (C11), 16.00 (C15H₃), 15.76 (C13H₃).

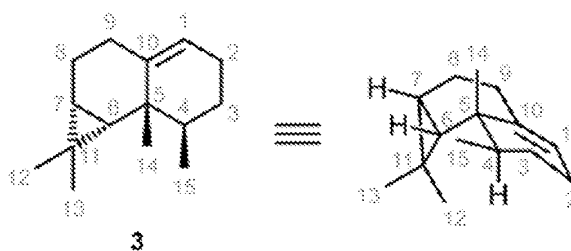
15

The NMR data are consistent with those from literature.

1-Aristolene (3)

rel-(1aR,7R,7aR,7bS)-1,1,7,7a-Tetramethyl-1a,2,4,5,6,7,7a,7b-octahydro-1H-cyclopropa[a]-naphthalene Also known as: Caralene, β-Gurjurenene

20



¹H NMR (400.1 MHz, CDCl₃) δ 5.266 (1H, dt, J = 4.6, 2.3 Hz, C1H), 2.314 – 2.179 (1H, m, C9H_AH_B), 2.035 – 1.917 (2H, m, C2H₂), 1.997 (1H, dddd, J = 14.1, 9.8, 7.1, 1.5 Hz, C8H_AH_B), 1.770 (1H, ddq, J = 10.7, 6.7, 4.4 Hz, C4H), 1.761 (1H, ddd, J = 13.0, 5.8, 1.4 Hz, C9H_AH_B), 1.496 – 1.386 (2H, m, C3H₂), 1.415 (1H, tdd, J = 13.8, 5.8, 3.5 Hz, C8H_AH_B), 1.092 (3H, s, C14H₃), 1.039

25

(3H, s, C12H₃), 1.004 (3H, s, C13H₃), 0.993 (3H, d, J = 6.9 Hz, C15H₃), 0.762 (1H, td, J = 9.6, 3.6 Hz, C7H), 0.581 (1H, d, J = 9.1 Hz, C6H).

¹³C{¹H} NMR (100.6 MHz, CDCl₃) δ 144.18 (C10), 120.32 (C1H), 36.79 (C5),
5 36.69 (C4H), 33.49 (C6H), 29.94 (C9H₂), 29.88 (C12H₃), 27.22 (C3H₂), 25.71 (C2H₂), 22.99 (C14H₃), 20.83 (C8H₂), 19.59 (C7H), 18.52 (C11), 16.50 (C13H₃), 16.10 (C15H₃).

The NMR data are consistent with those from literature. Assignments are based
10 on comparison with literature values.

It will be appreciated that many further modifications and permutations of various aspects of the described embodiments are possible. Accordingly, the described aspects are intended to embrace all such alterations, modifications,
15 and variations that fall within the spirit and scope of the appended claims.

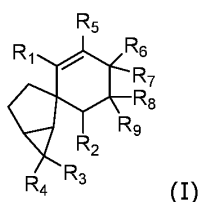
Throughout this specification and the claims which follow, unless the context requires otherwise, the word "comprise", and variations such as "comprises" and "comprising", will be understood to imply the inclusion of a stated integer
20 or step or group of integers or steps but not the exclusion of any other integer or step or group of integers or steps.

Throughout this specification and the claims which follow, unless the context requires otherwise, the phrase "consisting essentially of", and variations such as "consists essentially of" will be understood to indicate that the recited
25 element(s) is/are essential i.e. necessary elements of the invention. The phrase allows for the presence of other non-recited elements which do not materially affect the characteristics of the invention but excludes additional unspecified elements which would affect the basic and novel characteristics of the method
30 defined.

The reference in this specification to any prior publication (or information derived from it), or to any matter which is known, is not, and should not be taken as an acknowledgment or admission or any form of suggestion that that prior publication (or information derived from it) or known matter forms part of
5 the common general knowledge in the field of endeavour to which this specification relates.

Claims

1. A compound of Formula (I) or a salt, solvate or stereoisomer thereof:



- 5 wherein

R₁, R₂, R₃ and R₄ are independently selected from H, oxo, optionally substituted alkyl, optionally substituted alkenyl; and

R₅, R₆, R₇, R₈ and R₉ are independently selected from H, oxo, optionally substituted alkyl, optionally substituted alkenyl.

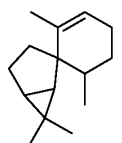
10

2. The compound according to claim 1, wherein R₁, R₂, R₃ and R₄ are independently optionally substituted C₁-C₄ alkyl.

3. The compound according to claim 1 or 2, wherein R₅, R₆, R₇, R₈ and R₉

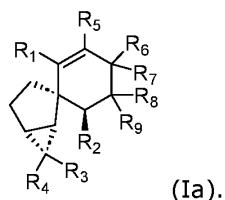
15 are independently selected from H, or optionally substituted C₁-C₄ alkyl.

4. The compound according to any one of claims 1 to 3, wherein the compound of Formula (I) is

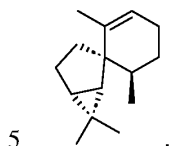


20

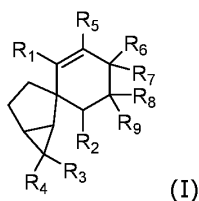
5. The compound according to any one of claims 1 to 3, wherein the compound of Formula (I) is a compound of Formula (Ia):



6. The compound according to any one of claims 1 to 5, wherein the compound of Formula (I) is



7. A biosynthetic method of preparing a compound of Formula (I) or a salt, solvate or stereoisomer thereof:



10 wherein

R₁, R₂, R₃ and R₄ are independently selected from H, oxo, optionally substituted alkyl, optionally substituted alkenyl; and

R₅, R₆, R₇, R₈ and R₉ are independently selected from H, oxo, optionally substituted alkyl, optionally substituted alkenyl;

15 the method comprising:

a) fermenting a cellular organism having a sesquiterpene synthetase (STS) enzyme, wherein the STS enzyme is characterised by SEQ 7 and/or SEQ 11 .

8. The biosynthetic method according to claim 7, wherein the method
20 further comprises transfecting a clonal construct comprising the STS enzyme into the cellular organism.

9. The biosynthetic method according to claim 8, wherein the clonal construct comprises a vector backbone with a pET28a and/or p15A origin.
- 5 10. The biosynthetic method according to any one of claims 7 to 9, wherein the cellular organism is a single celled organism.
11. The biosynthetic method according to any one of claims 7 to 10, wherein the cellular organism is an engineered single celled organism modified to
10 overexpress at least one mevalonate pathway gene.
12. The biosynthetic method according to any one of claims 7 to 11, wherein the cellular organism is *E. coli* BL21.
- 15 13. The biosynthetic method according to any one of claims 7 to 12, wherein the compound of Formula (I) is characterised by a yield of more than 60 mg/L.
14. The biosynthetic method according to any one of claims 7 to 13, wherein the compound of Formula (I) is characterised by a yield of more than 80 mg/L.
20
15. The biosynthetic method according to any one of claims 7 to 14, wherein the method is characterised by an impurity selected from 9-aristolene and 1-aristolene.
- 25 16. The biosynthetic method according to claim 15, wherein the method is characterised by a ratio of the compound of Formula (I) to the impurity of about 2:1 to about 6:1.