(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
22 August 2019 (22.08.2019)

WIPO | PCT

(10) International Publication Number
WO 2019/161141 A1

(51) International Patent Classification:
C12P 5/00 (2006.01) *C12P 1/04* (2006.01)
C12P 1/02 (2006.01) *C11D 3/50* (2006.01)

(21) International Application Number:
PCT/US2019/018122

(22) International Filing Date:
14 February 2019 (14.02.2019)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
62/630,640 14 February 2018 (14.02.2018) US

(71) Applicant: **GINKGO BIOWORKS** [US/US]; 27 Drydock Avenue, 8th Floor, Boston, MA 02210 (US).

(72) Inventors: **RIDLEY, Christian**; 27 Drydock Avenue, 8th Floor, Boston, MA 02210 (US). **WANG, Jue**; 5267 11th Avenue NE Apt. C, Seattle, WA 98105 (US). **MARR, Scott**; 27 Drydock Avenue, 8th Floor, Boston, MA 02210 (US).

(74) Agent: **JOHNSTONE, Oona, M.**; Wolf, Greenfield & Sacks, P.C., 600 Atlantic Avenue, Boston, MA 02210-2206 (US).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

Published:
— *with international search report (Art. 21(3))*
— *before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments (Rule 48.2(h))*
— *with sequence listing part of description (Rule 5.2(a))*

(54) Title: CHIMERIC TERPENE SYNTHASES

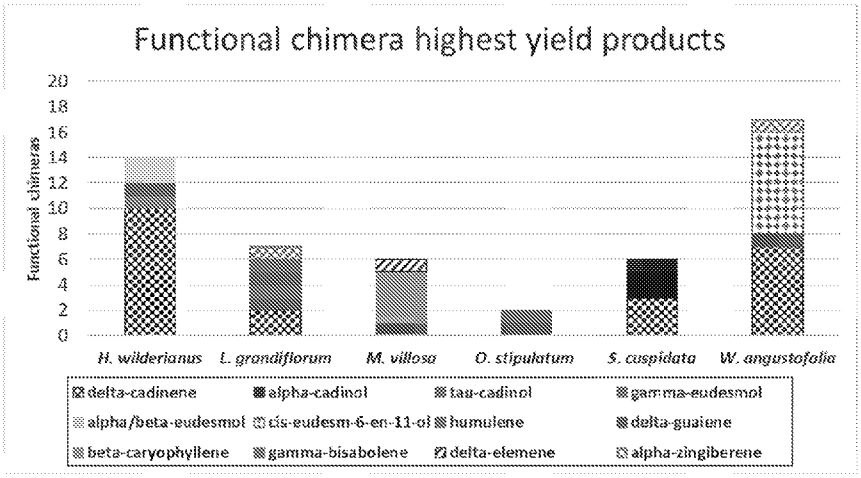


FIG. 7

(57) Abstract: Described herein are chimeric terpene synthases, methods for making chimeric terpene synthases, and methods for making terpenes using the same.

WO 2019/161141 A1

CHIMERIC TERPENE SYNTHASES

RELATED APPLICATIONS

This application claims the benefit under 35 U.S.C. § 119(e) of U.S. Provisional Application Serial No. 62/630,640, entitled "CHIMERIC TERPENE SYNTHASES" filed on February 14, 2018, which is herein incorporated by reference in its entirety.

FIELD OF THE INVENTION

The disclosure relates to chimeric terpene synthases, methods for making chimeric terpene synthases, and methods for making terpenes using the same.

BACKGROUND

Terpenes are a diverse class of organic compounds built from five carbon building blocks and encompass at least 400 distinct structural families. Given their structural diversity, terpenes have numerous roles including acting as pheromones, anti-oxidants, and anti-microbial agents. Although terpene synthases produce terpenes in both prokaryotes and eukaryotes, the wide array of terpene isomers often hinder high yield extractions from naturally occurring sources. Furthermore, the structural complexity of terpenes often limits *de novo* chemical synthesis.

SUMMARY

Aspects of the disclosure relate to chimeric terpene synthases comprising an amino acid sequence at least 90% identical to an amino acid selected from the group consisting of: SEQ ID NOs: 1-52. In some embodiments, the chimeric terpene synthase comprises an amino acid sequence at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% identical to an amino acid selected from the group consisting of: SEQ ID NOs: 1-52. In some embodiments, the chimeric terpene synthase comprises an amino acid sequence identical to an amino acid selected from the group consisting of: SEQ ID NOs: 1-52.

Further aspects of the disclosure relate to nucleic acid molecules encoding a chimeric terpene synthase described herein. In some embodiments, a nucleic acid molecule comprises a sequence that is at least 90% identical to a nucleotide sequence selected from the group

consisting of SEQ ID NOs: 67-118. In some embodiments, a nucleic acid molecule comprises a sequence selected from the group consisting of SEQ ID NOs: 67-118.

Further aspects of the disclosure relate to vectors comprising a nucleic acid molecule described herein. In some embodiments, the vector is a viral vector, a vector for transient expression, or a vector for inducible expression. In some embodiments, the vector is a lentiviral vector, a retroviral vector, an adenoviral vector, an adeno-associated vector, a galactose-inducible vector, or a doxycycline-inducible vector.

Further aspects of the disclosure relate to host cells comprising a nucleic acid described herein, or a vector described herein.

In some embodiments, the host cell is a fungal cell. In some embodiments, the cell is a yeast cell. In some embodiments, the cell is a *Saccharomyces, Pichia, Kluyveromyces, Hansenula*, or *Yarrowia* cell. In some embodiments, the cell is a *Saccharomyces cerevisiae* cell.

In some embodiments, the host cell is a plant cell.

In some embodiments, the host cell is a bacteria cell.

Further aspects of the disclosure relate to nucleic acid molecules encoding a chimeric terpene synthase, wherein at least 10% of the nucleic acid molecule sequence, or the amino acid sequence, is derived from a rare or extinct plant. In some embodiments, at least 40% of the nucleic acid molecule sequence, or the amino acid sequence, is derived from a rare or extinct plant.

In some embodiments, at least 20%, at least 30%, at least 40%, at least 50%, at least 60%, at least 70%, at least 80%, or at least 90% of the nucleic acid molecule sequence, or the amino acid sequence, is derived from a rare or extinct plant. In some embodiments, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, or at least 95% of the nucleic acid molecule sequence, or the amino acid sequence, is derived from a rare or extinct plant.

In some embodiments, the chimeric terpene synthase is a chimeric sesquiterpene synthase. In some embodiments, the rare or extinct plant is selected from the group consisting of: *Hibiscadelphus wilderianus*, *Leucadendron grandiflorum*, *Macrostylis villosa*, *Orbexilum stipulatum*, *Shorea cuspidate*, and *Wendlandia angustifolia*.

Further aspects of the disclosure relate to nucleic acid molecules encoding a chimeric terpene synthase. In some embodiments, at least 10% of the nucleic acid molecule sequence, or the amino acid sequence is derived from a rare or extinct plant. In some embodiments, at least

20%, at least 30%, at least 40%, at least 50%, at least 60%, at least 70%, at least 80%, or at least 90% of the nucleic acid molecule sequence is derived from a rare or extinct plant. In some embodiments, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, or at least 95% of the nucleic acid molecule sequence is derived from a rare or extinct plant.

In some embodiments, the nucleic acid molecule further comprises a TATA box sequence.

Further aspects of the disclosure relate to methods of producing one or more sesquiterpenes, wherein the method comprises culturing a host cell described herein under conditions suitable for producing the one or more sesquiterpenes.

Further aspects of the disclosure relate to compositions comprising one or more sesquiterpenes produced by the methods described herein.

In one embodiment, at least one of the one or more sesquiterpenes is an aroma compound.

Further aspects of the disclosure relate to methods of producing a perfume, wherein the method comprises: culturing a host cell described herein under conditions suitable for producing the one or more sesquiterpenes; and extracting the one or more sesquiterpenes.

Each of the limitations of the compositions and methods described herein may encompass various described embodiments. It is, therefore, anticipated that each of the limitations of the invention involving any one element or combinations of elements can be included in each aspect of the invention. This present disclosure is not limited in its application to the details of construction and the arrangement of components set forth in the following description or illustrated in the drawings.

BRIEF DESCRIPTION OF DRAWINGS

The accompanying drawings are not intended to be drawn to scale. The drawings are illustrative only and are not required for enablement of the disclosure. For purposes of clarity, not every component may be labeled in every drawing. In the drawings:

FIG. 1 is a series of pictures depicting structures of identified sesquiterpenes produced using sesquiterpene synthases (SQTSs) containing rare sequences from *H. wilderianus*.

FIG. 2 is a series of pictures depicting structures of identified sesquiterpenes produced using SQTSs containing rare sequences from *L. grandiflorum*.

FIG. 3 is a series of pictures depicting structures of sesquiterpenes produced using SQTSs containing rare sequences from *M. villosa*.

FIG. 4 is a series of pictures depicting structures of sesquiterpenes produced using SQTSs containing rare sequences from *O. stipulatum*.

FIG. 5 is a series of pictures depicting structures of identified sesquiterpenes produced using SQTSs containing rare sequences from *S. cuspidata*.

FIG. 6 is a series of pictures depicting structures of identified sesquiterpenes produced using SQTSs containing rare sequences from *W. angustifolia*.

FIG. 7 is a graph showing chimera product distribution versus plant species. The chimeras are categorized based on the sesquiterpene produced in highest yield.

FIGS. 8A-8F include a series of pictures depicting species of rare plants. FIG. 8A depicts *Hibiscadelphus wilderianus* (from Radlkofer *et al.*, New and Noteworthy Hawaiian Plants. Hawaiian Board of Agriculture and Forestry Botanical Bulletin. 1911;(1):1-15). FIG. 8B depicts *Leucadendron grandiflorum* (from Salisbury *et al.*, The Paradisus Londinensis or Coloured Figures of Plants Cultivated in the Vicinity of the Metropolis. 1805; (Volume 1, part 2): 105). FIG. 8C depicts *Macrostylis villosa* subsp. Villosa (from "Red List of South African Plants: Macrostylis villosa subsp. villosa," 2007). FIG. 8D depicts *Orbexilum stipulatum* (from Short, "Orbexilum stipulatum collected at Falls of the Ohio," 1840 from The Philadelphia Herbarium at the Academy of Natural Sciences). FIG. 8E depicts *Shorea cuspidata* (from "Kew Royal Botanical Gardens: Shorea cuspidata specimen K000700460," 1962). FIG. 8F depicts *Wendlandia angustifolia* (from "Kew Royal Botanical Gardens: Wendlandia angustifolia K000030921," collection date not recorded).

FIG. 9 is a series of pictures depicting selected gas chromatography–mass spectrometry (GC/MS) chromatograms from *H. wilderianus* chimera screening data (Table 4).

FIG. 10 is a series of pictures depicting selected GC/MS chromatograms from *L. grandiflorum* chimera screening data (Table 5).

FIG. 11 is a series of pictures depicting selected GC/MS chromatograms from *L. grandiflorum* chimera screening data (Table 5).

FIG. 12 is a series of pictures depicting selected GC/MS chromatograms from *M. villosa* chimera screening data (Table 6).

FIG. 13 is a series of pictures depicting selected GC/MS chromatograms from *S. cuspidata* chimera screening data (Table 8).

FIG. 14 is a series of pictures depicting selected GC/MS chromatograms from *W. angustifolia* chimera screening data (Table 9).

FIG. 15 is a series of pictures depicting selected GC/MS chromatograms from *W. angustifolia* chimera screening data (Table 9).

DETAILED DESCRIPTION

Although terpenes are widely used in the fragrance industry, purification of terpenes from natural sources and *de novo* chemical synthesis often have high production costs and low yield. This disclosure is premised, in part, on the unexpected finding that chimeric terpene synthases comprising a portion of a terpene synthase sequence from at least one rare or extinct plant can be leveraged to produce a diversity of sesquiterpenes. Accordingly, provided herein are chimeric terpene synthases, methods for making chimeric terpene synthases, and methods for making terpenes using the described chimeric terpene synthases. In some embodiments, the chimeric terpene synthases are chimeric sesquiterpene synthases.

This invention is not limited in its application to the details of construction and the arrangement of components set forth in the description. The invention is capable of other embodiments and of being practiced or of being carried out in various ways. Additionally, the phraseology and terminology used herein is for the purpose of description and should not be regarded as limiting. The use of terms such as "including," "comprising," "having," "containing," "involving," and/or variations thereof herein, is meant to encompass the items listed thereafter and equivalents thereof as well as additional items.

Chimeric Terpene Synthases

Aspects of the present disclosure relate to chimeric terpene synthases comprising fragments (*e.g.*, sequences) from at least two terpene synthases, wherein at least one of the two or more terpene synthases is from a rare or extinct plant. For example, the sequence of a chimeric terpene synthase may comprise one or more fragments (*e.g.*, one or more portions of the total sequence) from at least two, at least three, at least four, at least five, at least six, at least seven, at least eight, at least nine or at least ten terpene synthases. It should be appreciated that

chimeric terpene synthases described herein can be synthetic. Accordingly, chimeric terpene synthases, including synthetic chimeric terpene synthases, described herein comprise sequences derived from more than one terpene synthase, wherein at least one of the terpene synthases is from a rare or extinct plant. In some embodiments, the chimeric terpene synthases are chimeric sesquiterpene synthases.

Terpene synthases are enzymes that catalyze the formation of terpenes from isoprenoid diphosphate substrates. At least two types of terpene synthases have been characterized: classic terpene synthases and isoprenyl diphosphate synthase-type terpene synthases. Classic terpene synthases are found in prokaryotes (*e.g.*, bacteria) and in eukaryotes (*e.g.*, plants, fungi and amoebae), while isoprenyl diphosphate synthase-type terpene synthases have been found in insects (see, *e.g.*, Chen *et al.*, Terpene synthase genes in eukaryotes beyond plants and fungi: Occurrence in social amoebae. *Proc Natl Acad Sci U S A.* 2016;113(43):12132-12137, which is hereby incorporated by reference in its entirety for this purpose). Several highly conserved structural motifs have been reported in classic terpene synthases, including an aspartate-rich "DDxx(x)D/E" motif and a "NDxxSxxxD/E"(SEQ ID NO: 55) motif, which have both been implicated in coordinating substrate binding (see, *e.g.*, Starks *et al.*, Structural basis for cyclic terpene biosynthesis by tobacco 5-epi-aristolochene synthase. *Science.* 1997 Sep 19;277(5333):1815-20; and Christianson *et al.*, Unearthing the roots of the terpenome. *Curr Opin Chem Biol.* 2008 Apr;12(2):141-50, each of which is hereby incorporated by reference in its entirety for this purpose).

Terpene synthases may be classified by the type of terpenes they produce. As used herein, unless otherwise indicated, terpenes are organic compounds comprising isoprene (*i.e.*, C_5H_8) units and derivatives thereof. For example, terpenes include pure hydrocarbons with the molecular formula $(C_5H_8)_n$, in which n represents the number of isoprene subunits. Terpenes also include oxygenated compounds (often referred to as terpenoids). Terpenes are structurally diverse compounds and, for example, may be cyclic (*e.g.*, monocyclic, multi-cyclic, homocyclic and heterocyclic compounds) or acyclic (*e.g.*, linear and branched compounds). In some embodiments, a terpene may have an odor. As used herein, an aroma compound refers to a compound that has an odor. Any methods known in the art, including mass spectrometry (*e.g.*, gas chromatography-mass spectrometry (GC/MS, shown in Example 2 below), may be used to identify a terpene of interest.

Terpene synthases may include, for example, monoterpene synthases, diterpene synthases, and sesquiterpene synthases. Certain non-limiting examples of monoterpene synthases and sesquiterpene synthases may be found, for example, in Degenhardt *et al.*, Monoterpene and sesquiterpene synthases and the origin of terpene skeletal diversity in plants. *Phytochemistry*. 2009 Oct-Nov;70(15-16):1621-37, which is hereby incorporated by reference in its entirety for this purpose.

Monoterpene synthases catalyze the formation of 10-carbon monoterpenes. Generally, monoterpene synthases use geranyl diphosphate (GPP) as a substrate. Non-limiting examples of monoterpene synthases include Myrcene synthase (UniProtKb Identifier: O24474), (R)-limonene synthase (UniprotKB Identifier: Q2XSC6), (E)-beta-ocimene synthase (UniProtKB Identifier: Q5CD81) and Limonene synthase (UniProtKB Identifier: Q9FV72). Non-limiting examples of monoterpenes include, but are not limited to, limonene, sabinene, thujene, carene, borneol, eucalyptol and camphene.

Diterpene synthases promote the formation of 20-carbon diterpenes. Generally, diterpene synthases use geranylgeranyl diphosphate as a substrate. Non-limiting examples of diterpene synthases include cis-abienol synthase (UniProtKB identifier: H8ZM73), sclareol synthase (UniProtKB identifier: K4HYB0) and abietadiene synthase (Q38710). See, *e.g.*, Gong *et al.*, Diterpene synthases and their responsible cyclic natural products. *Nat Prod Bioprospect*. 2014;4(2):59-72, which is hereby incorporated by reference in its entirety for this purpose. Non-limiting examples of diterpenes include, but are not limited to, cembrene and sclareol.

Sesquiterpene synthases catalyze the formation of 15-carbon sesquiterpenes. Generally, sesquiterpene synthases convert farnesyl diphosphate (FDP) into sesquiterpenes. Non-limiting examples of sesquiterpene synthases include (+)-delta-cadinene synthase (UniProtKB Identifier: Q9SAN0), UniProtKB Identifier: A0A067FTE8, Beta-eudesmol synthase (UniProtKB Identifier: B1B1U4), (+)-delta-cadinene synthase isozyme XC14 (UniProtKB Identifier: Q39760), (+)-delta-cadinene synthase isozyme XC1 (UniProtKB Identifier: Q39761), (+)-delta-cadinene synthase isozyme A (UniProtKB Identifier: Q43714), Sesquiterpene synthase 2 (UniProtKB Identifier: Q9FQ26), Putative delta-guaiene synthase (UniProtKB Identifier: A0A0A0QUT9), Delta-guaiene synthase 1 (UniProtKB Identifier: D0VMR6), Alpha-zingiberene synthase (UniProtKB Identifier: Q5SBP4), (Z)-gamma-bisabolene synthase 1 (UniProtKB Identifier: Q9T0J9), A0A067D5M4, Delta-elemene synthase (UniProtKB Identifier: A0A097ZIE0),

ShoBecSQTS1, A0A068UHT0, terpene synthase (UniProtKB Identifier: G5CV47), A0A068VE40 and A0A068VI46.

In some embodiments, a sesquiterpene synthase is an alpha-guaiene synthase. As used herein, an alpha-guaiene synthase is capable of catalyzing the formation of alpha-guaiene. In some embodiments, an alpha-guaiene synthase uses (2E,6E)-farnesyl diphosphate as a substrate. Non-limiting examples of alpha-guaiene synthases include UniProtKB Identifier: D0VMR6, UniProtKB Identifier: D0VMR7, UniProtKB Identifier: D0VMR8, UniProtKB Identifier: Q49SP3. As disclosed herein, an alpha-guaiene synthase may comprise a sequence that is at least 50% (*e.g.*, at least 55%, 60%, 65%, 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more than 99%, including all values in between) identical to SEQ ID NO: 17, 22, or 29. In certain embodiments, an alpha-guaiene synthase comprises SEQ ID NO: 17, 22, or 29. In certain embodiments an alpha-guaiene synthase consists of SEQ ID NO: 17, 22, or 29.

As used herein, unless otherwise indicated, sesquiterpenes include sesquiterpene hydrocarbons and sesquiterpene alcohols (sesquiterpenols). Non-limiting examples of sesquiterpenes include but are not limited to, delta-cadinene, epi-cubenol, tau-cadinol, alpha-cadinol, gamma-selinene, 10-epi-gamma-eudesmol, gamma-eudesmol, alpha/beta-eudesmol, juniper camphor, 7-epi-alpha-eudesmol, cryptomeridiol isomer 1, cryptomeridiol isomer 2, cryptomeridiol isomer 3, humulene, alpha-guaiene, delta-guaiene, zingiberene, beta-bisabolene, beta-farnesene, beta-sesquiphellandrene, cubenol, alpha-bisabolol, alpha-curcumene, trans-nerolidol, gamma, bisabolene, beta-caryophyllene, trans-Sesquisabinene hydrate, delta-elemene, cis-eudesm-6-en-11-ol, daucene, isodaucene, trans-bergamotene, alpha-zingiberene, sesquisabinene hydrate, and 8-Isopropenyl-1,5-dimethyl-1,5-cyclodecadiene.

The present disclosure also encompasses chimeric terpene synthases that are multi-functional (*e.g.*, capable of producing more than one sesquiterpene). In some embodiments, a chimeric terpene synthase is capable of producing delta-cadinene and alpha-cadinol. In some embodiments, a chimeric terpene synthase is capable of producing delta-cadinene, tau-cadinol, and alpha-cadinol. In some embodiments, a chimeric terpene synthase is capable of producing alpha-guaiene and delta-guaiene. In some embodiments, the chimeric terpene synthase is capable of producing beta-caryophyllene and humulene.

In some embodiments, a chimeric terpene synthase (*e.g.*, a chimeric sesquiterpene synthase) of the present disclosure comprises an amino sequence at least 50% (*e.g.*, at least 55%, 60%, 65%, 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more than 99%, including all values in between) identical to a sequence selected from the group consisting of SEQ ID NOs: 1-52. In some embodiments, the chimeric terpene synthase comprises an amino acid sequence provided in SEQ ID NOs: 1-52.

In some embodiments, a chimeric terpene synthase comprises one or more sequences provided in SEQ ID NOs: 119-357.

The term "sequence identity," as known in the art, refers to a relationship between the sequences of two polypeptides or polynucleotides, as determined by sequence comparison (alignment). In the art, identity also means the degree of sequence relatedness between two sequences as determined by the number of matches between strings of two or more residues (*e.g.*, nucleic acid or amino acid residues). Identity measures the percent of identical matches between the smaller of two or more sequences with gap alignments (if any) addressed by a particular mathematical model or computer program (*e.g.*, "algorithms").

Identity of related polypeptides can be readily calculated by any of the methods known to one of ordinary skill in the art. The "percent identity" of two sequences (*e.g.*, nucleic acid or amino acid sequences) may, for example, be determined using the algorithm of Karlin and Altschul *Proc. Natl. Acad. Sci. USA* 87:2264-68, 1990, modified as in Karlin and Altschul *Proc. Natl. Acad. Sci. USA* 90:5873-77, 1993. Such an algorithm is incorporated into the NBLAST® and XBLAST® programs (version 2.0) of Altschul et al., *J. Mol. Biol.* 215:403-10, 1990. BLAST® protein searches can be performed, for example, with the XBLAST program, score=50, wordlength=3 to obtain amino acid sequences homologous to the protein molecules of the invention. Where gaps exist between two sequences, Gapped BLAST® can be utilized, for example, as described in Altschul *et al.*, *Nucleic Acids Res.* 25(17):3389-3402, 1997. When utilizing BLAST® and Gapped BLAST® programs, the default parameters of the respective programs (*e.g.*, XBLAST® and NBLAST®) can be used, or the parameters can be adjusted appropriately as would be understood by one of ordinary skill in the art.

Another local alignment technique which may be used, for example, is based on the Smith-Waterman algorithm (Smith, T.F. & Waterman, M.S. (1981) "Identification of common

molecular subsequences." *J. Mol. Biol.* 147:195-197). A general global alignment technique which may be used, for example, is the Needleman–Wunsch algorithm (Needleman, S.B. & Wunsch, C.D. (1970) "A general method applicable to the search for similarities in the amino acid sequences of two proteins." *J. Mol. Biol.* 48:443-453), which is based on dynamic programming. More recently, a Fast Optimal Global Sequence Alignment Algorithm (FOGSAA) was developed that purportedly produces global alignment of nucleic acid and amino acid sequences faster than other optimal global alignment methods, including the Needleman–Wunsch algorithm.

The present disclosure also encompasses compositions comprising one or more terpenes (*e.g.*, sesquiterpenes) produced by any one of the chimeric terpene synthases (*e.g.*, sesquiterpene synthases) described herein. In some embodiments, the composition comprises at least one terpene (*e.g.*, sesquiterpene) that is an aroma compound. In some embodiments, the composition is a perfume (*e.g.*, comprising a single fragrance or a mixture of fragrances). In some embodiments, the composition further comprises a fixative (*i.e.*, stabilizer) to reduce volatility of the composition. Non-limiting examples include fixatives include resinoids (*e.g.*, benzoin, olibanum, storax, labdanum, myrrh and tolu balsam) and benzyl benzoate. In some embodiments, the composition further comprises ethyl alcohol. In some embodiments, the composition further comprises distilled water.

In certain embodiments, a terpene synthase (*e.g.*, sesquiterpene synthase) of the present disclosure produces a terpene (*e.g.*, sesquiterpene) composition that comprises at least 1%, at least 2%, at least 3%, at least 4%, at least 5%, at least 6%, at least 7%, at least 8%, at least 9%, at least 10%, at least 15%, at least 20%, at least 30%, at least 40%, at least 50%, at least 60%, at 70%, at least 80%, at least 90%, at least 95%, or 100% including any values in between of a particular terpene, such as a sesquiterpene. Non-limiting examples of sesquiterpenes include delta-cadinene, epi-cubenol, tau-cadinol, alpha-cadinol, gamma-selinene, 10-epi-gamma-eudesmol, gamma-eudesmol, alpha/beta-eudesmol, juniper camphor, 7-epi-alpha-eudesmol, cryptomeridiol isomer 1, cryptomeridiol isomer 2, cryptomeridiol isomer 3, humulene, alpha-guaiene, delta-guaiene, zingiberene, beta-bisabolene, beta-farnesene, beta-sesquiphellandrene, cubenol, alpha-bisabolol, alpha-curcumene, trans-nerolidol, gamma, bisabolene, beta-caryophyllene, trans-Sesquisabinene hydrate, delta-elemene, cis-eudesm-6-en-11-ol, daucene, isodaucene, trans-bergamotene, alpha-zingiberene, sesquisabinene hydrate, and 8-Isopropenyl-

1,5-dimethyl-1,5-cyclodecadiene. As a non-limiting example, a terpene synthase may be heterologously expressed in a host cell, the sesquiterpenes produced by the recombinant host cell may be extracted, and the types of sesquiterpenes in the composition may be determined using gas chromatography–mass spectrometry. In some embodiments, a terpene synthase may be recombinantly expressed and is purified. In some embodiments, the sesquiterpenes produced by a purified terpene synthase may be extracted and the types of sesquiterpenes in the composition may be determined using gas chromatography–mass spectrometry.

In certain embodiments, an alpha-guaiene synthase is capable of producing a sesquiterpene composition that comprises at least 1%, at least 2%, at least 3%, at least 4%, at least 5%, at least 6%, at least 7%, at least 8%, at least 9%, at least 10%, at least 15%, at least 20%, at least 30%, at least 40%, at least 50%, at least 60%, at 70%, at least 80%, at least 90%, at least 95%, or 100% including any values in between of alpha-guaiene. In some embodiments, an alpha-guaiene synthase is capable of producing a sesquiterpene composition that comprises between 1% to 10%, between 5% to 20%, between 15% to 20%, between 16% and 20%, between 17% and 20%, between 18% and 20%, between 19% and 20%, between 20% and 25%, between 20% and 24%, between 20% and 23%, between 20% and 22%, between 20% and 21%, between 20% and 30%, between 30% and 40%, between 40% and 50%, between 50% and 60%, between 60% and 70%, between 70% and 80% ,between 80% and 90%, or between 90% and 100%, including any values in between alpha-guaiene.

Rare and extinct plants

At least one portion of the sequence of the chimeric terpene synthases disclosed herein is derived from a rare or extinct plant. As used herein, the term “rare plant” or “rare plants” encompasses plants that are uncommon, scarce, infrequently encountered, endangered (*e.g.*, threatened), vulnerable, only available in private collections, not found in the endemic location, only available in cultivation, and/or extinct. In some embodiments, a rare plant is a plant that is infrequently encountered (*e.g.*, only encountered in a few locations such as 1, 2, 3, 4, or 5 locations). In some embodiments, a rare plant is an extinct plant. As used herein, an extinct plant refers to a species of plant: having no living members; classified as having no living members; or predicted by one of ordinary skill in the art to have no living members. As a non-limiting example, the International Union for Conservation of Nature (IUCN) Red list of

Threatened Species may be used to determine the conservation status of a plant and identify rare plants. For example, plants classified as extinct, extinct in the wild, critically endangered, endangered, vulnerable, and near threatened on the IUCN Red List may be considered rare plants.

Non-limiting examples of rare plants include *Leucadendron grandiflorum, Shorea cuspidata, Macrostylis villosa, Orbexilum stipulatum, Myrcia skeldingii, Nesiota Elliptica, Macrostylis villosa, Wendlandia angustofola, Erica Pyramidalis, Stenocarpus dumbeenis, Pradosia glaziovii, Crassula subulata, Hibiscadelphus wilderianus,* and *Erica foliacea.*

In some embodiments, the rare plant may be *Hibiscadelphus wilderianus.* The *Hibiscadelphus* genus belongs to the tribe Hibisceae (Malvaceae) and members of the genus often have petals that form a tubular structure in which the lower petals are often shorter than the upper three petals (see, *e.g.*, Oppenheimer *et al.*, A new species of Hibiscadelphus Rock (Malvaceae, Hibisceae) from Maui, Hawaiian Islands; *PhytoKeys*, 2014;(39):65-75, which is hereby incorporated by reference in its entirety). The *Hibiscadelphus* genus is endemic to Hawaii and at least eight species have been described. Four of these species are extinct (including *Hibiscadelphus bombycinus*, *Hibiscadelphus crucibracteatus*, *Hibiscadelphus wilderianus,* and *Hibiscadelphus woodii*), two of these species only persist in cultivation (*Hibiscadelphus giffardianus* and *Hibiscadelphus hualalaiensis*), and two are extant in the wild (*Hibiscadelphus distans* and *Hibiscadelphus stellatus*).

Hibiscadelphus wilderianus is an extinct tree species last observed at an elevation of 2,600 feet in 1910 on the lava fields of Auwahi on the island of Maui in Hawaii (see, *e.g.*, Radlkofer *et al.*, New and Noteworthy Hawaiian Plants; Hawaiian Board of Agriculture and Forestry Botanical Bulletin, 1911;(1):1-15; "The IUCN Red List of Threatened Species: Hibiscadelphus wilderianus," World Conservation Monitoring Centre, 1998, each of which is hereby incorporated by reference in its entirety). A description in Latin of *Hibiscadelphus wilderianus* can be found in the Radlkofer *et al.* original report. A photo of a tree branch with leaves and fruit was included in the original Radlkofer *et al.* report and is reproduced in FIG. 8A.

In some embodiments, the rare plant may be *Leucadendron grandiflorum. Leucadendron* is a dioecious genus that belongs to the Proteaceae family and is endemic to South Africa. Species in the *Leucadendron* genus include evergreen shrubs and often have cone-shaped infructescences (seed heads). There are at least 80 species in the *Leucadendron* genus including

L. album, *L. arcuatum*, *L. argenteum*, *L. barkerae*, *L. bonum*, *L. brunioides*, *L. burchellii*, *L. cadens*, *L. chamelaea*, *L. cinereum*, *L. comosum*, *L. concavum*, *L. conicum*, *L. coniferum*, *L. cordatum*, *L. coriaceum*, *L. corymbosum*, *L. cryptocephalum*, *L. daphnoides*, *L. diemontianum*, *L. discolor*, *L. dregei*, *L. dubium*, *L. elimense*, *L. ericifolium*, *L. eucalyptifolium*, *L. flexuosum*, *L. floridum*, *L. foedum*, *L. galpinii*, *L. gandogeri*, *L. glaberrimum*, *L. globosum*, *L. grandiflorum*, *L. gydoense*, *L. immoderatum*, *L. lanigerum*, *L. laureolum*, *L. laxum*, *L. levisanus*, *L. linifolium*, *L. loeriense*, *L. loranthifolium*, *L. macowanii*, *L. meridianum*, *L. meyerianum*, *L. microcephalum*, *L. modestum*, *L. muirii*, *L. nervosum*, *L. nitidum*, *L. nobile*, *L. olens*, *L. orientale*, *L. osbornei*, *L. platyspermum*, *L. pondoense*, *L. procerum*, *L. pubescens*, *L. pubibracteolatum*, *L. radiatum*, *L. remotum*, *L. roodii*, *L. rourkei*, *L. rubrum*, *L. salicifolium*, *L. salignum*, *L. sericeum*, *L. sessile*, *L. sheilae*, *L. singular*, *L. sorocephalodes*, *L. spirale*, *L. spissifolium*, *L. stellare*, *L. stelligerum*, *L. strobilinum*, *L. teretifolium*, *L. thymifolium*, *L. tinctum*, *L. tradouwense*, *L. uliginosum*, *L. verticillatum*, *and L. xanthoconus.*

Leucadendron grandiflorum is also known commonly as Wynberg Conebush and was last observed in 1806 in Clapham, South Africa. Recorded sightings of *Leucadendron grandiflorum* have occurred on Wynberg Mountain and this species may have existed on the south slopes of Wynberg hill on moister granite soils (see, *e.g.*, T. Rebelo, "Wynberg Conebush - extinct for 200 years," iSpot, 25 July 2015, which is hereby incorporated by reference in its entirety). *Leucadendron grandiflorum* has been described and depicted in Salisbury *et al.*, The Paradisus Londinensis or Coloured Figures of Plants Cultivated in the Vicinity of the Metropolis. 1805; (Volume 1, part 2): 105; see www –dot- biodiversitylibrary.org-backslash-ia/mobot31753000575172#page/248/mode/1up, the contents of each of which is hereby incorporated by reference in its entirety. No modern collections of *Leucadendron grandiflorum* have been recorded, and it is considered that this species was likely scarce or extinct by the early 1800s (see, *e.g.*, T. Rebelo, "Wynberg Conebush - extinct for 200 years," iSpot, 25 July 2015; Catalogue of Life: Leucadendron grandiflorum (Salisb.) R. Br., 20 December 2017). Sister species include *L. globosum* and *L. elimense*. FIG. 8B depicts *Leucadendron grandiflorum.*

In some embodiments, the rare plant may be *Macrostylis villosa*. The *Macrostylis* genus belongs to the Rutaceae family and includes at least ten species (*e.g.*, *Macrostylis barbigera*, *Macrostylis cassiopoides*, *Macrostylis cauliflora*, *Macrostylis crassifolia*, *Macrostylis decipiens*,

Macrostylis hirta, *Macrostylis ramulosa*, *Macrostylis squarrosa*, *Macrostylis tenuis,* and *Macrostylis villosa*).

There are two recognized subspecies of *Macrostylis villosa*, *M. villosa* (Thunb.) Sond. subsp. minor and *M. villosa* (Thunb.) Sond. subsp. villosa. *M. villosa* (Thunb.) Sond. subsp. minor is classified as extinct as its habitat was converted to agriculture and extensive searches have failed to relocate surviving plants. It was previously found on the Western Cape in South Africa and inhabited gravel and clay soil on slopes (see, *e.g.*, "Red List of South African Plants: Macrostylis villosa subsp. minor," 2005, which is hereby incorporated by reference in its entirety). *M. villosa* (Thunb.) Sond. subsp. villosa is considered endangered due to population loss from urban expansion, foreign plant invasions and conversion of habitat to agriculture. A picture of *M. villosa* (Thunb.) Sond. subsp. villosa is reproduced in FIG. 8C (see, *e.g.*, "Red List of South African Plants: Macrostylis villosa subsp. villosa," 2007, which is hereby incorporated by reference in its entirety).

In some embodiments, the rare plant may be *Orbexilum stipulatum (Psoralea stipulata).* *Orbexilum* belongs to the Fabaceae family and members of this genus often have characteristic pod walls that are rugose and free from hair. *Orbexilum* also may be distinguished by its "scarcely accrescent calyx" (see, *e.g.*, Turner, Revision of the genus Orbexilum (Fabaceae: Psoraleeae). Lundellia. 2008; (11):1–7, which is hereby incorporated by reference in its entirety). *Orbexilum* species include *O. chiapasanum, O. gracile, O. lupinellum, O. macrophyllum, O. melanocarpum, O. oliganthum, O. onobrychis, O. pedunculatum, O. simplex, O. stipulatum, and O. virgatum.*

O. stipulatum, also known as the "Largestipule Leather-root" or as the "Falls-of-the-Ohio Scurfpea" was only found on Rock Island in Kentucky. The last recorded observation of *O. stipulatum* was in 1881, prior to resurfacing and flooding of this island. Despite many searches of similar habitats, including intensive searches in 1998, on both the Kentucky and Indiana shores of the Ohio River, this species has not been relocated. Therefore, this species has been classified as extinct (see, *e.g.*, NatureServe Explorer: Orbexilum stipulatum - (Torr. & Gray) Rydb., 2016 and Baskin *et al.* described above, which is each hereby incorporated by reference in its entirety).

O. stipulatum was a perennial herb and had leaves that were divided into 3 leaflets, each about 2 cm in length. The species had a persistent appendage at the base of the leaves and

was also described as having a corolla tube that did not extend beyond the calyx. It is likely that this plant bloomed in late May to mid-June, but seeds have not been observed in nature (see *e.g.*, "NatureServe Explorer: Orbexilum stipulatum - (Torr. & Gray) Rydb.," 2016; and Baskin *et al.*, Geographical origin of the specimens of Orbexilum stipulatum (T. & G.) Rydb. (Psoralea stipulata T. & G.). Castanea. 1986;(51): 207-210, each of which is hereby incorporated by reference in its entirety). A picture of *O. stipulatum* may be found in Short, "Orbexilum stipulatum collected at Falls of the Ohio," 1840 from The Philadelphia Herbarium at the Academy of Natural Sciences is reproduced in FIG. 8D.

In some embodiments, the rare plant may be *Shorea cuspidata*. *Shorea* is a genus in the Dipterocarpaceae family and includes many rainforest trees endemic to southeast Asia. Many *Shorea* species are angiosperms (flowering plants). Non-limiting examples of *Shorea* species may include *Shorea affinis*, *Shorea congestiflora*, *Shorea cordifolia*, *Shorea disticha*, *Shorea megistophylla*, *Shorea trapezifolia*, *Shorea zeylanica*, *Shorea acuminatissima*, *Shorea alutacea*, *Shorea angustifolia*, *Shorea bakoensis*, *Shorea balanocarpoides*, *Shorea chaiana*, *Shorea collaris*, *Shorea cuspidata*, *Shorea faguetiana*, *Shorea faguetioides*, *Shorea gibbosa*, *Shorea hopeifolia*, *Shorea iliasii*, *Shorea induplicata*, *Shorea kudatensis*, *Shorea laxa*, *Shorea longiflora*, *Shorea longisperma*, *Shorea macrobalanos*, *Shorea mujongensis*, *Shorea multiflora*, *Shorea obovoidea*, *Shorea patoiensis*, *Shorea peltata*, *Shorea polyandra*, *Shorea richetia*, *Shorea subcylindrica*, *Shorea tenuiramulosa*, and *Shorea xanthophylla*.

S. cuspidata is a tree endemic to Malaysia that is currently classified as extinct on the IUCN Red List ("The IUCN Red List: Shorea cuspidata," 1998, which is incorporated in its entirety by reference), although there have been a few recorded sightings of *S. cuspidata* subsequent to this classification in Bako National Park, Lambir National Park, and the Semenggoh Arboretum (Ashton, Shorea cuspidata. Tree Flora of Sabah and Sarawek. 2004;(5):246-247; Ling et al., Diversity of the tree flora in Semenggoh Arboretum, Sarawak, Borneo. Gardens' Bulletin Singapore. 2012; (64):139-169, which is each incorporated by reference in its entirety). *Shorea cuspidata* may be considered a rare plant. *Shorea cuspidata* has been characterized as a medium-sized tree with flowers second and pale lime-yellow petals (see, *e.g.*, Ashton, Man. Dipt. Brun. 1968: f. 10, pl. 14 (stem-base)). A picture of a *Shorea cuspidata* specimen is reproduced in FIG. 8E ("Kew Royal Botanical Gardens: Shorea cuspidata specimen K000700460," 1962, which is hereby incorporated by reference in its entirety).

In some embodiments, the rare plant may be *Wendlandia angustifolia*. *Wendlandia* is a genus of flowering plants that belongs to the Rubiaceae family. Non-limiting examples of *Wendlandia* species may include *Wendlandia aberrans, Wendlandia acuminata, Wendlandia amocana, Wendlandia andamanica, Wendlandia angustifolia, Wendlandia appendiculata, Wendlandia arabica, Wendlandia arborescens, Wendlandia augustini, Wendlandia basistaminea, Wendlandia bicuspidata, Wendlandia bouvardioides, Wendlandia brachyantha, Wendlandia brevipaniculata, Wendlandia brevituba, and Wendlandia buddleacea.*

W. angustifolia is a plant native to India that is currently classified as extinct in the IUCN Red List (see "The IUCN Red List: Wendlandia angustifolia," 1998, which is hereby incorporated by reference in its entirety). Subsequent to this classification, *W. angustifolia* was reportedly observed in Kalakkad Mundantharai Tiger Reserve in India (Viswanathan *et al.*, Rediscovery of Wendlandia Angustifolia Wight Ex Hook.f. (Rubiaceae), from Tamil Nadu, a Species Presumed Extinct. Journal of The Bombay Natural History Society. 2000 97(2):311-313, which is hereby incorporated by reference in its entirety). *W. angustifolia* may be considered a rare plant. *W. angustifolia* has been described as a shrub or tree with ternately whorled and linear-lanceolate leaves (see, *e.g.*, Viswanathan *et al.*, Rediscovery of Wendlandia Angustifolia Wight Ex Hook.f. (Rubiaceae), from Tamil Nadu, a Species Presumed Extinct, Journal of The Bombay Natural History Society. 2000; 97(2):311-313, which is hereby incorporated by reference in its entirety). A picture of a specimen is reproduced in FIG. 8F ("Kew Royal Botanical Gardens: Wendlandia angustifolia K000030921," collection date not recorded), which is hereby incorporated by reference in its entirety.

Methods of producing chimeric terpene synthases and terpenes

Also described herein are nucleic acid molecules encoding chimeric terpene synthases. In some embodiments, at least 10% (*e.g.*, at least 11%, 12%, 13%, 14%, 15%, 16%, 17%, 18%, 19%, 20%, 21%, 22%, 23%, 24%, 25%, 26%, 27%, 28%, 29%, 30%, 31%, 32%, 33%, 34%, 35%, 36%, 37%, 38%, 39%, 40%, 41%, 42%, 43%, 44%, 45%, 46%, 47%, 48%, 49%, 50%, 51%, 52%, 53%, 54%, 55%, 56%, 57%, 58%, 59%, 60%, 61%, 62%, 63%, 64%, 65%, 66%, 67%, 68%, 69%, 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%,

99%, or more than 99%, including all values in between) of the nucleic acid molecule encoding such a chimeric terpene synthase may be derived from a rare or extinct plant.

In some instances, a nucleic acid molecule encoding a chimeric terpene synthase comprises a nucleotide sequence that is at least 50% (*e.g.*, at least 55%, 60%, 65%, 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more than 99%, including all values in between) identical to a sequence selected from the group consisting of SEQ ID NOs: 67-118. In some instances, a nucleic acid molecule encoding a chimeric terpene synthase comprises a nucleotide sequence that is identical to a sequence selected from the group consisting of SEQ ID NOs: 67-118. In some instances, a nucleic acid molecule encoding a chimeric terpene synthase further comprises the nucleotide sequence TATA (TATA box sequence). In some instances, a nucleic acid molecule encoding a chimeric terpene synthase comprises the nucleotide sequence TATA (TATA box sequence) that is located N-terminal to a sequence selected from the group consisting of SEQ ID NOs: 67-118. In some instances, a nucleic acid molecule encoding a chimeric terpene synthase comprises a nucleotide sequence that encodes for a sequence set forth in SEQ ID NOs:119-357.

In some embodiments, at least 10% (*e.g.*, at least 11%, 12%, 13%, 14%, 15%, 16%, 17%, 18%, 19%, 20%, 21%, 22%, 23%, 24%, 25%, 26%, 27%, 28%, 29%, 30%, 31%, 32%, 33%, 34%, 35%, 36%, 37%, 38%, 39%, 40%, 41%, 42%, 43%, 44%, 45%, 46%, 47%, 48%, 49%, 50%, 51%, 52%, 53%, 54%, 55%, 56%, 57%, 58%, 59%, 60%, 61%, 62%, 63%, 64%, 65%, 66%, 67%, 68%, 69%, 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or more than 99%, including all values in between) of the amino acid sequence of the chimeric terpene synthase (*e.g.*, a chimeric sesquiterpene synthase) may be derived from a rare or extinct plant. In some instances, a chimeric terpene synthase comprises one or more sequences set forth in SEQ ID NOs:119-357.

Also described herein are chimeric terpene synthases that are capable of producing alpha-guaiene. In some embodiments, at least 10% (e.g., at least 11%, 12%, 13%, 14%, 15%, 16%, 17%, 18%, 19%, 20%, 21%, 22%, 23%, 24%, 25%, 26%, 27%, 28%, 29%, 30%, 31%, 32%, 33%, 34%, 35%, 36%, 37%, 38%, 39%, 40%, 41%, 42%, 43%, 44%, 45%, 46%, 47%, 48%, 49%, 50%, 51%, 52%, 53%, 54%, 55%, 56%, 57%, 58%, 59%, 60%, 61%, 62%, 63%, 64%,

65%, 66%, 67%, 68%, 69%, 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or more than 99%, including all values in between) of the nucleic acid molecule encoding such a chimeric terpene synthase may be derived from a rare or extinct plant.

In some embodiments, at least 10% (e.g., at least 11%, 12%, 13%, 14%, 15%, 16%, 17%, 18%, 19%, 20%, 21%, 22%, 23%, 24%, 25%, 26%, 27%, 28%, 29%, 30%, 31%, 32%, 33%, 34%, 35%, 36%, 37%, 38%, 39%, 40%, 41%, 42%, 43%, 44%, 45%, 46%, 47%, 48%, 49%, 50%, 51%, 52%, 53%, 54%, 55%, 56%, 57%, 58%, 59%, 60%, 61%, 62%, 63%, 64%, 65%, 66%, 67%, 68%, 69%, 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or more than 99%, including all values in between) of the amino acid sequence of the chimeric terpene synthase that is capable of producing alpha-guaiene may be derived from a rare or extinct plant.

In some instances, construction of the chimeras may include sequence (*e.g.*, nucleic acid sequence and/or amino acid sequence) alignments between at least two terpene synthases of interest. For example, sequence alignment analysis may be used to identify fragments (*e.g.*, domains) of a particular terpene synthase to include in a chimeric terpene synthase. In some embodiments, the chimeric terpene synthase is a chimeric sesquiterpene synthase. Non-limiting examples of analyses may include the types described in the blastn-mapdamage and tblastn pipelines described in Example 2.

In some embodiments, a chimeric terpene synthase coding sequence comprises a mutation at 1, 2, 3, 4, 5, or more positions corresponding to a reference chimeric terpene synthase coding sequence. In some embodiments, the chimeric terpene synthase coding sequence comprises a mutation in 1, 2, 3, 4, 5, or more codons of the coding sequence relative to a reference chimeric terpene synthase coding sequence. As will be understood by one of ordinary skill in the art, a mutation within a codon may or may not change the amino acid that is encoded by the codon due to degeneracy of the genetic code. In some embodiments, the one or more mutations in the coding sequence do not alter the amino acid sequence of the chimeric terpene synthase relative to the amino acid sequence of a reference chimeric terpene synthase.

In some embodiments, the one or more mutations in a chimeric terpene synthase sequence alter the amino acid sequence of the chimeric terpene synthase relative to the amino

acid sequence of a reference chimeric terpene synthase. In some embodiments, the one or more mutations alter the amino acid sequence of the chimeric terpene synthase relative to the amino acid sequence of a reference chimeric terpene synthase and alter (enhance or reduce) an activity of the chimeric terpene synthase relative to the reference chimeric terpene synthase.

The skilled artisan will also realize that mutations in a chimeric terpene synthase coding sequence may result in conservative amino acid substitutions to provide functionally equivalent variants of the foregoing polypeptides, *e.g.*, variants that retain the activities of the polypeptides. As used herein, a "conservative amino acid substitution" refers to an amino acid substitution that does not alter the relative charge or size characteristics or functional activity of the protein in which the amino acid substitution is made.

In some instances, an amino acid is characterized by its R group (see, e.g., Table 1). For example, an amino acid may comprise a nonpolar aliphatic R group, a positively charged R group, a negatively charged R group, a nonpolar aromatic R group, or a polar uncharged R group. Non-limiting examples of an amino acid comprising a nonpolar aliphatic R group include alanine, glycine, valine, leucine, methionine, and isoleucine. Non-limiting examples of an amino acid comprising a positively charged R group includes lysine, arginine, and histidine. Non-limiting examples of an amino acid comprising a negatively charged R group include aspartic acid and glutamic acid. Non-limiting examples of an amino acid comprising a nonpolar, aromatic R group include phenylalanine, tyrosine, and tryptophan. Non-limiting examples of an amino acid comprising a polar uncharged R group include serine, threonine, cysteine, proline, asparagine, and glutamine.

Variants can be prepared according to methods for altering polypeptide sequence known to one of ordinary skill in the art such as are found in references which compile such methods, *e.g.*, Molecular Cloning: A Laboratory Manual, J. Sambrook, et al., eds., Fourth Edition, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, 2012, or Current Protocols in Molecular Biology, F.M. Ausubel, et al., eds., John Wiley & Sons, Inc., New York, 2010.

Non-limiting examples of functionally equivalent variants of polypeptides may include conservative amino acid substitutions in the amino acid sequences of proteins disclosed herein. Conservative substitutions of amino acids include substitutions made amongst amino acids within the following groups: (a) M, I, L, V; (b) F, Y, W; (c) K, R, H; (d) A, G; (e) S, T; (f) Q, N; and (g) E, D. In some embodiments, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19,

20 or more than 20 residues can be changed when preparing variant polypeptides. In some embodiments, amino acids are replaced by conservative amino acid substitutions.

Table 1. Non-limiting Examples of Conservative Amino Acid Substitutions

Original Residue	R Group Type	Conservative Amino Acid Substitutions
Ala	nonpolar aliphatic R group	Cys, Gly, Ser
Arg	positively charged R group	His, Lys
Asn	polar uncharged R group	Asp, Gln, Glu
Asp	negatively charged R group	Asn, Gln, Glu
Cys	polar uncharged R group	Ala, Ser
Gln	polar uncharged R group	Asn, Asp, Glu
Glu	negatively charged R group	Asn, Asp, Gln
Gly	nonpolar aliphatic R group	Ala, Ser
His	positively charged R group	Arg, Tyr, Trp
Ile	nonpolar aliphatic R group	Leu, Met, Val
Leu	nonpolar aliphatic R group	Ile, Met, Val
Lys	positively charged R group	Arg, His
Met	nonpolar aliphatic R group	Ile, Leu, Phe, Val
Pro	polar uncharged R group	
Phe	nonpolar aromatic R group	Met, Trp, Tyr
Ser	polar uncharged R group	Ala, Gly, Thr
Thr	polar uncharged R group	Ala, Asn, Ser
Trp	nonpolar aromatic R group	His, Phe, Tyr, Met
Tyr	nonpolar aromatic R group	His, Phe, Trp
Val	nonpolar aliphatic R group	Ile, Leu, Met, Thr

Amino acid substitutions in the amino acid sequence of a polypeptide to produce a chimeric terpene synthase (*e.g.*, chimeric sesquiterpene synthase) variant having a desired property and/or activity can be made by alteration of the coding sequence of the chimeric terpene synthase (*e.g.*, chimeric sesquiterpene synthase). Similarly, conservative amino acid substitutions in the amino acid sequence of a polypeptide to produce functionally equivalent variants of the polypeptide typically are made by alteration of the coding sequence of the chimeric terpene synthase (*e.g.*, chimeric sesquiterpene synthase).

Mutations (*e.g.,* substitutions) can be made in a nucleotide sequence by a variety of methods known to one of ordinary skill in the art. For example, mutations can be made by PCR-directed mutation, site-directed mutagenesis according to the method of Kunkel (Kunkel, *Proc.*

Nat. Acad. Sci. U.S.A. 82: 488-492, 1985), or by chemical synthesis of a gene encoding a polypeptide.

Any suitable method, including circular permutation (Yu and Lutz, *Trends Biotechnol.* 2011 Jan;29(1):18-25), may be used to produce variants. In circular permutation, the linear primary sequence of a polypeptide can be circularized (*e.g.*, by joining the N-terminal and C-terminal ends of the sequence) and the polypeptide can be severed ("broken") at a different location. Thus, the linear primary sequence of the new polypeptide may have low sequence identity (*e.g.*, less than 80%, less than 75%, less than 70%, less than 65%, less than 60%, less than 55%, less than 50%, less than 45%, less than 40%, less than 35%, less than 30%, less than 25%, less than 20%, less than 15%, less than 10%, less or less than 5%, including all values in between) as determined by linear sequence alignment methods (*e.g.*, Clustal Omega or BLAST). Topological analysis of the two proteins, however, may reveal that the tertiary structure of the two polypeptides is similar or dissimilar. Without being bound by a particular theory, a variant polypeptide created through circular permutation of a reference polypeptide and with a similar tertiary structure as the reference polypeptide can share similar functional characteristics (*e.g.*, enzymatic activity, enzyme kinetics, substrate specificity or product specificity). In some instances, circular permutation may alter the secondary structure, tertiary structure or quaternary structure and produce an enzyme with different functional characteristics (*e.g.*, increased or decreased enzymatic activity, different substrate specificity, or different product specificity). See, *e.g.*, Yu and Lutz, *Trends Biotechnol.* 2011 Jan;29(1):18-25.

It should be appreciated that in a protein that has undergone circular permutation, the linear amino acid sequence of the protein would differ from a reference protein that has not undergone circular permutation. However, one of ordinary skill in the art would be able to readily determine which residues in the protein that has undergone circular permutation correspond to residues in the reference protein that has not undergone circular permutation by, for example, aligning the sequences and detecting conserved motifs, and/or by comparing the structures or predicted structures of the proteins, *e.g.*, by homology modeling.

Aspects of the present disclosure relate to the recombinant expression of genes encoding enzymes, functional modifications and variants thereof, as well as uses relating thereto.

A nucleic acid encoding any of the chimeric terpene synthases described herein may be incorporated into any appropriate vector through any method known in the art. For example, the

vector may be an expression vector, including but not limited to a viral vector (*e.g.*, a lentiviral, retroviral, adenoviral, or adeno-associated viral vector), any vector suitable for transient expression, or any vector for inducible expression (*e.g.*, a galactose-inducible or doxycycline-inducible vector). A non-limiting example of a vector for expression of a chimeric terpene synthase (*e.g.*, a chimeric sesquiterpene synthase) is described in Example 2 below.

In some embodiments, a vector replicates autonomously in the cell. A vector can contain one or more endonuclease restriction sites that are cut by a restriction endonuclease to insert and ligate a nucleic acid containing a gene described herein to produce a recombinant vector that is able to replicate in a cell. Vectors are typically composed of DNA, although RNA vectors are also available. Cloning vectors include, but are not limited to: plasmids, fosmids, phagemids, virus genomes and artificial chromosomes. As used herein, the terms "expression vector" or "expression construct" refer to a nucleic acid construct, generated recombinantly or synthetically, with a series of specified nucleic acid elements that permit transcription of a particular nucleic acid in a host cell (*e.g.*, microbe), such as a yeast cell. In some embodiments, the nucleic acid sequence of a gene described herein is inserted into a cloning vector such that it is operably joined to regulatory sequences and, in some embodiments, expressed as an RNA transcript. In some embodiments, the vector contains one or more markers, such as a selectable marker as described herein, to identify cells transformed or transfected with the recombinant vector.

In some embodiments, a vector is capable of integrating into the genome of a host cell.

A coding sequence and a regulatory sequence are said to be “operably joined” or “operably linked” when the coding sequence and the regulatory sequence are covalently linked and the expression or transcription of the coding sequence is under the influence or control of the regulatory sequence. If the coding sequence is to be translated into a functional protein, the coding sequence and the regulatory sequence are said to be operably joined or linked if induction of a promoter in the 5’ regulatory sequence transcribes the coding sequence and if the nature of the linkage between the coding sequence and the regulatory sequence does not (1) result in the introduction of a frame-shift mutation, (2) interfere with the ability of the promoter region to direct the transcription of the coding sequence, or (3) interfere with the ability of the corresponding RNA transcript to be translated into a protein. Thus, a promoter region is operably joined or linked to a coding sequence if the promoter region transcribes the coding sequence and the transcript can be translated into the protein or polypeptide of interest.

In some embodiments, the nucleic acid encoding any of the proteins described herein is under the control of regulatory sequences (*e.g.,* enhancer sequences). In some embodiments, a nucleic acid is expressed under the control of a promoter. The promoter can be a native promoter, *e.g.,* the promoter of the gene in its endogenous context, which provides normal regulation of expression of the gene. Alternatively, a promoter can be a promoter that is different from the native promoter of the gene, *e.g.,* the promoter is different from the promoter of the gene in its endogenous context. As used herein, a "heterologous promoter" or "recombinant promoter" is a promoter that is not naturally or normally associated with or that does not naturally or normally control transcription of a DNA sequence to which it is operably joined or linked. In some embodiments, a nucleotide sequence is under the control of a heterologous promoter.

In some embodiments, the promoter is a eukaryotic promoter. Non-limiting examples of eukaryotic promoters include TDH3, PGK1, PKC1, TDH2, PYK1, TPI1, AT1, CMV, EF1a, SV40, PGK1 (human or mouse), Ubc, human beta actin, CAG, TRE, UAS, Ac5, Polyhedrin, CaMKIIa, GAL1, GAL10, TEF1, GDS, ADH1, CaMV35S, Ubi, H1, U6, as would be known to one of ordinary skill in the art (see, *e.g.*, Addgene website: blog.addgene.org/plasmids-101-the-promoter-region). In some embodiments, the promoter is a prokaryotic promoter (*e.g.*, bacteriophage or bacterial promoter). Non-limiting examples of bacteriophage promoters include Pls1con, T3, T7, SP6, and PL. Non-limiting examples of bacterial promoters include Pbad, PmgrB, Ptrc2, Plac/ara, Ptac, Pm.

In some embodiments, the promoter is an inducible promoter. As used herein, an "inducible promoter" is a promoter controlled by the presence or absence of a molecule. Non-limiting examples of inducible promoters include chemically-regulated promoters and physically-regulated promoters. For chemically-regulated promoters, the transcriptional activity can be regulated by one or more compounds, such as alcohol, tetracycline, galactose, a steroid, a metal, or other compounds. For physically-regulated promoters, transcriptional activity can be regulated by a phenomenon such as light or temperature. Non-limiting examples of tetracycline-regulated promoters include anhydrotetracycline (aTc)-responsive promoters and other tetracycline-responsive promoter systems (*e.g.*, a tetracycline repressor protein (tetR), a tetracycline operator sequence (tetO) and a tetracycline transactivator fusion protein (tTA)). Non-limiting examples of steroid-regulated promoters include promoters based on the rat

glucocorticoid receptor, human estrogen receptor, moth ecdysone receptors, and promoters from the steroid/retinoid/thyroid receptor superfamily. Non-limiting examples of metal-regulated promoters include promoters derived from metallothionein (proteins that bind and sequester metal ions) genes. Non-limiting examples of pathogenesis-regulated promoters include promoters induced by salicylic acid, ethylene or benzothiadiazole (BTH). Non-limiting examples of temperature/heat-inducible promoters include heat shock promoters. Non-limiting examples of light-regulated promoters include light responsive promoters from plant cells. In certain embodiments, the inducible promoter is a galactose-inducible promoter. In some embodiments, the inducible promoter is induced by one or more physiological conditions (*e.g.*, pH, temperature, radiation, osmotic pressure, saline gradients, cell surface binding, or concentration of one or more extrinsic or intrinsic inducing agents). Non-limiting examples of an extrinsic inducer or inducing agent include amino acids and amino acid analogs, saccharides and polysaccharides, nucleic acids, protein transcriptional activators and repressors, cytokines, toxins, petroleum-based compounds, metal containing compounds, salts, ions, enzyme substrate analogs, hormones or any combination thereof.

In some embodiments, the promoter is a constitutive promoter. As used herein, a "constitutive promoter" refers to an unregulated promoter that allows continuous transcription of a gene. Non-limiting examples of a constitutive promoter includes CP1, CMV, EF1a, SV40, PGK1, Ubc, human beta actin, CAG, Ac5, polyhedrin, TEF1, GDS, CaM35S, Ubi, H1, and U6.

Other inducible promoters or constitutive promoters known to one of ordinary skill in the art are also contemplated herein.

The precise nature of the regulatory sequences needed for gene expression may vary between species or cell types, but generally include, as necessary, 5' non-transcribed and 5' non-translated sequences involved with the initiation of transcription and translation respectively, such as a TATA box, capping sequence, CAAT sequence, and the like. In particular, such 5' non-transcribed regulatory sequences will include a promoter region which includes a promoter sequence for transcriptional control of the operably joined or linked gene. Regulatory sequences may also include enhancer sequences or upstream activator sequences. The vectors disclosed herein may include 5' leader or signal sequences. The regulatory sequence may also include a terminator sequence. In some embodiments, a terminator sequence marks the end of a gene in DNA during transcription. The choice and design of one or more appropriate vectors suitable for

inducing expression of one or more genes described herein in a heterologous organism is within the ability and discretion of one of ordinary skill in the art.

Expression vectors containing the necessary elements for expression are commercially available and known to one of ordinary skill in the art (see, *e.g.*, Sambrook et al., Molecular Cloning: A Laboratory Manual, Fourth Edition, Cold Spring Harbor Laboratory Press, 2012).

Any suitable host cell may be used to produce any of the chimeric terpene synthases disclosed herein, including eukaryotic cells or prokaryotic cells. Suitable host cells include fungal cells (*e.g.*, yeast cells) and bacteria cells (*e.g.*, *E. coli* cells). Non-limiting examples of genera of yeast for expression include *Saccharomyces* (*e.g.*, *S. cerevisiae*), *Pichia*, *Kluyveromyces* (*e.g.*, *K. lactis*), *Hansenula* and *Yarrowia.* In some embodiments, the yeast strain is an industrial polyploid yeast strain. Other non-limiting examples of fungal cells include cells obtained from *Aspergillus* spp., *Penicillium* spp., *Fusarium* spp., *Rhizopus* spp., *Acremonium* spp., *Neurospora* spp., *Sordaria* spp., *Magnaporthe* spp., *Allomyces* spp., *Ustilago* spp., *Botrytis* spp., and *Trichoderma* spp.

The term "cell," as used herein, may refer to a single cell or a population of cells, such as a population of cells belonging to the same cell line or strain. Use of the singular term "cell" should not be construed to refer explicitly to a single cell rather than a population of cells.

A vector encoding any of the chimeric terpene synthases (*e.g.*, chimeric sesquiterpene synthases) described herein may be introduced into a suitable host cell using any method known in the art. Non-limiting examples of yeast transformation protocols are described in Example 2 below and in Gietz *et al.*, Yeast transformation by the LiAc/SS Carrier DNA/PEG method. *Methods Mol Biol.* 2006;313:107-20, which is hereby incorporated by reference in its entirety for this purpose. Host cells may be cultured under any conditions suitable as would be understood by one of ordinary skill in the art. For example, any media, temperature, and incubation conditions known in the art may be used. For host cells carrying an inducible vector, cells may be cultured with an appropriate inducible agent to promote expression.

Any of the cells disclosed herein can be cultured in media of any type (rich or minimal) and any composition prior to, during, and/or after contact and/or integration of a nucleic acid. The conditions of the culture or culturing process can be optimized through routine experimentation as understood by one of ordinary skill in the art. In some embodiments, the selected media is supplemented with various components. In some embodiments, the

concentration and amount of a supplemental component is optimized. In some embodiments, other aspects of the media and growth conditions (*e.g.,* pH, temperature, etc.) are optimized through routine experimentation. In some embodiments, the frequency that the media is supplemented with one or more supplemental components, and the amount of time that the cell is cultured is optimized.

Culturing of the cells described herein can be performed in culture vessels known and used in the art. In some embodiments, an aerated reaction vessel (*e.g.*, a stirred tank reactor) is used to culture the cells. In some embodiments, a bioreactor or fermentor is used to culture the cell. Thus, in some embodiments, the cells are used in fermentation. As used herein, the terms "bioreactor" and "fermentor" are interchangeably used and refer to an enclosure, or partial enclosure, in which a biological, biochemical and/or chemical reaction takes place, involving a living organism or part of a living organism. A "large-scale bioreactor" or "industrial-scale bioreactor" is a bioreactor that is used to generate a product on a commercial or quasi-commercial scale. Large scale bioreactors typically have volumes in the range of liters, hundreds of liters, thousands of liters, or more.

In some embodiments, a bioreactor comprises a cell (*e.g.*, a yeast cell) or a cell culture (*e.g.,* a yeast cell culture), such as a cell or cell culture described herein. In some embodiments, a bioreactor comprises a spore and/or a dormant cell type of an isolated microbe (*e.g.,* a dormant cell in a dry state).

Non-limiting examples of bioreactors include: stirred tank fermentors, bioreactors agitated by rotating mixing devices, chemostats, bioreactors agitated by shaking devices, airlift fermentors, packed-bed reactors, fixed-bed reactors, fluidized bed bioreactors, bioreactors employing wave induced agitation, centrifugal bioreactors, roller bottles, and hollow fiber bioreactors, roller apparatuses (for example benchtop, cart-mounted, and/or automated varieties), vertically-stacked plates, spinner flasks, stirring or rocking flasks, shaken multi-well plates, MD bottles, T-flasks, Roux bottles, multiple-surface tissue culture propagators, modified fermentors, and coated beads (*e.g.,* beads coated with serum proteins, nitrocellulose, or carboxymethyl cellulose to prevent cell attachment).

In some embodiments, the bioreactor includes a cell culture system where the cell (*e.g.*, yeast cell) is in contact with moving liquids and/or gas bubbles. In some embodiments, the cell or cell culture is grown in suspension. In other embodiments, the cell or cell culture is attached

to a solid phase carrier. Non-limiting examples of a carrier system includes microcarriers (*e.g.*, polymer spheres, microbeads, and microdisks that can be porous or non-porous), cross-linked beads (*e.g.,* dextran) charged with specific chemical groups (*e.g.,* tertiary amine groups), 2D microcarriers including cells trapped in nonporous polymer fibers, 3D carriers (*e.g.*, carrier fibers, hollow fibers, multicartridge reactors, and semi-permeable membranes that can comprising porous fibers), microcarriers having reduced ion exchange capacity, encapsulation cells, capillaries, and aggregates. In some embodiments, carriers are fabricated from materials such as dextran, gelatin, glass, or cellulose.

In some embodiments, industrial-scale processes are operated in continuous, semi-continuous or non-continuous modes. Non-limiting examples of operation modes are batch, fed batch, extended batch, repetitive batch, draw/fill, rotating-wall, spinning flask, and/or perfusion mode of operation. In some embodiments, a bioreactor allows continuous or semi-continuous replenishment of the substrate stock, for example a carbohydrate source and/or continuous or semi-continuous separation of the product, from the bioreactor.

In some embodiments, the bioreactor or fermentor includes a sensor and/or a control system to measure and/or adjust reaction parameters. Non-limiting examples of reaction parameters include biological parameters (*e.g.*, growth rate, cell size, cell number, cell density, cell type, or cell state, etc.), chemical parameters (*e.g.,* pH, redox-potential, concentration of reaction substrate and/or product, concentration of dissolved gases, such as oxygen concentration and CO_2 concentration, nutrient concentrations, metabolite concentrations, concentration of an oligopeptide, concentration of an amino acid, concentration of a vitamin, concentration of a hormone, concentration of an additive, serum concentration, ionic strength, concentration of an ion, relative humidity, molarity, osmolarity, concentration of other chemicals, for example buffering agents, adjuvants, or reaction by-products), physical/mechanical parameters (*e.g.*, density, conductivity, degree of agitation, pressure, and flow rate, shear stress, shear rate, viscosity, color, turbidity, light absorption, mixing rate, conversion rate, as well as thermodynamic parameters, such as temperature, light intensity/quality, etc.). Sensors to measure the parameters described herein are well known to one of ordinary skill in the relevant mechanical and electronic arts. Control systems to adjust the parameters in a bioreactor based on the inputs from a sensor described herein are well known to one of ordinary skill in the art in bioreactor engineering.

Terpenes produced by any of the host cells disclosed herein may be extracted using any method known in the art. A non-limiting example of a method for sesquiterpene extraction is provided in Example 2. Any of the terpenes produced from the methods, compositions, or host cells described herein may be used in a suitable composition for topical application to, for example, skin, hair, clothing, or articles in a home (*e.g.*, a perfume). As used herein, the term "perfume" is any fragrance formulation suitable for application to the hair, skin, or clothing of a person or an article in a home. This term includes, but is not limited to: an eau de cologne, eau de toilette, eau de parfum, perfume extract or extrait. In addition to comprising one or more terpenes of the application, such a perfume may include, for example, one or more natural oils, fixatives, emollients, or solvents.

Examples of natural oils which may be used in perfume formulations include, but are not limited to: amyris oil; angelica seed oil; angelica root oil; aniseed oil; valerian oil; basil oil; bay oil; mugwort oil; benzoin resin; bergamot oil; birch tar oil; bitter almond oil; savory oil; bucco-leaf oil; cabreuva oil; cade oil; calamus oil; camphor oil; cananga oil; cardamom oil; cascarilla oil; cassia oil; castoreum absolute; cedar-leaf oil; cedarwood oil; cistus oil; citronella oil; lemon oil; copaiba balsam oil; coriander oil; costus root oil; cumin oil; cypress oil; davana oil; dill oil; dillseed oil; elemi oil; tarragon oil; eucalyptus citriodora oil; eucalyptus oil; fennel oil; fir oil; galbanum oil; geranium oil; grapefruit oil; guaiac wood oil; gurjun balsam oil; helichrysum oil; ginger oil; iris root oil; calamus oil; blue chamomile oil; Roman chamomile oil; carrot-seed oil; cascarilla oil; pine-needle oil; spearmint oil; caraway oil; labdanum oil; lavandin oil; lavender oil; lemongrass oil; lovage oil; lime oil (*e.g.*, distilled or pressed lime oil); linaloe oil: litsea cubeba oil; bay leaf oil; mace oil; marjoram oil; mandarin oil; massoi bark oil; ambrette oil; clary sage oil; myristica oil; myrrh oil; myrtle oil; clove leaf oil; clove flower oil; neroli oil; olibanum oil; opopanax oil; orange oil; origanum oil; palmar osa oil; patchouli oil; perilla oil; Peru balsam oil; parsley leaf oil; parsley seed oil; petitgrain oil; peppermint oil; pepper oil; pimento oil; pine oil; pennyroyal oil; rosewood oil; rose oil; rosemary oil; Dalmatian sage oil; Spanish sage oil; sandalwood oil; celery seed oil; spike lavender oil; Japanese aniseed oil; styrax oil; tagetes oil; fir-needle oil; tea-tree oil; turpentine oil; thyme oil; tuberose absolute; vanilla extract; violet leaf absolute; verbena oil; vetiver oil; juniper oil; wine-lees oil; wormwood oil; wintergreen oil; ylang oil; hyssop oil; civet absolute; cinnamon leaf oil; cinnamon bark oil; as well as fractions thereof or constituents isolated therefrom; and combinations thereof.

Other examples of compounds which may be used in perfume formulations may include: wood moss absolute; beeswax absolute; cassia absolute; eau de brouts absolute; oakmoss absolute; galbanum resin; helichrysum absolute; iris root absolute; jasmine absolute; labdanum absolute; labdanum resin; lavandin absolute; lavender absolute; mimosa absolute; tincture of musk; myrrh absolute; olibanum absolute; orange blossom absolute; rose absolute; Tolu balsam; Tonka absolute; as well as fractions thereof or constituents isolated therefrom; and combinations thereof.

As used herein, the term "emollient" means a fatty or oleaginous substance which increases tissue moisture content (and may, for example, render skin softer and more pliable). Emollients for use with the instant compounds and methods may include any appropriate animal fats/oils, vegetable oils, and/or waxes. As a non-limiting set of examples, an emollient for use with the instant compositions and methods may be of natural or synthetic origin and may include: cold-pressed almond oil, jojoba oil, sunflower oil, olive oil, hazelnut oil, avocado oil, safflower oil, grapeseed oil, coconut oil, wheat germ oil, apricot kernel oil, natural waxes and "butters" (*e.g.*, unrefined beeswax, shea butter, jojoba butter, and/or cocoa butter), Schercemol™ LL Ester, Schercemol™ 1818 Ester, butylene glycol, capric/caprylic triglyceride, ceteareth-20, one or more fatty alcohols (*e.g.*, cetearyl alcohol, cetyl alcohol, and/or coconut fatty acids), one or more silicones (*e.g.*, cyclomethicone, dimethicone, and/or cyclopentasiloxane), emulsifying wax, petroleum jelly, fatty acids, glyceryl stearate, hydrogenated oils, isopropyl myristate , mineral oil, octyl palmitate, paraffin, squalene, stearic acid, palmitoyl proline, or magnesium palmitoyl glutamate.

As used herein, the term "fixative" means a compound used to equalize the vapor pressures (and thus the volatilities) of one or more compounds in the perfume. As a non-limiting set of examples, a fixative for use with the compounds and perfumes described herein may be: dipropylene glycol, diethyl phthalate, Hedione®, Abalyn™ D-E Methyl Ester of Rosin, Jojoba (such as Floraesters K-100 Jojoba or Floraesters K-20W Jojoba), Sepicide LD, and/or Foralyn™ 5020-F CG Hydrogenated Rosinate.

As used herein, the term "solvent" is the diluent used to create a perfume. As a non-limiting example, the solvent may be an alcohol (*e.g.*, an ethyl alcohol), 1,2-hexanediol, 1,2-heptanediol, a neutral smelling oil (*e.g.*, fractionated coconut oil or jojoba oil), or one or more volatile silicones. As a non-limiting example, Perfumers' Alcohol (a type of ethyl alcohol) may

be used. Perfumers' Alcohol is prepared from 200 proof ethyl alcohol which may contain very small amounts of butyl alcohol, denatonium benzoate (Britex), and/or hexylene glycol. Various grades of Perfumers' Alcohol are available including SDA 40B 200 Proof and SDA-B 200 proof.

Additional compounds or fragrance materials for use in the perfume composition according to the disclosure may include any compounds which are customarily used in the field.

The present invention is further illustrated by the following Examples, which in no way should be construed as limiting. The entire contents of all of the references (including literature references, issued patents, published patent applications, and co-pending patent applications) cited throughout this application are hereby expressly incorporated by reference.

EXAMPLES

Example 1. Functional characterization of chimeric terpene synthases.

Genomic DNA from 12 extinct plant samples were sequenced (Table 2). Sesquiterpene synthase (SQTS) fragments were recovered from seven plants (Table 11), but gaps in the sequencing prevented reconstruction of full-length genes. A library comprising 2,738 terpene synthase chimeras (containing sequence from sesquiterpene synthases from extant plants to fill the sequence gaps) was screened. The expression of 52 SQTS chimeras (sequences provided in Table 10) from six rare plants (Table 2) led to the production of sesquiterpenes in the screening strain. Methods and materials for each of the procedures described in this Example may be found in Example 2.

Table 2. Rare Plants that were Sequenced (The plants from which functional sesquiterpene chimeras were reconstructed are shown bold face and underlined.)

Family	Genus	Species	Continent	Location	Year Extinct
Crassulaceae	*Crassula*	*subulata*	AFRICA	South Africa	1900
Ericaceae	*Erica*	*pyramidalis*	AFRICA	South Africa	1910
Malvaceae	***Hibiscadelphus***	***wilderianus***	**OCEANIA**	**Hawaii**	**1910**
Proteaceae	***Leucadendron***	***grandiflorum***	**AFRICA**	**South Africa**	**1806**

Rutaceae	***Macrostylis***	***villosa***	**AFRICA**	**South Africa**	**1980**
Myrtaceae	*Myrcia*	*skeldingii*	AMERICA	Jamaica	1972
Rhamnaceae	*Nesiota*	*elliptica*	AFRICA	St. Helena	2003
Fabaceae	***Orbexilum***	***stipulatum***	**AMERICA**	**Kentucky**	**1881**
Sapotaceae	*Pradosia*	*glaziovii*	AMERICA	Brazil	1997
Dipterocarpaceae	***Shorea***	***cuspidata***	**ASIA**	**Malaysia**	**1996**
Proteaceae	*Stenocarpus*	*dumbeensis*	OCEANIA	New Caledonia	1905
Rubiaceae	***Wendlandia***	***angustifolia***	**ASIA**	**India**	**1997**

The terpenes produced by the functional SQTS chimeras were identified initially based on gas chromatography-mass spectrometry (GC/MS) data. In some cases, authentic standards or essential oils containing characterized sesquiterpenes were available to confirm mass spectrum- and retention time-based identifications. In other cases, standards were not available and structural identifications were made based on mass spectral analysis alone. The different methods used to identify the structures are detailed in Table 3, and the specific methods used to identify each sesquiterpene are indicated in Tables 4-9. In some cases, products were identified only as "sesquiterpene" or "sesquiterpenol." In one case, a mass spectrum was recovered but did not yield a match in the NIST/internal database. This sesquiterpenol was identified in the product tables as an "unidentified sesquiterpenol" and additional characterization may be used to determine its structure.

Fourteen SQTS chimeras derived from *Hibiscadelphus wilderianus* produced 1 or more sesquiterpenes (FIG. 1, Table 3). Seven SQTS chimeras derived from *Leucadendron grandiflorum* also produced sesquiterpenes (FIG. 2, Table 5), as did six SQTS chimeras from *Macrostylis villosa* (FIG. 3, Table 6), two from *O. stipulatum* (FIG. 4, Table 7), six from *Shorea cuspidata* (FIG. 5, Table 8), and seventeen from *Wendlandia angustifolia* (FIG. 6, Table 9). The SQTSs were found to produce one to nine different terpenes. The product profiles of the plant SQTS chimeras were different when the functional SQTS chimeras were grouped by the terpenes produced in highest yield (FIG. 7). Delta-cadinene synthases were the most numerous group of functional chimeras at a total of 22 and were derived from four of the plants. 10 of the 14 of the synthases from *H. wilderianus* were of this variety. Alpha-cadinol was frequently detected as a minor product of the delta-cadinene synthases; however, three SQTS chimeras from *S. cuspidata*

yielded more alpha-cadinol than delta-cadinene. These six SQTS chimeras derived *S. cuspidata* produced a very similar product mixture (Table 8, FIG. 13).

The screening of the 2,738-member chimeric sesquiterpene synthase library resulted in the successful expression of 52 functional chimeric sesquiterpene synthases (SQTSs). Fourteen synthases were derived from *H. wilderianus*, a tree which went extinct in Hawaii over 100 years ago. Cadinene, cadinol, and eudesmol-type sesquiterpenes were produced by these chimeras. A few active chimeras were also generated from *O. stipulatum*, a plant that went extinct in Kentucky in the 1800s. Two guaienes and gamma-bisabolene were produced by these synthases. Seven functional SQTS chimeras were constructed from *L. grandiflorum*, a plant that went extinct over 200 years ago. Diverse sesquiterpene and sesquiterpenol structures were produced by these chimeras, along with those derived from three other plants.

Table 3. The six methods used to identify the sesquiterpenes produced by the sesquiterpene synthases.

Method #	Description	Notes
1	Mass spectrum and retention time matched to authentic standard	High confidence in structure and stereochemistry.
2	Mass spectrum and retention time matched to previously characterized compounds in essential oils from plants.	High confidence in structure and stereochemistry.
3	Poor mass spectrum obtained due to low titer; retention time and chimera product profile were consistent with authentic standards or components in essential oils	Fairly high confidence in structure and stereochemistry.
4	Strong mass spectrum match to compound in NIST/internal database	Fairly high confidence in structure, could be an isomer.
5	Poor mass spectrum obtained due to low titer; retention time and chimera product profiles matched to terpenes identified using method # 4	Fairly high confidence in structure, could be an isomer.
6	Poor mass spectrum obtained due to low yields, best (closest) identification possible with NIST/internal database	Lower confidence based on the mass spectral data available.

Table 4. Functional sesquiterpene synthase chimeras derived from *H. wilderianus* sequences and their associated products.

Chimera name	% rare sequence	Terpene identification	Identification Method[1]	% composition[2]
HibWilSQTS117	49%	delta-cadinene	3	100%
HibWilSQTS118	50%	delta-cadinene	3	100%
HibWilSQTS120	46%	delta-cadinene	3	13%
		epi-cubenol	5	3%
		sesquiterpenol	6	2%
		tau-cadinol	2	82%
HibWilSQTS121	50%	delta-cadinene	2	99%
		alpha-cadinol	3	1%
HibWilSQTS123	47%	delta-cadinene	2	99%
		alpha-cadinol	3	1%
HibWilSQTS124	48%	delta-cadinene	2	98%
		alpha-cadinol	3	2%
HibWilSQTS126	44%	delta-cadinene	2	97%
		alpha-cadinol	3	3%
HibWilSQTS19	12%	gamma-selinene	4	1%
		10-epi-gamma-eudesmol	2	2%
		gamma-eudesmol	2	49%
		alpha/beta-eudesmol[3]	4	22%
		juniper camphor	6	1%
		7-epi-alpha-eudesmol	4	1%
		cryptomeridiol isomer 1	4	1%
		cryptomeridiol isomer 2	4	2%
		cryptomeridiol isomer 3	4	21%
HibWilSQTS34	13%	sesquiterpene	6	6%
		10-epi-gamma-eudesmol	3	15%
		gamma-eudesmol	3	27%
		alpha/beta-eudesmol[3]	5	52%
HibWilSQTS52	12%	delta-cadinene	2	60%
		tau-cadinol	3	9%
		alpha/beta-eudesmol[3]	4	31%
HibWilSQTS54	13%	delta-cadinene	2	99%
		alpha-cadinol	2	1%
HibWilSQTS55	12%	delta-cadinene	3	71%
		tau-cadinol	3	6%
		alpha-cadinol	3	23%

HibWilSQTS63	12%	sesquiterpene	6	11%
		delta-cadinene	2	29%
		sesquiterpenol	6	15%
		sesquiterpenol	6	5%
		tau-cadinol	3	10%
		alpha-cadinol	3	30%
HibWilSQTS90	25%	sesquiterpene	6	40%
		alpha/beta-eudesmol[3]	5	60%

[1] The structure identification ranking key is defined in Table 3, with lower numbers indicating a higher degree of confidence. [2] The composition of total sesquiterpenes from each chimera was a rough estimate based on a common ion count (m/z 204.2). The ratio of metabolites may have been different in the production strains and it is possible other minor metabolites were detected when samples were prepared. Representative GC/MS chromatograms for the chimeras with **bold font** can be found in FIG. 9. [3] Co-eluted under these run conditions. The peak was partially resolved under longer run conditions, about 6/4 alpha/beta-eudesmol.

Table 5. Functional sesquiterpene synthase chimeras derived from *L. grandiflorum* sequences and their associated products.

Chimera name	% rare sequence	Terpene identification	Identification Method[1]	% composition[2]
LeuGraSQTS335	14%	sesquiterpene	6	1%
		10-epi-gamma-eudesmol	3	1%
		gamma-eudesmol	3	49%
		alpha/beta-eudesmol[3]	5	23%
		cryptomeridiol isomer 2	5	1%
		cryptomeridiol isomer 3	5	25%
LeuGraSQTS345	12%	Humulene	3	100%
LeuGraSQTS365	11%	alpha-guaiene	3	20%
		delta-guaiene	3	80%
LeuGraSQTS377	14%	delta-cadinene	3	98%
		alpha-cadinol	3	2%
LeuGraSQTS379	12%	delta-cadinene	3	98%
		alpha-cadinol	3	2%
LeuGraSQTS385	13%	Zingiberene	4	55%
		beta-bisabolene	2	19%
		beta-farnesene	1	6%

		beta-sesquiphellandrene	2	6%
		Cubenol	5	5%
		alpha-bisabolol	1	4%
		alpha-curcumene	5	3%
		trans-nerolidol	1	2%
LeuGraSQTS393	10%	gamma-bisabolene	4	100%

[1] The structure identification ranking key is defined in Table 3, with lower numbers indicating a higher degree of confidence.

[2] The composition of total sesquiterpenes from each chimera was a rough estimate based on a common ion count (m/z 204.2). The ratio of metabolites may have been different in the production strains and other minor metabolites may have been detected when samples were prepared. Representative GC/MS chromatograms for the chimeras with **bold font** can be found in FIG. 10 and FIG.11.

[3] Co-eluted under these run conditions. The peak was partially resolved under longer run conditions, about 6/4 alpha/beta-eudesmol.

Table 6. Functional sesquiterpene synthase chimeras derived from *M. villosa* sequences and their associated products.

Chimera name	% rare sequence	Terpene identification	Identification Method[1]	% composition[2]
MacVolSQTS1139	14%	alpha-guaiene	3	19%
		delta-guaiene	3	81%
MacVolSQTS2198	62%	beta-caryophyllene	1	85%
		Humulene	1	15%
MacVolSQTS2202	69%	beta-caryophyllene	1	86%
		Humulene	1	14%
MacVolSQTS2222	69%	beta-caryophyllene	1	86%
		Humulene	1	14%
MacVolSQTS2251	65%	beta-caryophyllene	1	87%
		Humulene	1	13%
MacVolSQTS2274	38%	unknown sesquiterpene	6	16%
		trans-Sesquisabinene hydrate	5	14%
		delta-elemene	6	34%
		unknown sesquiterpene	6	16%

[1] The structure identification ranking key is defined in Table 3, with lower numbers indicating a higher degree of confidence.

[2] The composition of total sesquiterpenes from each chimera was a rough estimate based on a common ion count (m/z 204.2). The ratio of metabolites may have been different in the production strains and other minor metabolites may have been detected when samples were prepared. Representative GC/MS chromatograms for the chimeras with **bold font** can be found in FIG. 12.

Table 7. Functional sesquiterpene synthase chimeras derived from *O. stipulatum* sequences and their associated products.

Chimera name	% rare sequence	Terpene identification	Identification Method[1]	% composition[2]
OrbStiSQTS1368	10%	gamma-bisabolene	5	100%
OrbStiSQTS1414	42%	alpha-guaiene	3	21%
		delta-guaiene	3	79%

[1] The structure identification ranking key is defined in Table 3, with lower numbers indicating a higher degree of confidence.

[2] The composition of total sesquiterpenes from each chimera was a rough estimate based on a common ion count (m/z 204.2). The ratio of metabolites may have been different in the production strains and other minor metabolites may have been detected when samples were prepared.

Table 8. Functional sesquiterpene synthase chimeras derived from *S. cuspidata* sequences and their associated products.

Chimera name	% rare sequence	Terpene identification	Identification Method[1]	% composition[2]
ShoCusSQTS154	38%	delta-cadinene	3	41%
		Sesquiterpene	6	41%
		alpha-cadinol	3	18%
ShoCusSQTS155	35%	delta-cadinene	3	41%
		Sesquiterpene	6	41%
		alpha-cadinol	3	18%

ShoCusSQTS156	36%	alpha-cadinol	2	34%
		delta-cadinene	2	25%
		beta-caryophyllene	1	10%
		tau-cadinol	2	10%
		Sesquiterpene	6	10%
		Sesquiterpene	6	7%
		Humulene	1	4%
ShoCusSQTS157	38%	alpha-cadinol	3	59%
		Sesquiterpene	6	25%
		tau-cadinol	3	16%
ShoCusSQTS160	36%	alpha-cadinol	3	33%
		Sesquiterpene	6	32%
		delta-cadinene	3	5%
ShoCusSQTS161	37%	delta-cadinene	3	36%
		alpha-cadinol	3	34%
		Sesquiterpene	6	12%
		tau-cadinol	3	10%
		beta-caryophyllene	3	5%
		Sesquiterpene	6	3%

[1] The structure identification ranking key is defined in Table 3, with lower numbers indicating a higher degree of confidence.

[2] The composition of total sesquiterpenes from each chimera was a rough estimate based on a common ion count (m/z 204.2). The ratio of metabolites may have been different in the production strains and it is possible other minor metabolites were detected when samples were prepared. Representative GC/MS chromatograms for the chimeras with **bold font** can be found in FIG. 13.

Table 9. Functional sesquiterpene synthase chimeras derived from *W. angustifolia* sequences and their associated products.

Chimera name	% rare sequence	Terpene identification	Identification Method[1]	% composition[2]
WenAngSQTS1007	81%	cis-eudesm-6-en-11-ol	4	100%
WenAngSQTS1086	80%	Daucene	5	5%
		isodaucene	5	6%
		sesquiterpene	6	4%
		cis-eudesm-6-en-11-ol	4	85%

WenAngSQTS267	11%	gamma-eudesmol	3	66%
		alpha/beta-eudesmol[3]	5	15%
		cryptomeridiol isomer 3	5	19%
WenAngSQTS302	17%	sesquiterpene	6	2%
		trans-bergamotene	4	5%
		alpha-zingiberene	4	56%
		sesquisabinene hydrate	4	20%
		beta-sesquiphellandrene	2	7%
		trans-nerolidol	1	2%
		sesquiterpenol	6	4%
		sesquiterpenol	6	4%
WenAngSQTS738	46%	Sesquiterpene	6	6%
		sesquiterpene	6	7%
		delta-cadinene	2	36%
		unidentified sesquiterpenol	4	27%
		tau-cadinol	3	15%
		alpha-cadinol	3	9%
WenAngSQTS760	43%	Sesquiterpene	6	9%
		Sesquiterpene	6	4%
		Sesquiterpene	6	6%
		delta-cadinene	2	41%
		sesquiterpenol	6	22%
		tau-cadinol	3	11%
		alpha/beta-eudesmol[3]	5	7%
WenAngSQTS780	41%	sesquiterpene	6	9%
		sesquiterpene	6	3%
		sesquiterpene	6	6%
		delta-cadinene	2	40%
		sesquiterpenol	6	24%
		tau-cadinol	3	11%
		alpha/beta-eudesmol[3]	5	7%
WenAngSQTS793	75%	Daucene	5	3%
		beta-farnesene	1	2%
		8-Isopropenyl-1,5-dimethyl-1,5-cyclodecadiene	4	5%
		sesquiterpene	6	3%
		cis-eudesm-6-en-11-ol	4	87%

WenAngSQTS805	42%	sesquiterpene	6	5%
		sesquiterpene	6	6%
		delta-cadinene	2	39%
		unidentified sesquiterpenol	4	27%
		tau-cadinol	3	15%
		alpha-cadinol	3	8%
WenAngSQTS826	47%	delta-cadinene	3	42%
		sesquiterpenol	6	36%
		tau-cadinol	3	22%
WenAngSQTS829	74%	cis-eudesm-6-en-11-ol	5	100%
WenAngSQTS843	45%	delta-cadinene	3	53%
		sesquiterpenol	6	47%
WenAngSQTS848	84%	cis-eudesm-6-en-11-ol	5	100%
WenAngSQTS849	75%	Daucene	4	3%
		beta-farnesene	1	1%
		isodaucene	4	8%
		sesquiterpene	6	2%
		cis-eudesm-6-en-11-ol	4	86%
WenAngSQTS864	81%	Daucene	5	2%
		8-Isopropenyl-1,5-dimethyl-1,5-cyclodecadiene	4	5%
		sesquiterpene	6	3%
		cis-eudesm-6-en-11-ol	4	90%
WenAngSQTS925	80%	sesquiterpene	6	3%
		sesquiterpene	6	8%
		sesquiterpene	6	3%
		cis-eudesm-6-en-11-ol	5	86%
WenAngSQTS960	81%	delta-cadinene	2	99%
		alpha-cadinol	3	1%

[1] The structure identification ranking key is defined in Table 3, with lower numbers indicating a higher degree of confidence. [2] The composition of total sesquiterpenes from each chimera was a rough estimate based on a common ion count (m/z 204.2). The ratio of metabolites may have been different in the production strains and it is possible other minor metabolites were detected when samples were prepared. Representative GC/MS chromatograms for the chimeras with **bold font** can be found in Appendix FIG. 14 and FIG. 15.

[3] Co-eluted under these run conditions. The peak was partially resolved under longer run conditions, about 6/4 alpha/beta-eudesmol.

Table 10. Amino acid (AA) and nucleic acid sequences of sesquiterpene chimeras.

Chimera Name	Rare DNA source	Extant scaffold Uniprot #	% Rare DNA	Chimera AA sequence (beginning after the first encoded amino acid M)	Chimera Nucleic Acid Sequence
HibWilS QTS117	*Hibiscadelphus wilderianus*	Q9SAN0	49%	ASQASQVLASPHPAISS ENRPKADFHPGIWGDM FIICPDTDIDAATELQYE ELKAQVRKMIMEPVDD SNQKLPFIDAVQRLGVS YHFEKEIEDELENIYRD TNNNDADTDLYTTALR FRLLREHGFDISCDAFN KLKDEEGNFKASLTSD VPGLLELYEASYLRVH GEDILDEAISFATAQLT LALPTLHHPLSEQVGH ALKQSIRRGLPRVEARN FISIYQDLESHNKALLQ FAKIDFNMLQLLHRKE LSEICRWWKDLDFTRK LPFARDRVVEGYFWIM GVYFEPQYSLGRKMLT KVIAMASIVDDTYDSFA TYDELIPYTDAIERWDI KCMNQLPNYMQISYKA LLDVYEEMEQLLADKG RQYRVEYAKKAMIRLV QAYLLEAKWTHLNYKP TFEEFRDNALPTSGYA MLAITAFVGMGEVITPE TFEWAASDPKIIKASTII CRFMDDIAEHKFNHRR EDDCSAIECYMEQYKV TAQEAYDEFNKHIESS WKDVNEEFLKPTEMPT PVLCRSLNLARVMDVL YREGDGYTHVGKAAK GGITSLLIDPIQI (SEQ ID NO: 1)	atggccagtcaggcttcacaagttttagcatctcc ccacccagctatatcctctgaaaaccggccaaag gctgatttccatcctggtatctggggcgacatgttt attatctgtccagatacggacattgatgccgctac agagctgcaatatgaagaattgaaagcgcaagtc cgcaagatgatcatggaaccagtagacgattcta atcaaaagctaccattcattgacgctgttcaaagg ctcggagtgagctaccactttgaaaaagaaattga agacgaacttgaaaacatctaccgtgataccaata acaacgacgcagacactgatctatacactaccgc cttgagattcagattattgagagagcatggttttgat atttcctgcgatgctttcaacaagttgaaagacga agaaggtaatttcaaggcttcgttgacttctgacgt tcctggtttgttagaactctatgaggcttcctacttg agagtccacggtgaagatatcctagatgaagcca tatctttcgctactgctcagttaaccttggctttgcc aactttgcatcacccgctttcagagcaagttggtc acgcattgaagcaaagtatcagaagaggcctgc caagagttgaagccagaaactttatctctatttacc aagatttagaatcccacaataaggctttgttgcaat tcgccaaaattgactttaacatgttacaattgctaca taggaaggagctcagcgaaatttgtagatggtgg aaagatcttgattttaccagaaagttacctttcgctc gtgaccgtgtcgtcgaaggttatttctggattatgg gagtttacttcgaaccacaatatagcttgggtaga aagatgttgaccaaggttattgctatggcttctatc gtcgatgatacatacgattccttcgctacttacgac gaattgataccatatactgacgccatcgaaagatg ggacatcaagtgtatgaatcagctgccaaactata tgcaaatttcgtacaaagcgttattggatgtatacg aggaaatggaacaattgcttgcagataaaggtcg acagtacagagtggaatacgctaagaaagctatg attcggttggtgcaagcatatttgttagaagcgaa gtggacccatttaaactacaagccaactttcgaag aatttagagacaatgctttgccgacatctgggtatg ccatgctagctataaccgcgttcgttggtatgggt gaagttatcacgccagaaacctttgaatgggctg cttctgacccaaagattattaaggcctccactatca tctgccgctttatggatgatatcgctgagcataagt tcaaccacagaagggaggatgactgttccgctat

					tgaatgttacatggagcaatacaaagtcacagctc aagaagcatacgacgaatttaacaagcacataga atcgtcttggaaggacgttaatgaagagttcttga aaccaactgaaatgcctactccggtactgtgtaga agtttgaacctagccagagtcatggatgttttgtac agagaaggtgacggttatactcatgttggaaaag ccgctaagggtggtataacatcacttcttatcgatc ccattcaaatctaa (SEQ ID NO: 67)
HibWilS QTS118	*Hibiscadelphus wilderianus*	Q9SAN0	50%	ASQASQVLASPHPAISS ENRPKADFHPGIWGDM FIICPDTDIDAATELQYE ELKAQVRKMIMEPVDD SNQKLPFIDAVQRLGVS YHFEKEIEDELENIYRD TNNNDADTDLYTTALR FRLLREHGFDISCEAFN KLKDEEGNFKASLTSD VRGLLELYQASYMRIH GEDILDEAISFTTAQLTL ALPTLDPPLSEQVGHAL KQSIRRGLPRVEARNFI SIYQDLESHNKALLQFA KIDFNMLQLLHRKELSE ICRWWKDLDFTRKLPF ARDRVVEGYFWIMGV YFEPQYSLGRKMLTKVI AMASIVDDTYDSFATY DELIPYTDAIERWDIKC MNQLPNYMQISYKALL DVYEEMEQLLADKGR QYRVEYAKKAMIRLVQ AYLLEAKWTHLNYKPT FEEFRDNALPTSGYAM LAITAFVGMGEVITPET FEWAASDPKIIKASTIIC RFMDDIAEHKFNHRRE DDCSAIECYMKQYGAT AQEAYDEFNKHIESSW KDVNEEFLKPTEMPTP VLCRSLNLARVMDVLY REGDGYTHVGKAAKG GITSLLIDPIQI (SEQ ID NO: 2)	atggccagtcaggcttcacaagttttagcatctcc ccacccagctatatcctctgaaaaccggccaaag gctgatttccatcctggtatctggggcgacatgttt attatctgtccagatacggacattgatgccgctac agagctgcaatatgaagaattgaaagcgcaagtc cgcaagatgatcatggaaccagtagacgattcta atcaaaagctaccattcattgacgctgttcaaagg ctcggagtgagctaccactttgaaaaagaaattga agacgaacttgaaaacatctaccgtgataccaata acaacgacgcagacactgatctatacactaccgc cttgagattcagattattgagagagcatggttttgat atttcctgcgaagctttcaacaagttgaaagacga agagggtaatttcaaggcttcgttgacttctgatgtt agaggtttgttagaactctatcaggcttcctacatg agaatccacggtgaagatattcttgatgaagccat atctttcaccactgctcaattaaccttggctttgcct actttggatcccccattgtcagagcaagtcggtca tgccctaaagcagagtataagaagaggcctacc aagagttgaagccagaaactttatctctatttacca agacttggaatcccacaataaggctttattgcaatt cgctaaaattgactttaacatgttacaattgctacat aggaaggagctcagcgaaatctgtcgttggtgga aagatcttgattttactagaaagttgcctttcgcacg ggaccgtgtcgttgaaggttatttctggattatggg agtttacttcgaaccacaatatagcttgggtagaa agatgttgaccaaggttattgctatggcttctatcgt cgatgatacatacgattccttcgctacatacgacg aattgatcccatatactgacgccattgaaagatgg gacatcaagtgtatgaatcaactgccaaactatat gcaaatttcgtacaaagcattattggatgtatacga ggaaatggaacaattgcttgcggataaaggtcgg cagtacagagtggaatacgctaagaaagctatga ttcgattggtacaagcatatttattagaagcgaagt ggactcacttgaactacaagccaaccttcgaaga atttagagacaatgctttaccgacatctgggtatgc tatgcttgctataaccgcgttcgttggtatgggtga agtcatcacgccagaaacttttgaatgggccgctt ctgacccgaagattatcaaggcttccactatcatct gccgctttatggatgatatcgctgagcataagttca accacagaagggaggatgactgttccgctattga atgttacatgaagcaatacggtgcaaccgcccaa gaggcatacgacgaatttaacaaacacatagaat cgtcttggaaggacgttaatgaagagttcttgaaa

					ccaactgaaatgcctactccagtgctgtgtagaag tttgaaccttgctagagtcatggatgttttgtacaga gaaggtgacggttatactcatgtcgggaaagccg ctaagggtggtataacctcattgctaattgatccca ttcaaatctaa (SEQ ID NO: 68)
HibWilS QTS120	*Hibiscadelphus wilderianus*	Q9SAN0	46%	ASQASQVLASPHPAISS ENRPKADFHPGIWGDM FIICPDTDIDAATELQYE ELKAQVRKMIMEPVDD SNQKLPFIDAVQRLGVS YHFEKEIEDELENIYRD TNNNDADTDLYTTALR FRLLREHGFDISCDAFN KLKDEEGNFKASLTSD VPGLLELYEASYLRVH GEDILDEAISFATAQLT LALPTLHHPLSEQVGH ALKQSIRRGLPRVEARN FISIYQDLESHNKALLQ FAKIDFNMLQLLHRKE LSEICRWWKDLDFTRK LPFARDRVVEGYFWIM GVYFEPQYSLGRKMLT KVIAMASIVDDTYDSFA TYDELIPYTDAIERWDI KCMNQLPNYMQISYKA LLDVYEEMEQLLADKG RQYRVEYAKKAMIRLV QAYLLEAKWTHLNYKP TFEEFRDNALPTSGYA MLAITAFVGMGEVITPE TFEWAASDPKIIKASTII CRFMDDIAEHKFNHRR EDDCSAIECYMKQYGA TAQEAYDEFNKHIESS WKDVNEEFLKPTEMPT PVLCRSLNLARVMDVL YREGDGYTHVGKAAK GGITSLLIDPIQI (SEQ ID NO: 3)	atggccagtcaggcttcacaagttttagcatctcc ccacccagctatatcctctgaaaaccggccaaag gctgatttccatcctggtatctggggcgacatgttt attatctgtccagatacggacattgatgccgctac agagctgcaatatgaagaattgaaagcgcaagtc cgcaagatgatcatggaaccagtagacgattcta atcaaaagctaccattcattgacgctgttcaaagg ctcggagtgagctaccactttgaaaaagaaattga agacgaacttgaaaacatctaccgtgataccaata acaacgacgcagacactgatctatacactaccgc cttgagattcagattattgagagagcatggttttgat atttcctgcgatgctttcaacaagttgaaagacga agaaggtaatttcaaggcttcgttgacttctgacgt tcctggtttgttagaactctatgaggcttcctacttg agagtccacggtgaagatatcctagatgaagcca tatctttcgctactgctcagttaaccttggctttgcc aactttgcatcacccgctttcagagcaagttggtc acgcattgaagcaaagtatcagaagaggcctgc caagagttgaagccagaaactttatctctatttacc aagatttagaatcccacaataaggctttgttgcaat tcgccaaaattgactttaacatgttacaattgctaca taggaaggagctcagcgaaatttgtagatggtgg aaagatcttgattttaccagaaagttacctttcgctc gtgaccgtgtcgtcgaaggttatttctggattatgg gagtttacttcgaaccacaatatagcttgggtaga aagatgttgaccaaggttattgctatggcttctatc gtcgatgatacatacgattccttcgctacttacgac gaattgataccatatactgacgccatcgaaagatg ggacatcaagtgtatgaatcagctgccaaactata tgcaaatttcgtacaaagcgttattggatgtatacg aggaaatggaacaattgcttgcagataaaggtcg acagtacagagtggaatacgctaagaaagctatg attcggttggtgcaagcatatttgttagaagcgaa gtggacccatttaaactacaagccaactttcgaag aatttagagacaatgctttgccgacatctgggtatg ccatgctagctataaccgcgttcgttggtatgggt gaagttatcacgccagaaaccttgaatgggctg cttctgacccaaagattattaaggcctccactatca tctgccgctttatggatgatatcgctgagcataagt tcaaccacagaagggaggatgactgttccgctat tgaatgttacatgaagcaatacggtgcaacagctc aagaggcatacgacgaatttaacaaacacataga atcgtcttggaaggacgtcaatgaagagttcttga aaccaactgaaatgcctactccggtactgtgtaga agtttgaacctagccagagtcatggatgttttgtac agagaaggtgacggttatactcatgttgggaaag

					ccgctaagggtggtataacatcacttcttatcgatc ccattcaaatctaa (SEQ ID NO: 69)
HibWilS QTS121	*Hibiscadelphus wilderianus*	Q9SAN0	50%	ASQASQVLASPHPAISS ENRPKADFHPGIWGDM FIICPDTDIDAATELQYE ELKAQVRKMIMEPVDD SNQKLPFIDAVQRLGVS YHFEKEIEDELENIYRD TNNNDADTDLYTTALR FRLLREHGFDISCEAFN KLKDEEGNFKASLTSD VRGLLELYQASYMRIH GEDILDEAISFTTAQLTL ALPTLDPPLSEQVGHAL KQSIRRGLPRVEARNFI SIYQDLESHNKSLLEFA KIDFNLLQLLHRKELSEI CRWWKDLDFTRKLPFA RDRVVEGYFWIMGVYF EPQYSLGRKMLTKVIA MASIVDDTYDSFATYD ELIPYTDAIERWDIKCM NQLPNYMQISYKALLD VYEEMEQLLADKGRQ YRVEYAKKAMIRLVQA YLLEAKWTHLNYKPTF EEFRDNALPTSGYAML AITAFVGMGEVITPETF EWAASDPKIIKASTIICR FMDDIAEHKFNHRRED DCSAIECYMEQYKVTA QEAYDEFNKHIESSWK DVNEEFLKPTEMPTPVL CRSLNLARVMDVLYRE GDGYTHVGKAAKGGIT SLLIDPIQI (SEQ ID NO: 4)	atggcctcacaggcttcccaagttttagcatctcct cacccagctatatcttccgaaaaccgtccaaagg ctgatttccatccaggtatctggggcgacatgttta ttatctgtccagatacagacattgatgccgctacc gagttgcaatatgaagaattgaaagcccaagtca gaaagatgatcatggaaccagttgacgattctaat caaaagttgcctttcattgacgctgtccaaagattg ggtgtttcataccactttgaaaaagaaattgaaga cgaattagaaaacatctacagagatactaataaca acgacgcagacactgatttgtacaccactgccttg agattcagattattgcgtgagcatggttttgatattt cttgcgaagctttcaacaagttgaaagacgaaga gggtaatttcaaggcttccttaacctctgatgtcag aggtttgttggaattgtatcaggcttcctacatgag aatccacggtgaagatattttggatgaagctatatc tttcacaactgctcaattaactttagctttaccaactt tggatcctccattgtctgagcaagttggtcatgcct tgaagcagtcaatacgtagaggtttgccaagagtt gaagccagaaactttatctctatttaccaagacttg gaatcccacaataagtctttattagaatttgctaaaa ttgatttcaacttattgcaattgttacacagaaagga gttgtccgaaatctgtagatggtggaaagacttgg attttaccagaaagttaccttcgctagagatcgtgt cgttgaaggttatttctggatcatgggtgtctacttc gaaccacaatactccttgggtagaaagatgttgac caaagttattgctatggcctctattgttgacgatact tatgactcatttgcaacctacgacgaattgatacca tatacagacgctattgaaagatgggatatcaagtg tatgaaccaattgccaaattatatgcaaatatcttac aaggctttgttagacgtttacgaggaaatggaaca attgttggctgataagggtagacaatatagagtcg agtacgcaaaaaaagccatgatcagattggttca ggcctacttattagaggctaagtggacccatttga actacaagcctacttttgaagagttcagagacaat gctttaccaacctccggttatgccatgttggctatc actgcattcgttggtatgggtgaagtcattacacca gaaacttttgaatgggctgcctctgatccaaagatt attaaggcttctactatcatctgccgtttcatggatg atattgctgaacacaaattcaaccacagaagaga ggacgattgttccgctattgaatgttacatggaaca atacaaggttacagcccaagaagcatacgacga atttaacaagcatatcgaatcatcttggaaggacg ttaatgaagaatttttaaagcctaccgaaatgccaa caccagtcttgtgtagatctttgaacttggccagag ttatggatgtcttgtaccgtgaaggtgatggttata ctcatgtcggtaaggctgctaaaggtggtatcacc tccttgttgatcgaccctattcaaatttaa (SEQ ID NO: 70)

HibWilS QTS123	*Hibiscadelphus wilderianus*	Q9SAN0	47%	ASQASQVLASPHPAISS ENRPKADFHPGIWGDM FIICPDTDIDAATELQYE ELKAQVRKMIMEPVDD SNQKLPFIDAVQRLGVS YHFEKEIEDELENIYRD TNNNDADTDLYTTALR FRLLREHGFDISCDAFN KLKDEEGNFKASLTSD VPGLLELYEASYLRVH GEDILDEAISFATAQLT LALPTLHHPLSEQVGH ALKQSIRRGLPRVEARN FISIYQDLESHNKSLLEF AKIDFNLLQLLHRKELS EICRWWKDLDFTRKLP FARDRVVEGYFWIMGV YFEPQYSLGRKMLTKVI AMASIVDDTYDSFATY DELIPYTDAIERWDIKC MNQLPNYMQISYKALL DVYEEMEQLLADKGR QYRVEYAKKAMIRLVQ AYLLEAKWTHLNYKPT FEEFRDNALPTSGYAM LAITAFVGMGEVITPET FEWAASDPKIIKASTIIC RFMDDIAEHKFNHRRE DDCSAIECYMEQYKVT AQEAYDEFNKHIESSW KDVNEEFLKPTEMPTP VLCRSLNLARVMDVLY REGDGYTHVGKAAKG GITSLLIDPIQI (SEQ ID NO: 5)	atggcctcacaggcttcccaagttttagcatctcct cacccagctatatcttccgaaaaccgtccaaagg ctgatttccatccaggtatctggggcgacatgttta ttatctgtccagatacagacattgatgccgctacc gagttgcaatatgaagaattgaaagcccaagtca gaaagatgatcatggaaccagttgacgattctaat caaaagttgcctttcattgacgctgtccaaagattg ggtgtttcataccactttgaaaaagaaattgaaga cgaattagaaaacatctacagagatactaataaca acgacgcagacactgatttgtacaccactgccttg agattcagattattgcgtgagcatggttttgatattt cttgcgatgctttcaacaagttgaaagacgaagaa ggtaatttcaaggcttccttaacctctgacgtccca ggtttgttggaattgtatgaggcttcctacttaaga gttcacggtgaagatatcttggatgaagctatatct ttcgccactgctcagttaaccttggctttaccaactt tgcatcacccattgtctgagcaagttggtcacgca ttgaagcaatcaatcagaagaggtttgccaagag ttgaagctagaaactttatctctatttaccaagattta gaatcccacaataagtctttattagaatttgccaaa attgatttcaacttgttgcaattgttacaccgtaagg agttgtccgaaatatgtagatggtggaaagactta gattttacaagaaagttacctttcgctagagataga gtcgttgaaggttatttctggattatgggtgtctactt cgaaccacaatactccttgggtagaaagatgttga ccaaagttattgctatggcttctatcgttgacgatac ttatgactcatttgccacttacgacgaattgatccct tatacagacgctattgaacgttgggatatcaagtgt atgaaccagttgccaaattatatgcaaatatcttac aaggctttgttagacgtttacgaggaaatggaaca attgttggctgataagggtagacaatatagagtcg agtacgccaaaaaagcaatgattagattggttcag gcctacttattagaggctaagtggacccatttgaa ctacaagcctacatttgaagagttcagagacaatg ctttaccaacttccggttatgccatgttggctataac cgcattcgttggtatgggtgaagtcattaccccag aaacttttgaatgggccgcttctgatccaaagatta tcaaggcttctactatcatctgccgtttcatggatga tattgccgaacataaattcaaccacagaagagag gacgattgttccgctattgaatgttacatggaacaa tacaaggttacagcccaagaagcttacgacgaat ttaacaagcacatcgaatcatcttggaaggacgtc aatgaagaatttttgaagcctaccgaaatgccaac tccagtcttgtgtagatctttgaacttggcaagagtt atggatgtcttgtacagagaaggtgatggttatact catgtcggtaaggctgctaaaggtggtatcacctc cttgttgatcgaccctattcaaatttaa (SEQ ID NO: 71)
HibWilS QTS124	*Hibiscadelphus*	Q9SAN0	48%	ASQASQVLASPHPAISS ENRPKADFHPGIWGDM FIICPDTDIDAATELQYE	atggcctcacaggcttcccaagttttagcatctcct cacccagctatatcttccgaaaaccgtccaaagg ctgatttccatccaggtatctggggcgacatgttta

	wilderianus			ELKAQVRKMIMEPVDD SNQKLPFIDAVQRLGVS YHFEKEIEDELENIYRD TNNNDADTDLYTTALR FRLLREHGFDISCEAFN KLKDEEGNFKASLTSD VRGLLELYQASYMRIH GEDILDEAISFTTAQLTL ALPTLDPPLSEQVGHAL KQSIRRGLPRVEARNFI SIYQDLESHNKSLLEFA KIDFNLLQLLHRKELSEI CRWWKDLDFTRKLPFA RDRVVEGYFWIMGVYF EPQYSLGRKMLTKVIA MASIVDDTYDSFATYD ELIPYTDAIERWDIKCM NQLPNYMQISYKALLD VYEEMEQLLADKGRQ YRVEYAKKAMIRLVQA YLLEAKWTHLNYKPTF EEFRDNALPTSGYAML AITAFVGMGEVITPETF EWAASDPKIIKASTIICR FMDDIAEHKFNHRRED DCSAIECYMKQYGATA QEAYDEFNKHIESSWK DVNEEFLKPTEMPTPVL CRSLNLARVMDVLYRE GDGYTHVGKAAKGGIT SLLIDPIQI (SEQ ID NO: 6)	ttatctgtccagatacagacattgatgccgctacc gagttgcaatatgaagaattgaaagcccaagtca gaaagatgatcatggaaccagttgacgattctaat caaaagttgcctttcattgacgctgtccaaagattg ggtgtttcataccactttgaaaaagaaattgaaga cgaattagaaaacatctacagagatactaataaca acgacgcagacactgatttgtacaccactgccttg agattcagattattgcgtgagcatggttttgatattt cttgcgaagctttcaacaagttgaaagacgaaga gggtaatttcaaggcttccttaacctctgatgtcag aggtttgttggaattgtatcaggcttcctacatgag aatccacggtgaagatattttggatgaagctatatc tttcacaactgctcaattaactttagctttaccaactt tggatcctccattgtctgagcaagttggtcatgcct tgaagcagtcaatacgtagaggtttgccaagagtt gaagccagaaactttatctctatttaccaagacttg gaatcccacaataagtctttattagaatttgctaaaa ttgatttcaacttattgcaattgttacacagaaagga gttgtccgaaatctgtagatggtggaaagacttgg attttaccagaaagttacctttcgctagagatcgtgt cgttgaaggttatttctggatcatgggtgtctacttc gaaccacaatactccttgggtagaaagatgttgac caaagttattgctatggcctctattgttgacgatact tatgactcatttgcaacctacgacgaattgatacca tatacagacgctattgaaagatgggatatcaagtg tatgaaccaattgccaaattatatgcaaatatcttac aaggctttgttagacgtttacgaggaaatggaaca attgttggctgataagggtagacaatatagagtcg agtacgcaaaaaaagccatgatcagattggttca ggcctacttattagaggctaagtggacccatttga actacaagcctacttttgaagagttcagagacaat gctttaccaacctccggttatgccatgttggctatc actgcattcgttggtatgggtgaagtcattacacca gaaacttttgaatgggctgcctctgatccaaagatt attaaggcttctactatcatctgccgtttcatggatg atattgctgaacacaaattcaaccacagaagaga ggacgattgttccgctattgaatgttacatgaaaca atacggtgctacagcccaagaagcatacgacga atttaacaagcatatcgaatcatcttggaaggacg ttaatgaagaatttttaaagcctaccgaaatgccaa caccagtcttgtgtagatctttgaacttggcaagag ttatggatgtcttgtaccgtgaaggtgatggttata ctcatgtcggtaaggctgctaaaggtggcatcac ctccttgttgatcgaccctattcaaatttaa (SEQ ID NO: 72)
HibWilSQTS126	*Hibiscadelphus wilderianus*	Q9SAN0	44%	ASQASQVLASPHPAISS ENRPKADFHPGIWGDM FIICPDTDIDAATELQYE ELKAQVRKMIMEPVDD SNQKLPFIDAVQRLGVS YHFEKEIEDELENIYRD	atggcctcacaggcttcccaagttttagcatctcct cacccagctatatcttccgaaaaccgtccaaagg ctgatttccatccaggtatctggggcgacatgttta ttatctgtccagatacagacattgatgccgctacc gagttgcaatatgaagaattgaaagcccaagtca gaaagatgatcatggaaccagttgacgattctaat

				TNNNDADTDLYTTALR FRLLREHGFDISCDAFN KLKDEEGNFKASLTSD VPGLLELYEASYLRVH GEDILDEAISFATAQLT LALPTLHHPLSEQVGH ALKQSIRRGLPRVEARN FISIYQDLESHNKSLLEF AKIDFNLLQLLHRKELS EICRWWKDLDFTRKLP FARDRVVEGYFWIMGV YFEPQYSLGRKMLTKVI AMASIVDDTYDSFATY DELIPYTDAIERWDIKC MNQLPNYMQISYKALL DVYEEMEQLLADKGR QYRVEYAKKAMIRLVQ AYLLEAKWTHLNYKPT FEEFRDNALPTSGYAM LAITAFVGMGEVITPET FEWAASDPKIIKASTIIC RFMDDIAEHKFNHRRE DDCSAIECYMKQYGAT AQEAYDEFNKHIESSW KDVNEEFLKPTEMPTP VLCRSLNLARVMDVLY REGDGYTHVGKAAKG GITSLLIDPIQI (SEQ ID NO: 7)	caaaagttgcctttcattgacgctgtccaaagattg ggtgtttcataccactttgaaaaagaaattgaaga cgaattagaaaacatctacagagatactaataaca acgacgcagacactgatttgtacaccactgccttg agattcagattattgcgtgagcatggttttgatattt cttgcgatgctttcaacaagttgaaagacgaagaa ggtaatttcaaggcttccttaacctctgacgtccca ggtttgttggaattgtatgaggcttcctacttaaga gttcacggtgaagatatcttggatgaagctatatct ttcgccactgctcagttaaccttggctttaccaactt tgcatcacccattgtctgagcaagttggtcacgca ttgaagcaatcaatcagaagaggtttgccaagag ttgaagctagaaactttatctctatttaccaagattta gaatcccacaataagtctttattagaatttgccaaa attgatttcaacttgttgcaattgttacaccgtaagg agttgtccgaaatatgtagatggtggaaagactta gattttacaagaaagttacctttcgctagagataga gtcgttgaaggttatttctggattatgggtgtctactt cgaaccacaatactccttgggtagaaagatgttga ccaaagttattgctatggcttctatcgttgacgatac ttatgactcatttgccacttacgacgaattgatccct tatacagacgctattgaacgttgggatatcaagtgt atgaaccagttgccaaattatatgcaaatatcttac aaggctttgttagacgtttacgaggaaatggaaca attgttggctgataagggtagacaatatagagtcg agtacgccaaaaaagcaatgattagattggttcag gcctacttattagaggctaagtggacccatttgaa ctacaagcctacatttgaagagttcagagacaatg ctttaccaacttccggttatgccatgttggctataac cgcattcgttggtatgggtgaagtcattaccccag aaacttttgaatgggccgcttctgatccaaagatta tcaaggcttctactatcatctgccgtttcatggatga tattgccgaacataaattcaaccacagaagagag gacgattgttccgctattgaatgttacatgaaacaa tacggtgctacagcccaagaagcatacgacgaa tttaacaagcacatcgaatcatcttggaaggacgt taatgaagaatttttgaagcctaccgaaatgccaa ctccagtcttgtgtagatctttgaacttggccagag ttatggatgtcttgtacagagaaggtgatggttata ctcatgtcggtaaggctgctaaaggtggcatcac ctccttgttgatcgaccctattcaaatttaa (SEQ ID NO: 73)
HibWilS QTS19	*Hibiscadelphus wilderianus*	A0A067FTE8	12%	SIQVPQISSQNAKSQVM RRTANFHPSVWGDRFA NYTAEDKMNHARDLK ELKALKEEVGRKLLAT AGPIQLNLIDAIQRLGV GYHFERELEQALQHLY NEKYSDDDTEDDLYRIS LRFRLLRQHGYNVSCD KFNMFKDDKGNFKESL ASDALGMLSLYEAAHL	atgtccatacaggttccccaaatttcttcgcaaaat gcaaagtcacaagtaatgcgtagaaccgccaact ttcatccatctgtgtggggagacagattcgctaact acacggctgaggataaaatgaaccacgctcgcg acttgaaggaacttaaagcgttaaaggaagaagt tggtagaaagctgttggccacagctggcccaatt caactcaatctaatcgatgctatccaaagattgggt gtcggttatcacttcgaacgagaattggaacaag ctttgcaacatttatacaacgagaagtatagcgat

				GVHGEDILDEAIAFTTT HLKSVATHLSNPLKAQ VRHALRQPLHRGLPRL EHRRYISIYQDDASHYK ALLTLAKLDFNLVQSL HKKELCEISRWWKDLD FARKLPFARDRMVECY FWILGVYFEPNYSLARR ILTKVIAMTSIIDDIYDV YGTPEELKLFTEVIERW DESSMDQLPEYMQTFF GALLDLYNEIEKEIANE GWSYRVQYAKEAMKI LVEGYYDESKWFHENY IPKMEEYMRVALVTSG YTMLTTVSFLGMDNIV TKETFDWVFSRPKIIRA SEIIGRFMDDIKSHKFEQ ERGHCASAVECYMREH GVSEEEACSELKKQVD NAWKDINHEMIFSETSK AVPMSVLTRVLNLTRVI DVVYKEGDGYTHVGN EMKQNVAALLIDQVPI (SEQ ID NO: 8)	gacgacactgaagatgatttgtacaggatttctctg agatttagattgttaagacagcacggttacaatgtc tcctgcgacaaattcaacatgtttaaggatgacaa aggtaacttcaaggaaagtttggcttctgatgcctt gggtatgctctccttatacgaagcggctcatttgg gcgttcacggtgaagatatcttagacgaagctatt gcatttaccactactcatctaaagtccgtcgctact cacttatctaatcctctaaaggcccaagttcgtcat gccttgagacaaccgcttcacagaggtttgccaa gattggaacacagaaggtatatcagcatttaccag gatgacgcttctcattacaaagctttgttgacccttg cgaagttggatttcaatctagttcaatcattgcaca aaaaggagctatgtgagatctccagatggtggaa ggatttagacttcgctcgtaagttgccttttgctaga gatagaatggtcgaatgttatttctggatcttgggt gtgtatttcgaaccaaactactcactggcccggag aatattgaccaaagttattgctatgacttctattattg atgacatctatgacgtttacgggacaccagaaga attgaagttgttcactgaagtaatcgaacgttggg acgaatcgtcaatggaccaactaccagaatacat gcaaacgtttttcggtgctcttttagatttatacaatg agatagaaaaggaaattgccaacgaaggttggtc ttacagagtccaatatgcaaaagaagctatgaag attttagttgagggttactacgatgaatctaagtggt tccatgaaaactacataccaaagatggaggaata tatgcgggtagcattagttaccagcggatacaca atgttgactaccgtcagttttctggggatggacaa cattgttactaaggagacatttgattgggttttctcc agacctaaaatcataagagcatcagaaattatcg gtagattcatggacgatattaaatctcacaaattcg aacaggaaagaggtcactgtgcgtccgctgtcg aatgttatatgagggaacatggcgtgtctgaaga ggaagcttgcagtgagctcaagaagcaagtcga taacgcctggaaggacatcaaccacgaaatgatt ttctccgaaacttctaaggctgttcctatgagcgtg ctaaccagagttttgaacttgacgagagttattgat gtcgtctacaaggaaggtgatggttatactcatgt gggtaatgaaatgaaacaaaacgttgctgctctttt gatcgaccaagtcccaatttaa (SEQ ID NO: 74)
HibWilS QTS34	*Hibiscadelphus wilderianus*	B1B1U4	13%	EKQSLTFDGDEEAKIDR KSSKYHPSIWGDYFIQN SSLTHAKESTQRMIKRV EELKVQVKSMFKDTSD LLQLMNLINSIQMLGLD YHFENEIDEALRLIYEV DDKSYGLYETSLRFQLL RQHGYHVDGEEAFNM LKDEEGNFKASLTSDVP GLLELYQASYMRIHGE DILDEAISFTTAQLTLAL PTLDPPLSAQVSLFLELP	atggaaaagcagtccttgacatttgatggcgacg aggaagcaaaaatagatcgtaagtcgtcaaagta ccatcctagtatttggggtgactatttcatccaaaat tccagcttaacccacgccaaagaatctactcaaa ggatgatcaagagagttgaagaactaaaggtaca agtcaaatctatgttcaaggacacttctgatttgttg caactgatgaacttaattaactctattcaaatgctag gacttgactaccactttgaaaatgaaatcgatgag gctctccgcttgatctatgaagttgacgataagtca tacggtctgtacgaaacgagcttgagattccagtt gttgagacaacatggttaccacgtggatggtgaa

				LCRRNKILLARKYILIY QEDAMRNNVILELAKL NFNLLQSLYQEELKKISI WWNDLAFAKSLSFTRD RVVEGYYWVLTIYFEP QHSRARVICSKVFAFLS IMDDIYDNYGILEECTL LTEAIKRWNPQAIDGLP EYLKDYYLKLLKTFEEF EDELELNEKYRMLYLQ DEVKALAISYLQEAKW GIERHVPSLDEHLHNSL ISSGSSTVICASFVGMG EVATKEVFDWLSSFPK VVEACCVIGRLLNDIRS HELEQGRDHTASTVES YMKEHDTNVDVACEK LREIVEKAWKDLNNES LNPTKVPRLMIERIVNL SKSNEEIYKYNDTYTNS DTTMKDNISLVLVESC DYFNK (SEQ ID NO: 9)	gaagctttcaacatgcttaaagacgaagagggta actttaaggcgtccttgacctctgatgttccaggttt attggaattatatcaagctagctacatgagaataca tggtgaagatattttggatgaagccattagtttcact accgctcaattgactttagctcttcccaccttagac ccgccattgtcggcacaagtctctttgttcttggag ctaccattatgcagaagaaacaagattttgcttgcc agaaaatacatcttgatatatcaagaagatgctatg cgtaataatgttattctcgagttggctaagcttaact ttaacttattgcaatccttgtaccaagaagaactga agaaaatctctatctggtggaatgacttagcttttg caaagtctttatctttcactagagatagagtcgttga aggttattactgggtcctaaccatctacttcgaacc acagcactcccgagctagggtcatttgttcaaaag tttttgccttttgtccattatggatgacatttatgaca actatggaatccttgaagaatgtacattattaacag aagctattaagagatggaacccacaagccatcga cgggttgcctgaatacctaaaagactattacttga agttgttgaagactttcgaggaatttgaagatgagt tggaattgaatgagaagtacagaatgctgtatttg caagatgaagttaaagctctggctatctcatactta caagaggccaagtggggtattgaaagacacgta ccatcgttagatgagcatcttcacaattctttgataa gttccggctcttcgactgtgatttgtgctagcttcgt tggtatgggtgaagttgccacgaaggaagtcttc gattggttgtcctctttcccaaaggttgtcgaagctt gttgtgtcatcggtaggctcttgaacgatattcgttc ccatgaattagagcagggcagagaccacacgg cttccactgttgaatcttacatgaaggaacacgac accaatgtggacgttgcctgcgaaaagttgagag aaatcgtcgaaaaggcgtggaaagatctgaaca acgaatctctaaaccctactaaggttccaagattg atgatagaaagaatagtaaacttgtcaaagtccaa cgaagaaatttacaaatacaacgacacctacact aattctgatactacaatgaaggacaatattagtcta gtattggttgagtcctgtgattatttcaacaaataa (SEQ ID NO: 75)
HibWilS QTS52	*Hibiscadelphus wilderianus*	Q39760	12%	ASQVSQMPSSSPLSSNK DEMRPKADFQPSIWGD LFLNCPDKNIDAETEKR HQQLKEEVRKMIVAPM ANSTQKLAFIDSVQRLG VSYHFTKEIEDELENIY HNNNDAENDLYTTSLR FRLLREHGFNVSCDVF NKFKDEQGNFKSSVTS DVRGLLELYQASYLRV HGEDILDEAISFTTNHLS LAVASLDYPLSEEVSHA LKQSIRRGLPRVEARHY LSVYQDIESHNKVLLEF AKIDFNMVQLLHRKEL	atggccagtcaggtttcacaaatgccttcctcttct ccactatccagcaacaaagatgagatgagacca aaggctgactttcaaccctcgatatggggcgattt gttcctgaattgcccagacaagaacattgatgctg aaaccgaaaagcgtcatcaacaattgaaagaag aagtcagaaagatgatcgtggcaccaatggctaa ttctacacaaaagttggctttcattgactctgttcag aggcttggagtatcctaccactttactaaagaaatt gaggatgaattagaaaacatctatcacaacaataa cgacgcagaaaacgatttgtacacgacttcccta agattcagattattgagagaacatggtttcaatgtc tcttgtgacgtttttaacaagtttaaggatgagcaa ggtaatttcaagtcaagtgttacctctgacgtccgc ggtctcttggaattataccaagcgtcgtatttgaga

				SEISRWWKDLDFQRKL PYARDRVVEGYFWISG VYFEPQYSLGRKMLTK VIAMASIVDDTYDSYA TYEELIPYTKAIERWDI KCIDELPEYMKPSYKAL LDVYEEMEQLVAKHG RQYRVEYAKNAMIRLA QSYLVEARWTLQNYKP SFEEFKANALPTCGYA MLAITSFVGMGDIVTPE TFKWAANDPKIIQASTII CRFMDDVAEHKFEQER GHCASAVECYMREHG VSEEEACSELKKQVDN AWKDINHEMIFSETSKA VPMSVLTRVLNLTRVM DVLYREGDGYTYVGK AAKGGITSLLIEPVAL (SEQ ID NO: 10)	gttcacggtgaagatatcttggacgaagctatttc gttcacaactaatcatctctctttggccgttgcttcct tagattaccctctgtctgaagaggtctctcacgcttt gaagcaaagcataagacgtggtcttccaagagta gaagccagacactatttgagcgtttaccaagatat cgaatctcataacaaagtcttgttagaatttgctaa gattgacttcaacatggttcaattgctacataggaa agagctaagtgaaatttcaagatggtggaaagat ctcgattttcaaagaaagttaccttatgcacgcgac cgtgtagtcgaaggttacttctggatctccggggtt tacttcgaaccacaatacagcttgggtagaaagat gttgactaaggttattgctatggcttctatcgttgat gatacctatgactcctacgccacctacgaggaatt gatcccatatactaaggccattgaaagatgggac atcaagtgtatagacgaactgccagaatatatgaa gcctagttacaaagctttattggatgtctatgagga aatggaacaattggtcgccaaacacggtcgaca gtacagagtggaatacgctaagaatgctatgattc gattggcgcaatcctacttggttgaagcgagatg gactcttcaaaactacaagccatctttcgaagaatt taaggccaatgctttaccgacatgtggatatgctat gctagctataaccagcttcgttggtatgggtgatat tgtcacgccagaaacttttaaatgggctgcaaatg acccgaagattatccaggcttctactatcatctgcc gatttatggatgatgtagctgagcataagttcgaa caagaaaggggcactgtgcttccgctgtcgagt gttacatgagagaacacggtgtgtcagaagaag aggcatgttctgaattgaaaaagcaagtcgacaa cgcctggaaggacattaaccatgaaatgatttttc ggaaacctccaaagctgtcccaatgtcggttctca ctagagttcttaacttgactagagttatggacgtatt gtacagagaaggtgatggttatacatatgttggta aggctgcaaagggcggtatcacctctttattgatt gaaccagttgccttgtaa (SEQ ID NO: 76)
HibWilS QTS54	*Hibiscadelphus wilderianus*	Q39761	13%	ASQVSQMPSSSPLSSNK DEMRPKADFQPSIWGD LFLNCPDKNIDAETEKR HQQLKEEVRKMIVAPM ANSTQKLAFIDSVQRLG VSYHFTKEIEDELENIY HNNNDAENDLYTTSIRF RLLREHGYHVDGEEAF NMLKDEEGNFKASLTS DVPGLLELYQASYMRI HGEDILDEAISFTTAQL TLALPTLDPPLSEEVSH ALKQSIRRGLPRVEARH YLSVYQDIESHNKALLE FAKIDFNMLQFLHRKEL SEICRWWKDLDFQRKL PYARDRVVEGYFWISG	atggccagtcaggtttcacaaatgccttcctcttct ccactatccagcaacaaagatgagatgagacca aaggctgactttcaaccctcgatatggggcgatttt gttcctgaattgcccagacaagaacattgatgctg aaaccgaaaagcgtcatcaacaattgaaagaag aagtcagaaagatgatcgtggcaccaatggctaa ttctacacaaaagttggctttcattgactctgttcag aggcttggagtatcctaccactttactaaagaaatt gaggatgaattagaaaacatctatcacaacaataa cgacgcagaaaacgatttgtacacgacttccata agattcagattattgagagaacatggttaccacgt cgatggtgaggaagccttcaacatgctcaaggac gaagaaggtaattttaaggcttctttgacctcagac gttcctggtttgttagaactatatcaagcctcataca tgcgaatccatggtgaagatattttggacgaagcg atctcttttactactgctcaattaaccttggctttgcc

				VYFEPQYSLGRKMLTK VIAMASIVDDTYDSYA TYEELIPYTNAIERWDI KCIDEIPEYMKPSYKAL LDVYEEMVQLVAEHG RQYRVEYAKNAMIRLA QSYLVEAKWTLQNYKP SFEEFKANALPTCGYA MLAITSFVGMGDIVTPE TFKWAASDPKIIQASTII CRFMDDVAEHKFKHRR EDDCSAIECYMEEYGV TAQEAYDVFNKHVESA WKDLNQEFLKPTEMPT EVLNRSLNLARVMDVL YREGDGYTYVGKAAK GGITSLLIEPIAL (SEQ ID NO: 11)	aaccctggatccaccgctctctgaagaggtcagt cacgcgctaaagcaaagtattagaagaggtttac cacgtgtagaagctagacattatctgtccgtttacc aagacatcgaatctcacaataaagctctattggaa tttgccaagattgatttcaacatgttgcagttcctcc acagaaaggaactttcagaaatatgtcgttggtgg aaagatttggacttccaacgcaagttaccatatgct agagatcgcgttgtcgagggttacttctggatcag cggagtttactttgagccacaatacagtttgggtc ggaagatgttaactaaagttattgctatggcttctat tgtcgatgacacatatgactcctacgccacctacg aagaattaatcccttatactaacgccatcgaaaga tgggacattaagtgtatcgatgaaattccggaata catgaaaccatcttacaaagctttgcttgacgtcta cgaagaaatggtacaattggttgctgagcatggta ggcaatacagagttgaatatgcaaagaatgccat gattagattggctcaatcttacttggtggaagcaaa gtggacgttgcaaaattacaaacctagctttgagg aatttaaggcgaacgctctgcccacctgtgggtat gccatgctggcaattacttccttcgttggtatgggc gacattgtcactcctgaaacattcaaatgggctgc atccgatccaaagatcattcaagcttcgacgataa tctgtcgattcatggatgatgtcgctgagcacaag ttcaagcacaggagagaagatgactgttctgcca tagaatgttacatggaagaatacggtgttaccgcc caggaggcttacgatgtcttcaacaagcacgttg aatccgcgtggaaagatttgaaccaagaatttctc aagccaactgaaatgccaacagaggtgttgaac agatcacttaacctcgctcgtgttatggacgtattg tatagagaaggtgatggttatacttacgttggtaag gctgctaagggcggtatcacctctttattgatcgaa ccaatcgctttgtaa (SEQ ID NO: 77)
HibWilS QTS55	*Hibiscadelphus wilderianus*	Q43714	12%	ASQASQVLASPHPAISS ENRPKADFHPGIWGDM FIICPDTDIDAATELQYE ELKAQVRKMIMEPVDD SNQKLPFIDAVQRLGVS YHFEKEIEDELENIYRD TNNNDADTDLYTTALR FRLLREHGFDISCDAFN KFKDEAGNFKASLTSD VQGLLELYEASYMRVH GEDILDEAISFTTAQLTL ALPTLHHPLSEQVGHA LKQSIRRGLPRVEARNF ISIYQDLESHNKSLLQF AKIDFNLLQLLHRKELS EICRWWKDLDFTRKLP FARDRVVEGYFWIMGV YFEPQYSLGRKMLTKVI AMASIVDDTYDSYATY DELIPYTNAIERWDIKC	atggccagtcaggcttcacaagttttagcatctcc ccacccagctatatcctctgaaaaccggccaaag gctgatttccatcctggtatctggggcgacatgttt attatctgtccagatacggacattgatgccgctac agagctgcaatatgaagaattgaaagcgcaagtc cgcaagatgatcatggaaccagtagacgattcta atcaaaagctaccattcattgacgctgttcaaagg ctcggagtgagctaccactttgaaaaagaaattga agacgaacttgaaaacatctaccgtgataccaata acaacgacgcagacactgatctatacactaccgc cttgagattcagattattgagagagcatggttttgat atttcctgcgatgctttcaacaagttcaaagacgaa gctggtaatttcaaggcttcgttgacttctgacgttc aaggtttgttggaattgtatgaggcctcctacatga gagtccacggtgaagatatcctagatgaagctat atcttttaccactgctcagttaaccttggctttaccta ctttgcatcacccgttgtcagagcaagttggtcac gcactcaagcagagtatcagaagaggcctgcca agagttgaagccagaaactttatctctatttaccaa

				MNQLPNYMKISYKALL NVYEEMEQLLANQGR QYRVEYAKKAMIRLVQ AYLLEAKWTHQNYKPT FEEFRDNALPTSGYAM LAITAFVGMGEVITPET FKWAASDPKIIKASTIIC RFMDDIAEHKFEQERG HCASAVECYMREHGVS EEEACSELKKQVDNAW KDINHEMIFSETSKAVP MSVLTRVLNLTRVMDV LYREGDGYTHVGKAA KGGITSLLIDPIQI (SEQ ID NO: 12)	gatttggaatcccacaataagtccttgttacaattc gctaaaattgactttaacctttacaattgctccata ggaaggaactcagcgaaatttgtagatggtggaa agatcttgatttcactagaaagttgccttttgcacgt gaccgtgtcgtcgaaggttatttctggattatggga gtttacttcgaaccacaatatagcttgggtagaaa gatgttgaccaaggttattgctatggcttctatcgtc gatgatacatacgattcttacgctacatatgacgaa ttgataccatatactaacgccatcgaaagatggga catcaagtgtatgaatcaactgccaaactacatga agattagttacaaagcattattgaatgtatatgagg agatggaacaattgcttgcgaatcaaggtcgaca gtacagagtggaatacgctaagaaagctatgatt cggttggtgcaagcctacttattagaagcgaagtg gactcatcaaaactacaagccaaccttcgaagaa tttagagacaatgctttgccgacatcagggtatgct atgctagctataaccgcgttcgttggtatgggtga agttatcacgccagaaacttttaaatgggccgctt ctgacccaaagattattaaggcttccactatcatct gccgctttatggatgatatcgctgagcataagttc gagcaagaaaggggcactgtgcttccgctgtc gaatgttacatgagagaacacggtgtctcagaag aagaggcctgttctgaattgaaaaagcaggtcga caacgcctggaaggatattaaccatgagatgattt ttagtgaaacatccaaagctgtcccaatgagtgtt ctaaccagagttttgaaccttactagagttatggac gtattgtacagagaaggtgatggttatacgcatgt cggtaaggctgcaaagggtggtatcacctctttgt tgattgaccccattcaaatctaa (SEQ ID NO: 78)
HibWilS QTS63	*Hibiscadelphus wilderianus*	Q9FQ26	12%	AASFANKCRPLANFHP TVWGYHFLYYNPEITN QEKIEVDEYKETIRKML VEAPEGSEQKLVLIDA MQRLGVAYHFHNEIET SIQNIFDAPKQNNDDNL HIVSLRFRLVRQQGHY MSSDVFKQFTNQDGKF KETLTNDVQGLLSLYE ASHLRVRNEEILEEALT FTTTHLESIVSNLSNKN NSLKVEVSEALSQPIRM TLPRIGARKYISIYENND AHNHLLLKFAKLDFNM LQKFHQRELSDLTRWW KDLDFANKIPYARDRL VECYFWILGVYFEPKYS RARKMMTKVLKMTSII DDTFDAYANFDELVPF NDAIQRWDANAIDSIPP YMRPIYQALLDIYGEM DQVLSKEGKLDRVYYA KYEMKKLVRAYFKESQ	atggccgcatcatttgctaacaaatgtagaccttta gctaatttccacccaactgtttggggttaccatttct tgtattacaacccagagataaccaatcaggaaaa gatcgaagtcgatgaatacaaggaaacaattcgt aagatgttggttgaagcccctgaagggtccgagc aaaaattggtcttaatcgacgctatgcaaagattg ggtgttgcatatcactttcataacgaaattgaaacc tctattcaaaatatcttcgatgctccaaagcaaaac aacgacgataacttgcacattgtctctttaagattc agattggtccgtcaacagggtcattacatgtcctct gacgtttttaagcaattcactaaccaagatggtaaa ttcaaggaaaccttgactaatgatgtccaaggtttg ttgtcattatatgaagcttctcacttgagagttagaa atgaagaaatattagaggaagctttgacttttacca caactcatttggaatccatcgtttctaacttatcaaa caaaaataactctttaaaggttgaagtttctgaagc tttgtcccaaccaatcagaatgactttgccaagaat tggtgccagaaagtacatttccatatacgaaaaca atgacgcccacaaccatttgttgttaaagttcgcta agttggattttaatatgttacaaaagttccaccaaa gagaattgtccgacttgaccagatggtggaaaga

				WLNDDNHIPKYEEHME NAIVTVGYMMGATNC LVGMEEFISKETFEWL MSEPVIVRASSLIGRAM DDIVGHEVEQERGHCA SAVECYMREHGVSEEE ACSELKKQVDNAWKDI NHEMIFSETSKAVPMSV LTRVLNLTRVIDTLYQE EDEYTNAKGKLKNMIH SILIESVKI (SEQ ID NO: 13)	cttggactttgctaacaagatcccatatgctagag atcgtttagtcgagtgctatttttggattttgggtgttt acttcgaacctaaatactctcgtgctagaaagatg atgaccaaggtcttgaaaatgacatctattattgat gatacttttgatgcttacgccaatttcgacgaattg gttccattcaatgacgccatccaaagatgggacg ctaacgcaatcgattctattccaccatacatgcgtc caatctaccaggccttgttagatatatatggtgaaa tggaccaagttttatccaaagagggtaagttggat agagtctactatgctaagtatgagatgaaaaagtt ggtcagagcctactttaaggaatctcaatggttaa acgacgataatcatatacctaagtatgaagaacac atggaaaacgctattgttactgtcggttacatgatg ggtgctacaaactgtttggttggtatggaggaattt atctcaaaagaaaccttcgaatggttgatgtcaga accagttattgttagagcatcttccttgataggtag agcaatggatgatatcgtcggtcacgaggttgaa caagaacgtggtcattgtgcttcagcagtcgaatg ttacatgagagagcatggtgtttctgaagaagaag cttgctccgaattaaagaagcaagttgacaacgct tggaaggacattaaccacgagatgatcttctctga aacttctaaagctgtcccaatgtctgtcttaaccag agttttaaacttgacaagagttattgatactttgtac caggaagaagatgaatacaccaacgctaagggt aaattaaaaaatatgatccactccatcttgattgagt cagtcaagatctaa (SEQ ID NO: 79)
HibWilS QTS90	*Hibiscadelphus wilderianus*	B1B1U4	25%	EKQSLTFDGDEEAKIDR KSSKYHPSIWGDYFIQN SSLTHAKESTQRMIKRV EELKVQVKSMFKDTSD LLQLMNLINSIQMLGLD YHFENEIDEALRLIYEV DDKSYGLYETSLRFQLL RQHGYHVDGEEAFNM LKDEEGNFKASLTSDVP GLLELYQASYMRIHGE DILDEAISFTTAQLTLAL PTLDPPLSAQVSLFLELP LCRRNKILLARKYILIY QEDAMRNNVILELAKL NFNLLQSLYQEELKKISI WWNDLAFAKSLSFTRD RVVEGYYWVLTIYFEP QHSRARVICSKVFAFLS IMDDIYDNYGILEECTL LTEAIKRWNPQAIDGLP EYLKDYYLKLLKTFEEF EDELELNEKYRMLYLQ DEVKALAISYLQEAKW GIERHVPSLDEHLHNSL ISSGSSTVICASFVGMG EVATKEVFDWLSSFPK VVEACCVIGRLLNDIRS	atggaaaagcagtctttgacatttgatggtgacga ggaagcaaaaatagatcgtaagtcatccaagtac catccttctatttggggcgactatttcatccaaaatt cctctttaacccacgccaaagaatctactcaaaga atgatcaagagagttgaagaattgaaggtccaag ttaaatcaatgttcaaggacacttccgatttattgca attgatgaacttaattaactctattcaaatgttgggtt tggactaccactttgaaaatgaaatcgatgaggct ttgagattgatctatgaagtcgacgataagtcctac ggtttgtacgaaacatcattaagattccagttgttaa gacaacatggttaccacgttgatggtgaagaagc tttcaacatgttgaaggatgaggaaggtaactttaa agcttctttaacctccgacgttccaggtttgttaga gttgtatcaagcctcttacatgcgtattcatggtga agatatattggatgaagctatttcattcactaccgct caattaactttggctttgccaactttagacccaccat tgtccgcacaagtctctttgttcttggagttgccatt atgcagaagaaacaagattttgttggccagaaaat acatcttgatatatcaagaagatgctatgcgtaata atgttattttggagttagccaagttgaactttaactta ttgcaatctttataccaagaagaattgaagaaaatc tctatctggtggaatgacttagcttttgctaagtcttt atctttcaccagagatagagtcgttgaaggttatta ctgggtcttgactatctacttcgaacctcagcactc cagagccagagttatttgttccaaagtttttgcttttt

				HEFEQERGHCASAVEC YMREHGVSEEEACSEL KKQVDNAWKDINHEMI FSETSKAVPMSVLTRVL NLTRGNEEIYKYNDTY TNSDTTMKDNISLVLVE SCDYFNK (SEQ ID NO: 14)	tgtctattatggatgacatttatgacaactatggtat cttggaagaatgtacattattaaccgaagctattaa gagatggaacccacaagcaatcgacggtttgcc agaatacttgaaagactattacttgaagttgttaaa gactttcgaggaatttgaagatgaattagaattgaa tgagaagtacagaatgttgtatttgcaagatgaag ttaaagctttggctatctcctacttacaagaggcca agtggggtattgaaagacacgtcccttcattagat gagcatttgcacaattctttgatatcctctggttcttc cactgtcatttgtgcttcattcgttggtatgggtgaa gttgctaccaaggaagtcttcgattggttgtcctctt tcccaaaggttgtcgaagcctgttgtgttatcggta gattgttgaacgatattcgttcccatgaatttgagc aggaaagaggtcactgcgcttccgctgttgaatgt tacatgagagaacacggtgtctctgaagaagaag cctgctcagaattgaagaagcaagttgacaacgc atggaaagatataaaccatgaaatgatattctctga aacatctaaggccgttcctatgtcagtcttgacca gagttttgaacttgacccgtggtaatgaagaaatct acaagtacaacgatacttatactaattcagacacc accatgaaagacaacatctccttggtcttggttga atcttgtgactatttcaacaagtaa (SEQ ID NO: 80)
LeuGraS QTS335	*Leucadendron grandiflorum*	A0A067FTE8	14%	SIQVPQISSQNAKSQVM RRTANFHPSVWGDRFA NYTAEDKMNHARDLK ELKALKEEVGRKLLAT AGPIQLNLIDAIQRLGV GYHFERELEQALQHLY NEKYSDDDTEDDLYRIS LRFRLLRQHGYNVSCD AFNRFKDTKGSFKEDLI KDVNSMLCLYEATHLR VHGEDILDEALGFTTSQ LKSILPKLKPLLASQVM HALKQPLHRGLPRLEH RRYISIYQDDASHYKAL LTLAKLDFNLVQSLHK KELCEISRWWKDLDFA RKLPFARDRMVECYFW ILGVYFEPNYSLARRILT KVIAMTSIIDDIYDVYG TPEELKLFTEVIERWDE SSMDQLPEYMQTFFGA LLDLYNEIEKEIANEGW SYRVQYAKEAMKILVE GYYDESKWFHENYIPK MEEYMRVALVTSGYT MLTTVSFLGMDNIVTK ETFDWVFSRPKIIRASEI IGRFMDDIKSHKFEQER GHAASAVECYMKQHG LSEQEVCEELYRQVSN	atgtccatacaggttccccaaatttcttcgcaaaat gcaaagtcacaagtaatgcgtagaaccgccaact ttcatccatctgtgtggggagacagattcgctaact acacggctgaggataaaatgaaccacgctcgcg acttgaaggaacttaaagcgttaaaggaagaagt tggtagaaagctgttggccacagctggcccaatt caactcaatctaatcgatgctatccaaagattgggt gtcggttatcacttcgaacgagaattggaacaag ctttgcaacatttatacaacgagaagtatagcgat gacgacactgaagatgatttgtacaggatttctctg agatttagattgttaagacagcacggttacaatgtc tcctgcgacgccttcaacagatttaaagataccaa gggtagtttcaaggaagacttgatcaaagatgtta actctatgctctgtttatacgaagcaactcatttgcg ggttcacggtgaagatattttggacgaagctttgg gatttacaacttcccaactaaagtccatcttaccta agttaaaaccattgctggcttctcaagtcatgcatg ccttgaagcaaccgctacaccgtggtttgccaag actcgaacacagaaggtatattagcatttaccagg atgacgcttctcattacaaagccttgttgactcttgc gaagttggatttcaatctagttcaatcattacacaa aaaggagctctgtgagatctccagatggtggaag gatttagacttcgctcgtaagttgccttttgctagag atagaatggtcgaatgttatttctggatcttgggtgt gtatttcgaaccaaactactcactggctagaagaa tattgaccaaagttattgctatgacctctattatcgat gacatttatgacgtttacggcactccagaagaatt gaagctattcactgaagtaatcgaacgttgggac

				AWKDINEECLNPTAVP MPLLMRALNLARVIDV VYKEGDGYTHVGNEM KQNVAALLIDQVPI (SEQ ID NO: 15)	gaatcgtcaatggaccaactgccagaatacatgc aaacgtttttcggtgctttgttagatttatacaatgag atagaaaaggaaattgcaaacgaaggttggtctt acagagtccagtatgcgaaagaagctatgaagat tttggttgagggttactacgatgaatctaagtggtt ccatgaaaattacatacccaagatggaggaatat atgcgggtagccttagttaccagcgggtacacaa tgttgactaccgtcagttttctggggatggacaac atcgttactaaggagacatttgattgggttttctcca gacctaagataatccgagccagtgaaattattggt agattcatggacgatatcaaatctcataagtttgaa caagagagaggtcacgctgcaagcgctgtcgaa tgttatatgaagcaacacggtctctcagaacaaga agtctgtgaagaactttacagacaagtctccaacg cttggaaggacatcaatgaggaatgcttgaatcc aaccgctgttccaatgccattgttgatgagagcac taaacttggcacgcgtaatcgacgtagtttataaa gaaggtgacggttacactcacgttggtaacgaaa tgaagcaaaacgtggctgctctacttattgatcaa gtaccaatctaa (SEQ ID NO: 81)
LeuGraS QTS345	*Leucadendron grandiflorum*	A0A0A0QUT9	12%	SAAQVSPAPVPAHNAA ASKEEVRRSAGYHPSF WGEFFLTHTSEYAKKD DKIQKQHEELKQEVKG MLVDATTEPTKKLELID AILRLGVGYHFEDEIQA ELERIHRLGDLDCDLYN TCIWFRVLRGQGFTVS AEEFNKFKNSDGNFKE DLINDVSGMLCLYEAT HLRVHGEDILDEALEFT TTRLKSILPDLEPPLATQ VMHALELPYHKGMQR LEARQYIPIYEADMTKN ISLLHFAKLDFNLLQAL HQSEIREITRWWKDLDF KTRLPYARDRLVECYF WILGVQYEPQYSMSRL FLTKVISLASVFDDTYD IYGTFEELKLLTDAIER WEIEATDSLPSYMQILY RALLDVFDEYKDKLIN VQGKDYCLYYGKEAM KGLIRSYHTEAVSFHTG YVQNFEEYLDNSAVSS GYPMLTVEALIGMGHP YATKEALDWALKVPR VIKASSDICRLVDDLRT YKVEEERGDAPSGVHC YMRDYNVSEEEACSKI EEMIDLAWKAINEEMQ KPGHLPLPILLPALNFTR MMEVLYQNIDGYTNSG	atgtccgcagcgcaagtcagtcctgctccagttcc agcccacaatgctgctgcttctaaggaagaggtg cgtagatcggccggatatcatccatcattctgggg tgaatttttccttactcacacaagcgaatacgctaa aaaggacgataagattcagaaacaacatgaaga attgaagcaagaggttaagggcatgctagtagat gctacgaccgaacccactaaaaagttagaattga tagacgccatcctgagattgggtgtcggttaccac tttgaagatgagattcaagctgaattggaaaggat ccacagactcggtgacttagattgcgacttgtata acacctgtatttggttcagagttcttagaggtcaag gttttactgtctctgctgaagaatttaacaagttcaa aaattccgacggaaacttcaaggaagatttgatca atgacgtttctggtatgttgtgtttatacgaagccac ccatttgcgggttcacggtgaggatattttggatga agcgctcgaatttactaccacacgtttaaagtctat cttaccagacttggaaccgccattggctactcaag taatgcacgcactagaactaccttaccataagggt atgcagagattggaagcccgacaatacattccaa tctatgaagccgatatgactaaaaacatcagcttgt tgcatttcgctaagcttgatttcaacctgttacaggc tctccaccaatccgaaatcagagagataacccgc tggtggaaagatcttgactttaaaactagattgcca tatgctagagatcgcttagtcgaatgttacttctgg attctaggcgttcaatacgagccacaatacagtat gtctcggttgtttttaaccaaggttatttcattggctt ctgtcttcgatgacacatatgacatttacggtacctt cgaagaattaaagttgttgactgacgccatagaaa gatgggagatcgaagcaacagattccttgccgtc ttacatgcaaattttatatcgcgctttgctggacgtc ttcgatgaatacaaggataaattgattaacgttcaa

				GRTKDRITSLLVHPITI (SEQ ID NO: 16)	gggaaggactattgtttgtattacggtaaagaagc gatgaagggtttgattcgtagctaccacactgaag ctgtgtcgtttcataccggctatgttcagaatttcga ggaatacttagacaactccgcagtttcctctggtta cccaatgctgacggttgaagctttgattggtatgg gacacccttacgctactaaggaagctttagattgg gcattgaaggtgccaagagttatcaaggctagttc agacatctgtagattagtcgatgacttaaggacgt acaaggtcgaggaggaaagaggtgatgctccct cgggggtccattgctacatgagagactataatgtc tcagaagaagaagcatgttctaagatcgaagaaa tgatcgatctggcctggaaagctataaacgaaga aatgcaaaagccaggtcatctaccactaccaatct tgttgcctgccttgaacttcactagaatgatggag gtcctttaccaaaatattgatggttatacaaattccg gtggtagaaccaaggacagaatcacctctttgttg gttcacccaattactatttaa (SEQ ID NO: 82)
LeuGraS QTS365	*Leucadendron grandiflorum*	D0VMR6	11%	SSAKLGSASEDVNRRD ANYHPTVWGDFFLTHS SNFLENNDSILEKHEEL KQEVRNLLVVETSDLPS KIQLTDEIIRLGVGYHFE TEIKAQLEKLHDHQLH LNFDLLTTSVWFRLLR GHGFSISSDIFNKFKNSD GNFKEDLINDVSGMLC LYEATHLRVHGEDILDE ALEFTTTRLKSILPDLEP PLNECVRDALHIPYHRN VQRLAARQYIPQYDAE PTKIESLSLFAKIDFNML QALHQRELREASRWW KEFDFPSKLPYARDRIA EGYYWMMGAHFEPKF SLSRKFLNRIIGITSLIDD TYDVYGTLEEVTLFTE AVERWDIEAVKDIPKY MQVIYTGMLGIFEDFK DNLINARGKDYCIDYAI EVFKEIVRSYQREAEYF HTGYVPSYDEYMENSII SGGYKMFIILMLIGRGE FELKETLDWASTIPEMV EASSLIARYIDDLQTYK AEEERGETVSAVRCYM REFGVSEEQACKKMRE MIEIEWKRLNKTTLEAD EISSSVVIPSLNFTRVLE VMYDKGDGYSDSQGV TKDRIAALLRHAIEI (SEQ ID NO: 17)	atgtcctcagcaaaattgggttctgcttctgaagat gtcaaccgtagagacgctaattaccatccaaccg tttggggagatttctttttaacacactcctctaacttc ttggagaacaatgactcaatattggaaaagcacg aagaattgaagcaagaggttagaaacttattggtc gttgaaacttctgacttgccttccaagattcagttga ctgatgaaattatcagattaggtgtcggttatcatttt gagaccgaaatcaaagcccaattagaaaagttgc acgatcatcaattgcacttgaacttcgacttgttga ccacatctgtttggttcagattattgagaggtcacg gtttttccatttcttccgacatcttcaataagttcaaa aattcagatggtaactttaaggaagatttaatcaac gacgtttctggtatgttgtgcttgtacgaagctactc atttgcgtgtccacggtgaagatattttagacgaa gccttggaatttactactaccagattgaagtctattt tgccagatttagaaccaccattaaatgaatgtgtca gagacgctttgcatattccttatcacagaaacgttc aacgtttggctgcaagacaatacataccacagta cgatgccgaaccaacaaaaatcgagtctttgtcat tattcgctaagattgatttcaacatgttgcaagcttt gcatcaaagagaattgagagaggcttccagatg gtggaaagaatttgacttcccttctaagttaccatat gccagagatcgtatcgctgaaggttactactggat gatgggtgcccactttgaaccaaagttctcattgtc tcgtaagttcttaaacagaatcattggtatcacttctt taattgatgacacctatgatgtttacggtactttgga ggaagttactttgtttaccgaagctgttgaaagatg ggacattgaagctgtcaaggacattccaaaatac atgcaagtcatctatacaggtatgttaggtatatttg aagatttcaaagacaacttgataaatgctagaggt aaggattactgtatcgactatgcaatcgaggttttc aaagaaatcgttagatcctaccaaagagaagctg aatatttccacaccggttacgttccatcctacgatg

					aatacatggaaaactctattatatctggtggttaca agatgttcattatcttaatgttaatcggtagaggag aatttgagttgaaggaaactttggactgggcttcc actattcctgaaatggtcgaggcatcttccttgatc gctcgttatattgacgacttgcaaacctataaagct gaagaagagagaggagaaaccgtctccgcagt cagatgttacatgcgtgaatttggtgtttcagaaga acaagcctgtaagaagatgagagagatgatcga aattgaatggaagagattgaataaaacaactttag aagctgacgaaatttcttcatctgtcgttattccatc attgaacttcaccagagttttggaggtcatgtacga taagggtgatggttactctgattcccaaggtgttac taaagaccgtatcgccgctttattgagacacgcca tcgaaatctaa (SEQ ID NO: 83)
LeuGraS QTS377	*Leucadendron grandiflorum*	Q39760	14%	ASQVSQMPSSSPLSSNK DEMRPKADFQPSIWGD LFLNCPDKNIDAETEKR HQQLKEEVRKMIVAPM ANSTQKLAFIDSVQRLG VSYHFTKEIEDELENIY HNNNDAENDLYTTSLR FRLLREHGFNVSCDAF NRFKDTKGSFKEDLIKD VNSMLCLYEATHLRVH GEDILDEALGFTTSQLK SILPKLKPLLASQVMHA LKQPLRRGLPRVEARH YLSVYQDIESHNKVLLE FAKIDFNMVQLLHRKE LSEISRWWKDLDFQRK LPYARDRVVEGYFWIS GVYFEPQYSLGRKMLT KVIAMASIVDDTYDSY ATYEELIPYTKAIERWD IKCIDELPEYMKPSYKA LLDVYEEMEQLVAKHG RQYRVEYAKNAMIRLA QSYLVEARWTLQNYKP SFEEFKANALPTCGYA MLAITSFVGMGDIVTPE TFKWAANDPKIIQASTII CRFMDDVAEHKFKHRR EDDCSAIECYMEEYGV TAQEAYDVFNKHVESA WKDVNKEFLKPTEMPT EVLNRSLNLARVMDVL YREGDGYTYVGKAAK GGITSLLIEPVAL (SEQ ID NO: 18)	atggccagtcaggtttcacaaatgccttcctcttct ccactatccagcaacaaagatgagatgagacca aaggctgactttcaaccctcgatatggggcgattt gttcctgaattgcccagacaagaacattgatgctg aaaccgaaaagcgtcatcaacaattgaaagaag aagtcagaaagatgatcgtggcaccaatggctaa ttctacacaaaagttggctttcattgactctgttcag aggcttggagtatcctaccactttactaaagaaatt gaggatgaattagaaaacatctatcacaacaataa cgacgcagaaaacgatttgtacacgacttcccta agattcagattattgagagaacatggtttcaatgtc tcttgtgacgcctttaacagatttaaggataccaaa ggttcattcaaggaagacttgatcaaggatgttaat tccatgttgtgtttatacgaagcgactcaccttcga gttcatggtgaggatattttggacgaagctttgggt ttcacaacctctcaactcaaatcaatcttacctaagt taaagccattgctggcttcgcaagtcatgcacgct ttgaagcaaccgctaagacgtggtttgccaagag ttgaagccagacactatttgagcgtttaccaagat attgaatctcataacaaagtcttgttggaatttgcta agatcgacttcaacatggttcaacttctccatagga aggagctcagtgaaattagtagatggtggaaaga tttagacttccaacgtaaattgccatacgctagaga tcgcgttgtcgaaggttatttttggattagtggggta tacttcgaaccgcaatattccctgggtagaaagat gttaactaaggttattgccatggcttctatcgtcga cgatacctacgattcttacgcaacttatgaggaatt aatcccatacaccaaagctatagaaagatgggat ataaagtgtatagacgaattgcctgagtatatgaa gccatcatacaaggctttgttggacgtgtacgaag aaatggaacagttagttgccaaacacggtcggca atacagagttgaatatgctaagaatgctatgatcc ggctagcccaatcttatctggtcgaggctagatgg actctacaaaactacaagccttccttcgaagaattt aaggctaacgcattgccaacttgtggttacgctat gttggcgatcacttctttcgttggtatgggcgacat tgttaccccagaaacatttaagtgggccgcgaac

					gatccaaagattattcaagcttcaacgataatctgc cggtttatggatgacgtcgccgaacacaagttca aacataggagggaagacgattgttctgctatcga gtgttatatggaagaatacggagtaactgcccag gaggcctacgacgtcttcaataagcacgtggaat cagcttggaaggatgttaataaggaatttttgaag cccaccgagatgcctacggaagtgctgaacaga tctttgaacctcgcaagagttatggatgtcttgtac agagaaggtgatggttatacttatgtgggtaaggc tgctaaaggtgggattacctccctattgatcgaac cagtcgctttataa (SEQ ID NO: 84)
LeuGraS QTS379	*Leucadendron grandiflorum*	Q39761	12%	ASQVSQMPSSSPLSSNK DEMRPKADFQPSIWGD LFLNCPDKNIDAETEKR HQQLKEEVRKMIVAPM ANSTQKLAFIDSVQRLG VSYHFTKEIEDELENIY HNNNDAENDLYTTSIRF RLLREHGYNVSCDIFNK FKNSDGNFKEDLINDVS GMLCLYEATHLRVHGE DILDEALEFTTTRLKSIL PDLEPPLATQVMHALK QSIRRGLPRVEARHYLS VYQDIESHNKALLEFA KIDFNMLQFLHRKELSE ICRWWKDLDFQRKLPY ARDRVVEGYFWISGVY FEPQYSLGRKMLTKVIA MASIVDDTYDSYATYE ELIPYTNAIERWDIKCID EIPEYMKPSYKALLDV YEEMVQLVAEHGRQY RVEYAKNAMIRLAQSY LVEAKWTLQNYKPSFE EFKANALPTCGYAMLA ITSFVGMGDIVTPETFK WAASDPKIIQASTIICRF MDDVAEHKFKHRRED DCSAIECYMEEYGVTA QEAYDVFNKHVESAW KDLNQEFLKPTEMPTE VLNRSLNLARVMDVLY REGDGYTYVGKAAKG GITSLLIEPIAL (SEQ ID NO: 19)	atggccagtcaggtttcacaaatgccttcctcttct ccactatccagcaacaaagatgagatgagacca aaggctgactttcaaccctcgatatggggcgattt gttcctgaattgcccagacaagaacattgatgctg aaaccgaaaagcgtcatcaacaattgaaagaag aagtcagaaagatgatcgtggcaccaatggctaa ttctacacaaaagttggctttcattgactctgttcag aggcttggagtatcctaccactttactaaagaaatt gaggatgaattagaaaacatctatcacaacaataa cgacgcagaaaacgatttgtacacgacttccata agattcagattattgagagaacatggttacaatgtc tcttgtgacatctttaacaagttcaagaatagcgat ggtaacttcaaggaagacttgattaatgatgtttca ggtatgctctgtttatatgaagcgacccacttgcga gttcatggtgaggatatcttagacgaagctttgga atttacaactactcgcctaaaatctattttgcctgac ttagaaccacccctggccacccaagtcatgcacg ctttgaagcaaagcatcagacgtggtcttccaaga gttgaagccagacactacttgagtgtttatcaagat attgaatctcataacaaagctttgttggaatttgcta agattgatttcaacatgttacaattcctacatagga aggagctatcggaaatctgtagatggtggaaaga tctcgattttcaaagaaagttaccttacgcacggg accgtgtcgtcgaaggttatttctggatttccgggg tttacttcgaaccacaatacagtttgggtagaaag atgttgactaaggttattgctatggcttctatcgtcg atgacacctacgattcttacgccacctatgaggaa ttgataccatatactaacgccatcgaaagatggga catcaagtgtatagacgagatcccagaatacatg aagccttcgtataaagctttattggatgtatacgag gaaatggtgcaattggttgccgaacacggtagac agtacagagtggaatacgctaagaatgctatgatt cgccttgcgcaatcctacttggttgaagcgaaatg gactctccaaaactacaagccatctttcgaagaat ttaaggccaatgctttaccgacatgcggatatgct atgctagctatcaccagcttcgttggtatgggtgat attgtcacgccagaaacttttaaatgggctgcatct gacccaaagattattcaggcttccactatcatctgt aggttcatggatgatgttgctgaacataagtttaag cacagaagagaagacgactgttcagctattgaat

					gttacatggaagaatacggcgtcaccgcgcaag aagcctacgacgtattcaacaaacacgtcgagtc ggcatggaaggatctgaaccaagaatttctaaaa cccactgagatgccaacagaagttctcaacagaa gtttgaacttggctagagtaatggacgttttgtata gagagggtgatggttatacttatgttggtaaagcc gctaagggtggcattacctcattgcttatcgagcc aatcgctttgtaa (SEQ ID NO: 85)
LeuGraS QTS385	*Leucadendron grandiflorum*	Q5SBP4	13%	ESRRSANYQASIWETNF TNSPLLSKLQNELSVAH LEELKLEVKQLIWSTKD PLFLLKFIDSIQRLGVAY HFEEEIKESLHLVYLEE RNGDHQHYKEKGLHFT ALRFRILRQDGYHVPQ DVFSSFMNKAGDFEES LSKDTKGLVSLYEASY LSMEGETILDMAKDFSS HHLHKMVEDATDKRV ANQIIHSLEMPLHRRVQ KLEAIWFIQFYECGSDA NPTLVELAKLDFNMVQ ATYQEELKRLSRWYEE TGLQEKLSFARHRLAE AFLWSMGIIPEGHFGYG RMHLMKIGAYITLLDDI YDVYGTLEELQVLTEII ERWDINLLDQLPEYMQ IFFLYMFNSTNELAYEIL RDQGINVISNLKGLWV ELSQCYFKEATWFHNG YTPTTEEYLNVACISAS GPVILFSGYFTTTNPINK HELQSLERHAHSLSMIL RLADDLGTSSDEMKRG DVPKAIQCFMNDTGCC EEEARQHVKRLIDAEW KKMNKDILMEKPFKNF CPTAMNLGRISMSFYE HGDGYGGPHSDTKKK MVSLFVQPMNITI (SEQ ID NO: 20)	atggaaagtaggcgttcagcaaattatcaggcttc catatgggagacaaactttactaactctccacttta tctaagttgcaaaatgaactgtcggtcgcccatct cgaagaattgaaactagaggtgaagcaattaatct ggagcacgaaggatcccttattcctttgaaattca ttgactccattcaaagattgggcgttgcttaccactt tgaagaagaaatcaaggaatctttgcacctggtct acctggaagagcgaaacggtgatcatcaacact ataaggaaaaaggattgcatttcaccgctttgaga ttcagaatattgagacaggacggttaccacgtacc acaagatgtttttttcttcattcatgaataaggctggt gactttgaagaaagtttatccaaagacactaaggg tttggtctctttgtacgaagcctcctacctctctatg gaaggtgaaaccattttggatatggccaaggattt ctcctctcaccatttacacaagatggttgaagatgc tactgacaaaagagttgctaaccaaatcattcata gcttggagatgcctttgcatagaagagttcaaaag ctagaggctatctggttcatccaattttatgaatgc ggttccgacgccaacccgaccttggtcgaattgg cgaaattagattttaatatggtgcaagctacttacc aagaagaattaaagcgtctatctaggtggtacga ggaaaccggtctccaagaaaagttgtctttcgctc gtcacagattggctgaagctttcttgtggtctatgg gcattattcctgaaggtcatttcggatatggcagaa tgcaccttatgaagatcggtgcatacattaccttatt ggatgatatttatgacgtttatggtactttggaagaa ttgcaagtattgacagaaatcatcgaaagatggg atattaacctttggaccagttgccagaatacatgc aaatattcttcctctacatgtttaactctacaaatgaa ctagcttacgaaatcttaagagaccaaggtattaat gtcatatccaaccttaaaggtctttgggtcgaactg tcacaatgttatttcaaagaagccacgtggttccac aacggttataccccaaccactgaggaatacctaa acgttgcttgtatttcagcgtccggtccagttatctt gttttcgggatactttactactacaaatccaatcaac aagcatgaattgcaatctttagaaagacacgctca ctccttaagtatgatcttaagactagcggatgacct aggtacttcttcggatgagatgaagcggggtgat gttcctaaggctattcaatgtttcatgaacgacacg gggtgttgcgaagaagaagccagacagcacgtt aagagattgattgacgcagaatggaagaagatg aataaggatatcttgatggagaagccatttaaaaa cttctgtccaactgcaatgaatttaggccgtatcag

					tatgtctttctacgagcacggtgacggttacggcg gtccacattctgataccaaaaagaagatggtctcg ttgtttgttcaacccatgaatattaccatttaa (SEQ ID NO: 86)
LeuGraS QTS393	*Leucadendron grandiflorum*	Q9T0J9	10%	ESQTTFKYESLAFTKLS HCQWTDYFLSVPIDESE LDVITREIDILKPEVMEL LSSQGDDETSKRKVLLI QLLLSLGLAFHFENEIK NILEHAFRKIDDITGDE KDLSTISIMFRVFRTYG HNLPSSIFNKFKNSDGN FKEDLINDVSGMLCLY EATHLRVHGEDILDEAL EFTTTRLKSILPGGTCRP HILRLIRNTLYLPQRWN MEAVIAREYISFYEQEE DHDKMLLRLAKLNFKL LQLHYIKELKSFIKWW MELGLTSKWPSQFRERI VEAWLAGLMMYFEPQ FSGGRVIAAKFNYLLTI LDDACDHYFSIHELTRL VACVERWSPDGIDTLE DISRSVFKLMLDVFDDI GKGVRSEGSSYHLKEM LEELNTLVRANLDLVK WARGIQVPSFEEHVEV GGIALTSYATLMYSFV GMGETAGKEAYEWVR SRPRLIKSLAAKGRLMD DITDFDSDMSNGFAAN AINYYMKQFVVTKEEA ILECQRMIVDINKTINEE LLKTTSVPGRVLKQAL NFGRLLELLYTKSDDIY NCSEGKLKEYIVTLLID PIRL (SEQ ID NO: 21)	atggaatcacagactacattcaaatatgagtcttta gcatttaccaagttgtcccattgccaatggactgat tacttcttgtctgttccaatagacgaatccgaattgg acgtcatcaccagagaaattgatattttaaagcctg aggttatggaattgttatcttcacaaggtgatgacg aaacatctaagcgtaaagtcttgttgatccaattgtt gttatctttgggattagcctttcacttcgaaaacga gattaagaatatcttggaacacgctttcagaaaga ttgatgacatcactggtgacgaaaaggatttgtcc accatttccataatgtttagagttttcagaacttacg gtcataacttgccatcctctatctttaataaattcaaa aactcagatggtaatttcaaggaagacttgataaa cgatgtttctggtatgttgtgtttatacgaagctact cacttgagagtccatggtgaagacattttagatga agctttagagtttaccactacccgtttgaagtctatc ttgccaggtggtacttgtagacctcacattttaaga ttgattagaaacactttatatttgccacaaagatgg aacatggaagccgtcatcgctcgtgaatacatatc cttttacgaacaagaggaagaccacgataagatg ttattgagattggctaagttgaatttcaaattgttaca gttgcattatattaaggaattgaagtcattcatcaaa tggtggatggaattgggtttaacatctaaatggcc atctcaatttagagagcgtatcgttgaagcctggtt agctggtttgatgatgtactttgaaccacaattctcc ggtggtagagttattgcagctaagttcaactatttat tgaccattttggatgatgcttgtgatcactacttctc aattcatgaattgaccagattggtcgcttgtgttga aagatggtctccagacggtatcgatacattggag gacatctcccgttctgtctttaagttaatgttggatgt ttttgacgatatcggtaagggtgttagatccgaag gttcttcctatcacttgaaagaaatgttggaagaatt aaatactttagttagagcaaatttggacttggttaaa tgggccagaggtatccaagtcccatctttcgaag agcatgttgaggttggtggtattgctttaacatccta cgccactttgatgtactctttcgtcggaatgggtga aaccgctggtaaggaagcctacgaatgggttcgt tccagacctcgtttgataaagtctttggcagctaaa ggtagattgatggacgacattactgattttgattca gatatgtctaacggtttcgctgctaacgcaattaac tattacatgaagcaattcgtcgttaccaaggaaga agccatcttagaatgccagagaatgatcgtcgac atcaacaagaccattaatgaagagttgttaaaaac tacatctgttcctggtagagtcttgaagcaagcttt gaacttcggtagattattggaattgttgtacactaa atctgacgacatctataattgttccgaaggtaagtt

					aaaggaatacattgttactttgttgatcgatccaata agattgtaa (SEQ ID NO: 87)
MacVolSQTS1139	*Macrostylis villosa*	D0VMR6	14%	SSAKLGSASEDVNRRD ANYHPTVWGDFFLTHS SNFLENNDSILEKHEGL EQKIRTMLISPTDTISKK LSLIDAVQRLGVAYHFE KEIEDEIEKLSCKEYND GNDLQTVALRFRLLRQ QGYFVSCDVFKRFKNT KGEFETEDARTLWCLY EATHLRVDGEDILEEAI QFSRKKLEALLPELSFP LNECVRDALHIPYHRN VQRLAARQYIPQYDAE PTKIESLSLFAKIDFNML QALHQRELREASRWW KEFDFPSKLPYARDRIA EGYYWMMGAHFEPKF SLSRKFLNRIIGITSLIDD TYDVYGTLEEVTLFTE AVERWDIEAVKDIPKY MQVIYTGMLGIFEDFK DNLINARGKDYCIDYAI EVFKEIVRSYQREAEYF HTGYVPSYDEYMENSII SGGYKMFIILMLIGRGE FELKETLDWASTIPEMV EASSLIARYIDDLQTYK AEEERGETVSAVRCYM REFGVSEEQACKKMRE MIEIEWKRLNKTTLEAD EISSSVVIPSLNFTRVLE VMYDKGDGYSDSQGV TKDRIAALLRHAIEI (SEQ ID NO: 22)	atgtcctcagcaaaattgggttctgcttctgaagat gtcaaccgtagagacgctaattaccatccaaccg tttggggagatttctttttaacacactcctctaacttc ttggagaacaatgactcaatattggaaaagcacg aaggtttggaacaaaagattagaactatgttaatct ctcctaccgatactatctccaagaaattatctttgat tgacgccgttcagagattgggtgtcgcttatcatttt gagaaggaaattgaagatgaaatcgaaaagttat catgtaaagagtacaacgacggtaatgacttgca aaccgtcgccttgagattcagattattgagacaac aaggttatttcgtttcctgcgatgtttttaagcgtttc aagaacactaagggtgaatttgagactgaagatg ctagaacattgtggtgtttatacgaagctactcactt gagagttgacggtgaagatattttggaagaagct atccaattctctcgtaagaaattagaagcattgttg ccagaattatcctttccattgaatgaatgtgttagag atgccttgcatatcccataccacagaaacgtccag agattggctgcacgtcaatatataccacaatacga cgctgagcctaccaagattgaatccttatctttgttc gctaagattgactttaatatgttgcaggccttgcac caaagagaattgagagaagcttccagatggtgg aaggagttcgattttccatctaaattgccttatgccc gtgatagaatcgctgaaggttactactggatgatg ggtgctcatttcgaaccaaaattttctttgtctcgta agttcttaaacagaatcattggtataacctccttaat tgatgatacttatgacgtctacggtactttagaaga agttaccttgttcaccgaagccgttgaaagatggg atattgaggctgtcaaagacatcccaaagtacatg caagttatatacacaggtatgttaggtattttcgaa gatttcaaagacaatttgattaacgccagaggtaa ggattattgcatcgattacgctatcgaagttttcaa ggagattgtcagatcttaccaaagagaagcagaa tactttcacactggttacgttccatcttatgacgaat acatggaaaactcaattatctcaggtggttacaaa atgtttataatcttgatgttaatcggtagaggtgagt tcgaattgaaagaaaccttagattgggcttcaact attccagaaatggtcgaagcttcttccttgatagct agatacatcgacgatttgcaaacatacaaggccg aagaagaacgtggtgaaacagtttcagcagtcag atgttacatgagagagtttggtgtttctgaggaaca agcttgtaagaagatgagagaaatgattgagatc gaatggaagagattgaacaagactaccttggaag ctgacgaaatttcttcttccgttgttattccatctttga actttactagagtcttggaagtcatgtatgacaagg gagacggttattctgattcccaaggtgttaccaag gatcgtattgctgctttgttaagacacgccattgag atataa (SEQ ID NO: 88)

MacVolSQTS2198	*Macrostylis villosa*	A0A067D5M4	62%	RDLKSVLSSKESTKADVNRRSSNYHPSIWGDHFINVSSNEKYTNTEVEKRFETLKAEIEKLLVSNNTAWKTLEEIVAIVNQLQRLGLAYHFENQIKEALQSIYDSHVNGNCDVNYDHNNDLYIVALRFRLLRQHGYKVSADIFKKFRDEKGEFKAMLTNDAKGLLCLYEASYLRVQGENILEEACEFSRKHLKSLLSHLSTSLAEQVKHSLEIPLHRGMPRLEARHYISIYEEDNSSRNELILELAKLDFNLLQALHRRELGEISRWWKDIDFATKLPFARDRLVECYFWILGVYFEPKYSITRKFMTKVIAIASVIDDIYDVYGTLEELKLFTHAIERWETVAANELPKYMQVCYFALLDVFKEMEDKLVNKGLLYSMPCAKEAVKGLVRAYFVEAEWFNANYMPTFEEYMENSTMSSGYPMLAVEALIGIEDATISKEAFDWAISVPKIIRSCALIARLVDDIHTYKVEQERGDAPSSVECYMQQYDVSEEEACNRIKGMVEIEWMNINEEIQDPNHPPLQWLLPSLNLARMMVVLYQNGDNYTNSSGKTKDRIASLLVDPLPM (SEQ ID NO: 23)	atgcgtgacttgaaatccgtcttatcttcaaaggaatctacaaaggcagatgttaatagaagatcctctaactatcaccttccatctggggtgatcatttcattaacgtttcttcaaatgagaagtacactaacactgaagtcgaaaaaagatttgaaaccttgaaggccgaaatagaaaagttgttagtttctaacaacaccgcttggaagaccttggaggaaattgtcgctatcgttaatcagttgcaaagattagggttggcttaccacttcgaaaaccaaatcaaagaagccttgcaatccatttatgactctcatgtcaacggtaattgcgacgttaattacgatcacaacaacgatttgtacatagtcgctttaagatttcgtttgttgagacaacacggttataaagtctctgctgacattttcaagaagtttagagatgaaaagggtgaatttaaggctatgttaacaaatgacgccaaaggtttgttgtgtttatacgaagcatcctatttgagagttcaaggtgaaaatatcttagaagaggcttgtgaattttctcgtaagcatttgaagtcattattgtctcacttgtccacctcatggctgagcaagttaagcactctttggaaatccccattacatagaggtatgccaagattggaagctagacattacatttctatttacgaggaagataactcctctcgtaatgaattgatattagagttggcaaagttggacttcaacttgttgcaggccttacacagaagagaattggggtgaaatttctcgttggtggaaagatattgatttcgctactaaattgccattcgccagagacagattagttgaatgttacttctggatcttgggtgtttatttttgaacctaaatactccatcactagaaagttcatgactaaggttatcgctattgcttccgtcatcgatgatatatacgacgtttatggtaccttggaggaattgaagttgttcactcatgctattgaaagatgggaaactgtcgctgccaacgaattaccaaagtacatgcaagtttgttacttttgctttgttagacgtctttaaggaaatggaagataaattagtcaataaaggtttgttatactccatgccatgtgcaaaggaggctgttaaaggtttggttagagcttacttcgttgaggctgaatggttcaacgctaactatatgccaaccttcgaagaatatatggaaaactcaactatgtcctctggttatccaatgttggctgtcgaagctttgatcggtattgaagacgcaactatttcaaaggaagccttcgattgggcaatatctgttcctaaaattatccgttcatgcgcattgatcgccagattggtcgatgacattcacacctacaaggtcgaacaagagagaggtgatgccccatcttccgtcgaatgttacatgcaacaatacgacgtttctgaggaagaagcctgtaatagaattaagggtatggttgaaattgaatggatgaatataaacgaggaaatccaggatccaaaccacccacctttacaatggttgttgccatctttgaacttagctcgtatgatggtcgttttgtaccaaaatggtgacaactatacaaactcctccggtaaaaccaaggatagaattgcttccttgttggtcgacccctttgccaatgtaa (SEQ ID NO: 89)

MacVolSQTS2202	*Macrostylis villosa*	A0A067D5M4	69%	RDLKSVLSSKESTKADVNRRSSNYHPSIWGDHFINVSSNEKYTNTEVEKRFETLKAEIEKLLVSNNTAWKTLEEIVAIVNQLQRLGLAYHFENQIKEALQSIYDSHVNGNCDVNYDHNNDLYIVALRFRLLRQHGYKVSADIFKKFKDEKGEFKDMIRNDARGLLCLYEASHLRVKGEDILEEATEFSRKHLKSLLPQLSTSLAEQVKHSLEIPLHRGMPRLEARHYISIYEENNSSRNELLLELAKLDFNLLQALHRRELGDISRWWKDIDFATKLPFARDRLVECYFWILGVYFEPKYSITRKFMTKVIAIASVIDDIYDVYGTLEELKLFTHAIERWETVAANELPKYMQVCYFALLDVFKEMEDKLVNKGLLYSMPCAKEAVKGLVRAYFVEAEWFNANYMPTFEEYMENSTMSSGYPMLAVEALIGIEDATISKEAFDWAISVPKIIRSCALIARLVDDIHTYKVEQERGDAPSSVQCYVQQYGVSEEEACNKIKGMVEIEWMNINEEIQDPNHPPLQWLLPSLNLARMMVVLYQNGDNYTNSSGKTKDRIASLLVDPLPM (SEQ ID NO: 24)	atgcgtgacttgaaatccgtcttatcttcaaaggaatctacaaaggcagatgttaatagaagatcctctaactatcaccttccatctggggtgatcatttcattaacgtttcttcaaatgagaagtacactaacactgaagtcgaaaaaagatttgaaaccttgaaggccgaaatagaaaagttgttagtttctaacaacaccgcttggaagaccttggaggaaattgtcgctatcgttaatcagttgcaaagattagggttggcttaccacttcgaaaaccaaatcaaagaagccttgcaatccatttatgactctcatgtcaacggtaattgcgacgttaattacgatcacaacaacgatttgtacatagtcgctttaagatttcgtttgttgagacaacacggttataaagtctctgctgacattttcaagaagtttaaagatgaaaagggtgaatttaaggatatgatcagaaatgacgccagaggtttattgtgtttatacgaagcatcccatttgagagttaagggtgaagatattttagaagaggctactgaattttctcgtaagcacttgaagtcattgttaccacaattgtccacatcattggctgagcaagttaagcactctttggaaatccccattacatagaggtatgccaagattggaagctagacattacatttctatttatgaggaaaacaactcctctcgtaatgaattgttgttagagttggcaaagttggacttcaacttgttgcaggctttacacagaagagaatttgggtgatatttctcgttggtggaaagacatcgatttcgccactaaattgccattcgccagagacagattagttgaatgttacttctggatcttgggtgtttattttgaacctaaatactccattactagaaaattcatgaccaaggttatcgctatagcttctgtcatcgatgatatatacgacgtttacggtaccttggaagaattgaagttgttcactcatgctattgagcgttgggaaactgtcgctgctaatgaattaccaaagtatatgcaagtttgttactttgctttgttagacgtctttaaggaaatggaagataaattagtcaataaaggtttgttatactccatgccatgtgcaaaggaggctgttaagggtttggttagagcctacttcgttgaggctgaatggttcaacgctaactatatgccaaccttcgaagaatatatggaaaactcaactatgtcctctggttatcctatgttggctgtcgaagctttgatcggtattgaagacgcaactatttcaaaggaagccttcgattgggcaatatccgttccaaaaattatcagatcttgtgcattgatcgccagattggtcgatgacattcacacctacaaggtcgaacaagagagaggtgatgccccatcttctgtccaatgctacgttcaacaatacggtgtctccgaagaagaagcctgtaataaaattaagggtatggttgagattgaatggatgaatataaacgaagaaatccaggatccaaaccacccacctttacaatggttgttgccatctttgaacttagctcgtatgatggttgttttgtaccaaaatggtgacaactacacaaactcctccggtaaaaccaaggatagaattgcttccttgttggtcgacccctttgccaatgtaa (SEQ ID NO: 90)

MacVolSQTS2222	*Macrostylis villosa*	A0A067D5M4	69%	RDLKSVLSSKESTKADVNRRSSNYHPSIWGDHFINVSSNEKYTNTEVEKRFETLKAEIEKLLVSNNTAWKTLEEIVAIVNQLQRLGLAYHFENQIKEALQSIYDSHVNGNCDVNYDHNNDLYIVALRFRLLRQHGYKVSADIFKKFKDEKGEFKDMIRNDARGLLCLYEASHLRVKGEDILEEATEFSRKHLKSLLPQLSTSLAEQVKHSLEIPLHRGMPRLEARHYISIYEENNSSRNELLLELAKLDFNLLQALHRRELGDISRWWKDIDFATKLPFARDRLVECYFWILGVYFEPKYSITRKFMTKVIAIASVIDDIYDVYGTLEELKLFTHAIERWETVAANELPKYMQVCYFALLDVFKEMEDKLVNKGLLYSMPCAKEAVKGLVKAYFVEAKWFHAKYVPTFEEYMENSTMSSGYPMLAVEALVGLEDMAITKRALDWAISVPKIIRSCALIARLDDDVHTYKVEQERGDAPSSVQCYMQQYDVSEEEACNRIKGMVETAWMEINGEIQDTNHLPLQWLLPSLNLARMMVVLYQNGDNYTNSSGKTKDRIASLLVDPLPM (SEQ ID NO: 25)	atgcgtgacttgaaatccgtcttatcttcaaaggaatctacaaaggcagatgttaatagaagatcctctaactatcacccttccatctggggtgatcatttcattaacgtttcttcaaatgagaagtacactaacactgaagtcgaaaaaagatttgaaaccttgaaggccgaaatagaaaagttgttagtttctaacaacaccgcttggaagaccttggaggaaattgtcgctatcgttaatcagttgcaaagattagggttggcttaccacttcgaaaaccaaatcaaagaagccttgcaatccatttatgactctcatgtcaacggtaattgcgacgttaattacgatcacaacaacgatttgtacatagtcgctttaagatttcgtttgttgagacaacacggttataaagtctctgctgacattttcaagaagtttaaagatgaaaagggtgaatttaaggatatgatcagaaatgacgccagaggtttattgtgtttatacgaagcatcccatttgagagttaagggtgaagatattttagaagaggctactgaattttctcgtaagcacttgaagtcattgttaccacaattgtccacatcattggctgagcaagttaagcactctttggaaatccccattacatagaggtatgccaagattggaagctagacattacatttctatttatgaggaaaacaactcctctcgtaatgaattgttgttagagttggcaaagttggacttcaacttgttgcaggctttacacagaagagaatttgggtgatatttctcgttggtggaaagacatcgatttcgccactaaattgccattcgccagagacagattagttgaatgttacttctggatcttgggtgtttattttgaacctaaatactccattactagaaaattcatgaccaaggttatcgctatagcttctgtcatcgatgatatatacgacgtttacggtaccttggaagaattgaagttgttcactcatgctattgagcgttgggaaactgtcgctgctaatgaattaccaaagtatatgcaagtttgttactttgctttgttagacgtctttaaggaaatggaagataaattagtcaataaaggtttgttatactccatgccatgtgcaaaggaggctgttaagggtttggttaaggcctacttcgttgaggctaagtggttccacgctaagtatgtcccaaccttcgaagaatatatggaaaactcaactatgtcctctggttatcctatgttggctgttgaagctttggttggtttagaagacatggccattacaaagagagctttggattgggcaatatccgttccaaaaattatcagatcatgtgcattgatcgccagattggacgatgacgttcacacttacaaggtcgaacaagagagaggtgatgccccatcttctgtccaatgctacatgcaacaatacgacgtctccgaagaagaagcatgtaatcgtattaagggtatggttgaaactgcttggatggaaatcaacggtgagatccaggataccaaccacttgccattacaatggttgttgccatctttgaacttagctagaatgatggtcgttttgtaccaaaatggtgacaactacaccaactcctccggtaaaaccaaggatagaattgcctctttgttggtcgacccttgcctatgtaa (SEQ ID NO: 91)

MacVolSQTS2251	*Macrostylis villosa*	A0A067D5M4	65%	RDLKSVLSSKESTKADVNRRSSNYHPSIWGDHFINVSSNEKYTNTEVEKRFETLKAEIEKLLVSNNTAWKTLEEIVAIVNQLQRLGLAYHFENQIKEALQSIYDSHVNGNCDVNYDHNNDLYIVALRFRLLRQHGYKVSADIFKKFKDEKGEFKDMIRNDARGLLCLYEASHLRVKGEDILEEATEFSRKHLKSLLPQLSTSLAEQVKHSLEIPLHRGMPRLEARHYISIYEENNSSRNELLLELAKLDFNLLQALHRRELGDISRWWKDIDFATKLPFARDRLVECYFWILGVYFEPKYSITRKFMTKVIAIASVIDDIYDVYGTLEELKLFTHAIERWETVAANELPKYMQVCYFALLDVFKEMEDKLVNKGLLYSMPCAKEAVKGLVKAYFVEAKWFHAKYVPTFEEYMENSTMSSGYPMLAVEALVGLEDMAITKRALDWAISVPKIIRSCALIARLDDDVHTYKVEQERGDAPSSVECYMQQYDVSEEEACNRIKGMVEIEWMNINEEIQDPNHPPLQWLLPSLNLARMMVVLYQNGDNYTNSSGKTKDRIASLLVDPLPM (SEQ ID NO: 26)	atgcgtgacttgaaatccgtcttatcttcaaaggaatctacaaaggcagatgttaatagaagatcctctaactatcacccttccatctggggtgatcatttcattaacgtttcttcaaatgagaagtacactaacactgaagtcgaaaaaagatttgaaaccttgaaggccgaaatagaaaagttgttagtttctaacaacaccgcttggaagaccttggaggaaattgtcgctatcgttaatcagttgcaaagattagggttggcttaccacttcgaaaaccaaatcaaagaagccttgcaatccatttatgactctcatgtcaacggtaattgcgacgttaattacgatcacaacaacgatttgtacatagtcgctttaagatttcgtttgttgagacaacacggttataaagtctctgctgacattttcaagaagtttaaagatgaaaagggtgaatttaaggatatgatcagaaatgacgccagaggtttattgtgtttatacgaagcatcccatttgagagttaagggtgaagatattttagaagaggctactgaattttctcgtaagcacttgaagtcattgttaccacaattgtccacatcattggctgagcaagttaagcactctttggaaatccccattacatagaggtatgccaagattggaagctagacattacatttctatttatgaggaaaacaactcctctcgtaatgaattgttgttagagttggcaaagttggacttcaacttgttgcaggctttacacagaagagaattgggtgatatttctcgttggtggaaagacatcgatttcgccactaaattgccattcgccagagacagattagttgaatgttacttctggatcttgggtgtttattttgaacctaaatactccattactagaaaattcatgaccaaggttatcgctatagcttctgtcatcgatgatatatacgacgtttacggtaccttggaagaattgaagttgttcactcatgctattgagcgttgggaaactgtcgctgctaatgaattaccaaagtatatgcaagtttgttactttgctttgttagacgtctttaaggaaatggaagataaattagtcaataaaggtttgttatactccatgccatgtgcaaaggaggctgttaagggtttggttaaggcctacttcgttgaggctaagtggttccacgctaagtatgtcccaaccttcgaagaatatatggaaaactcaactatgtcctctggttatcctatgttggctgttgaagctttggttggtttagaagacatggccattacaaagagagctttggattgggcaatatccgttccaaaaattatcagatcatgtgcattgatcgccagattggacgatgacgttcacacttacaaggtcgaacaagagagaggtgatgccccatcttctgtcgaatgctacatgcaacaatacgacgtctccgaagaagaagcatgtaatcgtattaagggtatggttgagattgaatggatgaacataaacgaagaaatccaggatccaaaccacccaccttacaatggttgttgccatctttgaacttagctagaatgatggtcgttttgtaccaaaatggtgacaactacaccaactcctccggtaaaaccaaggatagaattgcttctttgttggtcgacccttgccaatgtaa (SEQ ID NO: 92)

MacVolSQTS2274	*Macrostylis villosa*	A0A097ZIE0	38%	SFAVSASPAKFIQNVEKDSTRRSANFHPSIWGDHFLQYTCDSQEPDDDGSVKHQQLKEEIRKMLTAETKLSQKLDLIDAIQRLGVAYHFESEIDEILGRVHQAYQESDLCVNENDGLYYISLQFRLLRENGYRISADVFNKFRDIDGNFKPSLARNVRGMLSLYEATHLRVHGENILDEAHAFATSHLESIATHQISSPLAEQVKHALFQPIHKGVQRLEARNYMPFYQEEASHNEALLTFAKLDFNKLQKLHQKELSEITRWWKELDFAHNLPFTIRDRIAECYFWAVAVYFEPQYSLGRRMLAKVFPMTSIIDDIYDVYGKFEELELFTSAIERWDISAIDELPEYMKLCYRALLDVYSEAEKDLASQGKLYHLHYAKEAMKNQVKNYFFEAKWCHQNYIPSVDEYMTVASVTSGYPMLSTTSFVGMGDIVTKESFEWSLTNPRVIRASSVAARLMNDMVSHKFEQSREHVASSIECYMKQYGATEEETCNELRKQVSNAWKDINEECLCPTAVPMPLIVRILNLTRFLDVVYRFEDGYTHSGVVLKDFVASLLINPVSI (SEQ ID NO: 27)	atgtccttcgcagtttcagcctctcctgctaaatttatacagaatgtcgagaaggattctaccagacgttctgctaacttccacccatccatctggggtgaccatttttgcaatacacttgcgactcacaagaaccagatgatgacgggtctgttaagcatcaacaattaaaggaagaaattagaaaaatgttgacagctgaaactaagttgtcccagaagttagatttgattgacgccatccaaagattgggtgtcgcttatcacttcgaatctgaaatcgatgagattttaggtagagttcaccaagcttaccaagaatcagacttgtgtgtcaacgaaaatgacggtttgtattacatttctttgcaattcagattattgcgtgaaaacggttacagaatatctgccgatgtctttaacaaattcagagatatcgatggtaattttaagccatccttggctagaaacgttagaggtatgttatccttgtatgaagccacccatttgcgtgttcacggtgaaaacattttggacgaagctcacgctttcgcaacttctcatttagaatctattgccacccaccaaatctcttccccattggctgagcaggtcaagcatgctttgttccaaccaattcacaaaggtgttcaaagattagaagcaagaaattacatgcctttctatcaagaagaagcttcccacaacgaggctttgttaacatttgctaagttggactttaacaagttgcaaaagttgcatcagaaagaattgtctgaaatcactcgttggtggaaggaattagatttcgctcacaatttgccatttactattagagatagaatcgcagaatgtttacttctgggctgttgcagtttacttcgagccacaatattccttaggtagacgtatgttggccaaagtttttcctatgacctctataattgacgatatctacgacgtctacggtaaattcgaagaattagaattgttcacctcagctatcgaaagatgggatatctctgctatcgatgagttaccagagtatatgaagttgtgttacagagccttgttagatgtctactctgaagccgaaaaggacttagcatcccaaggtaagttgtatcacttgcattacgccaaagaagctatgaagaatcaggttaagaactacttttttcgaggctaagtggtgccatcaaaactatattccatctgttgatgaatacatgaccgttgcttccgtcacttcaggttacccaatgttgtccactacttcttttgtcggtatgggtgatattgttacaaaggaatccttcgaatggtctttgaccaatcctagagttatcagagcttcctctgttgctgctagattaatgaatgacatggtctcacacaagttcgaacaatctcgtgaacacgtcgcttcttcaatagaatgttacatgaaacaatacggtgcaactgaggaagaaacctgtaacgagttgagaaaacaagtttctaacgcttggaaggatattaacgaagaatgtttatgtccaacagccgtcccaatgcctttgatagtcagaattttaaatttgactagattcttggacgttgtttatcgttttgaagacggttacacccattccggtgtcgtcttgaaggactttgttgcctctttgttgattaacccagtttccatctaa (SEQ ID NO: 93)
OrbStiSQTS1368	*Orbexilum stipulatum*	Q9T0J9	10%	ESQTTFKYESLAFTKLSHCQWTDYFLSVPIDESELDVITREIDILKPEVMEL	atggaatcacagactacattcaaatatgagtcttagcatttaccaagttgtcccattgccaatggactgattacttcttgtctgttccaatagacgaatccgaattgg

				LSSQGDDETSKRKVLLI QLLLSLGLAFHFENEIK NILEHAFRKIDDITGDE KDLSTISIMFRVFRTYG HNLPAEVFERFKDQHG NFKASLSSDVEGMLSL YEASFLDYEGEDILDEA KAFTSFHLRGALAGGT CRPHILRLIRNTLYLPQR WNMEAVIAREYISFYE QEEDHDKMLLRLAKLN FKLLQLHYIKELKSFIK WWMELGLTSKWPSQF RERIVEAWLAGLMMYF EPQFSGGRVIAAKFNYL LTILDDACDHYFSIHEL TRLVACVERWSPDGID TLEDISRSVFKLMLDVF DDIGKGVRSEGSSYHL KEMLEELNTLVRANLD LVKWARGIQVPSFEEH VEVGGIALTSYATLMY SFVGMGETAGKEAYE WVRSRPRLIKSLAAKG RLMDDITDFDSDMSNG FAANAINYYMKQFVVT KEEAILECQRMIVDINK TINEELLKTTSVPGRVL KQALNFGRLLELLYTK SDDIYNCSEGKLKEYIV TLLIDPIRL (SEQ ID NO: 28)	acgtcatcaccagagaaattgatattttaaagcctg aggttatggaattgttatcttcacaaggtgatgacg aaacatctaagcgtaaagtcttgttgatccaattgtt gttatctttgggattagcctttcacttcgaaaacga gattaagaatatcttggaacacgctttcagaaaga ttgatgacatcactggtgacgaaaaggatttgtcc accatttccataatgtttagagttttcagaacttacg gtcataacttgccagctgaagtctttgaaagattca aagaccaacacggtaatttcaaagcttctttgtcat ccgatgttgaaggtatgttgtctttatacgaagcct ctttcttggactatgaaggtgaagatattttagatga agctaaggcctttacttcttttcatttgcgtggtgctt tggctggtggtacctgtagacctcacatcttaagat tgatcagaaacactttatacttgccacaaagatgg aacatggaggccgtcatagctcgtgaatatatctc cttttacgaacaagaggaagaccacgataagatg ttattgagattagctaagttgaatttcaagttgttaca gttgcattacattaaggaattgaaatcattcatcaa gtggtggatggaattgggtttaacatctaaatggc catctcaatttagagagcgtattgttgaagcttggtt agctggtttgatgatgtacttcgaaccacaattctc cggtggtagagttattgcagccaagtttaactattt gttaaccattttggatgatgcttgtgatcactatttct caatccatgaattgactagattggtcgcttgtgttg aaagatggtctccagacggtatcgataccttgga ggacatctcccgttctgtctttaagttaatgttggat gtttttgacgatattggtaaaggtgttagatccgaa ggttcttcctaccacttgaaagaaatgttggaaga attaaatacattagttagagcaaacttggacttggtt aaatgggccagaggtatccaagtcccatctttcga agagcatgttgaggttggtggtattgctttaacatc ctacgcaactttgatgtactctttcgtcggaatggg tgaaactgctggtaaggaagcatacgaatgggtt cgttcaagacctcgtttgataaagtctttggccgct aagggtagattgatggacgacatcactgattttga ttccgatatgtctaacggtttcgctgctaacgcaatt aactattacatgaagcagttcgtcgttacaaagga agaagccatcttagaatgccaaagaatgattgtcg acatcaataagaccatcaatgaagagttgttaaaa actacctctgttccaggtagagtcttgaaacaagc tttgaacttcggtagattattggaattgttgtatacta agtccgacgacatttacaactgttctgaaggtaaa ttaaaggaatacatagttactttgttgattgatccaa taagattgtaa (SEQ ID NO: 94)
OrbStiS QTS1414	*Orbexilum stipulatum*	A0A067 FTE8	43%	SIQVPQISSQNAKSQVM RRTANFHPSVWGDRFA NYTAEDKMNHARDLK ELKALKEEVGRKLLAT AGPIVKLELVDDVKRL GIGYRFEKEIVEALHRC FISSERFTHRNLHQTAL SFRLLRECGYDVTCDK	atgtccatacaggttccccaaatttcttcgcaaaat gcaaagtcacaagtaatgcgtagaaccgccaact ttcatccatctgtgtggggagacagattcgctaact acacggctgaggataaaatgaaccacgctcgcg acttgaaggaacttaaagcgttaaaggaagaagt tggtagaaagctgttggccacagctggcccaatt gttaagctagagttggtcgatgatgtcaaaagact

				FNKFTNKEGKFNSKLG ENIKGMIDLYEASQLGI AGEYILAEAGEFSGLVL KEKVACINNNPLKAQV RHALRQPLHRGLPRLE HRRYISIYQDDASHYKA LLTLAKLDFNLVQSLH KKELCEISRWWKDLDF ARKLPFARDRMVECYF WILGVYFEPQYSVPRRT TTKVIGLCSVIDDMYD AYGTIDELELFTNAIER LDTSTMDQLPEYMQTF FGALLDLYNEIEKEIAN EGWSYRVQYAKEAMK ILVEGYYDESRWLKCN HAPTMEEYMKVRGVSS GYPLLITISFIGMEDTTE EILTWATSEPMIIRASVI VCRLMDDIKSHKFEQE RGHAASAVECYMKQH GLSEQEVCEELYRQVS NAWKDINEECLNPTAV PMPLLMRALNLARVID VVYKEGDGYTHVGNE MKQNVAALLIDQVPI (SEQ ID NO: 29)	cgggatcggttatagattcgaaaaggaaatcgttg aagctttacaccgttgctttattagttccgaaagatt cactcataggaatttgcaccaaaccgccttgagct tcagattgttacgggaatgtggttacgacgtcactt gtgataagtttaataagttcactaacaaagagggt aagtttaactcaaagttgggtgaaaatatcaaggg tatgatagacttgtatgaagctagccaacttggtat tgctggtgaatacatcttggctgaagcaggtgaat tttcgggcttagttctaaaagaaaaggttgcttgtat taacaataacccattgaaagcgcaggtcagacat gccctaagacaacctctgcacagaggtctcccaa gattagaacacaggagatacatctctatttaccaa gatgacgcttctcactataaggctttgttgaccctg gccaagttggatttcaacttggttcaatccctccat aagaaagagctttgcgaaatttccagatggtgga aagatcttgacttcgctcggaagttacctttgcac gtgaccgtatggtcgaatgttatttctggatcttgg gagtttacttcgaaccacaatacagtgtaccaaga agaactaccactaaggttattggtttgtgttctgtca tcgatgatatgtacgatgcttacggtacaattgacg aattagagctttttactaacgccatcgaaagattgg acacctctactatggatcagctaccagaatatatg caaactttctttggtgctttattggatttgtataacga gatcgaaaaagaaatcgcaaatgaaggttggtcc taccgagtgcaatacgctaaggaagctatgaaaa ttttggtggaaggatactatgatgaaagcagatgg ttgaagtgtaaccacgccccaaccatggaagaat acatgaaggtccgtggtgttagttctggttaccctc tcttgataaccatatctttcataggtatggaggaca ctactgaagagatcttaacatgggctacatctgaa cctatgattatcagagccagtgtcattgtttgtagat tgatggacgacattaaatcccataagtttgagcaa gagaggggcatgctgcgagcgctgtagaatgc tatatgaagcaacacggtctatcagaacaagaag tttgtgaagaactttacagacaggtctctaatgcat ggaaggacatcaatgaagaatgtttgaacccgac cgctgttccaatgccattgttaatgagagcgctga acttggctcgcgtcattgacgtagtttataaagaag gtgacggctacacccacgttggtaatgaaatgaa gcaaaacgtagctgctctcctaatcgatcaagtac caatctaa (SEQ ID NO: 95)
ShoCusS QTS154	*Shorea cuspidata*	ShoBecS QTS1	38%	ALQDSEVPSSILNATAG NRPTASYHPTLWGEKF LVVSTQSTSGSMKNEPT TQGEYDELKQQVTKML TDATTNDPSKKLHLID MVQRLGIAYHFEIEIEN ALEKINLGDANYFEYD LYTIALGFRLLRQQGIK VSSEIFKKFMDEKGKFK EDVVNDVLGMLNLYE AAHLRLRGEDILDEAL	atggctttgcaggattcagaagtcccttcttccatat taaacgccactgctggtaatcgtccaaccgcatct taccatccaacattgtggggagagaaattcttagtt gtttccactcaatctacctctggttccatgaagaac gaaccaactacacaaggtgaatatgacgaattga agcaacaagtcaccaagatgttgactgatgctac cactaacgacccatccaaaaagttgcacttgatcg atatggttcaaagattaggtattgcctaccactttga gattgaaatcgaaaatgctttggaaaagattaactt aggtgacgctaactacttcgaatatgacttgtaca

				AFTTSHLESMATKVSPL LAEQIAHALNCPIQKGL PRIEARHYISLYSRETHF ASSNAALLRFAKIDFN MVQALHQKEISGITKW WKNLDFSTKLPYARDR IVECYFWIMGAYFEPK YSLARTFLTKVIAMTSI LDDTYDNYGTNKELEL LTKCIERWDIDVIDQLP EYMKLVYQALLNVYSE MEAKVAKEGRSYAIDY AKESMKKTMKAYLDE AKWRQEDYVPPIEEYM QVARISSAYPMLITNSF VGMGEVATKEAFDWIS NDPKILKASTTICRLMD DITSHEFEQTRDHVASG VECYMKQYGVSREETV KLFREDVANAWKDINE GFMKPAIFPMPILTVVL NFARVMDFLYKDGDN YTNSHMLKDYITSLLV NPLLI (SEQ ID NO: 30)	ccatcgctttgggttttagattgttgagacaacagg gtattaaagtctcatctgaaatcttcaagaagtttat ggatgagaaaggtaagttcaaagaagacgttgtt aatgatgtcttaggtatgttgaacttatacgaagca gcccatttgagattaagaggtgaagatatcttgga cgaggctttagccttcactacctcccacttggaatc tatggctacaaaggtttctcctttgttggctgaaca aatagcccatgctttaaattgcccaattcaaaagg gtttaccaagaattgaagccagacactatatctcat tgtactcccgtgaaactcactttgcttcttctaacgc tgcattgttgagattcgctaaaattgacttcaacatg gttcaagctttgcaccagaaggagatctctggtatt acaaagtggtggaaaaatttggatttctcaactaa gttgccatacgctagagacagaatcgtcgaatgtt atttttggatcatgggtgcttactttgaacctaagtat tccttggctagaacttttttgaccaaggttatagcaa tgacctctatattagatgatacatacgataactacg gtactaataaggaattggagttgttaactaaatgta ttgaacgttgggacatcgacgttattgatcaattac cagaatatatgaagttggtctaccaagcattgttga acgtttactcagaaatggaagccaaagtcgctaa ggagggtcgttcttacgccattgactatgctaagg aatccatgaaaaagaccatgaaggcatacttgga tgaagctaaatggagacaagaagactacgttcct ccaatagaagaatatatgcaagtcgctagaatttc ctctgcctacccaatgttaatcactaattccttcgtt ggtatgggtgaagttgctaccaaagaggcattcg attggatttccaatgacccaaagattttgaaggctt ctactactatatgtagattgatggatgatatcacttc tcatgaatttgaacaaacaagagaccatgttgcct ctggtgtcgaatgttatatgaaacaatacggtgttt cacgtgaagaaaccgttaagttattcagagagga tgtcgctaacgcttggaaagacattaacgagggtt tcatgaagcctgctatattcccaatgccaatcttga ctgttgttttgaactttgccagagtcatggatttctta tacaaggatggtgacaactatactaattctcatatg ttgaaggattacattacatcattgttggtcaatccat tattaatctaa (SEQ ID NO: 96)
ShoCusS QTS155	*Shorea cuspidata*	ShoBecS QTS1	35%	ALQDSEVPSSILNATAG NRPTASYHPTLWGEKF LVVSTQSTSGSMKNEPT TQGEYDELKQQVTKML TDATTNDPSKKLHLID MVQRLGIAYHFEIEIEN ALEKINLGDANYFEYD LYTIALGFRLLRQQGIK VSSEIFKKFMDEKGKFK EDVVNDVLGMLNLYE AAHLRLRGEDILDEAL AFTTSHLESMATKVSPL LAEQIAHALNCPIQKGL PRIEARHYISLYSRETHF	atggcattgcaggattctgaagtcccttcctcaata ttaaacgccaccgctggtaatagaccaactgcttc ttatcacccaacattgtggggagagaagttcttgg ttgtttccactcaatctacctcaggttctatgaaaaa cgaaccaaccactcaaggtgaatacgacgaatta aagcaacaagtcacaaagatgttgactgatgcca ctactaatgacccatccaaaaagttgcatttaatcg atatggttcaacgtttgggtattgcttaccactttga aattgagatcgaaaacgctttggaaaaaataaact taggtgacgctaattatttcgaatacgatttgtacac cattgctttaggttttagattgttgagacaacaaggt atcaaggtctcttctgagattttcaagaaatttatgg acgaaaagggtaagttcaaagaagatgttgtcaa

				ASSNAALLRFAKIDFN MVQALHQKEISGITKW WKNLDFSTKLPYARDR IVECYFWIMGAYFEPK YSLARTFLTKVIAMTSI LDDTYDNYGTNKELEL LTKCIERWDIDVIDQLP EYMKLVYQALLNVYSE MEAKVAKEGRSYAIDY AKESMKKTMKAYLDE AKWRQEDYVPTIEEYM QVALISSAYPMLITNSF VGMGEVATKEAFDWIS NNPKMLKASTIICRLMD DITSHEFEQTRDHVASG VECYMKQYGVSREETV KLFREDVANAWKDINE GFMKPAIFPMPILTVVL NFARVMDFLYKDGDN YTNSHMLKDYITSLLV NPLLI (SEQ ID NO: 31)	cgatgttttgggtatgttgaacttgtacgaagcagc tcatttaagattaagaggtgaagacatcttggacg aagccttggccttcacaacctcccacttagagtca atggctactaaggtctctcctttgttggctgaacaa attgcccatgctttgaactgcccaatccaaaaggg tttaccacgtattgaagcaagacactatatttctttat actccagagaaactcacttcgcttcctctaatgctg ctttgttgagatttgctaagatcgatttcaatatggtt caagccttgcatcagaaggaaatatcaggtataa ccaaatggtggaagaacttggacttttccactaaa ttaccatatgctagagatcgtattgttgaatgttactt ctggatcatgggtgcttactttgaaccaaagtattc tttagcaagaacattcttgaccaaagtcattgcaat gacctctatcttagacgatacttacgacaactacg gtactaacaaggaattggagttgttgactaagtgt atcgaaagatgggatattgatgttatcgaccagtta cctgagtatatgaagttggtttatcaagctttgttaa atgtttactctgaaatggaagctaaggtcgccaaa gaaggtcgttcctacgccattgactacgcaaaag aatctatgaagaaaaccatgaaagcctacttgga cgaggctaagtggagacaagaagattacgtccct accattgaagaatatatgcaagttgcattaatatca tccgcttatccaatgttgattacaaactcattcgtcg gtatgggtgaggtcgctactaaggaagcttttgac tggatctccaataacccaaagatgttgaaggcttc tactattatatgtagattgatggatgatatcacttcc catgaatttgaacagaccagagaccacgttgcct ctggtgttgaatgttacatgaaacaatacggtgtct ccagagaagaaaccgttaagttgttcagagaaga tgttgctaacgcttggaaggacatcaatgaaggtt tcatgaagccagcaatcttcccaatgcctatcttga ctgttgtcttgaattttgccagagttatggacttttgt acaaggatggtgataactatactaactctcatatgt taaaagactacattacctcattattggttaatccatt attgatttaa (SEQ ID NO: 97)
ShoCusS QTS156	*Shorea cuspidata*	ShoBecS QTS1	36%	ALQDSEVPSSILNATAG NRPTASYHPTLWGEKF LVVSTQSTSGSMKNEPT TQGEYDELKQQVTKML TDATTNDPSKKLHLID MVQRLGIAYHFEIEIEN ALEKINLGDANYFEYD LYTIALGFRLLRQQGIK VSSEIFKKFMDEKGKFK EDVVNDVLGMLNLYE AAHLRLRGEDILDEAL AFTTSHLESMATKVSPL LAEQIAHALNCPIQKGL PRIEARHYISLYSRETHF ASSNAALLRFAKIDFN MVQALHQKEISGITKW WKNLDFSTKLPYARDR	atggctttacaggactccgaggttccttcatctatat tgaacgccaccgctggtaatcgtccaactgcatct tatcatccaacattgtggggtgaaaaattcttggtc gtttctactcaatccacctctgggtccatgaagaac gaaccaactacccaaggtgaatacgatgaattaa agcaacaagtcacaaagatgttgactgatgctac cactaatgacccatctaaaaagttgcacttgattga catggttcaaagattaggtatcgcctaccactttga aattgagatcgaaaacgctttggaaaagattaact taggtgatgctaattatttcgaatacgatttgtacac tatagccttgggttttagattattgagacaacaggg tatcaaggtttcatctgaaatcttcaaaaagttcatg gacgagaaaggtaagtttaaggaagacgtcgtta acgatgtcttgggtatgttaaacttgtatgaagctg cccatttgagattgcgtggtgaagacattttagatg aggctttggcttttaccacatcccacttagaatcaa

				IVECYFWIMGAYFEPK YSLARTFLTKVIAMTSI LDDTYDNYGTNKELEL LTKCIERWDIDVIDQLP EYMKLVYQALLNVYSE MEAKVAKEGRSYAIDY AKESMKKTMKAYLDE AKWRQEDYVPPIEEYM QVARISSGYPMLITNSL VGMGEVATKEAFDLIS NDPKMLKASTTICRLM DDITSHEFEQTRDHVAS GVECYMKQYGVSREET VKLFREDVANAWKDIN EGFMKPAIFPMPILTVV LNFARVMDFLYKDGD NYTNSHMLKDYITSLL VNPLLI (SEQ ID NO: 32)	tggcaactaaggtttcacctttgttggctgaacaaa tcgcccacgctttaaattgcccaattcaaaaaggtt tgccaagaatagaagccagacattacatttctttgt actccagagaaacccacttcgcttcttctaacgca gcattgttgcgtttcgctaagatcgactttaatatgg ttcaagcattgcatcagaaagagatttccggtatta ctaagtggtggaagaatttagatttctctacaaaatt gccatatgctagagatagaatcgtcgaatgttactt ctggattatgggtgcttattttgaaccaaagtactct ttggccagaacctttttaaccaaagtcattgctatg acttctatcttagatgacacatacgacaattacggt actaacaaggaattggaattgttaaccaagtgtatt gaaagatgggatatagatgttatcgatcaattgcct gaatacatgaagttagtttatcaagctttgttgaac gtctactccgaaatggaggctaaggtcgctaagg aaggtcgttcctatgccatcgattacgctaaggaa tccatgaaaaagactatgaaagcctatttggacga agctaagtggagacaagaggactacgttccacct atcgaagagtacatgcaagttgcaagaatttcttc cggttatccaatgttaattaccaactccttggttggt atgggtgaagtcgccactaaagaagccttcgattt gatttctaacgacccaaaaatgttgaaggcttcca ccactatatgtagattgatggacgatatcacttctc acgaatttgaacaaactagagatcacgtcgcttca ggtgttgaatgttatatgaagcaatacggtgtttctc gtgaggaaaccgttaagttattcagagaagacgt cgctaacgcatggaaggacattaatgagggtttc atgaagccagcaatctttccaatgccaatcttgact gtcgtcttaaacttcgctagagttatggacttttgta caaagatggtgataattacacaaactctcatatgtt aaaggattacatcacttcattgttggtcaacccttg ttgatttaa (SEQ ID NO: 98)
ShoCusS QTS157	*Shorea cuspidata*	ShoBecS QTS1	38%	ALQDSEVPSSILNATAG NRPTASYHPTLWGEKF LVVSTQSTSGSMKNEPT TQGEYDELKQQVTKML TDATTNDPSKKLHLID MVQRLGIAYHFEIEIEN ALEKINLGDANYFEYD LYTIALGFRLLRQQGIK VSSEIFKKFMDEKGKFK EDVVNDVLGMLNLYE AAHLRLRGEDILDEAL AFTTSHLESMATKVSPL LAEQIAHALNCPIQKGL PRIEARHYISLYSRETHF ASSNAALLRFAKIDFN MVQALHQKEISGITKW WKNLDFSTKLPYARDR IVECYFWIMGAYFEPK YSLARTFLTKVIAMTSI LDDTYDNYGTNKELEL	atggccttacaggactccgaagttccatcatctatt ttgaacgctactgctggtaatagacctacagcatc ttaccatccaaccttgtggggagagaagtttttggt cgtttccactcaatctacctccggttctatgaaaaa cgaaccaactacacaaggtgaatatgatgaatta aagcaacaagtcaccaagatgttgactgatgcta ctaccaacgacccatctaaaaagttgcacttaata gatatggttcaacgtttgggtatcgcctaccacttc gagattgaaatcgaaaatgctttagaaaaaattaa cttgggtgacgctaactacttcgaatatgatttgta cactatcgcattaggttttagattgttgagacaaca gggtattaaggtctcctcagaaattttcaagaagtt catggatgaaaaaggtaagtttaaggaggacgtt gtcaatgacgttttaggtatgttgaacttgtatgaag ctgctcatttacgtttgagaggtgaagatatcttgg acgaagccttggctttcactacatcacacttggaat ctatggctaccaaggtttccccattgttggccgag caaatagcacatgccttaaattgtcctattcaaaaa ggtttgccaagaatcgaagctagacactacatctc

				LTKCIERWDIDVIDQLP EYMKLVYQALLNVYSE MEAKVAKEGRSYAIDY AKESMKKTMKAYLDE AKWRQEDYVPPMDEY MQVALISCGYPMLITNS FVGMGEVATKEAFDWI SNDPKILKASTTICRLM DDITSHEFEQTRDHVAS GVECYMKQYGVSREET VKLFREDVANAWKDIN EGFMKPAIFPMPILTVV LNFARVMDFLYKDGD NYTNSHMLKDYITSLL VNPLLI (SEQ ID NO: 33)	tttatactctcgtgaaactcactttgcttcctctaacg ctgccttgttgagattcgctaagattgattttaatatg gttcaagccttgcaccagaaagaaatctctggtat caccaagtggtggaagaatttggacttctccacca agttgccatatgctagagacagaattgtcgaatgc tacttctggataatgggtgcatattttgaacctaagt actctttagctagaacttttttgactaaagttattgct atgacatcaattttggatgatacttacgataactac ggtactaacaaagaattagaattattgaccaagtg tatcgagagatgggacattgacgtcattgaccaat taccagaatacatgaagttggtttatcaagctttgtt gaacgtctactccgagatggaagcaaaggttgcc aaggaaggtcgttcttatgctatagattatgctaaa gaatctatgaaaaagacaatgaaggcatacttgg acgaagctaagtggagacaagaggattatgttcc tccaatggatgaatacatgcaagttgctttgatatc ctgtggttacccaatgttgatcaccaactctttcgtt ggtatgggtgaagtcgctaccaaagaagcctttg attggatctctaatgacccaaagattttgaaagcat ctaccactatctgtagattaatggatgacattacct cccatgagttcgaacagacaagagatcacgttgc ttcaggtgtcgaatgttatatgaagcaatacggtgt ttctcgtgaagaaactgttaaattattcagagagga tgttgctaacgcttggaaagacattaatgaaggttt catgaagcctgctattttcccaatgccaattttgac cgtcgtcttgaatttcgctagagtcatggatttttat acaaggacggtgataactacacaaactcacatat gttgaaagattacatcacttcattattagttaatccat tgttgatataa (SEQ ID NO: 99)
ShoCusS QTS160	*Shorea cuspidata*	ShoBecS QTS1	36%	ALQDSEVPSSILNATAG NRPTASYHPTLWGEKF LVVSTQSTSGSMKNEPT TQGEYDELKQQVTKML TDATTNDPSKKLHLID MVQRLGIAYHFEIEIEN ALEKINLGDANYFEYD LYTIALGFRLLRQQGIK VSSEIFKKFMDEKGKFK EDVVNDVLGMLNLYE AAHLRLRGEDILDEAL AFTTSHLESMATKVSPL LAEQIAHALNCPIQKGL PRIEARHYISLYSRETHF ASSNAALLRFAKIDFN MVQALHQKEISGITKW WKNLDFATMLPYARD RIVECYFWIMGVYFEPK YSLARTFLTKVIAMTSI LDDTYDNYGTNKELEL LTKCIERWDIDVIDQLP EYMKLVYQALLNVYSE MEAKVAKEGRSYAIDY AKESMKKTMKAYLDE	atggcattacaggattcagaggtcccatcctctatt ttgaacgctactgccggtaatcgtcctaccgcttct taccacccaacattgtggggtgaaaagtttttagtt gtttccactcaatctacctccggctctatgaaaaac gaaccaaccactcaaggtgaatatgacgaattga agcaacaagtcactaagatgttgacagatgctact accaatgacccatctaaaaagttgcatttgatagat atggttcaaagattgggtattgcctaccacttcgaa atcgaaatcgaaaacgctttagaaaagattaattta ggtgacgctaactatttcgaatacgatttatacaca atcgctttgggttttagattgttgagacagcaaggt atcaaggtctcttcagagattttcaaaaagttcatg gatgagaaaggtaagtttaaggaagacgttgtca acgacgttttgggtatgttgaatttatatgaagcag cccatttgagattgcgtggtgaagatatattggac gaggctttagctttcactacctcccacttggaatct atggcaaccaaagtttccccattgttagctgaaca aattgcccacgctttgaactgtcctatccaaaagg gtttgccaagaattgaagccagacattacatatctt tgtattcaagagaaactcacttcgcttcttccaatg ctgctttattaagatttgctaagatcgattttaacatg gtccaagctttgcatcaaaaagagatctctggtatt

				AKWRQEDYVPTIEEYM QVALISSAYPMLITNSF VGMGEVATKEAFDWIS NNPKMLKASTIICRLMD DITSHEFEQTRDHVASG VECYMKQYGVSREETV KLFREDVANAWKDINE GFMKPAIFPMPILTVVL NFARVMDFLYKDGDN YTNSHMLKDYITSLLV NPLLI (SEQ ID NO: 34)	acaaagtggtggaagaacttggacttcgctactat gttaccatacgccagagatcgtattgttgaatgcta cttctggatcatgggtgtttattttgaaccaaagtac tccttagctagaaccttcttgaccaaagttattgca atgacttctattttagacgatacatacgacaactac ggtactaataaggaattggaattgttgactaaatgt attgaaagatgggacatcgatgtcattgatcaattg cctgagtatatgaagttggtttatcaggcattattga acgtctactcagaaatggaagctaaggttgccaa agagggtagatcctacgctattgattacgccaaa gaatctatgaagaagaccatgaaggcctatttgg acgaagctaagtggagacaagaagactacgtcc ctaccatcgaagaatatatgcaagtcgctttaatat cttcagcctacccaatgttaataactaattcatttgt cggtatgggtgaggttgccactaaggaagcttttg attggatctctaacaacccaaaaatgttaaaggctt ccactattatttgtagattgatggatgacatcacctc ccacgaatttgaacagacccgtgaccacgttgcc tctggtgttgaatgttatatgaagcaatacggtgttt cacgtgaggaaaccgtcaagttgttcagagaaga tgttgctaatgcttggaaagacatcaatgagggttt catgaagccagcaatcttcccaatgccaattttga ctgtcgttttgaacttcgcaagagttatggatttctta tataaggacggcgacaattacactaactctcatat gttgaaagactacatcacttctttgttggtcaaccc attgttaatataa (SEQ ID NO: 100)
ShoCusS QTS161	*Shorea cuspidata*	ShoBecS QTS1	37%	ALQDSEVPSSILNATAG NRPTASYHPTLWGEKF LVVSTQSTSGSMKNEPT TQGEYDELKQQVTKML TDATTNDPSKKLHLID MVQRLGIAYHFEIEIEN ALEKINLGDANYFEYD LYTIALGFRLLRQQGIK VSSEIFKKFMDEKGKFK EDVVNDVLGMLNLYE AAHLRLRGEDILDEAL AFTTSHLESMATKVSPL LAEQIAHALNCPIQKGL PRIEARHYISLYSRETHF ASSNAALLRFAKIDFN MVQALHQKEISGITKW WKNLDFATMLPYARD RIVECYFWIMGVYFEPK YSLARTFLTKVIAMTSI LDDTYDNYGTNKELEL LTKCIERWDIDVIDQLP EYMKLVYQALLNVYSE MEAKVAKEGRSYAIDY AKESMKKTMKAYLDE AKWRQEDYVPPIEEYM QVARISSGYPMLITNSL VGMGEVATKEAFDLIS	atggctttgcaagactctgaagtcccttcctcaattt taaacgcaaccgctggtaatagaccaacagcctc ttaccatccaacttttgtggggtgagaaatttttggtt gtttccactcagtctacctcaggttctatgaagaac gaaccaactacccaaggtgaatatgatgaattga agcaacaagtcactaagatgttaacagatgctact accaatgacccatccaaaaagttgcacttgataga tatggttcaacgtttgggtatcgcctaccacttcga aatcgagattgaaaacgctttagagaaaatcaact tgggcgacgctaattacttcgaatatgatttataca ccattgccttaggttttagattgttgagacaacaag gtattaaggtttcttccgaaattttcaagaagtttatg gatgaaaaaggtaagttcaaggaagacgtcgtta acgacgttttaggtatgttgaacttgtatgaagctg cccatttaagattgcgtggtgaagatatcttggatg aagctttagcattcacaacctctcacttggaatctat ggctactaaagtctctccattgttagctgagcagat cgcccacgctttgaattgccctatccaaaagggtt tgccaagaatagaagcaagacattacatttccttgt actcaagagaaacacacttcgcttcctctaacgct gctttgttaagatttgctaaaattgactttaatatggt tcaagccttacatcaaaaggagatttctggtatcac caagtggtggaagaacttggacttcgcaactatgt tgccatacgcaagagaccgtattgttgaatgttatt tctggatcatgggtgtctacttcgaacctaagtact

				NDPKMLKASTTICRLM DDITSHEFEQTRDHVAS GVECYMKQYGVSREET VKLFREDVANAWKDIN EGFMKPAIFPMPILTVV LNFARVMDFLYKDGD NYTNSHMLKDYITSLL VNPLLI (SEQ ID NO: 35)	cattggctagaacttttttaactaaagtcatagccat gacctccattttggatgacacctacgataactatg gtactaacaaggaattagagttgttaacaaagtgt atagaaagatgggacattgatgtcatcgatcaatt gcctgaatacatgaagttggtttaccaggctttgtt aaatgtctactcagaaatggaagctaaggttgcta aagaaggtcgttcttatgcaattgattacgcaaag gagtctatgaagaaaactatgaaagcttatttgga cgaagctaaatggagacaagaagactatgttcca ccaatcgaagaatatatgcaagtcgctagaatctc ttccggttacccaatgttgattactaactcattagtc ggtatgggtgaggttgccactaaggaagctttcg acttgatttctaatgatccaaagatgttaaaagcct ccactacaatctgtagattgatggacgacattactt ctcatgaatttgaacagacacgtgatcacgttgcc tctggtgtcgagtgctatatgaagcaatacggtgtt tccagagaagaaaccgtcaagttgtttagagaag acgttgctaacgcttggaaggatatcaatgaagg cttcatgaaaccagcaatctttccaatgccaattttg accgttgttttgaacttcgctagagtcatggacttct tgtataaggatggcgacaactacactaattcacat atgttgaaagattacataacttcattattagttaacc ctttattgatctaa (SEQ ID NO: 101)
WenAng SQTS1007	*Wendlandia angustifolia*	A0A068UHT0	81%	ASAQASLPSNNRQETV RPLADFPENIWADRIAP FTLDKQEYEMCQREIE MLKAEVASMLLATGKT MMQRFDFIDKIERLGVS HHFDIEIENQLQEFFNV YTNLGEYSAYDLSSAA LQFRLFRQHGFNISCGIF DQFIDAKGKFKESLCN DIRGLLSLYEAAHVRTH GDKILEEALAFTTTHMT SGGPHLDSSLAKQVKY ALEQPLHKGILRYEAW RYISIYEEDESNNKLLL RLAKLDYHLLQMSYKQ ELCEITRWGKGLESVSN FPYARDRFVECYFWAV GTLYEPQYSLARMTFA KVAALITMIDDIYDAYG TLDELQILTDSAERWD GSGVDQLSDYIRASYN TLLKFNKEVGEDLAKK QRTYAFDKYIEDWKQY MRTNFSQSRWFFTKEL PSFADYINNGAITIGAY LVASAAFLYMDSAKED VINWMSTNPKLVVAYS THSRLINDFGGHKFEKE RGSSTAIECYMKDHNV SEEEAANKFREMMEDA	atggcctcagcacaagcttccttaccttctaataac agacaggaaacagtccgtccattggctgacttcc cagagaacatctgggctgatagaattgccccattt accttggataagcaagaatacgaaatgtgtcaaa gagaaatagagatgttaaaagctgaagttgcttct atgttgttggcaactggtaagactatgatgcaaag attcgacttcattgataagatcgaaagattggggg tctcccaccattttgacattgaaatcgaaaatcaatt gcaagagtttttcaacgtttataccaacttaggtga atactctgcctatgatttgtcatctgctgccttgcag ttccgtttatttagacaacacggtttcaatatttcctg cggtattttcgaccaatttatcgacgctaaaggtaa gttcaaggaatctttatgtaacgatatcagaggttt gttgtctttgtacgaagctgctcatgttagaactca cggtgataaaattttggaagaagctttagctttcac cactactcacatgacctccggtggtccacatttag attcttcattggccaagcaagttaaatacgcattgg aacagccattgcataagggtatattgagatatgaa gcttggagatacatatctatctacgaagaggacg aatccaacaataagttattattgcgtttggctaagtt ggactatcacttgttacaaatgtcatacaagcaag agttgtgtgaaattacaagatggggtaaaggtttg gaatctgtctccaactttccttatgcccgtgacaga ttcgttgaatgttacttttgggctgtcggtactttgta cgaaccacaatactcattggctagaatgaccttcg ctaaggttgctgctttaattactatgatcgatgatatt tatgatgcctacggtaccttggacgaattgcaaat attaactgactctgccgaaagatgggatggttccg

				WKVMNEECLRPTTIPR DGLKMLLNIARVGETV YKHRIDGFTQPHAIEEH IRAMLVDFMSI (SEQ ID NO: 36)	gtgtcgatcagttgtctgactatattagagcttccta taatacattattgaaatttaataaggaggttggtga agatttggcaaaaaagcaacgtacctacgctttcg acaagtacatcgaagattggaaacaatacatgag aaccaacttctctcaatcaagatggtttttcactaag gagttgccatctttcgctgattacattaacaacggt gccatcacaatcggtgcatatttggttgcctctgct gctttcttatatatggactccgcaaaagaagatgtt atcaactggatgtccacaaaccctaagttggtcgt tgcttactccactcactctcgtttaattaatgactttg gtggtcacaagttcgaaaaggagagaggttcctc tactgctattgaatgctacatgaaggaccataatgt ctccgaagaagaagccgcaaacaagtttagaga aatgatggaggacgcttggaaggttatgaatgaa gaatgtttaagaccaactaccatccctagagacg ggttgaagatgttgttaaacatagccagagtcggt gaaactgtttacaagcatagaatcgatggttttacc caaccacatgctattgaagaacacataagagcca tgttggtcgatttcatgtctatttaa (SEQ ID NO: 102)
WenAng SQTS1086	*Wendlandia angustifolia*	A0A068UHT0	80%	ASAQASLPSNNRQETV RPLADFPENIWADRIAP FTLDKQEYEMCQREIE MLKAEVASMLLATGKT MMQRFDFIDKIERLGVS HHFDIEIENQLQEFFNV YTNLGEYSAYDLSSAA LQFRLFRQHGFNISCGIF DQFIDAKGKFKESLCN DIRGLLSLYEAAHVRTH GDKILEEALAFTTTHMT SGGPHLDSSLAKQVKY ALEQPLHKGILRYEAW RYISIYEEDESNNKLLL RLAKLDYHLLQMSYKQ ELCEITRWGKGLESVSN FPYARDRFVECYFWAV GTLYEPQYSLARMTFA KVAALITMIDDIYDAYG TLDELQILTDSAERWD GSGVDQLSDYIRASYN TLLKFNKEVGEDLAKK QRTYAFDKYIEDWKQY MRTNFSQSRWFFTKEL PSFADYINNGAITIGAY LVASAAFLYMDSAKED VINWMSTNPKLVVAYS THSRLINDFGGHKFDKE RGTGTAIECYMKDHNIS EEEAAKKFREMIENTW KVMNEECLRPIPIPRDT LKMLLNIARVGETVYK HRIDGFTQPHAIEEHIRA	atggcctcagcacaagcttccttaccttctaataac agacaggaaacagtccgtccattggctgacttcc cagagaacatctgggctgatagaattgccccattt accttggataagcaagaatacgaaatgtgtcaaa gagaaatagagatgttaaaagctgaagttgcttct atgttgttggcaactggtaagactatgatgcaaag attcgacttcattgataagatcgaaagattggggg tctcccaccattttgacattgaaatcgaaaatcaatt gcaagagtttttcaacgtttataccaacttaggtga atactctgcctatgatttgtcatctgctgccttgcag ttccgtttatttagacaacacggtttcaatatttcctg cggtattttcgaccaatttatcgacgctaaaggtaa gttcaaggaatctttatgtaacgatatcagaggttt gttgtctttgtacgaagctgctcatgttagaactca cggtgataaaattttggaagaagctttagctttcac cactactcacatgacctccggtggtccacatttag attcttcattggccaagcaagttaaatacgcattgg aacagccattgcataagggtatattgagatatgaa gcttggagatacatatctatctacgaagaggacg aatccaacaataagttattattgcgtttggctaagtt ggactatcacttgttacaaatgtcatacaagcaag agttgtgtgaaattacaagatggggtaaaggtttg gaatctgtctccaactttccttatgcccgtgacaga ttcgttgaatgttacttttgggctgtcggtactttgta cgaaccacaatactcattggctagaatgaccttcg ctaaggttgctgctttaattactatgatcgatgatatt tatgatgcctacggtaccttggacgaattgcaaat attaactgactctgccgaaagatgggatggttccg gtgtcgatcagttgtctgactatattagagcttccta taatacattattgaaatttaataaggaggttggtga agatttggcaaaaaagcaacgtacctacgctttcg

				MLVDFMSI (SEQ ID NO: 37)	acaagtacatcgaagattggaaacaatacatgag aaccaacttctctcaatcaagatggtttttcactaag gagttgccatctttcgctgattacattaacaacggt gccatcacaatcggtgcatatttggttgcctctgct gctttcttatatatggactccgcaaaagaagatgtt atcaactggatgtccacaaaccctaagttggtcgt tgcttactccactcactctcgtttaattaatgactttg gtggtcacaagttcgacaaggagagaggtaccg gtactgctattgaatgctacatgaaggaccataat atatccgaagaagaagccgcaaagaagtttaga gaaatgatcgagaacacctggaaggtcatgaatg aagaatgtttaagaccaattccaatccctagagac acattgaagatgttgttaaacatcgccagagttggt gaaactgtctacaagcatagaatcgatggttttact caaccacatgctattgaagaacacataagagctat gttggttgatttcatgtctatttaa (SEQ ID NO: 103)
WenAngSQTS267	*Wendlandia angustifolia*	G5CV47	11%	SLLEGNVNHENGIFRPE ANFSPSMWGNIFRDSSK DNQISEEVVEEIEALKE VVKHMIISTTSNAIEQK LELVDNLERLGLAYHF EGQINRLLSSAYNANHE DEGNHKRNKEDLYAA ALEFRIFRQHGFNVSSD CFNQFKDTKGKFKKTL LIDVKGMLSLYEAAHV REHGDDILEEALIFATF HLERITPNSLDSTLEKQ VGHALMQSLHRGIPRA EAHFNISIYEECGSSNEK LLRLAKLDYNLVQVLH KEELSELTKWWKDLDF ASKLSYVRDRMVECFF WTVGVYFEPQYSRARV MLAKCIAMISVIDDTYD SYGTLDELIIFTEVVDR WDISEVDRLPNYMKPI YISLLYLFNEYEREINEQ DRFNGVNYVKEAMKEI VRSYYIEAEWFIEGKIPS FEEYLNNALVTGTYYL LAPASLLGMESTSKRTF DWMMKKPKILVASAII GRVIDDIATYKIEKEKG QLVTGIECYMQENNLS VEKASAQLSEIAESAW KDLNKECIKTTTSNIPN EILMRVVNLTRLIDVVY KNNQDGYSNPKNNVKS VIEALLVNPINM (SEQ ID NO: 38)	atgtccttgttagaaggtaacgttaatcacgagaa cggaatatttagaccagaagctaatttctcaccttc tatgtggggtaacattttccgtgattcttccaaaga caaccaaatctctgaagaagtcgttgaagaaatc gaggcattgaaggaagtcgttaagcatatgattat ttctacaacctccaacgccatcgaacagaaattag agttggtcgataatttggaaagattgggtttggctt accacttcgaaggtcaaatcaacagattattatcat ctgcctataatgctaaccatgaagacgaaggtaa ccacaagagaaataaggaggacttgtacgcagc tgctttggaatttagaattttcagacaacatggtttta acgtttcctctgattgctttaatcaattcaaagatact aagggtaagttcaaaaagactttgttgattgatgtc aagggtatgttgtccttgtatgaagctgcccacgtt cgtgaacatggtgacgacatcttagaagaagcttt gatctttgctaccttccacttagaaagaattactcca aattctttggattccacattggaaaaacaagttggt cacgcattgatgcaatcattacacagaggtattcc aagagccgaagctcattttaacatatctatttacga agagtgtggttcttctaatgaaaagttgttaagattg gctaagttggactacaacttagtccaagtcttgca caaggaggaattatcagaattgaccaaatggtgg aaagatttagacttcgcttctaagttgtcctacgttc gtgatagaatggttgaatgtttttctggactgtcgg tgtttatttcgaaccacagtactccagagccagag ttatgttagctaagtgtattgctatgatctctgttatc gacgatacttacgattcctatggtaccttggacga gttaattatattcactgaagtcgttgatagatgggat atatccgaggtcgaccgtttgcctaactatatgaa accaatctacatttctttgttatacttgtttaacgaata tgaaagagaaattaacgaacaagaccgtttcaat ggtgttaactacgttaaggaagctatgaaggaaat cgtcagatcttattacatcgaggccgaatggttcat agaaggtaaaatcccatctttcgaagagtacttga

					acaatgcattggttacaggtacctattacttattggc cccagcatctttgttgggtatggaatccacctcaa agagaacttttgattggatgatgaagaagccaaa aattttggtcgcttctgctatcattggtagagttattg atgatattgctacttacaagatagaaaaggaaaag ggacagttagtcactggtattgaatgctacatgca agagaacaacttatcagttgaaaaggcctccgct caattgtctgaaatcgccgagtccgcttggaaag acttgaataaagaatgtatcaaaactaccacctcc aacattcctaacgaaatattgatgagagttgtcaac ttgacaagattaattgacgttgtctacaagaataat caagatggttattctaaccctaagaacaatgttaag tcagtcatcgaagctttgttggttaatccaatcaata tgtaa (SEQ ID NO: 104)
WenAng SQTS302	*Wendlandia angustifolia*	Q5SBP4	17%	ESRRSANYQASIWDDN FIQSLASPYAGEKYVSQ ANELKEQVKMMLDEE DMKLLDCLELVDNLER LGLAYHFEGQINRLLSS AYNANHEDEGNHKRN KEDLYAAALEFRIFRQH GFNVPQDVFSSFMNKA GDFEESLSKDTKGLVSL YEASYLSMEGETILDM AKDFSSHHLHKMVEDA TDKRVANQIIHSLEMPL HRRVQKLEAIWFIQFYE CGSDANPTLVELAKLD FNMVQATYQEELKRLS RWYEETGLQEKLSFAR HRLAEAFLWSMGIIPEG HFGYGRMHLMKIGAYI TLLDDIYDVYGTLEELQ VLTEIIERWDINLLDQLP EYMQIFFLYMFNSTNEL AYEILRDQGINVISNLK GLWVELSQCYFKEATW FHNGYTPTTEEYLNVA CISASGPVILFSGYFTTT NPINKHELQSLERHAHS LSMILRLADDLGTSSDE MKRGDVPKAIQCFMND TGCCEEEARQHVKRLI DAEWKKMNKDILMEK PFKNFCPTAMNLGRISM SFYEHGDGYGGPHSDT KKKMVSLFVQPMNITI (SEQ ID NO: 39)	atggaaagtaggcgttcagcaaattatcaggcttc catatgggatgacaactttattcaatctcttgcctct ccttacgctggagagaagtacgtctcgcaagcta acgaattgaaagaacaagtgaagatgatgttaga cgaagaggatatgaaactgttagattgcttggaat tggttgacaacttggaaagactaggcttggcttat cacttcgagggtcaaatcaatagactcttgagcag tgcctacaacgctaaccatgaagatgaaggtaat cacaagagaaataaggaagacttatacgcggcg gctttggagttcagaatttttagacaacatggtttca acgttccacaggacgtcttctcttcctttatgaataa ggccggtgattttgaagaatccctttctaaggatac aaaaggtttggtttcattgtatgaagcttcttacctat caatggaaggtgaaaccatcttagacatggctaa ggatttctcctctcaccatttacacaaaatggtcga agatgctactgataagcgagttgctaaccaaatca ttcacagccttgaaatgccattgcacagaagggta caaaaactcgaagcaatatggttcattcaattctac gaatgtggttctgacgccaaccccactttggtaga attggctaagttagacttcaacatggttcaagctac gtatcaagaagaactaaagagattgtcgagatgg tacgaagagaccggactgcaagaaaagttatcttt tgcacgtcatcgtttggccgaagctttttgtggtct atgggtatcattccagaaggccatttcggttacgg tagaatgcacttgatgaagatcggtgcctatattac tttattggatgatatttatgatgtctacggtaccttgg aagagttgcaagttctaactgaaatcatcgaacgt tgggacattaatttgttggaccagctgcctgagta catgcaaatcttcttttttatacatgttcaattccacaa acgaattagcttatgagatacttagagatcaagga attaatgttatctctaacctcaaagggttgtgggtc gaattgtcccagtgttatttttaaggaagcaacctgg tttcataacggttacactccaactacagaggaata cttgaacgttgcttgtattagtgcatctggtccagtg atccttttctccggttatttcaccacgactaacccga ttaataagcatgaattacaaagtttagaaagacac gctcattcactaagcatgattctgagattggctgac

					gaccttgggacctcatctgatgaaatgaaacggg gcgatgtgccaaaggccatccagtgctttatgaat gacactggttgttgtgaagaagaggcaagacaa cacgtcaaaagactcatagacgctgaatggaag aagatgaacaaggacatcttgatggaaaaaccct ttaagaacttctgtccaactgctatgaatttaggtag gataagcatgtccttttacgagcacggtgatggtt acggtggtccacactctgataccaaaaaaaagat ggttagcttgttcgttcaacctatgaacattaccatc taa (SEQ ID NO: 105)
WenAng SQTS738	*Wendlandia angustifolia*	A0A068VE40	46%	ASTEIAVPLNNQHESVR QLADFPENIWADRVAS FTLDKQGHDMCAKEIE MLKEEVMSMLLEEKP MMEKFNLIDNIERLGIS YHFGDKIEDQLQEYYD ACTNFEKHAECDLSIAA LQFRLFRQHGFNISCGIF DGFLDANGKFKESLCN DIKGLLSLYEAAHVRT HGDKILEEALFFTTTHL TREIPNVGSTLAKQVKY ALEQPLHKGIPRYEAW RYISIYEEDESSNKLLLR LAKLDYHLSQMLNKQ DLCEIIRWGKELDIISKV PYARDRIVECYFWAVA TYYEPQYSLARMTLTK ATVFAGMIDDTYDAYG TLDELKIFTEAVERWDS SGIDQLSDYMKAAYTL VLNFNKEVGEDLAKKQ RTYAFDKYIEEWKQYA RTSFTQSKWFLTNELPS FSDYLSNGMVTSTYYL LSAAAFLDMDSASEDVI NWMSTNPKLFVALTTH ARLANDVGSHKFEKER GSGTAIECYMKDYHVS EEEAMKKFEEMCDDA WKVMNEECLRSTTIPR EILKVILNLARTCEVVY KHRGDGFTDQRRIEAHI NAMLMDSVSI (SEQ ID NO: 40)	atggcctcaacagaaatcgcagttcctttgaataa ccagcacgagtccgtccgtcaattagctgacttcc cagaaaacatttgggctgatagagttgcttctttta ccttggataagcaaggtcatgacatgtgtgctaaa gaaatagaaatgttaaaggaagaagtcatgtctat gttgttggaggaaaagccaatgatggaaaaattc aacttgatcgataatattgaaagattaggcatctcc taccacttcggtgacaagattgaagatcaattaca agaatattacgacgcctgcactaactttgagaagc atgctgaatgtgatttgtcaatagctgccttgcaatt cagattgtttagacaacacggtttcaatatttcttgt ggtatctttgacggtttcttggatgcaaacggtaaa ttcaaggaatctttatgtaatgacattaagggtttgtt gtccttatacgaagccgctcatgttagaactcacg gtgataaaattttggaggaagctttgtttttttaccact actcatttgacccgtgaaatcccaaacgttggttct actttggctaagcaggtcaaatatgctttagagca accattgcacaagggtatcccaagatacgaagcc tggagatatatttcaatttacgaagaagacgaatct tccaacaagttgttattacgtttggcaaagttggatt accatttgtcccaaatgttgaacaaacaggacttgt gcgagatcattagatggggtaaggaattagacatt atttctaaggttccttatgctagagatagaatcgtc gaatgttacttctgggctgttgccacatattacgaa ccacaatactccttggctagaatgacattgaccaa agctactgtttttgctggtatgatcgatgatacctat gacgcttacggtactttagatgagttgaagatattc actgaagcagtcgaacgttgggactcttccggtat tgaccaattgtcagattacatgaaagcagcttaca ccttagtcttaaattttaacaaggaagttggtgaag atttagccaagaaacaaagaacttacgccttcgac aagtacatcgaagaatggaagcaatatgctagaa cctctttcacccaatctaagtggttcttgaccaatg agttgccatccttttctgattatttgtctaacggtatg gttacttcaacatactacttattgtctgccgctgcctt cttggacatggattccgcttctgaagacgtcataa attggatgtctaccaaccctaaattgttcgtcgcttt gacaactcacgctagattggccaacgacgttggt tctcataaatttgaaaaggaaagaggttcaggtac cgcaatagaatgttatatgaaggattaccacgtttc tgaggaagaagctatgaagaaattcgaggaaat

					gtgtgacgatgcttggaaggtcatgaacgaagaa tgcttgcgttccactacaatcccaagagagattttg aaggttattttgaacttggcaagaacttgtgaagtc gtttacaagcatcgtggtgatggcttcaccgatca aagaagaattgaagctcacatcaacgccatgtta atggactccgtttccatctaa (SEQ ID NO: 106)
WenAng SQTS760	*Wendlandia angustifolia*	A0A068VE40	43%	ASTEIAVPLNNQHESVR QLADFPENIWADRVAS FTLDKQGHDMCAKEIE MLKEEVMSMLLEEKP MMEKFNLIDNIERLGIS YHFGDKIEDQLQEYYD ACTNFEKHAECDLSIAA LQFRLFRQHGFNISCGIF DGFLDANGKFKESLCN DIKGLLSLYEAAHVRT HGDKILEEALFFTTTHL TREIPNVGSTLAKQVKH ALEQPLHRGIPRYEAYC FISIYEEDESNNKLLLRL AKLDYHLLQMSYKREL SEIIRWGKELDIISKVPY ARDRIVECYFWAVATY YEPQYSLARMTLTKAT VFAGMIDDTYDAYGTL DELKIFTEAVERWDSSG IDQLSDYMKAAYTLVL NFNKEVGEDLAKKQRT YAFDKYIEEWKQYART SFTQSKWFLTNELPSFS DYLSNGMVTSTYYLLS AATFLGMDGASEDVIN WMSTNPKLFVALTTHA RLANDVGSHKFEKERG SGTAIECYMKDYHVSE EEAMKKFEEMCDDAW KVMNEECLRSTTIPREI LKVILNLARTCEVVYK HRGDGFTDQRRIEAHIN AMLMDSVSI (SEQ ID NO: 41)	atggcctcaacagaaatcgcagttcctttgaataa ccagcacgagtccgtccgtcaattagctgacttcc cagaaaacatttgggctgatagagttgcttctttta ccttggataagcaaggtcatgacatgtgtgctaaa gaaatagaaatgttaaaggaagaagtcatgtctat gttgttggaggaaaagccaatgatggaaaaattc aacttgatcgataatattgaaagattaggcatctcc taccacttcggtgacaagattgaagatcaattaca agaatattacgacgcctgcactaactttgagaagc atgctgaatgtgatttgtcaatagctgccttgcaatt cagattgtttagacaacacggtttcaatatttcttgt ggtatctttgacggtttcttggatgcaaacggtaaa ttcaaggaatctttatgtaatgacattaagggtttgtt gtccttatacgaagccgctcatgttagaactcacg gtgataaaattttggaggaagctttgtttttaccact actcatttgacccgtgaaatcccaaacgttggttct actttggctaagcaggtcaaacacgctttagagca accattgcacagaggtatcccaagatatgaagcc tactgcttcatttcaatttatgaagaagacgaatcta acaacaagttgttattacgtttggcaaagttggatt accatttgttgcaaatgtcctacaaaagagaattgt ccgagatcattagatggggtaaggaattagacatt atttctaaggttccttatgctagagatagaatcgtc gaatgttacttttgggctgttgccacatattacgag ccacaatactccttggctagaatgacattgaccaa agctactgttttcgctggtatgatcgatgatacctat gacgcttacggtactttagacgaattgaagatattc actgaagcagtcgaacgttgggattcttccggtat tgaccaattgtcagattacatgaaagcagcttaca ccttagtcttaaattttaacaaggaagttggtgagg atttagccaagaaacaaagaacttacgccttcgac aagtacatcgaagaatggaagcaatatgctagaa cctctttcacccaatctaagtggttcttgaccaatg aattgccatccttttctgattatttgtctaacggtatg gttacttcaacatactacttattgtctgccgctacatt cttgggtatggacggtgcttctgaagacgtcataa attggatgtctactaaccctaaattgttcgtcgcttt gacaacccatgctagattggccaacgacgttggt tctcacaagtttgaaaaggaaagaggctccggta ctgcaatagaatgttatatgaaagattaccacgttt ctgaggaggaagctatgaagaaattcgaagaaat gtgtgacgatgcctggaaggtcatgaacgaaga atgcttgcgttctactaccatcccaagagagatttt gaaggttattttgaacttggccagaacctgtgaag

					tcgtttacaagcatcgtggtgatggtttcactgatc agagaagaattgaagctcacatcaacgctatgtta atggactccgtttccatctaa (SEQ ID NO: 107)
WenAng SQTS780	*Wendlandia angustifolia*	A0A068VE40	41%	ASTEIAVPLNNQHESVR QLADFPENIWADRVAS FTLDKQGHDMCAKEIE MLKEEVMSMLLEEKP MMEKFNLIDNIERLGIS YHFGDKIEDQLQEYYD ACTNFEKHAECDLSIAA LQFRLFRQHGFNISCGIF DGFLDANGKFKESLCN DIKGLLSLYEAAHVRT HGDKILEEALFFTTTHL TREIPNVGSTLAKQVKH ALEQPLHRGIPRYEAYC FISMYEEDESSNKLLLR LAKLDYHLSQMLNKQ DLCEIIRWGKELDIISKV PYARDRIVECYFWAVA TYYEPQYSLARMTLTK ATVFAGMIDDTYDAYG TLDELKIFTEAVERWDS SGIDQLSDYMKAAYTL VLNFNKEVGEDLAKKQ RTYAFDKYIEEWKQYA RTSFTQSKWFLTNELPS FSDYLSNGMVTSTYYL LSAATFLGMDGASEDV INWMSTNPKLFVALTT HARLANDVGSHKFEKE RGSGTAIECYMKDYHV SEEEAMKKFEEMCDDA WKVMNEECLRSTTIPR EILKVILNLARTCEVVY KHRGDGFTDQRRIEAHI NAMLMDSVSI (SEQ ID NO: 42)	atggcctcaacagaaatcgcagttcctttgaataa ccagcacgagtccgtccgtcaattagctgacttcc cagaaaacatttgggctgatagagttgcttctttta ccttggataagcaaggtcatgacatgtgtgctaaa gaaatagaaatgttaaaggaagaagtcatgtctat gttgttggaggaaaagccaatgatggaaaaattc aacttgatcgataatattgaaagattaggcatctcc taccacttcggtgacaagattgaagatcaattaca agaatattacgacgcctgcactaactttgagaagc atgctgaatgtgatttgtcaatagctgccttgcaatt cagattgtttagacaacacggtttcaatatttcttgt ggtatctttgacggtttcttggatgcaaacggtaaa ttcaaggaatctttatgtaatgacattaagggtttgtt gtccttatacgaagccgctcatgttagaactcacg gtgataaaattttggaggaagctttgtttttttaccact actcatttgacccgtgaaatcccaaacgttggttct actttggctaagcaggtcaaacacgctttagagca accattgcacagaggtatcccaagatatgaagcc tactgcttcatttcaatgtatgaagaagacgaatctt ccaacaagttgttattacgtttggcaaagttggatt accatttgtcccaaatgttgaacaaacaggacttgt gtgagatcattagatggggtaaggaattagacatt atttctaaggttccttatgctagagatagaattgtcg aatgttacttttgggctgttgccacatactacgaac cacaatattccttggctagaatgacattgaccaaa gctactgttttcgctggtatgatcgatgatacctatg acgcttacggtactttagatgagttgaagatattca ctgaagcagtcgaacgttgggactcttccggtatt gaccaattgtcagattacatgaaagcagcttacac cttagtcttaaattttaacaaggaagttggtgaaga tttagccaagaaacaaagaacttacgccttcgaca agtacatcgaagaatggaagcaatatgctagaac ctctttcacccaatctaagtggttcttgaccaatga gttgccatccttttctgattatttgtctaacggtatgg ttacttcaacatactacttattgtctgccgctacattc ttgggtatggacggtgcttctgaagatgtcataaat tggatgtctactaaccctaaattgttcgtcgctttga caacccatgctagattggccaacgacgttggttct cacaagtttgaaaaggaaagaggctccggtactg caatagaatgctatatgaaagattaccacgtttctg aggaagaagctatgaagaaattcgaggaaatgt gtgacgatgcctggaaggtcatgaacgaagaat gtttgcgttctactaccatcccaagagagattttga aggttattttgaacttggccagaacctgtgaagtc gtttacaagcatcgtggtgatggtttcactgaccaa agaagaatcgaagctcacattaacgctatgttaat

					ggactccgtttccatctaa (SEQ ID NO: 108)
WenAngSQTS793	*Wendlandia angustifolia*	A0A068UHT0	75%	ASAQASLPSNNRQETV RPLADFPENIWADRIAP FTLDKQEYEMCQREIE MLKAEVASMLLATGKT MMQRFDFIDKIERLGVS HHFDIEIENQLQEFFNV YTNLGEYSAYDLSSAA LQFRLFRQHGFNISCGIF DQFIDAKGKFKESLCN DIRGLLSLYEAAHVRTH GDKILEEALAFTTTHMT SGGPHLDSSLAKQVKY ALEQPLHKGILRYEAW RYISIYEEDESNNKLLL RLAKLDYHLLQMSYKQ ELCEITRWGKGLESVSN FPYARDRFVECYFWAV GTLYEPQYSLARMTFA KVAALITMIDDIYDAYG TLDELQILTDSAERWD GSGVDQLSDYIRASYN TLLKFNKEVGEDLAKK QRTYAFDKYIEDWKQY MRTSFTQSKWFLTNEL PSFADYISNGAITIGAYL IASAGFLDMDSALEDVI NWMSTNPKLMVAYST HSRLINDYGGHKFDKE RGSVTALDCYMKDYSV SEEEAAKKFREMCEDN WKVMNEECLRPTTIPR DGLKMLLNIARVGETV YKHRIDGFTQPHAIEEH IRAMLVDFMSI (SEQ ID NO: 43)	atggcctcagcacaagcttccttaccttctaataac agacaggaaacagtccgtccattggctgacttcc cagagaacatctgggctgatagaattgccccattt accttggataagcaagaatacgaaatgtgtcaaa gagaaatagagatgttaaaagctgaagttgcttct atgttgttggcaactggtaagactatgatgcaaag attcgacttcattgataagatcgaaagattggggg tctcccaccattttgacattgaaatcgaaaatcaatt gcaagagtttttcaacgtttataccaacttaggtga atactctgcctatgatttgtcatctgctgccttgcag ttccgtttatttagacaacacggtttcaatatttcctg cggtattttcgaccaatttatcgacgctaaaggtaa gttcaaggaatctttatgtaacgatatcagaggttt gttgtctttgtacgaagctgctcatgttagaactca cggtgataaaattttggaagaagctttagctttcac cactactcacatgacctccggtggtccacatttag attcttcattggccaagcaagttaaatacgcattgg aacagccattgcataagggtatattgagatatgaa gcttggagatacatatctatctacgaagaggacg aatccaacaataagttattattgcgtttggctaagtt ggactatcacttgttacaaatgtcatacaagcaag agttgtgtgaaattacaagatggggtaaaggtttg gaatctgtctccaactttccttatgcccgtgacaga ttcgttgaatgttacttttgggctgtcggtactttgta cgaaccacaatactcattggctagaatgaccttcg ctaaggttgctgctttaattactatgatcgatgatatt tatgatgcctacggtaccttggacgaattgcaaat attaactgactctgccgaaagatgggatggttccg gtgtcgatcagttgtctgactatattagagcttccta taatacattattgaaatttaataaggaggttggtga agatttggcaaaaaagcaacgtacctacgctttcg acaagtacatcgaagattggaaacaatacatgag aacctctttcactcaatcaaagtggtttttgactaac gagttgccatctttcgctgattacatttccaacggt gccatcacaatcggtgcatatttaattgcctctgcc ggtttttttggatatggattccgccttggaagacgtt attaactggatgtctaccaacccaaaattaatggtc gcttattccacccactcaagattgatcaatgattac ggtggtcacaagttcgacaaggaaagagggtca gttactgctttggattgctacatgaaggattactcc gtctctgaggaagaagctgcaaagaagttcaga gaaatgtgtgaagacaactggaaggttatgaatg aagaatgtttgagacctactacaattccaagagat ggtttgaagatgttgttaaacattgctagagtcggt gaaactgtttacaaacatagaatcgacggttttact caacctcatgcaatcgaggagcacattagagcca tgttagttgacttcatgtctatttaa (SEQ ID NO: 109)

WenAngSQTS805	*Wendlandia angustifolia*	A0A068VE40	42%	ASTEIAVPLNNQHESVRQLADFPENIWADRVASFTLDKQGHDMCAKEIEMLKEEVMSMLLEEKPMMEKFNLIDNIERLGISYHFGDKIEDQLQEYYDACTNFEKHAECDLSIAALQFRLFRQHGFNISCGIFDGFLDANGKFKESLCNDIKGLLSLYEAAHVRTHGDKILEEALFFTTTHLTREIPNVGSTLAKQVKYALEQPLHKGIPRYEAWRYISIYEEDESNNKLLLRLAKLDYHLLQMSYKRELSEIIRWGKELDIISKVPYARDRIVECYFWAVATYYEPQYSLARMTLTKATVFAGMIDDTYDAYGTLDELKIFTEAVERWDSSGIDQLSDYMKAAYTLVLNFNKEVGEDLAKKQRTYAFDKYIEEWKQYARTSFTQSKWFLTNELPSFSDYLSNGMVTSTYYLLSAATFLGMDGASEDVINWMSTNPKLFVALTTHARLANDVGSHKFEKERGSSTAIECYMKDYHVSEEEAMEKFEEMCDDAWKVMNEECLRSTTIPREILKVILNLARTCEVVYKHRGDGFTDQRRIEAHINAMLMDSVSI (SEQ ID NO: 44)	atggcctcaacagaaatcgcagttcctttgaataaccagcacgagtccgtccgtcaattagctgacttccagaaaacatttgggctgatagagttgcttctttaccttggataagcaaggtcatgacatgtgtgctaaagaaatagaaatgttaaaggaagaagtcatgtctatgttgttggaggaaaagccaatgatggaaaaattcaacttgatcgataatattgaaagattaggcatctcctaccacttcggtgacaagattgaagatcaattacaagaatattacgacgcctgcactaactttgagaagcatgctgaatgtgatttgtcaatagctgccttgcaattcagattgtttagacaacacggtttcaatatttcttgtggtatctttgacggtttcttggatgcaaacggtaaattcaaggaatctttatgtaatgacattaagggtttgttgtccttatacgaagccgctcatgttagaactcacggtgataaaattttggaggaagctttgtttttttaccactactcatttgacccgtgaaatcccaaacgttggttctactttggctaagcaggtcaaatatgctttagagcaaccattgcacaagggtatcccaagatacgaagcctggagatatatttcaatttacgaagaagacgaatctaacaacaagttgttattacgtttggcaaagttggattaccatttgttgcaaatgtcctacaaaagagaattgtccgagatcattagatggggtaaggaattagacattatttctaaggttccttatgctagagatagaatcgtcgaatgttatttctgggctgttgccacatactacgagccacaatactccttggctagaatgacattgaccaaagctactgtttttgctggtatgatcgatgatacctatgacgcttacggtactttagacgaattgaagatattcactgaagcagtcgaacgttgggattcttccggtattgaccaattgtcagattacatgaaagcagcttacaccttagtcttaaattttaacaaggaagttggtgaggatttagccaagaaacaaagaacttacgccttcgacaagtacatcgaagaatggaagcaatatgctagaacctctttcacccaatctaagtggttcttgaccaatgaattgccatccttttctgattatttgtctaacggtatggttacttcaacatactacttattgtctgccgctacattcttgggtatggacggtgcttctgaagacgtcataaattggatgtctactaaccctaaattgttcgtcgctttgacaacccacgctagattggccaacgacgttggttctcataaatttgaaaaggaaagaggctcctccactgcaatagaatgctatatgaaggattaccacgtttctgaggaggaagctatggaaaaattcgaagaaatgtgtgacgatgcctggaaggtcatgaacgaagaatgcttgcgttccactaccatcccaagagagatttgaaggttattttgaacttggccagaacctgtgaagtcgtttacaagcatcgtggtgatggtttcactgatcagagaagaattgaagctcacatcaacgctatgttaatggactcagtttccatctaa (SEQ ID NO: 110)
WenAngSQTS826	*Wendlandia*	A0A068VE40	47%	ASTEIAVPLNNQHESVRQLADFPENIWADRVASFTLDKQGHDMCAKEIE	atggcctcaacagaaatcgcagttcctttgaataaccagcacgagtccgtccgtcaattagctgacttccagaaaacatttgggctgatagagttgcttctttta

	angustifolia			MLKEEVMSMLLEEKP MMEKFNLIDNIERLGIS YHFGDKIEDQLQEYYD ACTNFEKHAECDLSIAA LQFRLFRQHGFNISCGIF DGFLDANGKFKESLCN DIKGLLSLYEAAHVRT HGDKILEEALFFTTTHL TREIPNVGSTLAKQVKY ALEQPLHKGIPRYEAW RYISIYEEDESNNKLLL RLAKLDYHLLQMSYKR ELSEIIRWGKELDIISKV PYARDRIVECYFWAVA TYYEPQYSLARMTLTK ATVFAGMIDDTYDAYG TLDELKIFTEAVERWDS SGIDQLSDYMKAAYTL VLNFNKEVGEDLAKKQ RTYAFDKYIEEWKQYA RTSFTQSKWFLTNELPS FADYLSNGMVTSTYYL LSAAALLDMDSALEDV INWMSTNPKFFVALTT HARLTNDVGSHKFEKE RGSGTAIECYMKDYHV SEEEAMKKFEEMCDDA WKVMNEECLRSTTIPR EILKVILNLARTCEVVY KHRGDGFTDQRRIEAHI NAMLMDSVSI (SEQ ID NO: 45)	ccttggataagcaaggtcatgacatgtgctaaa gaaatagaaatgttaaaggaagaagtcatgtctat gttgttggaggaaaagccaatgatggaaaaattc aacttgatcgataatattgaaagattaggcatctcc taccacttcggtgacaagattgaagatcaattaca agaatattacgacgcctgcactaactttgagaagc atgctgaatgtgatttgtcaatagctgccttgcaatt cagattgtttagacaacacggtttcaatatttcttgt ggtatctttgacggtttcttggatgcaaacggtaaa ttcaaggaatctttatgtaatgacattaagggtttgtt gtccttatacgaagccgctcatgttagaactcacg gtgataaaattttggaggaagctttgtttttttaccact actcatttgacccgtgaaatcccaaacgttggttct actttggctaagcaggtcaaatatgctttagagca accattgcacaagggtatcccaagatacgaagcc tggagatatatttcaatttacgaagaagacgaatct aacaacaagttgttattacgtttggcaaagttggat taccatttgttgcaaatgtcctacaaaagagaattg tccgagatcattagatggggtaaggaattagacat tatttctaaggttccttatgctagagatagaatcgtc gaatgttatttctgggctgttgccacatactacgag ccacaatactccttggctagaatgacattgaccaa agctactgtttttgctggtatgatcgatgatacctat gacgcttacggtactttagacgaattgaagatattc actgaagcagtcgaacgttgggattcttccggtat tgaccaattgtcagattacatgaaagcagcttaca ccttagtcttaaattttaacaaggaagttggtgagg atttagccaagaaacaaagaacttacgccttcgac aagtacatcgaagaatggaagcaatatgctagaa cctctttcacccaatctaagtggttcttgaccaatg aattgccatcctttgcagattatttgtctaacggtat ggttacttcaacatactacttattgtctgctgctgcc ttgttggacatggactccgctttagaagatgtcata aattggatgtctaccaaccctaaattcttcgtcgctt tgacaactcacgctagattgaccaacgacgttggt tctcataaatttgaaaaggaaagaggttccggtac tgcaatagaatgctatatgaaggattaccacgtttc tgaggaggaagctatgaagaaattcgaagaaat gtgtgacgatgcctggaaggtcatgaacgaaga atgcttgcgttctactacaatcccaagagagatttt gaaggttattttgaacttggccagaacctgtgaag tcgtttacaagcatcgtggtgatggcttcactgacc agagaagaattgaagctcacatcaacgccatgtt aatggactccgtttccatctaa (SEQ ID NO: 111)
WenAngSQTS829	*Wendlandia angustifolia*	A0A068UHT0	74%	ASAQASLPSNNRQETV RPLADFPENIWADRIAP FTLDKQEYEMCQREIE MLKAEVASMLLATGKT MMQRFDFIDKIERLGVS HHFDIEIENQLQEFFNV	atggccagtgcgcaagcatcattaccttccaataa cagacaggaaacagtccgtcccctagctgacttc ccagagaacatctgggctgataggattgctccatt taccctggataagcaagaatacgaaatgtgtcaa agagaaatagagatgttgaaagctgaagtggcct ctatgttgcttgccactggaaagactatgatgcaa

				YTNLGEYSAYDLSSAA LQFRLFRQHGFNISCGIF DQFIDAKGKFKESLCN DIRGLLSLYEAAHVRTH GDKILEEALAFTTTHMT SGGPHLDSSLAKQVKY ALEQPLHKGILRYEAW RYISIYEEDESNNKLLL RLAKLDYHLLQMSYKQ ELCEITRWGKGLESVSN FPYARDRFVECYFWAV GTLYEPQYSLARMTFA KVAALITMIDDIYDAYG TLDELQILTDSAERWD GSGVDQLSDYIRASYN TLLKFNKEVGEDLAKK QRTYAFDKYIEDWKQY MRTSFTQSKWFLTNEL PSFADYISNGAITIGAYL IASAGFLDMDSALEDVI NWMSTNPKLMVAYST HSRLINDYGGHKFDKE RGTGTAIECYMKDHNIS EEEAAKKFREMIENTW KVMNEECLRPIPIPRDT LKMLLNIARVGETVYK HRIDGFTQPHAIEEHIRA MLVDFMSI (SEQ ID NO: 46)	cgattcgacttcattgataagatcgaaagattggg cgtatcgcaccattttgacattgaaatcgaaaatca actccaagagtttttcaacgtttataccaacttgggt gaatacagcgcgtatgatctgtcatctgctgcattg cagttcagattatttagacaacacggtttcaatattt cctgcggtattttcgaccaatttatcgacgctaaag gtaagttcaaggaatctttatgtaacgatatccggg gtttgttgtctctctacgaagctgctcatgttagaac gcacggtgataaaattttggaagaagcattggctt ttactactacccatatgacttccggtggtccacacc tagactctagcttggctaagcaagtcaagtacgc gcttgagcaaccattacacaaggggattttgagat acgaagcttggcgttatatatccatctacgaagaa gacgaatctaataacaaacttctgttaagattggct aaactcgattatcatttgcttcaaatgtcctacaagc aggaattatgtgaaatcacgagatggggcaagg gtttagagtcagtttctaatttcccttacgctagaga tcgttttgttgaatgttatttctgggccgtaggaaca ttgtacgaaccgcaatacagtctagccagaatga cctttgctaaagttgctgccttgattactatgattga cgatatctacgatgcctatggtaccttggacgagtt acaaatattgaccgattctgctgaaagatgggatg gttcgggagtcgaccaattgtctgactatatacgc gctagttataacactttgttgaagttcaacaaggaa gtcggtgaggatttagccaaaaagcaaagaacgt acgcatttgacaaatacatcgaagattggaagca atacatgagaacttctttcacccagtccaagtggtt cctgaccaacgaactcccttccttcgctgactaca tttccaatggggctattacaattggtgcttacttgat cgccagcgcgggtttttggatatggattctgccct agaagacgttattaactggatgtctactaacccaa aattgatggtggcttattcaactcacagcagactta tcaatgattatggtggtcacaagttcgacaaggaa agagggacgggtacagctattgaatgctacatga aggatcataacatctctgaggaagaagctgcaaa gaagttcagagaaatgatcgagaacacttggaag gttatgaatgaagaatgtctacggccaattccaatt ccaagagatactctcaagatgctattgaacattgct agggtcggtgaaactgtttacaaacacagaatcg acggttttacccaaccacatgcaatcgaggaaca catcagggccatgttggtcgacttcatgtcaatttta a (SEQ ID NO: 112)
WenAng SQTS843	*Wendlandia angustifolia*	A0A068VE40	45%	ASTEIAVPLNNQHESVR QLADFPENIWADRVAS FTLDKQGHDMCAKEIE MLKEEVMSMLLEEKP MMEKFNLIDNIERLGIS YHFGDKIEDQLQEYYD ACTNFEKHAECDLSIAA LQFRLFRQHGFNISCGIF DGFLDANGKFKESLCN DIKGLLSLYEAAHVRT	atggcctcaacagaaatcgcagttcctttgaataa ccagcacgagtccgtccgtcaattagctgacttcc cagaaaacatttgggctgatagagttgcttctttta ccttggataagcaaggtcatgacatgtgtgctaaa gaaatagaaatgttaaaggaagaagtcatgtctat gttgttggaggaaaagccaatgatggaaaaattc aacttgatcgataatattgaaagattaggcatctcc taccacttcggtgacaagattgaagatcaattaca agaatattacgacgcctgcactaactttgagaagc

				HGDKILEEALFFTTTHL TREIPNVGSTLAKQVKH ALEQPLHRGIPRYEAYC FISIYEEDESNNKLLLRL AKLDYHLLQMSYKREL SEIIRWGKELDIISKVPY ARDRIVECYFWAVATY YEPQYSLARMTLTKAT VFAGMIDDTYDAYGTL DELKIFTEAVERWDSSG IDQLSDYMKAAYTLVL NFNKEVGEDLAKKQRT YAFDKYIEEWKQYART SFTQSKWFLTNELPSFS DYLSNGMVTSTYYLLS AAAFLDMDSASEDVIN WMSTNPKLFVALTTHA RLANDVGSHKFEKERG SGTAIECYMKDYNVSE EEALKKFEEMCEDTWK VMNEECLRSTTIPREIL KVILNLARTCEVVYKH RGDGFTDQRRIEAHINA MLMDSVSI (SEQ ID NO: 47)	atgctgaatgtgatttgtcaatagctgccttgcaatt cagattgtttagacaacacggtttcaatatttcttgt ggtatctttgacggtttcttggatgcaaacggtaaa ttcaaggaatctttatgtaatgacattaagggtttgtt gtccttatacgaagccgctcatgttagaactcacg gtgataaaattttggaggaagctttgtttttaccact actcatttgacccgtgaaatcccaaacgttggttct actttggctaagcaggtcaaacacgctttagagca accattgcacagaggtatcccaagatatgaagcc tactgcttcatttcaatttatgaagaagacgaatcta acaacaagttgttattacgtttggcaaagttggatt accatttgttgcaaatgtcctacaaaagagaattgt ccgagatcattagatggggtaaggaattagacatt atttctaaggttccttatgctagagatagaatcgtc gaatgttacttttgggctgttgccacatattacgag ccacaatactccttggctagaatgacattgaccaa agctactgttttcgctggtatgatcgatgatacctat gacgcttacggtactttagacgaattgaagatattc actgaagcagtcgaacgttgggattcttccggtat tgaccaattgtcagattacatgaaagcagcttaca ccttagtcttaaattttaacaaggaagttggtgagg atttagccaagaaacaaagaacttacgccttcgac aagtacatcgaagaatggaagcaatatgctagaa cctctttcacccaatctaagtggttcttgaccaatg aattgccatccttttctgattatttgtctaacggtatg gttacttcaacatactacttattgtctgccgctgcctt cttggacatggactccgcttctgaagatgtcataa attggatgtctaccaaccctaaattgttcgtcgcttt gacaactcatgctagattggccaacgacgttggtt ctcacaagtttgaaaaggaaagaggttcaggtac cgcaatagaatgttatatgaaagattacaacgtttc tgaggaggaagctttgaagaaattcgaagaaatg tgtgaagatacttggaaggtcatgaacgaagaat gcttgcgttccactacaatcccaagagagattttg aaggttattttgaacttggccagaacctgtgaagtc gtttacaagcatcgtggtgacggcttcactgatca gagaagaattgaagctcacatcaatgctatgttaa tggactccgtttccatctaa (SEQ ID NO: 113)
WenAng SQTS848	*Wendlandia angustifolia*	A0A068UHT0	84%	ASAQASLPSNNRQETV RPLADFPENIWADRIAP FTLDKQEYEMCQREIE MLKAEVASMLLATGKT MMQRFDFIDKIERLGVS HHFDIEIENQLQEFFNV YTNLGEYSAYDLSSAA LQFRLFRQHGFNISCGIF DQFIDAKGKFKESLCN DIRGLLSLYEAAHVRTH GDKILEEALAFTTTHMT SGGPHLDSSLAKQVKY ALEQPLHKGILRYEAW	atggcctcagcacaagcttccttaccttctaataac agacaggaaacagtccgtccattggctgacttcc cagagaacatctgggctgatagaattgccccattt accttggataagcaagaatacgaaatgtgtcaaa gagaaatagagatgttaaaagctgaagttgcttct atgttgttggcaactggtaagactatgatgcaaag attcgacttcattgataagatcgaaagattggggg tctcccaccattttgacattgaaatcgaaaatcaatt gcaagagtttttcaacgtttataccaacttaggtga atactctgcctatgatttgtcatctgctgccttgcag ttccgtttatttagacaacacggtttcaatatttcctg cggtattttcgaccaatttatcgacgctaaaggtaa

				RYISIYEEDESNNKLLL RLAKLDYHLLQMSYKQ ELCEITRWGKGLESVSN FPYARDRFVECYFWAV GTLYEPQYSLARMTFA KVAALITMIDDIYDAYG TLDELQILTDSAERWD GSGVDQLSDYIRASYN TLLKFNKEVGEDLAKK QRTYAFDKYIEDWKQY MRTNFSQSRWFFTKEL PSFADYINNGAITIGAY LVASAAFLYMDSAKED VINWMSTNPKLVVAYS THSRLINDFGGHKFDKE RGSGTALECYMKDYN VSEEEAANKFREMMED AWKVMNEDCLRPTSIP RDVSKVLLNVARAGEI VYKHRIDGFTEPHIIKD HIRATLVDFMAIN (SEQ ID NO: 48)	gttcaaggaatctttatgtaacgatatcagaggttt gttgtctttgtacgaagctgctcatgttagaactca cggtgataaaattttggaagaagctttagctttcac cactactcacatgacctccggtggtccacatttag attcttcattggccaagcaagttaaatacgcattgg aacagccattgcataagggtatattgagatatgaa gcttggagatacatatctatctacgaagaggacg aatccaacaataagttattattgcgtttggctaagtt ggactatcacttgttacaaatgtcatacaagcaag agttgtgtgaaattacaagatggggtaaaggtttg gaatctgtctccaactttccttatgcccgtgacaga ttcgttgaatgttacttttgggctgtcggtactttgta cgaaccacaatactcattggctagaatgaccttcg ctaaggttgctgctttaattactatgatcgatgatatt tatgatgcctacggtaccttggacgaattgcaaat attaactgactctgccgaaagatgggatggttccg gtgtcgatcagttgtctgactatattagagcttccta taatacattattgaaatttaataaggaggttggtga agatttggcaaaaaagcaacgtacctacgctttcg acaagtacatcgaagattggaaacaatacatgag aaccaacttctctcaatcaagatggtttttcactaag gagttgccatctttcgctgattacattaacaacggt gccatcacaatcggtgcatatttggttgcctctgct gctttcttatatatggactccgcaaaagaagatgtt atcaactggatgtccacaaaccctaagttggtcgt tgcttactccactcactctcgtttaattaatgactttg gtggtcacaagttcgacaaggagagaggttccg gtactgctttggaatgctacatgaaggactacaat gtctctgaagaagaagccgcaaacaagtttagag aaatgatggaggacgcttggaaggttatgaatga agactgtttaagaccaacttccatccctagagatgt ctccaaggttttgttaaacgtcgccagagctggtg aaattgtttacaagcatagaatcgatggttttaccg aaccacatatcattaaagatcacataagagccac cttggttgatttcatggctattaattaa (SEQ ID NO: 114)
WenAng SQTS849	*Wendlandia angustifolia*	A0A068UHT0	75%	ASAQASLPSNNRQETV RPLADFPENIWADRIAP FTLDKQEYEMCQREIE MLKAEVASMLLATGKT MMQRFDFIDKIERLGVS HHFDIEIENQLQEFFNV YTNLGEYSAYDLSSAA LQFRLFRQHGFNISCGIF DQFIDAKGKFKESLCN DIRGLLSLYEAAHVRTH GDKILEEALAFTTTHMT SGGPHLDSSLAKQVKY ALEQPLHKGILRYEAW RYISIYEEDESNNKLLL RLAKLDYHLLQMSYKQ ELCEITRWGKGLESVSN	atggccagtgcgcaagcatcattaccttccaataa cagacaggaaacagtccgtcccctagctgacttc ccagagaacatctgggctgataggattgctccatt taccctggataagcaagaatacgaaatgtgtcaa agagaaatagagatgttgaaagctgaagtggcct ctatgttgcttgccactggaaagactatgatgcaa cgattcgacttcattgataagatcgaaagattggg cgtatcgcaccattttgacattgaaatcgaaaatca actccaagagttttttcaacgtttataccaacttgggt gaatacagcgcgtatgatctgtcatctgctgcattg cagttcagattatttagacaacacggtttcaatattt cctgcggtattttcgaccaatttatcgacgctaaag gtaagttcaaggaatctttatgtaacgatatccggg gtttgttgtctctctacgaagctgctcatgttagaac gcacggtgataaaattttggaagaagcattggctt

				FPYARDRFVECYFWAV GTLYEPQYSLARMTFA KVAALITMIDDIYDAYG TLDELQILTDSAERWD GSGVDQLSDYIRASYN TLLKFNKEVGEDLAKK QRTYAFDKYIEDWKQY MRTSFTQSKWFLTNEL PSFADYISNGAITIGAYL IASAGFLDMDSALEDVI NWMSTNPKLMVAYST HSRLINDYGGHKFDKE RGSVTALDCYMKDYSV SEEEAAKKFREMIENT WKVMNEECLRPIPIPRD TLKMLLNIARVGETVY KHRIDGFTEPHIIKDHIR AMLVDFMAIN (SEQ ID NO: 49)	ttactactacccatatgacttccggtggtccacacc tagactctagcttggctaagcaagtcaagtacgc gcttgagcaaccattacacaaggggattttgagat acgaagcttggcgttatatatccatctacgaagaa gacgaatctaataacaaacttctgttaagattggct aaactcgattatcatttgcttcaaatgtcctacaagc aggaattatgtgaaatcacgagatggggcaagg gtttagagtcagtttctaatttcccttacgctagaga tcgttttgttgaatgttatttctgggccgtaggaaca ttgtacgaaccgcaatacagtctagccagaatga cctttgctaaagttgctgccttgattactatgattga cgatatctacgatgcctatggtaccttggacgagtt acaaatattgaccgattctgctgaaagatgggatg gttcgggagtcgaccaattgtctgactatatacgc gctagttataacactttgttgaagttcaacaaggaa gtcggtgaggatttagccaaaaagcaaagaacgt acgcatttgacaaatacatcgaagattggaagca atacatgagaacttctttcacccagtccaagtggtt cctgaccaacgaactcccttccttcgctgactaca tttccaatggggctattacaattggtgcttacttgat cgccagcgcgggtttttggatatggattctgccct agaagacgttattaactggatgtctactaacccaa aattgatggtggcttattcaactcacagcagactta tcaatgattatggtggtcacaagttcgacaaggaa agagggagcgttacagctttggattgctacatgaa ggattacagtgtctctgaggaagaagctgcaaag aagttcagagaaatgatcgaaaacacctggaag gttatgaatgaagaatgtctgcggccaattccaatt ccaagagatactctaaagatgctattgaacattgct agggtaggtgaaactgtttacaaacatagaatcg acggttttactgaaccacatataattaaggaccac atcagggcaatgttggtcgacttcatggctattaac taa (SEQ ID NO: 115)
WenAng SQTS864	*Wendlandia angustifolia*	A0A068UHT0	81%	ASAQASLPSNNRQETV RPLADFPENIWADRIAP FTLDKQEYEMCQREIE MLKAEVASMLLATGKT MMQRFDFIDKIERLGVS HHFDIEIENQLQEFFNV YTNLGEYSAYDLSSAA LQFRLFRQHGFNISCGIF DQFIDAKGKFKESLCN DIRGLLSLYEAAHVRTH GDKILEEALAFTTTHMT SGGPHLDSSLAKQVKY ALEQPLHKGILRYEAW RYISIYEEDESNNKLLL RLAKLDYHLLQMSYKQ ELCEITRWGKGLESVSN FPYARDRFVECYFWAV GTLYEPQYSLARMTFA KVAALITMIDDIYDAYG	atggcctcagcacaagcttccttaccttctaataac agacaggaaacagtccgtccattggctgacttcc cagagaacatctgggctgatagaattgccccattt accttggataagcaagaatacgaaatgtgtcaaa gagaaatagagatgttaaaagctgaagttgcttct atgttgttggcaactggtaagactatgatgcaaag attcgacttcattgataagatcgaaagattggggg tctcccaccattttgacattgaaatcgaaaatcaatt gcaagagtttttcaacgtttataccaacttaggtga atactctgcctatgatttgtcatctgctgccttgcag ttccgtttatttagacaacacggtttcaatatttcctg cggtattttcgaccaatttatcgacgctaaaggtaa gttcaaggaatctttatgtaacgatatcagaggttt gttgtctttgtacgaagctgctcatgttagaactca cggtgataaaattttggaagaagctttagctttcac cactactcacatgacctccggtggtccacatttag attcttcattggccaagcaagttaaatacgcattgg aacagccattgcataagggtatattgagatatgaa

				TLDELQILTDSAERWD GSGVDQLSDYIRASYN TLLKFNKEVGEDLAKK QRTYAFDKYIEDWKQY MRTNFSQSRWFFTKEL PSFADYINNGAITIGAY LVASAAFLYMDSAKED VINWMSTNPKLVVAYS THSRLINDFGGHKFDKE RGSVTALDCYMKDYSV SEEEAAKKFREMCEDN WKVMNEECLRPTTIPR DGLKMLLNIARVGETV YKHRIDGFTQPHAIEEH IRAMLVDFMSI (SEQ ID NO: 50)	gcttggagatacatatctatctacgaagaggacg aatccaacaataagttattattgcgtttggctaagtt ggactatcacttgttacaaatgtcatacaagcaag agttgtgtgaaattacaagatggggtaaaggtttg gaatctgtctccaactttccttatgcccgtgacaga ttcgttgaatgttacttttgggctgtcggtactttgta cgaaccacaatactcattggctagaatgaccttcg ctaaggttgctgctttaattactatgatcgatgatatt tatgatgcctacggtaccttggacgaattgcaaat attaactgactctgccgaaagatgggatggttccg gtgtcgatcagttgtctgactatattagagcttccta taatacattattgaaatttaataaggaggttggtga agatttggcaaaaaagcaacgtacctacgctttcg acaagtacatcgaagattggaaacaatacatgag aaccaacttctctcaatcaagatggtttttcactaag gagttgccatctttcgctgattacattaacaacggt gccatcacaatcggtgcatatttggttgcctctgct gctttcttatatatggactccgcaaaagaagatgtt atcaactggatgtccacaaaccctaagttggtcgt tgcttactccactcactctcgtttaattaatgactttg gtggtcacaagttcgacaaggagagaggttccgt tactgctttggactgctacatgaaggactactctgt ctccgaagaagaagccgcaaagaagtttagaga aatgtgtgaagacaattggaaggtcatgaatgaa gagtgtttaagaccaactaccatccctagagatgg gttgaagatgttgttaaacatagccagagttggtg aaactgtctacaagcatagaattgatggttttaccc aaccacatgctatcgaagaacacatcagagctat gttggttgatttcatgtctatttaa (SEQ ID NO: 116)
WenAng SQTS925	*Wendlandia angustifolia*	A0A068UHT0	80%	ASAQASLPSNNRQETV RPLADFPENIWADRIAP FTLDKQEYEMCQREIE MLKAEVASMLLATGKT MMQRFDFIDKIERLGVS HHFDIEIENQLQEFFNV YTNLGEYSAYDLSSAA LQFRLFRQHGFNISCGIF DQFIDAKGKFKESLCN DIRGLLSLYEAAHVRTH GDKILEEALAFTTTHMT SGGPHLDSSLAKQVKY ALEQPLHKGILRYEAW RYISIYEEDESNNKLLL RLAKLDYHLLQMSYKQ ELCEITRWGKGLESVSN FPYARDRFVECYFWAV GTLYEPQYSLARMTFA KVAALITMIDDIYDAYG TLDELQILTDSAERWD GSGVDQLSDYIRASYN TLLKFNKEVGEDLAKK QRTYAFDKYIEDWKQY	atggcctcagcacaagcttccttaccttctaataac agacaggaaacagtccgtccattggctgacttcc cagagaacatctgggctgatagaattgccccattt accttggataagcaagaatacgaaatgtgtcaaa gagaaatagagatgttaaaagctgaagttgcttct atgttgttggcaactggtaagactatgatgcaaag attcgacttcattgataagatcgaaagattggggg tctcccaccattttgacattgaaatcgaaaatcaatt gcaagagtttttcaacgtttataccaacttaggtga atactctgcctatgatttgtcatctgctgccttgcag ttccgtttatttagacaacacggtttcaatatttcctg cggtattttcgaccaatttatcgacgctaaaggtaa gttcaaggaatctttatgtaacgatatcagaggttt gttgtctttgtacgaagctgctcatgttagaactca cggtgataaaattttggaagaagctttagctttcac cactactcacatgacctccggtggtccacatttag attcttcattggccaagcaagttaaatacgcattgg aacagccattgcataagggtatattgagatatgaa gcttggagatacatatctatctacgaagaggacg aatccaacaataagttattattgcgtttggctaagtt ggactatcacttgttacaaatgtcatacaagcaag

				MRTNFSQSRWFFTKEL PSFADYINNGAITIGAY LVASAAFLYMDSAKED VINWMSTNPKLVVAYS THSRLINDFGGHKFDKE RGSVTALDCYMKDYSV SEEEAAKKFREMIENT WKVMNEECLRPIPIPRD TLKMLLNIARVGETVY KHRIDGFTEPHIIKDHIR AMLVDFMAIN (SEQ ID NO: 51)	agttgtgtgaaattacaagatggggtaaaggtttg gaatctgtctccaactttccttatgcccgtgacaga ttcgttgaatgttacttttgggctgtcggtactttgta cgaaccacaatactcattggctagaatgaccttcg ctaaggttgctgctttaattactatgatcgatgatatt tatgatgcctacggtaccttggacgaattgcaaat attaactgactctgccgaaagatgggatggttccg gtgtcgatcagttgtctgactatattagagcttccta taatacattattgaaatttaataaggaggttggtga agatttggcaaaaaagcaacgtacctacgctttcg acaagtacatcgaagattggaaacaatacatgag aaccaacttctctcaatcaagatggtttttcactaag gagttgccatctttcgctgattacattaacaacggt gccatcacaatcggtgcatatttggttgcctctgct gctttcttatatatggactccgcaaaagaagatgtt atcaactggatgtccacaaaccctaagttggtcgt tgcttactccactcactctcgtttaattaatgactttg gtggtcacaagttcgacaaggagagaggttccgt tactgctttggactgctacatgaaggactactctgt ctccgaagaagaagccgcaaagaagtttagaga aatgatcgaaaacacctggaaggtcatgaatgaa gagtgtttaagaccaattccaatccctagagacac attgaagatgttgttaaacatagccagagttggtg aaactgtctacaagcatagaattgatggttttactg aaccacatatcatcaaagatcacatcagagctatg ttggttgatttcatggctattaattaa (SEQ ID NO: 117)
WenAng SQTS960	*Wendlandia angustifolia*	A0A068VI46	81%	YEREIEMLKAEVESML LATGKTMMQRFDFIDK IERLGVSHHFDIEIENQL QEFFNVYTNFGEYSAY DLSSAALQFKQWCDHN RSLSCSITRGLLSLYEA AHVRTHGDKILEEALH LTSGESHLDSTLAKQV KCALEQPLHKGIPRYEA WRYISIYEEDESHNKLL LRLAKLDYHFLQISYRQ DLCEIIRWDSSGVDQLS DYIRAVGEELAKKQRT YAFGTFLGMDGASEDV INWMSTIPKLMFACSTH ARLINDFGGHKFDKER GTGTALECYMKDYNVS EEEAANKFREMMEDA WKVMNEECLRPTTIPR EILKMLLNIVRVGETTN KHRIDGFTQPHAIEEHIR AMLVDFMSV (SEQ ID NO: 52)	atgtatgagagagaaatcgaaatgttaaaggctg aagtcgaatctatgttgttggccaccggtaaaaca atgatgcagcgtttcgattttatagacaagattgaa agattgggcgtttcccaccatttcgatattgaaatc gagaaccaattacaagaatttttcaatgtttacacta acttcggtgaatactcagcttacgacttgtcttccg cagccttgcaatttaagcaatggtgtgaccacaat agatcattatcttgctctattactagaggtttgttatc cttgtatgaggctgctcatgtcagaacccacggtg ataagatcttggaagaagctttacacttgacttctg gtgaatcccatttggactccaccttggctaaacaa gttaaatgtgcattagaacaaccattgcacaaggg tatacctcgttacgaagcctggagatatatttctatc tacgaagaggatgaatcacataacaagttgttgtt gagattagctaaattggattatcacttcttacagatt tcttacagacaagatttgtgtgaaatcattcgttgg gactcatctggtgtcgaccaattatctgattacatc agagcagttggtgaggaattggctaagaagcaa agaacatacgctttcggtacttttttaggtatggatg gtgcctctgaagatgttattaactggatgtccacta tcccaaagttgatgttcgcttgctctacacatgcca gattgattaatgactttggtggtcataaattcgataa ggaaagaggtactggtaccgctttagagtgttata tgaaagactataacgtctccgaagaagaagccg

					ccaacaagtttagagaaatgatggaggacgcttg gaaagttatgaatgaagaatgtttgcgtccaacca ctattccaagagaaatattaaagatgttgttgaaca tcgtccgtgttggtgaaactactaataagcacaga atcgatggtttcacacagcctcacgctattgagga acacattagagctatgttggttgactttatgtccgtc taa (SEQ ID NO: 118)

Table 11. Non-limiting examples of sequence fragment(s) derived from rare plants.

Chimera Name	Ancient DNA Source	Fragments	SEQ ID NO
HibWilSQTS117	Hibiscadelphus wilderianus	LKDEEGNFKASLTSDVPGLLELYEASYLRVHGEDI LDEAISFA	119
		NKALLQFAKIDFNMLQLLHRKELSEICRWWKDLD FTRKLP	120
		DRVVEGYFWIMGVYFEPQYSLGRKMLTKVIAMA SIVDDTYDSFATYDELIPYTDAIER	121
		YMQISYKALLDVYEEMEQLLADKGRQYRVEY	122
		WTHLNYKPTFEEFRDNALPTSGYAMLAIT	123
		TFEWAASDPKIIKASTIICRFMDDIAE	124
		EDDCSAIECYMEQYKVTAQEAYDEFNKHIESSWK DVNEEFLK	125
HibWilSQTS118	Hibiscadelphus wilderianus	EAFNKLKDEEGNFKASLTSDVRGLLELYQASYMR IHGEDILDEAISFTTAQLTLALPTLDPP	126
		NKALLQFAKIDFNMLQLLHRKELSEICRWWKDLD FTRKLP	127
		DRVVEGYFWIMGVYFEPQYSLGRKMLTKVIAMA SIVDDTYDSFATYDELIPYTDAIER	128
		YMQISYKALLDVYEEMEQLLADKGRQYRVEY	129
		WTHLNYKPTFEEFRDNALPTSGYAMLAIT	130
		TFEWAASDPKIIKASTIICRFMDDIAE	131
		SAIECYMKQYGATAQEAYDEFNKHIESSWK	132
HibWilSQTS120	Hibiscadelphus wilderianus	LKDEEGNFKASLTSDVPGLLELYEASYLRVHGEDI LDEAISFA	133
		NKALLQFAKIDFNMLQLLHRKELSEICRWWKDLD FTRKLP	134
		DRVVEGYFWIMGVYFEPQYSLGRKMLTKVIAMA SIVDDTYDSFATYDELIPYTDAIER	135
		YMQISYKALLDVYEEMEQLLADKGRQYRVEY	136
		WTHLNYKPTFEEFRDNALPTSGYAMLAIT	137
		TFEWAASDPKIIKASTIICRFMDDIAE	138
		SAIECYMKQYGATAQEAYDEFNKHIESSWK	139
HibWilSQTS121	Hibiscadelphus wilderianus	EAFNKLKDEEGNFKASLTSDVRGLLELYQASYMR IHGEDILDEAISFTTAQLTLALPTLDPP	140
		LLEFAKIDFNLLQLLHRKELSEICRWWKD	141
		DRVVEGYFWIMGVYFEPQYSLGRKMLTKVIAMA SIVDDTYDSFATYDELIPYTDAIER	142
		YMQISYKALLDVYEEMEQLLADKGRQYRVEY	143
		WTHLNYKPTFEEFRDNALPTSGYAMLAIT	144
		TFEWAASDPKIIKASTIICRFMDDIAE	145
		EDDCSAIECYMEQYKVTAQEAYDEFNKHIESSWK DVNEEFLK	146

HibWilSQTS123	Hibiscadelphus wilderianus	LKDEEGNFKASLTSDVPGLLELYEASYLRVHGEDILDEAISFA	147
		LLEFAKIDFNLLQLLHRKELSEICRWWKD	148
		DRVVEGYFWIMGVYFEPQYSLGRKMLTKVIAMASIVDDTYDSFATYDELIPYTDAIER	149
		YMQISYKALLDVYEEMEQLLADKGRQYRVEY	150
		WTHLNYKPTFEEFRDNALPTSGYAMLAIT	151
		TFEWAASDPKIIKASTIICRFMDDIAE	152
		EDDCSAIECYMEQYKVTAQEAYDEFNKHIESSWKDVNEEFLK	153
HibWilSQTS124	Hibiscadelphus wilderianus	EAFNKLKDEEGNFKASLTSDVRGLLELYQASYMRIHGEDILDEAISFTTAQLTLALPTLDPP	154
		LLEFAKIDFNLLQLLHRKELSEICRWWKD	155
		DRVVEGYFWIMGVYFEPQYSLGRKMLTKVIAMASIVDDTYDSFATYDELIPYTDAIER	156
		YMQISYKALLDVYEEMEQLLADKGRQYRVEY	157
		WTHLNYKPTFEEFRDNALPTSGYAMLAIT	158
		TFEWAASDPKIIKASTIICRFMDDIAE	159
		SAIECYMKQYGATAQEAYDEFNKHIESSWK	160
HibWilSQTS126	Hibiscadelphus wilderianus	LKDEEGNFKASLTSDVPGLLELYEASYLRVHGEDILDEAISFA	161
		LLEFAKIDFNLLQLLHRKELSEICRWWKD	162
		DRVVEGYFWIMGVYFEPQYSLGRKMLTKVIAMASIVDDTYDSFATYDELIPYTDAIER	163
		YMQISYKALLDVYEEMEQLLADKGRQYRVEY	164
		WTHLNYKPTFEEFRDNALPTSGYAMLAIT	165
		TFEWAASDPKIIKASTIICRFMDDIAE	166
		SAIECYMKQYGATAQEAYDEFNKHIESSWK	167
HibWilSQTS19	Hibiscadelphus wilderianus	FEQERGHCASAVECYMREHGVSEEEACSELKKQVDNAWKDINHEMIFSETSKAVPMSVLTRVLNLTR	168
HibWilSQTS34	Hibiscadelphus wilderianus	GYHVDGEEAFNMLKDEEGNFKASLTSDVPGLLELYQASYMRIHGEDILDEAISFTTAQLTLALPTLDPPLS	169
HibWilSQTS52	Hibiscadelphus wilderianus	FEQERGHCASAVECYMREHGVSEEEACSELKKQVDNAWKDINHEMIFSETSKAVPMSVLTRVLNLTR	170
HibWilSQTS54	Hibiscadelphus wilderianus	GYHVDGEEAFNMLKDEEGNFKASLTSDVPGLLELYQASYMRIHGEDILDEAISFTTAQLTLALPTLDPPLSE	171
HibWilSQTS55	Hibiscadelphus wilderianus	FEQERGHCASAVECYMREHGVSEEEACSELKKQVDNAWKDINHEMIFSETSKAVPMSVLTRVLNLTR	172
HibWilSQTS63	Hibiscadelphus wilderianus	EQERGHCASAVECYMREHGVSEEEACSELKKQVDNAWKDINHEMIFSETSKAVPMSVLTRVLNLTR	173
HibWilSQTS90	Hibiscadelphus wilderianus	GYHVDGEEAFNMLKDEEGNFKASLTSDVPGLLELYQASYMRIHGEDILDEAISFTTAQLTLALPTLDPPLS	174
		FEQERGHCASAVECYMREHGVSEEEACSELKKQVDNAWKDINHEMIFSETSKAVPMSVLTRVLNLTRG	175
LeuGraSQTS335	Leucadendron grandiflorum	DAFNRFKDTKGSFKEDLIKDVNSMLCLYEATHLRVHGEDILDEALGFTTSQLKSILPKLKPLLASQVMHALKQPL	176
LeuGraSQTS345	Leucadendron grandiflorum	FNKFKNSDGNFKEDLINDVSGMLCLYEATHLRVHGEDILDEALEFTTTRLKSILPDLEPPLATQVMHA	177
LeuGraSQTS365	Leucadendron grandiflorum	IFNKFKNSDGNFKEDLINDVSGMLCLYEATHLRVHGEDILDEALEFTTTRLKSILPDLEPPL	178

LeuGraSQTS377	Leucadendron grandiflorum	DAFNRFKDTKGSFKEDLIKDVNSMLCLYEATHLR VHGEDILDEALGFTTSQLKSILPKLKPLLASQVMH ALKQPL	179
LeuGraSQTS379	Leucadendron grandiflorum	IFNKFKNSDGNFKEDLINDVSGMLCLYEATHLRV HGEDILDEALEFTTTRLKSILPDLEPPLATQVMHA	180
LeuGraSQTS385	Leucadendron grandiflorum	ETNFTNSPLLSKLQNELSVAHLEELKLEVKQLIWS TKDPLFLLKFIDSIQRLGVAYHFEEEIKESLHLVYL E	181
LeuGraSQTS393	Leucadendron grandiflorum	IFNKFKNSDGNFKEDLINDVSGMLCLYEATHLRV HGEDILDEALEFTTTRLKSILP	182
MacVolSQTS1139	Macrostylis villosa	EGLEQKIRTMLISPTDTISKKLSLIDAVQRLGVAYH FEKEIEDEIEKLSCKEYNDGNDLQTVALRFRLLRQ QGYFVSC	183
MacVolSQTS2198	Macrostylis villosa	LQRLGLAYHFENQIKEALQSI	184
		LSHLSTSLAEQVKHSLEIPLHRGMPRLEARHYISIY EEDNSS	185
		ELAKLDFNLLQALHRRELGEISRWWKDIDFATKL PFARDRLVECYFWILGVYFEPKYSITRKFMTKVIAI ASVIDDIYDVYGTLEELKLFTHAIERWETVAANEL PKYMQVCYFALLDVFKEMEDKLVNKGLLYSMPC AKEAVKGLVRAYFVEAEWFNANYMPTFEEYMEN STMSSGYPMLAVEALIGIEDATISKEAFDWAISVP KIIRSCALIARLVDDIH	186
		DAPSSVECYMQQYDVSEEEACNRIKGMVEIEW	187
		NLARMMVVLYQNGDNYTNSSGKTKDRIASLLV	188
MacVolSQTS2202	Macrostylis villosa	LQRLGLAYHFENQIKEALQSI	189
		KFKDEKGEFKDMIRNDARGLLCLYEASHLRVKGE DILEEATEFSRKHLKSLLPQLSTSLAEQVKHSLEIP LHRGMPRLEARHYISIYEENNSSRNELLLELAKLD FNLLQALHRRELGDISRWWKDIDFATKLPFARDR LVECYFWILGVYFEPKYSITRKFMTKVIAIASVIDD IYDVYGTLEELKLFTHAIERWETVAANELPKYMQ VCYFALLDVFKEMEDKLVNKGLLYSMPCAKEAV KGLVRAYFVEAEWFNANYMPTFEEYMENSTMSS GYPMLAVEALIGIEDATISKEAFDWAISVPKIIRSC ALIARLVDDIH	190
		KVEQERGDAPSSVQCYVQQ	191
		NLARMMVVLYQNGDNYTNSSGKTKDRIASLLV	192
MacVolSQTS2222	Macrostylis villosa	LQRLGLAYHFENQIKEALQSI	193
		KFKDEKGEFKDMIRNDARGLLCLYEASHLRVKGE DILEEATEFSRKHLKSLLPQLSTSLAEQVKHSLEIP LHRGMPRLEARHYISIYEENNSSRNELLLELAKLD FNLLQALHRRELGDISRWWKDIDFATKLPFARDR LVECYFWILGVYFEPKYSITRKFMTKVIAIASVIDD IYDVYGTLEELKLFTHAIERWETVAANELPKYMQ VCYFALLDVFKEMEDKLVNKGLLYSMPCAKEAV	194
		YVPTFEEYMENSTMSSGYPMLAVEALV	195
		DWAISVPKIIRSCALIA	196
		KVEQERGDAPSSVQCYMQQYDVSEEEACNRIKG MVETAWMEINGEIQDTNHL	197
		NLARMMVVLYQNGDNYTNSSGKTKDRIASLLV	198
MacVolSQTS2251	Macrostylis villosa	LQRLGLAYHFENQIKEALQSI	199
		KFKDEKGEFKDMIRNDARGLLCLYEASHLRVKGE DILEEATEFSRKHLKSLLPQLSTSLAEQVKHSLEIP LHRGMPRLEARHYISIYEENNSSRNELLLELAKLD	200

		FNLLQALHRRELGDISRWWKDIDFATKLPFARDR LVECYFWILGVYFEPKYSITRKFMTKVIAIASVIDD IYDVYGTLEELKLFTHAIERWETVAANELPKYMQ VCYFALLDVFKEMEDKLVNKGLLYSMPCAKEAV	
		YVPTFEEYMENSTMSSGYPMLAVEALV	201
		DWAISVPKIIRSCALIA	202
		DAPSSVECYMQQYDVSEEEACNRIKGMVEIEW	203
		NLARMMVVLYQNGDNYTNSSGKTKDRIASLLV	204
MacVolSQTS2274	Macrostylis villosa	KFIQNVEKDSTRRSANFHPSIWGDH	205
		DDGSVKHQQLKEEIRKMLTAETKLSQKLDLIDAIQ RLGVAYHFESEIDEIL	206
		SLARNVRGMLSLYEATHLRVHGENILDEA	207
		LEARNYMPFYQEEASHNEALLTFAKLDFNKLQKL HQKELSEITR	208
		FEQSREHVASSIECYMKQYGATEEETCNELRKQV SNAWKDINEECLCPTAVPMPLIVRILNLT	209
OrbStiSQTS1368	Orbexilum stipulatum	AEVFERFKDQHGNFKASLSSDVEGMLSLYEASFL DYEGEDILDEAKAFTSFHLRGAL	210
OrbStiSQTS1414	Orbexilum stipulatum	VKLELVDDVKRLGIGYRFEKEIVEALHRCFISSERF THRNLHQTALSFRLLRECGYDVT	211
		FNKFTNKEGKFNSKLGENIKGMIDLYEASQLGIAG EYILAEAGEFSGLVLKEKVACINN	212
		VYFEPQYSVPRRTTTKVIGLCSVIDDMYDAYGTID ELELFTNAIERLDTST	213
		RWLKCNHAPTMEEYMKVRGVSSGYPLLITISFIG MEDTTEEILTWATSEPMIIRASVIVCRLMDDI	214
ShoCusSQTS154	Shorea cuspidata	FMDEKGKFKEDVVNDVLGMLNLYEAAHLRLRGE DILDEALAFTTSHLE	215
		WWKNLDFSTKLPYARDRIVECYFWIMGAYFE	216
		SLARTFLTKVIAMTSILDDTYDNYG	217
		DYVPPIEEYMQVARISSAYPMLITNSFVGMGEVAT KEAFDWISNDPKILKASTTICRLMDD	218
		EFEQTRDHVASGVECYMKQYGVSREETVK	219
ShoCusSQTS155	Shorea cuspidata	FMDEKGKFKEDVVNDVLGMLNLYEAAHLRLRGE DILDEALAFTTSHLE	220
		WWKNLDFSTKLPYARDRIVECYFWIMGAYFE	221
		SLARTFLTKVIAMTSILDDTYDNYG	222
		YMQVALISSAYPMLITNSFVGMGEVATKEAFDWI SNNPKMLKASTII	223
		EFEQTRDHVASGVECYMKQYGVSREETVK	224
ShoCusSQTS156	Shorea cuspidata	FMDEKGKFKEDVVNDVLGMLNLYEAAHLRLRGE DILDEALAFTTSHLE	225
		WWKNLDFSTKLPYARDRIVECYFWIMGAYFE	226
		SLARTFLTKVIAMTSILDDTYDNYG	227
		DYVPPIEEYMQVARIS	228
		GYPMLITNSLVGMGEVATKEAFDLISNDPKMLKA ST	229
		EFEQTRDHVASGVECYMKQYGVSREETVK	230
ShoCusSQTS157	Shorea cuspidata	FMDEKGKFKEDVVNDVLGMLNLYEAAHLRLRGE DILDEALAFTTSHLE	231
		WWKNLDFSTKLPYARDRIVECYFWIMGAYFE	232
		SLARTFLTKVIAMTSILDDTYDNYG	233
		VPPMDEYMQVALISCGYPMLITNSFVGMGEVATK EAFDWISNDPKILKASTTICRLMDD	234

		EFEQTRDHVASGVECYMKQYGVSREETVK	235
ShoCusSQTS160	Shorea cuspidata	FMDEKGKFKEDVVNDVLGMLNLYEAAHLRLRGEDILDEALAFTTSHLE	236
		WWKNLDFATMLPYARDRIVECYFWIMGVYFEPKYSLARTFLTKVIAMTSILDDTYDNYG	237
		YMQVALISSAYPMLITNSFVGMGEVATKEAFDWISNNPKMLKASTII	238
		EFEQTRDHVASGVECYMKQYGVSREETVK	239
ShoCusSQTS161	Shorea cuspidata	FMDEKGKFKEDVVNDVLGMLNLYEAAHLRLRGEDILDEALAFTTSHLE	240
		WWKNLDFATMLPYARDRIVECYFWIMGVYFEPKYSLARTFLTKVIAMTSILDDTYDNYG	241
		DYVPPIEEYMQVARIS	242
		GYPMLITNSLVGMGEVATKEAFDLISNDPKMLKAST	243
		EFEQTRDHVASGVECYMKQYGVSREETVK	244
WenAngSQTS1007	Wendlandia angustofolia	SNNRQETVRPLADFPENIWADRIAPFT	245
		EMCQREIEMLKAEVASMLLATGKTMMQRFDFIDKIERLGVSHHFD	246
		IFDQFIDAKGKFKESLCNDIRGLLSLYEAAHVRTHGDKILEEALAFTTTHMTSGGPHLDSSLAKQVKYALEQPLHKGILRYEAWRYISIYEEDESNNKLLLRLAKLDYHLLQMSYKQEL	247
		RWGKGLESVSNFPYARDRFVECYFWAVGTLYEPQYSLARMTFAKVAA	248
		RWDGSGVDQLSDYIRASYNTLLKFNKEVGEDLAKKQRTYAFDKYIEDWKQYMRTNFSQSRWFFTKELPSFADYINNGAITIGAYLVASAAFLYMDSAKEDVINWMSTNPKLVVAYSTHSRLINDFGGHKFEKERGSSTAIECYMKDHNVSEEEAANKFREMMEDAWKVMNEECLRPTTI	249
		ETVYKHRIDGFTQPHAIEEHIRAMLVDFMSI	250
WenAngSQTS1086	Wendlandia angustofolia	SNNRQETVRPLADFPENIWADRIAPFT	251
		EMCQREIEMLKAEVASMLLATGKTMMQRFDFIDKIERLGVSHHFD	252
		IFDQFIDAKGKFKESLCNDIRGLLSLYEAAHVRTHGDKILEEALAFTTTHMTSGGPHLDSSLAKQVKYALEQPLHKGILRYEAWRYISIYEEDESNNKLLLRLAKLDYHLLQMSYKQEL	253
		RWGKGLESVSNFPYARDRFVECYFWAVGTLYEPQYSLARMTFAKVAA	254
		RWDGSGVDQLSDYIRASYNTLLKFNKEVGEDLAKKQRTYAFDKYIEDWKQYMRTNFSQSRWFFTKELPSFADYINNGAITIGAYLVASAAFLYMDSAKEDVINWMSTNPKLVVAYSTHSRLINDFGGHK	255
		KERGTGTAIECYMKDHN	256
		EMIENTWKVMNEECLRPIPIPRDTLKML	257
		ETVYKHRIDGFTQPHAIEEHIRAMLVDFMSI	258
WenAngSQTS267	Wendlandia angustofolia	LELVDNLERLGLAYHFEGQINRLLSSAYNANHEDEGNHKRNKEDLYAAALEFRIFRQHGFNV	259
WenAngSQTS302	Wendlandia angustofolia	YVSQANELKEQVKMMLDEEDMKLLDCLELVDNLERLGLAYHFEGQINRLLSSAYNANHEDEGNHKRNKEDLYAAALEFRIFRQHGFNVPQ	260
WenAngSQTS738	Wendlandia angustofolia	NNQHESVRQLADFPENIWADRV	261
		QGHDMCAKEIEMLKEEVMSMLLE	262

		STLAKQVKYALEQPLHKGIPRYEAWRYISIYEED	263
		LAKLDYHLSQMLNKQDLCEI	264
		RDRIVECYFWAVATYYEPQYSLARMT	265
		EVGEDLAKKQRTYAFDKYIE	266
		YARTSFTQSKWFLTNELPSFSDYL	267
		AAFLDMDSASEDVINWMSTNPKLFVALTTHARLA NDVGSHKFEKERGSGTAIECYMKDYHVSEEEAM KKFEEMCDDAWKVMNEE	268
WenAngSQTS760	Wendlandia angustofolia	NNQHESVRQLADFPENIWADRV	269
		QGHDMCAKEIEMLKEEVMSMLLE	270
		QVKHALEQPLHRGIPRYEAYCFISIYEEDESNNKLL LRLAKLDYHLLQMSYKRE	271
		RDRIVECYFWAVATYYEPQYSLARMT	272
		EVGEDLAKKQRTYAFDKYIE	273
		YARTSFTQSKWFLTNELPSFSDYL	274
		TFLGMDGASEDVINWMSTNPKLFVA	275
		KFEKERGSGTAIECYMKDYHVSEEEAMKKFEEMC DDAWKVMNEE	276
WenAngSQTS780	Wendlandia angustofolia	NNQHESVRQLADFPENIWADRV	277
		QGHDMCAKEIEMLKEEVMSMLLE	278
		QVKHALEQPLHRGIPRYEAYCF	279
		LAKLDYHLSQMLNKQDLCEI	280
		RDRIVECYFWAVATYYEPQYSLARMT	281
		EVGEDLAKKQRTYAFDKYIE	282
		YARTSFTQSKWFLTNELPSFSDYL	283
		TFLGMDGASEDVINWMSTNPKLFVA	284
		KFEKERGSGTAIECYMKDYHVSEEEAMKKFEEMC DDAWKVMNEE	285
WenAngSQTS793	Wendlandia angustofolia	SNNRQETVRPLADFPENIWADRIAPFT	286
		EMCQREIEMLKAEVASMLLATGKTMMQRFDFID KIERLGVSHHFD	287
		IFDQFIDAKGKFKESLCNDIRGLLSLYEAAHVRTH GDKILEEALAFTTTHMTSGGPHLDSSLAKQVKYA LEQPLHKGILRYEAWRYISIYEEDESNNKLLLRLA KLDYHLLQMSYKQEL	288
		RWGKGLESVSNFPYARDRFVECYFWAVGTLYEP QYSLARMTFAKVAA	289
		RWDGSGVDQLSDYIRASYNTLLKFNKEVGEDLAK KQRTYAFDKYIEDWKQYMRTSFTQSKWFLTNELP SFADY	290
		LDMDSALEDVINWMSTNPKLMVAY	291
		KFDKERGSVTALDCYMKDYSVSEEEAAKKFREM CEDNWKVMNEECLRPTTI	292
		ETVYKHRIDGFTQPHAIEEHIRAMLVDFMSI	293
WenAngSQTS805	Wendlandia angustofolia	NNQHESVRQLADFPENIWADRV	294
		QGHDMCAKEIEMLKEEVMSMLLE	295
		STLAKQVKYALEQPLHKGIPRYEAWRYISIYEEDE SNNKLLLRLAKLDYHLLQMSYKRE	296
		RDRIVECYFWAVATYYEPQYSLARMT	297
		EVGEDLAKKQRTYAFDKYIE	298
		YARTSFTQSKWFLTNELPSFSDYL	299
		TFLGMDGASEDVINWMSTNPKLFVA	300
		STAIECYMKDYHVSEEEAMEKFEEMCDDAWKVM NEE	301

WenAngSQTS826	Wendlandia angustofolia	NNQHESVRQLADFPENIWADRV	302
		QGHDMCAKEIEMLKEEVMSMLLE	303
		STLAKQVKYALEQPLHKGIPRYEAWRYISIYEEDE SNNKLLLRLAKLDYHLLQMSYKRE	304
		RDRIVECYFWAVATYYEPQYSLARMT	305
		EVGEDLAKKQRTYAFDKYIE	306
		YARTSFTQSKWFLTNELPSFADYLS	307
		AALLDMDSALEDVINWMSTNPKFFVALTTHARLT NDVGSHKFEKERGSGTAIECYMKDYHVSEEEAM KKFEEMCDDAWKVMNEE	308
WenAngSQTS829	Wendlandia angustofolia	SNNRQETVRPLADFPENIWADRIAPFT	309
		EMCQREIEMLKAEVASMLLATGKTMMQRFDFID KIERLGVSHHFD	310
		IFDQFIDAKGKFKESLCNDIRGLLSLYEAAHVRTH GDKILEEALAFTTTHMTSGGPHLDSSLAKQVKYA LEQPLHKGILRYEAWRYISIYEEDESNNKLLLRLA KLDYHLLQMSYKQEL	311
		RWGKGLESVSNFPYARDRFVECYFWAVGTLYEP QYSLARMTFAKVAA	312
		RWDGSGVDQLSDYIRASYNTLLKFNKEVGEDLAK KQRTYAFDKYIEDWKQYMRTSFTQSKWFLTNELP SFADY	313
		LDMDSALEDVINWMSTNPKLMVAY	314
		KERGTGTAIECYMKDHN	315
		EMIENTWKVMNEECLRPIPIPRDTLKML	316
		ETVYKHRIDGFTQPHAIEEHIRAMLVDFMSI	317
WenAngSQTS843	Wendlandia angustofolia	NNQHESVRQLADFPENIWADRV	318
		QGHDMCAKEIEMLKEEVMSMLLE	319
		QVKHALEQPLHRGIPRYEAYCFISIYEEDESNNKLL LRLAKLDYHLLQMSYKRE	320
		RDRIVECYFWAVATYYEPQYSLARMT	321
		EVGEDLAKKQRTYAFDKYIE	322
		YARTSFTQSKWFLTNELPSFSDYL	323
		AAFLDMDSASEDVINWMSTNPKLFVALTTHARLA NDVGSHK	324
		RGSGTAIECYMKDYNVSEEEALKKFEEMCEDTW KVMNEE	325
WenAngSQTS848	Wendlandia angustofolia	SNNRQETVRPLADFPENIWADRIAPFT	326
		EMCQREIEMLKAEVASMLLATGKTMMQRFDFID KIERLGVSHHFD	327
		IFDQFIDAKGKFKESLCNDIRGLLSLYEAAHVRTH GDKILEEALAFTTTHMTSGGPHLDSSLAKQVKYA LEQPLHKGILRYEAWRYISIYEEDESNNKLLLRLA KLDYHLLQMSYKQEL	328
		RWDGSGVDQLSDYIRASYNTLLKFNKEVGEDLAK KQRTYAFDKYIEDWKQYMRTNFSQSRWFFTKELP SFADYINNGAITIGAYLVASAAFLYMDSAKEDVIN WMSTNPKLVVAYSTHSRLINDFGGHKFDKERGSG TALECYMKDYNVSEEEAANKFREMMEDAWKVM NEDCLRPTSIPRDVSKVLLNVARAGEIVYKHRIDG FTEPHIIKDHIRATLVDFMAIN	329
		RWGKGLESVSNFPYARDRFVECYFWAVGTLYEP QYSLARMTFAKVAA	330
WenAngSQTS849		SNNRQETVRPLADFPENIWADRIAPFT	331

	Wendlandia angustofolia	EMCQREIEMLKAEVASMLLATGKTMMQRFDFID KIERLGVSHHFD	332
		IFDQFIDAKGKFKESLCNDIRGLLSLYEAAHVRTH GDKILEEALAFTTTHMTSGGPHLDSSLAKQVKYA LEQPLHKGILRYEAWRYISIYEEDESNNKLLLRLA KLDYHLLQMSYKQEL	333
		RWGKGLESVSNFPYARDRFVECYFWAVGTLYEP QYSLARMTFAKVAA	334
		RWDGSGVDQLSDYIRASYNTLLKFNKEVGEDLAK KQRTYAFDKYIEDWKQYMRTSFTQSKWFLTNELP SFADY	335
		LDMDSALEDVINWMSTNPKLMVAY	336
		KFDKERGSVTALDCYMKDYSVSEEEAAKKFREMI ENTWKVMNEECLRPIPIPRDTLKML	337
		EPHIIKDHIRAMLVDFMAI	338
WenAngSQTS864	Wendlandia angustofolia	SNNRQETVRPLADFPENIWADRIAPFT	339
		EMCQREIEMLKAEVASMLLATGKTMMQRFDFID KIERLGVSHHFD	340
		IFDQFIDAKGKFKESLCNDIRGLLSLYEAAHVRTH GDKILEEALAFTTTHMTSGGPHLDSSLAKQVKYA LEQPLHKGILRYEAWRYISIYEEDESNNKLLLRLA KLDYHLLQMSYKQEL	341
		RWGKGLESVSNFPYARDRFVECYFWAVGTLYEP QYSLARMTFAKVAA	342
		RWDGSGVDQLSDYIRASYNTLLKFNKEVGEDLAK KQRTYAFDKYIEDWKQYMRTNFSQSRWFFTKELP SFADYINNGAITIGAYLVASAAFLYMDSAKEDVIN WMSTNPKLVVAYSTHSRLINDFGGHKFDKERGSV TALDCYMKDYSVSEEEAAKKFREMCEDNWKVM NEECLRPTTI	343
		ETVYKHRIDGFTQPHAIEEHIRAMLVDFMSI	344
WenAngSQTS925	Wendlandia angustofolia	SNNRQETVRPLADFPENIWADRIAPFT	345
		EMCQREIEMLKAEVASMLLATGKTMMQRFDFID KIERLGVSHHFD	346
		IFDQFIDAKGKFKESLCNDIRGLLSLYEAAHVRTH GDKILEEALAFTTTHMTSGGPHLDSSLAKQVKYA LEQPLHKGILRYEAWRYISIYEEDESNNKLLLRLA KLDYHLLQMSYKQEL	347
		RWGKGLESVSNFPYARDRFVECYFWAVGTLYEP QYSLARMTFAKVAA	348
		RWDGSGVDQLSDYIRASYNTLLKFNKEVGEDLAK KQRTYAFDKYIEDWKQYMRTNFSQSRWFFTKELP SFADYINNGAITIGAYLVASAAFLYMDSAKEDVIN WMSTNPKLVVAYSTHSRLINDFGGHKFDKERGSV TALDCYMKDYSVSEEEAAKKFREMIENTWKVMN EECLRPIPIPRDTLKML	349
		EPHIIKDHIRAMLVDFMAI	350
WenAngSQTS960	Wendlandia angustofolia	EAFNKLKDEEGNFKASLTSDVRGLLELYQASYMR IHGEDILDEAISFTTAQLTLALPTLDPP	351
		NKALLQFAKIDFNMLQLLHRKELSEICRWWKDLD FTRKLP	352
		DRVVEGYFWIMGVYFEPQYSLGRKMLTKVIAMA SIVDDTYDSFATYDELIPYTDAIER	353
		YMQISYKALLDVYEEMEQLLADKGRQYRVEY	354
		WTHLNYKPTFEEFRDNALPTSGYAMLAIT	355

		TFEWAASDPKIIKASTIICRFMDDIAE	356
		EDDCSAIECYMEQYKVTAQEAYDEFNKHIESSWK DVNEEFLK	357

Example 2. Materials and methods for construction of terpene synthase chimeras.

Terpene synthases for Capture-seq and chimera scaffolding

Candidate sesquiterpene synthases (SQTSs) were designed by combining sequence fragments from rare flower genomes (Table 11) with "scaffold" SQTSs from sources including UniProt and GenBank.

For Capture-seq (targeted sequencing of terpene synthases), a subset of 5,171 terpene synthases (TPSs) were compiled from UniProt that had nucleotide sequences in EMBL/Genbank. Oligonucleotide chips were generated for enriching the flower DNA samples for TPS-homologous sequences, and then subjected first to Illumina sequencing. The Capture-seq libraries were also sequenced a second time at higher depth.

For SQTS chimera reconstruction, sequences closer to annotated SQTSs than annotated mono-, di-, or tri-terpene synthases were selected. This set of 1,521 putative SQTSs were used (in both nucleotide and peptide form) as query sequences for blastn and tblastn in the chimera construction pipeline below.

Chimera reconstruction

Two methods were used for constructing chimeric SQTSs: 1) the blastn-mapDamage pipeline, and 2) the tblastn pipeline.

blastn-mapdamage pipeline

Generally, the blastn-mapdamage pipeline conservatively detects fragments with high nucleotide similarity to the scaffolds resulting in chimeric terpene synthases (*e.g.*, chimeric sesquiterpene synthases) that are likely very close to the original enzyme sequences in the rare flowers. To detect mutations that may be artifacts of stereotypical rare DNA damage, bam-formatted Illumina read alignments were inputted into mapDamage software.

Specifically, the following steps were used to generate alignments of DNA fragments from each flower to various SQTS scaffolds:

1. Illumina reads (fastq files) from genomic capture-seq runs were combined and assembled by SPADES into longer contigs.
2. The 1521-set of SQTS scaffolds were used as queries in a blastn search with default parameters against the SPADES contigs. Relatively few scaffolds had hits, so all of the scaffolds with hits were chosen to serve as references for read alignment in the next step.
3. Combined reads from the sequencing runs were quality-trimmed (using bbduk) and pair-merged (using bbmerge) and aligned to chosen SQTS reference sequences using bwa mem. Results were reformatted to bam, sorted, and indexed.
4. mapDamage was run on the aligned reads. This resulted in a read alignment where SNPs resembling DNA damage were assigned low quality scores.
5. Read alignments were processed as follows: bases with quality < 25 were masked (changed to the reference); alignments were reformatted to fasta; SNPs with counts < 6 were masked; duplicate reads were removed; SNPs with frequency < 0.1 were masked; reads that were exact subsequences of other reads were removed; reads were translated in the frame of the reference; and subsequences were removed again. The quality and SNP frequency thresholds used for masking the alignment were determined empirically by looking at distributions of quality and SNP frequency.
6. Read alignments and SPADE contig alignments (after reference-frame translation) were combined and realigned using Clustal Omega. This was done because some contigs spanned regions of the scaffolds that the reads did not.

The alignments from the above steps were used to construct SQTS chimeras as follows:

1. The alignment was split into "independent subregions" such that each subregion did not contain any fragment (aligned read) overlapping with and differing from a fragment from another subregion (identical overlaps were allowed between subregions).
2. In each subregion, all possible combinations of "compatible fragments" were enumerated. Compatible fragments were defined as fragments that either overlapped identically (and therefore could be merged into a longer fragment) or did not overlap at all (and, *e.g.*, were assumed to come from the same haplotype). Fragment

combinations were "max-coverage" – that is, contained as many compatible fragments as possible. Each max-coverage fragment combination was considered to be a possible reconstruction of that region of the alignment, and was merged into a superfragment (which may have contained gaps) and saved.

3. Superfragments from each subregion were downsampled to 90% or 95% identity using a custom, iterative algorithm, and all possible combinations of downsampled superfragments from different subregions were combined. Regions that were shorter than a certain threshold are downsampled to a single sequence. Each combination of superfragments was merged into the scaffold to generate a chimera sequence. The downsampling parameters were varied slightly varied according to the sample and scaffold to allow >1 but <100 chimeras to be constructed in each case.

After running the above pipeline on each sample, a total of 1136 chimeras were generated. A significant fraction of the chimeras were constructed purely from aligned reads.

A total of 652 sesquiterpene synthase chimeras were created using these methods.

tblastn pipeline

Generally, the tblastn pipeline maximized the sensitivity of detecting fragments homologous to the SQTS scaffolds, and therefore cast a wide net for potentially usable sequences.

Specifically, the following steps were used to generate alignments of DNA fragments from each flower to various SQTS scaffolds:

1. The 1521-set of SQTS scaffolds were used as protein queries to tblastn to search all-frames translations of the SPADES contigs (described above).
2. Hits (aligned contigs) were filtered to a minimum of 40% identity to the scaffold and a minimum length that depends on hit identity by a heuristic function. The filtering criteria were chosen by inspecting plots of hit length versus identity across all samples.
3. Downsampling scaffolds was performed by hierarchically clustering the scaffolds by the number of identical residues to each hit. The scaffold in each cluster with the greatest number of identities across all of its hits was kept for chimera reconstruction.

Downsampling reduced the number of scaffolds by 20-fold. This step was skipped for samples in which fewer than 10 scaffolds have hits.

4. Certain scaffolds were always chosen as a cluster representation because they were previously identified as having activity and/or were known in the literature (even if another sequence had more identities to hits). These preferred scaffolds were not downsampled, and tblastn hits were kept for chimera construction.
5. The aligned portions of all contigs hitting a scaffold were realigned to the scaffold using Clustal Omega. Unaligned portions of contigs were discarded as likely representing introns. This alignment was then used for chimera construction.
6. Chimeras were constructed from aligned tblastn hits using the combinatorial compatible fragments method described above without downsampling in subregions. Both "max-coverage" (as many as possible compatible fragments in each set) and "min-coverage" (only one compatible fragment in each set) chimeras were generated. The min-coverage chimeras may avoid combining fragments from unrelated sequences.

The tblastn pipeline yielded 10,114 "max-coverage" chimeras and 2,624 "min-coverage" chimeras. Certain max-coverage chimeras were downsampled to 95% identity by CD-HIT. This resulted in 388 sequences (382 after removing sequences with ambiguous amino acids). Certain max coverage chimeras were filtered to a minimum rare DNA content of 60% and downsampled to 90% identity. This resulted in 1320 sequences. Certain min-coverage chimeras were filtered to a minimum rare DNA content of 10% and downsampled to 95% identity by CD-HIT.

Encoding and synthesis order

Each enzyme was codon-optimized twice: once using a yeast expression-weighted codon table, and once using a yeast expression-weighted codon table after removing codons with <10% frequency. A different random number was used as the seed for each encoding. Encodings for different enzymes were completely independent – no specific procedure was used to preserve codons at residues inherited by chimeras from scaffolds.

Sequences encoding the chimeric enzymes were cloned into the pESC-URA3 screening vector, driven by pGAL1 and terminated by tCYC1.

Chimera reconstruction aided by extant transcriptome

For one of the extinct flower species, *Shorea cuspidata*, transcriptome sequencing data was available on an extant relative *Shorea beccariana*. This made it possible to construct chimeras using SQTS scaffolds from a related flower. This was done in a 2-step process:

1. The *S. beccariana* (Sb) transcriptome data were assembled and mined for SQTS homologs. The data were downloaded from the data set SRR687302 from the NCBI SRA database. Assembly was done using Trinity, and ORFs were predicted via Transdecoder. BLAST was used to identify fragments homologous to a set of 1,500 curated SQTS sequences.
2. The identified Sb SQTSs or SQTS fragments were used as scaffold sequences in either the tblastn or blastn-mapDamage pipelines to reconstruct chimeras. If the scaffold was a fragment itself, it was in turn merged into the closest Uniprot-sourced SQTS sequence to generate a full-length chimera.

Screening Strain and Sesquiterpene Synthase Transformation

The chimeric sesquiterpene synthases were transformed into high copy pESC-URA3-derived expression vectors under the control of the galactose-inducible P(gal1) promoter (Sikorski *et al.*, A system of shuttle vectors and yeast host strains designed for efficient manipulation of DNA in Saccharomyces cerevisiae. *Genetics*. 1989 May;122(1):19-27, which is hereby incorporated by reference in its entirety for this purpose).

These vectors were transformed into a haploid *Saccharomyces cerevisiae* CEN.PK2 strain (MATa ura3-52 trp1-289 leu2-3_112 his3Δ1 MAL2-8C SUC2) that had been modified to increase sesquiterpene flux via integration of two copies of the catalytic region of HMG-CoA reductase 1 under control of convergent P(gal1) promoters at the homothallic switching endonuclease (YDL227C) locus on chromosome 4 (see SEQ ID NO: 53 shown below). See: Entian *et al.*, Yeast Genetic Strain and Plasmid Collections. *Methods in Microbiology*. 2007; (36): 629-666; tHMG1, Donald *et al.*, Effects of overproduction of the catalytic domain of 3-hydroxy-3-methylglutaryl coenzyme A reductase on squalene synthesis in Saccharomyces cerevisiae. *Appl Environ Microbiol*. 1997 Sep;63(9):3341-4; Özaydın *et al.*, Carotenoid-based phenotypic screen of the yeast deletion collection reveals new genes with roles in isoprenoid

production. Metab Eng. 2013 Jan;15:174-83, each of which is hereby incorporated by reference in its entirety). Competition for fanesyl pyrophosphate was reduced in these cells by replacing the Erg9 (Farnesyl-diphosphate farnesyl transferase) promoter with the methionine-repressible Met3 promoter as shown below in SEQ ID NO: 54 and incubating in media containing methionine (see: Ro *et al.*, Production of the antimalarial drug precursor artemisinic acid in engineered yeast. *Nature*. 2006 Apr 13;440(7086):940-3; and Asadollahi *et al.*, Production of plant sesquiterpenes in Saccharomyces cerevisiae: effect of ERG9 repression on sesquiterpene biosynthesis. *Biotechnol Bioeng*. 2008 Feb 15;99(3):666-77, each of which is hereby incorporated by reference in its entirety for this purpose). This strain with downregulated Erg9 and containing two copies of galactose-inducible tHMG1 on chromosome 4 was designated t119889.

The transformation of the chimeric sesquiterpene vectors into strain t119889 was performed employing the chemical transformation techniques demonstrated in Gietz *et al.*, Yeast transformation by the LiAc/SS Carrier DNA/PEG method. *Methods Mol Biol*. 2006;313:107-20, which is hereby incorporated by reference in its entirety for this purpose.

Sesquiterpene Production and Extraction

Transformant colonies were inoculated into 300 µl of SC-ura medium (Synthetic Complete with 2% dextrose, no uracil added) in 96 deep well plates. The plates were covered with Excel Scientific AeroSeal membranes (BS-25) and incubated for 48 hours at 30 °C in a shaking incubator. 30 µl of the cultures (1:15 dilution) were mixed into 420 µl of SC -ura induction medium containing 1.8% galactose and 0.2% raffinose as the carbon sources, yielding a starting optical density at 600 nm (OD_{600}) of approximately 0.1-0.2. A 0.88% dodecane overlay (4 µl) was added to each well and the plates were covered with AeroSeal membranes and incubated at 30°C in a shaking incubator for four days. 15 µl of each culture was removed to measure OD_{600} at the end of the four days. 350 µl of ethyl acetate (250 µM tridecane internal) was added to directly to each well and mixed (1:1 Extraction). The 96-well plates were then centrifuged and the ethyl acetate extractions were stored at -80 °C in glass vials until analysis by GC-MS.

Sesquiterpene structure identification

Ethyl Acetate samples (1.0uL) were injected into the Agilent/Gerstel 7890B GC System, where the GC inlet was set to 250C with a split ratio of 2:1. The capillary column was an Agilent DB-5MS (20m×0.18 mm×0.18μm) with carrier gas (helium) flow set to 1.5ml/min. The GC oven temperature was set to 100°C (hold for 0.10 min) with a ramp of 40°C/min to 155°C, where the ramp was then 15°C/min to 190°C and then finally the ramp was changed to 75°C/min to 280C (5-minute method). For a more comprehensive analysis of targets, the GC oven temperature was set to 100°C (hold for 2.0 min) with a 10°C/min ramp to 250°C (hold for 2.0 min) was utilized (20-minute method). The MS source and quadruple for both methods were set to 230°C and 180°C on the Agilent 5977B MSD (Etune), respectively. The mass scan range was set to 40-250 mz where spectra and linear retention index calculations were matched against the NIST MS database (2008 version), in addition to available standards and essential oils.

Peaks present in the extracted ion chromatogram (204.2mz parent mass) were identified in one of six ways (see Table 3). The authentic standards utilized in this screen for verification of products were beta-caryophyllene (Sigma-Aldrich catalog # W225207-SAMPLE-K), beta-farnesene (Sigma-Aldrich catalog # 73492-1ML-F), trans-nerolidol (Sigma-Aldrich catalog # 18143-100MG-F), and alpha-humulene (Sigma-Aldrich catalog # 53675-1ML). Sesquitperene rich essential oils used to aid structure identification were derived from the following plants: Rhodendron, Sweet Basil, Black Pepper, Citronella, Ylang, Balsam copaiba, and Patchouli.

ΔHO(YDL227C)::2xP(gal)-*tHMG1* integration on chromosome 4.

AGGGTTCGCAAGTCCTGTTTCTATGCCTTTCTCTTAGTAATTCACGAAATAAACCT
ATGGTTTACGAAATGATCCACGAAAATCATGTTATTATTTACATCAACATATCGCG
AAAATTCATGTCATGTCCACATTAACATCATTGCAGAGCAACAATTCATTTTCATAG
AGAAATTTGCTACTATCACCCACTAGTACTACCATTGGTACCTACTACTTTGAATTG
TACTACCGCTGGGCGTTATTAGGTGTGAAACCACGAAAAGTTCACCATAACTTCGA
ATAAAGTCGCGGAAAAAAGTAAACAGCTATTGCTACTCAAATGAGGTTTGCAGAAG
CTTGTTGAAGCATGATGAAGCGTTCTAAACGCACTATTCATCATTAAATATTTAAA
GCTCATAAAATTGTATTCAATTCCTATTCTAAATGGCTTTTATTTCTATTACAACTA
TTAGCTCGATGCACGAGCGCAACGCTCACAACGCTCGTCCAACGCCGGCGGACCT*ACG*
GATTAGAGCCGCCGAGCGGGTGACAGCCCTCCGAAGGAAGACTCTCCTCCGTGCGTCCTCG
TCTTCACCGGTCGCGTTCCTGAAACGCAGATGTGCCTCGCGCCGCACTGCTCCGAACAATAA
AGATTCTACAATACTAGCTTTTATGGTTATGAAGAGGAAAAATTGGCAGTAACCTGGCCCCACA
AACCTTCAAATGAACGAATCAAATTAACAACCATAGGATGATAATGCGATTAGTTTTTTAGCCTT
ATTTCTGGGGTAATTAATCAGCGAAGCGATGATTTTTGATCTATTAACAGATATATAAATGCAAA
AACTGCATAACCACTTTAACTAATACTTTCAACATTTTCGGTTTGTATTACTTCTTATTCAAATGT
*AATAAAAGTATCAACAAAAAATTGTTAATATACCTCTATACTTTAACGTCAAGGAGAAAAAACT*A
TA**ATGGCTGCAGACCAATTGGTGAAGACTGAAGTCACCAAGAAGTCTTTTACTGCT**

CCTGTACAAAAGGCTTCTACACCAGTTTTAACCAATAAAACAGTCATTTCTGGATC
GAAAGTCAAAAGTTTATCATCTGCGCAATCGAGCTCATCAGGACCTTCATCATCTA
GTGAGGAAGATGATTCCCGCGATATTGAAAGCTTGGATAAGAAAATACGTCCTTTA
GAAGAATTAGAAGCATTATTAAGTAGTGGAAATACAAAACAATTGAAGAACAAAGA
GGTCGCTGCCTTGGTTATTCACGGTAAGTTACCTTTGTACGCTTTGGAGAAAAAAT
TAGGTGATACTACGAGAGCGGTTGCGGTACGTAGGAAGGCTCTTTCAATTTTGGC
AGAAGCTCCTGTATTAGCATCTGATCGTTTACCATATAAAAATTATGACTACGACC
GCGTATTTGGCGCTTGTTGTGAAAATGTTATAGGTTACATGCCTTTGCCCGTTGGT
GTTATAGGCCCCTTGGTTATCGATGGTACATCTTATCATATACCAATGGCAACTAC
AGAGGGTTGTTTGGTAGCTTCTGCCATGCGTGGCTGTAAGGCAATCAATGCTGGC
GGTGGTGCAACAACTGTTTTAACTAAGGATGGTATGACAAGAGGCCCAGTAGTCC
GTTTCCCAACTTTGAAAAGATCTGGTGCCTGTAAGATATGGTTAGACTCAGAAGAG
GGACAAAACGCAATTAAAAAAGCTTTTAACTCTACATCAAGATTTGCACGTCTGCA
ACATATTCAAACTTGTCTAGCAGGAGATTTACTCTTCATGAGATTTAGAACAACTA
CTGGTGACGCAATGGGTATGAATATGATTTCTAAGGGTGTCGAATACTCATTAAAG
CAAATGGTAGAAGAGTATGGCTGGGAAGATATGGAGGTTGTCTCCGTTTCTGGTA
ACTACTGTACCGACAAAAAACCAGCTGCCATCAACTGGATCGAAGGTCGTGGTAA
GAGTGTCGTCGCAGAAGCTACTATTCCTGGTGATGTTGTCAGAAAAGTGTTAAAAA
GTGATGTTTCCGCATTGGTTGAGTTGAACATTGCTAAGAATTTGGTTGGATCTGCA
ATGGCTGGGTCTGTTGGTGGATTTAACGCACATGCAGCTAATTTAGTGACAGCTGT
TTTCTTGGCATTAGGACAAGATCCTGCACAAAATGTCGAAAGTTCCAACTGTATAA
CATTGATGAAAGAAGTGGACGGTGATTTGAGAATTTCCGTATCCATGCCATCCATC
GAAGTAGGTACCATCGGTGGTGGTACTGTTCTAGAACCACAAGGTGCCATGTTGG
ACTTATTAGGTGTAAGAGGCCCACATGCTACCGCTCCTGGTACCAACGCACGTCAA
TTAGCAAGAATAGTTGCCTGTGCCGTCTTGGCAGGTGAATTATCCTTATGTGCTGC
CCTAGCAGCCGGCCATTTGGTTCAAAGTTATATGACCCACAACAGGAAACCTGCTG
AACCAACAAAACCTAACAATTTGGACGCCACTGATATAAATCGTTTGAAAGATGGG
TCCGTCACCTGCATTAAATCCTAAGCTAGCTA*AGATCCGCTCTAACCGAAAAGGAAGG*
AGTTAGACAACCTGAAGTCTAGGTCCCTATTTATTTTTTTATAGTTATGTTAGTATTAAGAA
CGTTATTTATATTTCAAATTTTTCTTTTTTTTCTGTACAGACGCGTGTACGCATGTAACATT
*ATACTGAAAACCTTGCTTGAGAAGGTTTTGGGACGCTCGAAGATCCAGCT*CGGCCGTACG
AAAATCGTTATTGTCTTGAAGGTGAAATTTCTACTCTTATTAATGGTGAACGTTAAGCTG
ATGCTATGATGGAAGCTGATTGGTCTTAACTTGCTTGTCATCTTGCTAATGGTCATATGG
CTCGTGTTATTACTTAAGTTATTTGTACTCGTTTTGAACGTAATGCTAATGATCATCTTAT
GGAATAATAGTGAACGGCCG*agctggatcttcgagcgtcccaaaaccttctcaagcaaggttttcagtataatgttacatgc*
gtacacgcgtctgtacagaaaaaaagaaaaatttgaaatataaataacgttcttaatactaacataactataaaaaaataaatagggacct
*agacttcaggttgtctaactccttcctttcggttagagcggatct*TAGCTAGCttaggatttaatgcaggtgacggacccatctttcaaa
cgatttatatcagtggcgtccaaattgttaggttttgttggttcagcaggtttcctgttgtgggtcatataactttgaaccaaatggccggctg
ctagggcagcacataaggataattcacctgccaagacggcacaggcaactattcttgctaattgacgtgcgttggtaccaggagcggta
gcatgtgggcctcttacacctaataagtccaacatggcaccttgtggttctagaacagtaccaccaccgatggtacctacttcgatggatg
gcatggatacggaaattctcaaatcaccgtccacttctttcatcaatgttatacagttggaactttcgacattttgtgcaggatcttgtcctaa
tgccaagaaaacagctgtcactaaattagctgcatgtgcgttaaatccaccaacagacccagccattgcagatccaaccaaattcttagc
aatgttcaactcaaccaatgcggaaacatcacttttaacacttttctgacaacatcaccaggaatagtagcttctgcgacgacactcttac
cacgaccttcgatccagttgatggcagctggtttttttgtcggtacagtagttaccagaaacggagacaacctccatatcttcccagccatac
tcttctaccatttgctttaatgagtattcgacacccttagaaatcatattcatacccattgcgtcaccagtagttgttctaaatctcatgaaga
gtaaatctcctgctagacaagtttgaatatgttgcagacgtgcaaatcttgatgtagagttaaaagctttttttaattgcgttttgtccctcttct

gagtctaaccatatcttacaggcaccagatcttttcaaagttgggaaacggactactgggcctcttgtcataccatccttagttaaaacagt tgttgcaccaccgccagcattgattgccttacagccacgcatggcagaagctaccaaacaaccctctgtagttgccattggtatatgataa gatgtaccatcgataaccaaggggcctataacaccaacgggcaaaggcatgtaacctataacatttttcacaacaagcgccaaatacgc ggtcgtagtcataatttttatatggtaaacgatcagatgctaatacaggagcttctgccaaaattgaaagagccttcctacgtaccgcaac cgctctcgtagtatcacctaatttttttctccaaagcgtacaaaggtaacttaccgtgaataaccaaggcagcgacctctttgttcttcaattg ttttgtatttccactacttaataatgcttctaattcttctaaaggacgtattttcttatccaagctttcaatatcgcgggaatcatcttcctcact agatgatgaaggtcctgatgagctcgattgcgcagatgataaacttttgactttcgatccagaaatgactgttttattggttaaaactggtg tagaagccttttgtacaggagcagtaaaagacttcttggtgacttcagtcttcaccaattggtctgcagccatTAT*agtttttctccttgac gttaaagtatagaggtatattaacaatttttgttgatacttttattacatttgaataagaagtaatacaaaccgaaaatgttgaaagtattagttaa agtggttatgcagtttttgcatttatatatctgttaatagatcaaaaatcatcgcttcgctgattaattaccccagaaataaggctaaaaaactaat cgcattatcatcctatggttgttaatttgattcgttcatttgaaggtttgtggggccaggttactgccaatttttcctcttcataaccataaaagctagt attgtagaatctttattgttcggagcagtgcggcgcgaggcacatctgcgtttcaggaacgcgaccggtgaagacgaggacgcacggagga gagtcttccttcggagggctgtcacccgctcggcggcttctaatccgt*AGGTCCGCCGGCGTTGGACGAGCGTTGTG AGCGTTGCGCTCGTGCATC**aatgtgtatattagtttaaaaagttgtatgtaataaaagtaaaatttaatattttggatgaaa aaaaccatttttagacttttttcttaactagaatgctggagtagaaatacgccatctcaagatacaaaaagcgttaccggcactgatttgttt caaccagtatatagattattattgggtcttgatcaactttcctcagacatatcagtaacagttatcaagctaaatatttacgcgaaagaaaa acaaatattttaattgtgatacttgtgaattttattttattaaggatacaaagttaagagaaaacaaaatttatatacaatataagtaatatt catatatatgtgatgaatgcagtcttaacgagaagacatggccttggtgacaactctcttcaaaccaacttcagcctttctcaattcatcag cagatgggtcttcgatttgcaaagcagcca** (SEQ ID NO: 53)

Upper case, bold: HO upstream homology sequence (SEQ ID NO: 56)
Upper case, italicized and underlined: P(gal1) (SEQ ID NO: 57)
Upper case, underlined and bold: *tHMG1* (SEQ ID NO: 58)
***Upper case, bold and italicized*:** CYC1 terminator (SEQ ID NO: 59)
***Lower case, bold and italicized*:** CYC1 terminator, reverse complement (SEQ ID NO: 60)
Lower case, underlined and bold: *tHMG1*, reverse complement (SEQ ID NO: 61)
Lower case, italicized and underlined: P(gal1), reverse complement (SEQ ID NO: 62)
Lower case, bold: HO downstream homology sequence (SEQ ID NO: 63)

P(met3) integration upstream of Erg9 with flanking genes included.
ATGTCCGGTAAATGGAGACTAGTGCTGACTGGGATAGGCAATCCAGAGCCTCAGT ACGCTGGCACCCGTCACAATGTAGGGCTATATATGCTGGAGCTGCTACGAAAGCG GCTTGGTCTGCAGGGGAGAACCTATTCCCCTGTGCCTAATACGGGCGGCAAAGTG CATTATATAGAAGACGAACATTGTACGATACTAAGATCGGATGGCCAGTACATGAA TCTAAGTGGAGAACAGGTGTGCAAGGTCTGGGCCCGGTACGCCAAGTACCAAGCC CGACACGTTGTTATTCATGACGAGTTAAGTGTGGCGTGTGGAAAAGTGCAGCTCA GAGCCCCCAGCACCAGTATTAGAGGTCATAATGGGCTGCGAAGTCTACTGAAATG CTCCGGAGGCCGTGTACCCTTTGCCAAATTGGCTATTGGAATCGGCAGAGAACCT GGGTCCCGCTCTAGAGACCCTGCGAGCGTCTCCCGCTGGGTTCTGGGAGCTCTAA CTCCGCAGGAACTACAAACCTTGCTTACACAGAGTGAACCTGCTGCCTGGCGTGCT CTGACTCAGTACATTTCATAGGTTTAACTTGATACTACTAGATTTTTTCTCTTCATTTAT AAAATTTTTGGTTATAATTGAAGCTTTAGAAGTATGAAAAAATCCTTTTTTTTCATTCTTT GCAACCAAAATAAGAAGCTTCTTTTATTCATTGAAATGATGAATATAAACCTAACAAAA GAAAAAGACTCGAATATCAAACATTAAAAAAAAATAAAAGAGGTTATCTGTTTTCCCAT TTAGTTGGAGTTTGCATTTTCTAATAGATAGAACTCTCAATTAATGTGGATTTAGTTTCT**

CTGTTCGTTTTTTTTTGTTTTGTTCTCACTGTATTTACATTTCTATTTAGTATTTAGTTATT
CATATAATCTTAACTTCTCGAGGAGCTCGATCTTGAAACTGAGTAAGATGCTCAGAATA
CCCGTCAAGATAAGAGTATAATGTAGAGTAATATACCAAGTATTCAGCATATTCTCCTC
TTCTTTTGTATAAATCACGGAAGGGATGATTTATAAGAAAAATGAATACTATTACACTT
CATTTACCACCCTCTGATCTAGATTTTCCAACGATATGTACGTAGTGGTATAAGGTGAGG
GGGTCCACAGATATAACATCGTTTAATTTAGTACTAACAGAGACTTTTGTCACAACTAC
ATATAAGTGTACAAATATAGTACAGATATGACACACTTGTAGCGCCAACGCGCATCCTA
CGGATTGCTGACAGAAAAAAAGGTCACGTGACCAGAAAAGTCACGTGTAATTTTGTAA
CTCACCGCATTCTAGCGGTCCCTGTCGTGCACACTGCACTCAACACCATAAACCTTAGC
AACCTCCAAAGGAAATCACCGTATAACAAAGCCACAGTTTTACAACTTAGTCTCTTATG
AAGTGTCTCTCTCTGTCGTAACAGTTGTGATATCGGAAGAAGAGAAAAGACGAAGAGC
AGAAGCGGAAAACGTATACACGTCACATATCACACACACACA**atgggaaagctattacaattggcat**
tgcatccggtcgagatgaaggcagctttgaagctgaagttttgcagaacaccgctattctccatctatgatcagtccacgtctccatatctc
ttgcactgtttcgaactgttgaacttgacctccagatcgtttgctgctgtgatcagagagctgcatccagaattgagaaactgtgttactctc
ttttatttgattttaagggctttggataccatcgaagacgatatgtccatcgaacacgatttgaaaattgacttgttgcgtcacttccacgag
aaattgttgttaactaaatggagtttcgacggaaatgcccccgatgtgaaggacagagccgttttgacagatttcgaatcgattcttattga
attccacaaattgaaaccagaatatcaagaagtcatcaaggagatcaccgagaaaatgggtaatggtatggccgactacatcttagatg
aaaattacaacttgaatgggttgcaaaccgtccacgactacgacgtgtactgtcactacgtagctggtttggtcggtgatggtttgacccgt
ttgattgtcattgccaagtttgccaacgaatctttgtattctaatgagcaattgtatgaaagcatgggtcttttcctacaaaaaaccaacatc
atcagagattacaatgaagatttggtcgatggtagatccttctggcccaaggaaatctggtcacaatacgctcctcagttgaaggacttca
tgaaacctgaaaacgaacaactggggttggactgtataaaccacctcgtcttaaacgcattgagtcatgttatcgatgtgttgacttatttg
gccggtatccacgagcaatccactttccaattttgtgccattccccaagttatggccattgcaaccttggctttggtattcaacaaccgtgaa
gtgctacatggcaatgtaaagattcgtaagggtactacctgctatttaattttgaaatcaaggactttgcgtggctgtgtcgagatttttgac
tattacttacgtgatatcaaatctaaattggctgtgcaagatccaaatttcttaaaattgaacattcaaatctccaagatcgaacagtttatg
gaagaaatgtaccaggataaattacctcctaacgtgaagccaaatgaaactccaattttcttgaaagttaaagaaagatccagatacga
tgatgaattggttccaacccaacaagaagaagagtacaagttcaatatggttttatctatcatcttgtccgttcttcttgggttttattatatat
acactttacacagagcgtga (SEQ ID NO: 54)

Uppercase, bold and underlined: Upstream sequence PTH1 (YHR189W) (SEQ ID NO: 64)
Uppercase and underlined: P(met3) (SEQ ID NO: 65)
Lowercase, bold and underlined: Erg9 (YHR190W) (SEQ ID NO: 66)

EQUIVALENTS

Those skilled in the art will recognize, or be able to ascertain using no more than routine experimentation, many equivalents to the specific embodiments of the invention described herein. Such equivalents are intended to be encompassed by the following claims.

All references, including patent documents, disclosed herein are incorporated by reference in their entirety, particularly for the disclosure referenced herein.

CLAIMS

What is claimed is:

1. A chimeric terpene synthase comprising an amino acid sequence at least 90% identical to an amino acid selected from the group consisting of: SEQ ID NOs: 1-52.

2. The chimeric terpene synthase of claim 1, wherein the chimeric terpene synthase comprises an amino acid sequence identical to an amino acid selected from the group consisting of: SEQ ID NOs: 1-52.

3. A nucleic acid molecule encoding a chimeric terpene synthase of claim 1 or 2.

4. A vector comprising the nucleic acid molecule of claim 3.

5. The vector of claim 4, wherein the vector is a viral vector, a vector for transient expression, or a vector for inducible expression.

6. The vector of claim 4 or claim 5, wherein the vector is a lentiviral vector, a retroviral vector, an adenoviral vector, an adeno-associated vector, a galactose-inducible vector, or a doxycycline-inducible vector.

7. A host cell comprising the nucleic acid molecule of claim 3.

8. A host cell that has been transformed with the vector of any one of claims 4-6 .

9. The host cell of claim 7 or claim 8, wherein the cell is a fungal cell.

10. The host cell of claim 9, wherein the fungal cell is a yeast cell.

11. The host cell of claim 10, wherein the yeast cell is a *Saccharomyces, Pichia, Kluyveromyces, Hansenula*, or *Yarrowia* cell.

12. The host cell of claim 11, wherein the cell is a *Saccharomyces cerevisiae* cell.

13. The host cell of claim 7 or claim 8, wherein the cell is a plant cell.

14. The host cell of claim 7 or claim 8, wherein the cell is a bacteria cell.

15. A nucleic acid molecule encoding a chimeric terpene synthase, wherein the nucleic acid molecule comprises a sequence that is at least 90% identical to a nucleotide sequence selected from the group consisting of: SEQ ID NOs: 67-118.

16. The nucleic acid molecule of claim 15, wherein the nucleic acid molecule comprises a sequence set forth in SEQ ID NO: 67-118.

17. The nucleic acid molecule of claim 15 or claim 16, wherein the nucleic acid molecule further comprises a TATA box sequence.

18. The nucleic acid molecule of claim 3, wherein at least 40% of the nucleic acid molecule sequence is derived from an extinct plant.

19. The chimeric terpene synthase, nucleic acid molecule, or host cell of any preceding claim, wherein the chimeric terpene synthase is a chimeric sesquiterpene synthase.

20. The chimeric terpene synthase, nucleic acid molecule, or host cell of claim 19, wherein the chimeric terpene synthase is an alpha-guaiene synthase.

21. The chimeric terpene synthase, nucleic acid molecule, or host cell of claim 20, wherein the alpha-guaiene synthase comprises a sequence that is at least 90% identical to SEQ ID NO: 17, 22, or 29.

22. The chimeric terpene synthase, nucleic acid molecule, or host cell of claim 21, wherein the alpha-guaiene synthase comprises SEQ ID NO: 17, 22, or 29.

23. The chimeric terpene synthase, nucleic acid molecule, or host cell of claim 18, wherein the alpha-guaiene synthase consists of SEQ ID NO: 17, 22, or 29.

24. The nucleic acid molecule of claim 18, wherein the extinct plant is selected from the group consisting of: *Hibiscadelphus wilderianus*, *Leucadendron grandiflorum*, *Macrostylis villosa*, *Orbexilum stipulatum*, *Shorea cuspidate*, and *Wendlandia angustifolia*.

25. A method of producing one or more sesquiterpenes, wherein the method comprises culturing a host cell comprising a nucleic acid of any one of claims 3, 15-18, or 24 under conditions suitable for producing the one or more sesquiterpenes.

26. A composition comprising one or more sesquiterpenes produced by the method of claim 25.

27. The composition of claim 26, wherein at least one of the one or more sesquiterpenes is an aroma compound.

28. A method of producing a perfume, wherein the method comprises:

culturing a host cell comprising a nucleic acid of any one of claims 3, 15-18, or 24 under conditions suitable for producing the one or more sesquiterpenes; and

extracting the one or more sesquiterpenes.

29. A chimeric terpene synthase that is capable of producing alpha-guaiene, wherein at least 40% of the amino acid sequence of the chimeric terpene synthase is derived from an extinct plant.

30. A nucleic acid molecule encoding a chimeric terpene synthase that is capable of producing alpha-guaiene, wherein at least 40% of the nucleic acid molecule sequence is derived from an extinct plant.

31. A host cell that expresses the nucleic acid of claim 30.

32. A host cell that expresses a chimeric terpene synthase that is capable of producing alpha-guaiene.

33. A method of producing alpha-guaiene comprising culturing the host cell of claim 31 or 32.

34. The method of claim 33, further comprising collecting the alpha-guaiene produced by the host cell.

35. The host cell of claim 31 or 32, wherein the cell is a fungal cell.

36. The host cell of claim 31 or 32, wherein the cell is a plant cell.

37. The host cell of claim 31 or 32, wherein the cell is a bacteria cell.

38. The host cell of any one of claims 31, 32, or 35, wherein the cell is a yeast cell.

39. The host cell of claim 38 wherein the yeast cell is a *Saccharomyces, Pichia, Kluyveromyces, Hansenula*, or *Yarrowia* cell.

40. The host cell of claim 39, wherein the yeast cell is a *Saccharomyces cerevisiae* cell.

41. A chimeric terpene synthase comprising a sequence selected from the group consisting of SEQ ID NOs: 119-357.

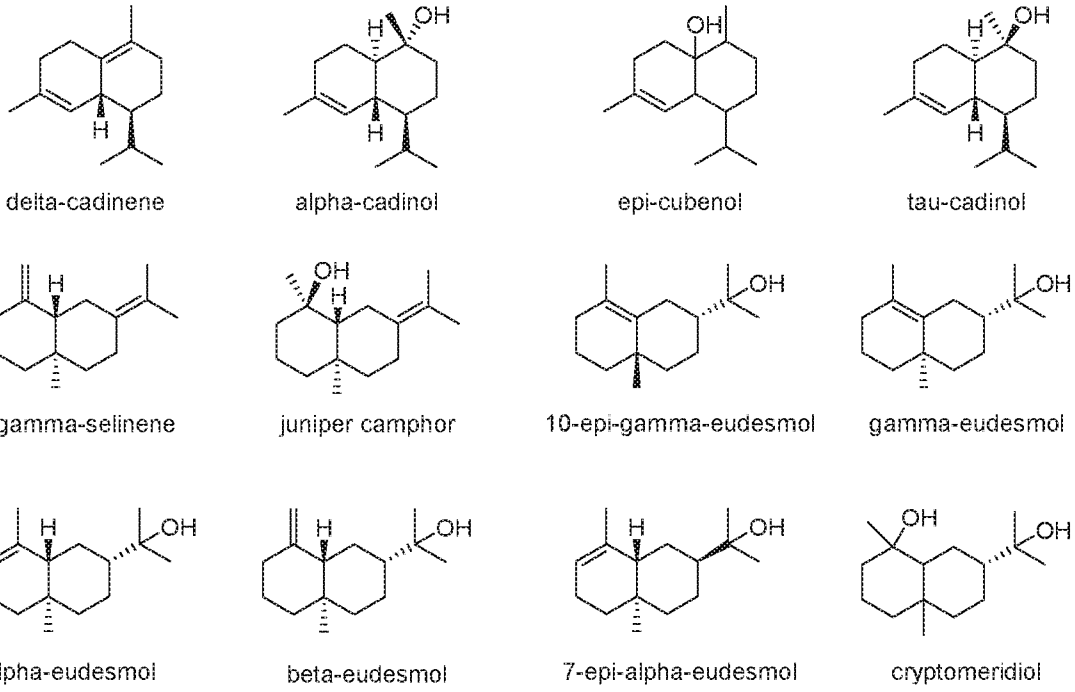

delta-cadinene
alpha-cadinol
epi-cubenol
tau-cadinol
gamma-selinene
juniper camphor
10-epi-gamma-eudesmol
gamma-eudesmol
alpha-eudesmol
beta-eudesmol
7-epi-alpha-eudesmol
cryptomeridiol

FIG. 1

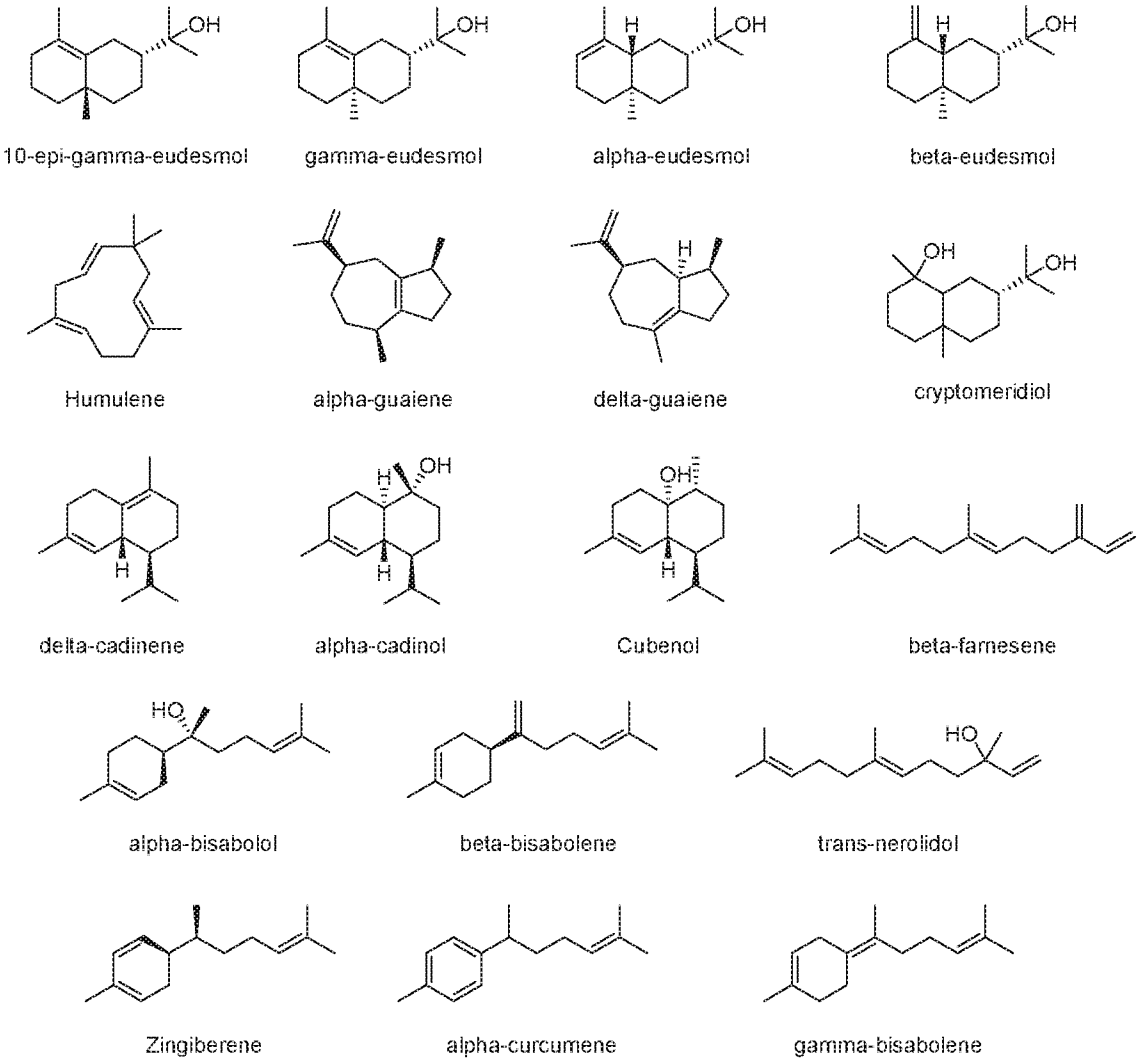

10-epi-gamma-eudesmol
gamma-eudesmol
alpha-eudesmol
beta-eudesmol
Humulene
alpha-guaiene
delta-guaiene
cryptomeridiol
delta-cadinene
alpha-cadinol
Cubenol
beta-farnesene
alpha-bisabolol
beta-bisabolene
trans-nerolidol
Zingiberene
alpha-curcumene
gamma-bisabolene

FIG. 2

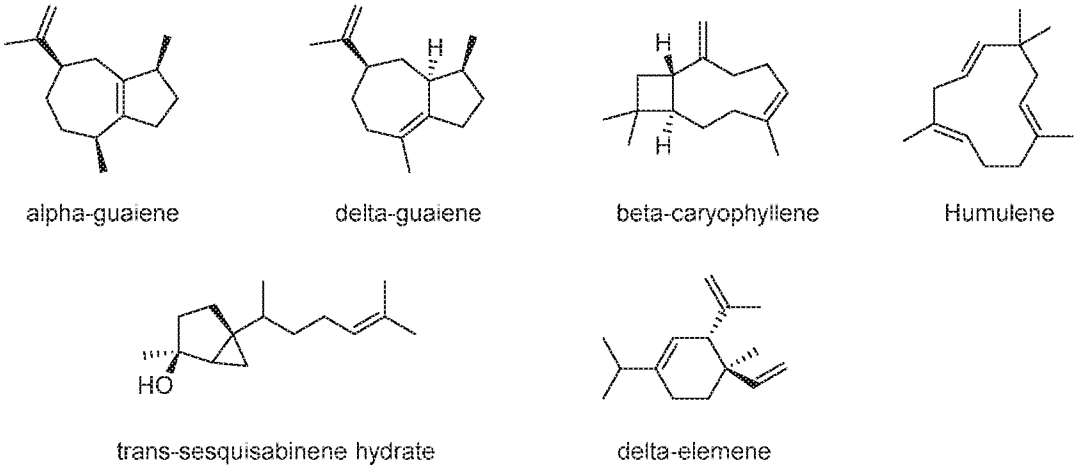

alpha-guaiene
delta-guaiene
beta-caryophyllene
Humulene
HO
trans-sesquisabinene hydrate
delta-elemene

FIG. 3

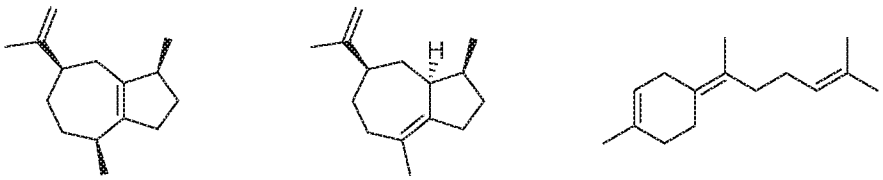

alpha-guaiene
delta-guaiene
gamma-bisabolene

FIG. 4

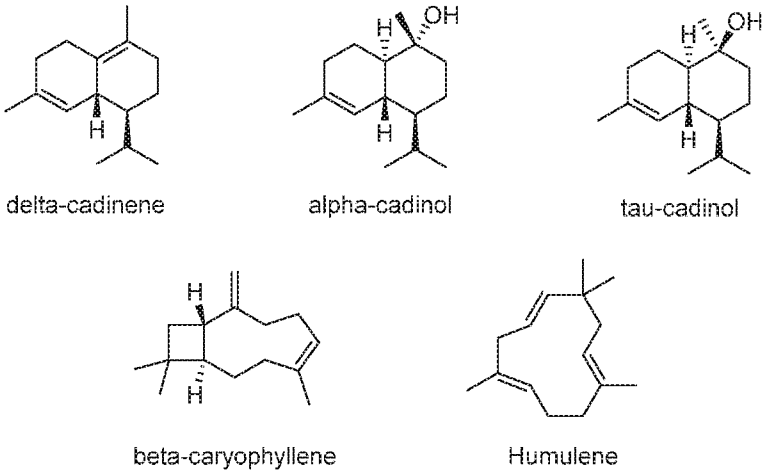
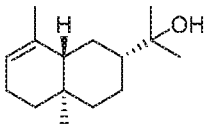
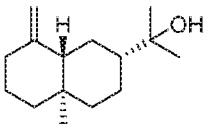

delta-cadinene
alpha-cadinol
tau-cadinol
beta-caryophyllene
Humulene

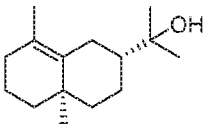
FIG. 5



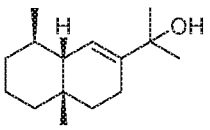
alpha-eudesmol



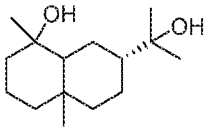
beta-eudesmol



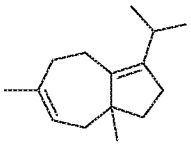
gamma-eudesmol



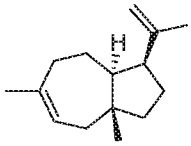
cis-eudesm-6-en-11-ol



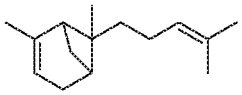
cryptomeridiol



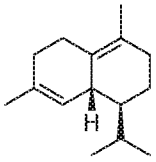
Daucene



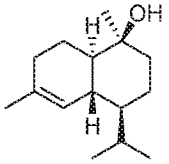
Isodaucene



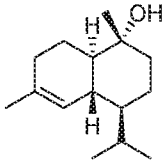
(alpha)-trans-bergamotene



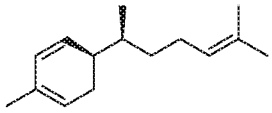
delta-cadinene



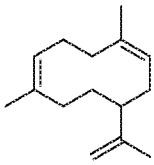
tau-cadinol



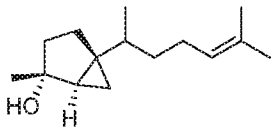
alpha-cadinol



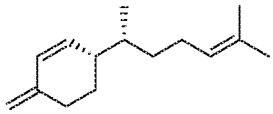
alpha-zingiberene



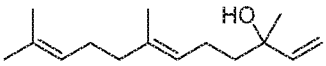
8-Isopropenyl-1,5-dimethyl-1,5-cyclodecadiene



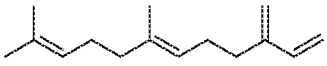
Sesquisabinene hydrate



beta-sesquiphellandrene



trans-nerolidol



beta-farnesene

FIG. 6

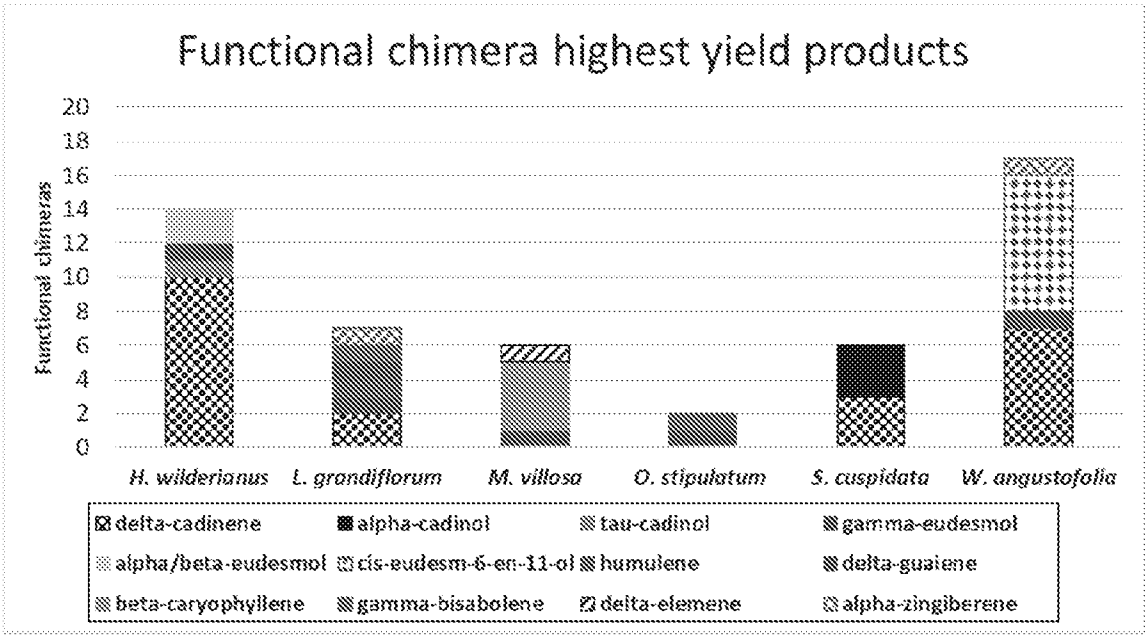

Functional chimera highest yield products
Functional chimeras
20
18
16
14
12
10
8
6
4
2
0
H. wilderianus
L. grandiflorum
M. villosa
O. stipulatum
S. cuspidata
W. angustofolia
delta-cadinene
alpha-cadinol
tau-cadinol
gamma-eudesmol
alpha/beta-eudesmol
cis-eudesm-6-en-11-ol
humulene
delta-guaiene
beta-caryophyllene
gamma-bisabolene
delta-elemene
alpha-zingiberene

FIG. 7

FIG. 8A

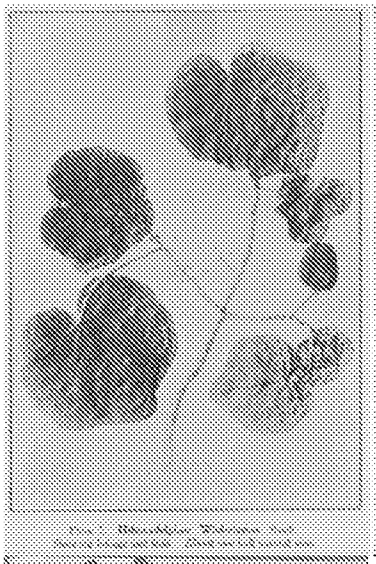


FIG. 8B



FIG. 8C



FIG. 8D



FIG. 8E

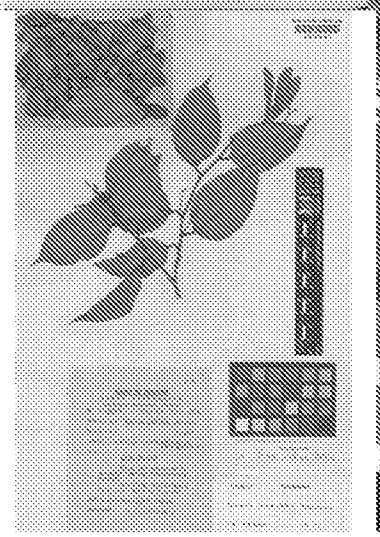
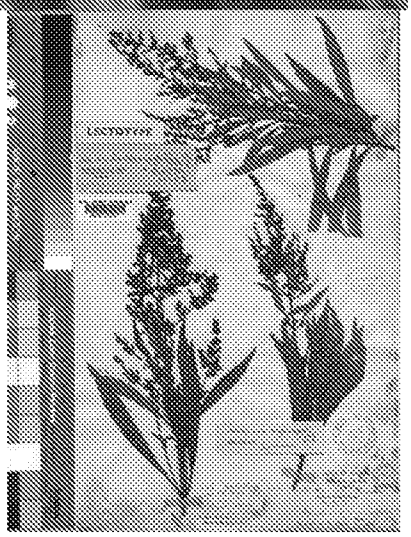


FIG. 8F



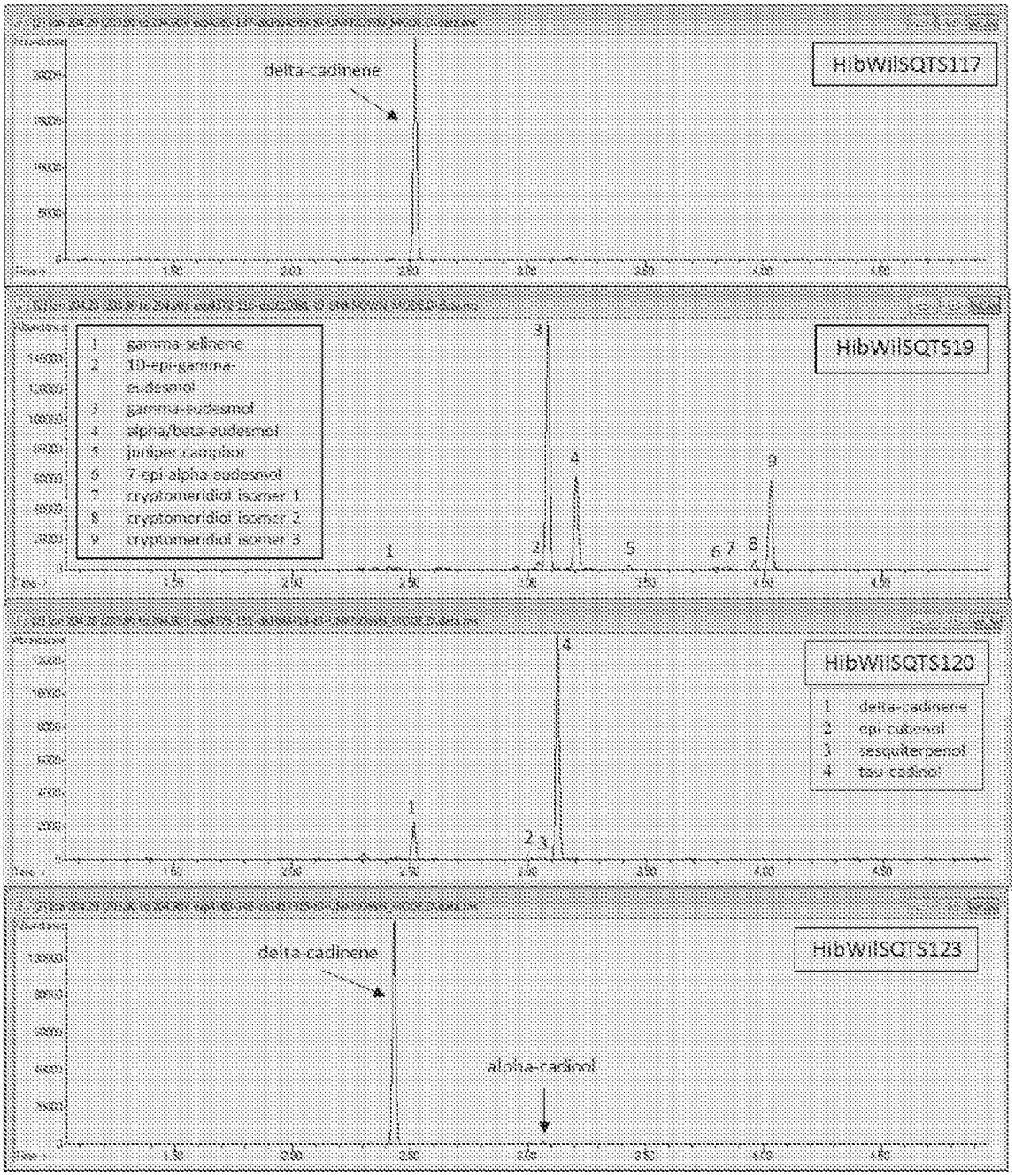

delta-cadinene
HibWilSQTS117
1 gamma-selinene
2 10-epi-gamma-eudesmol
3 gamma-eudesmol
4 alpha/beta-eudesmol
5 juniper camphor
6 7-epi-alpha-eudesmol
7 cryptomeridiol isomer 1
8 cryptomeridiol isomer 2
9 cryptomeridiol isomer 3
HibWilSQTS19
HibWilSQTS120
1 delta-cadinene
2 epi-cubenol
3 sesquiterpenol
4 tau-cadinol
delta-cadinene
alpha-cadinol
HibWilSQTS123

FIG. 9

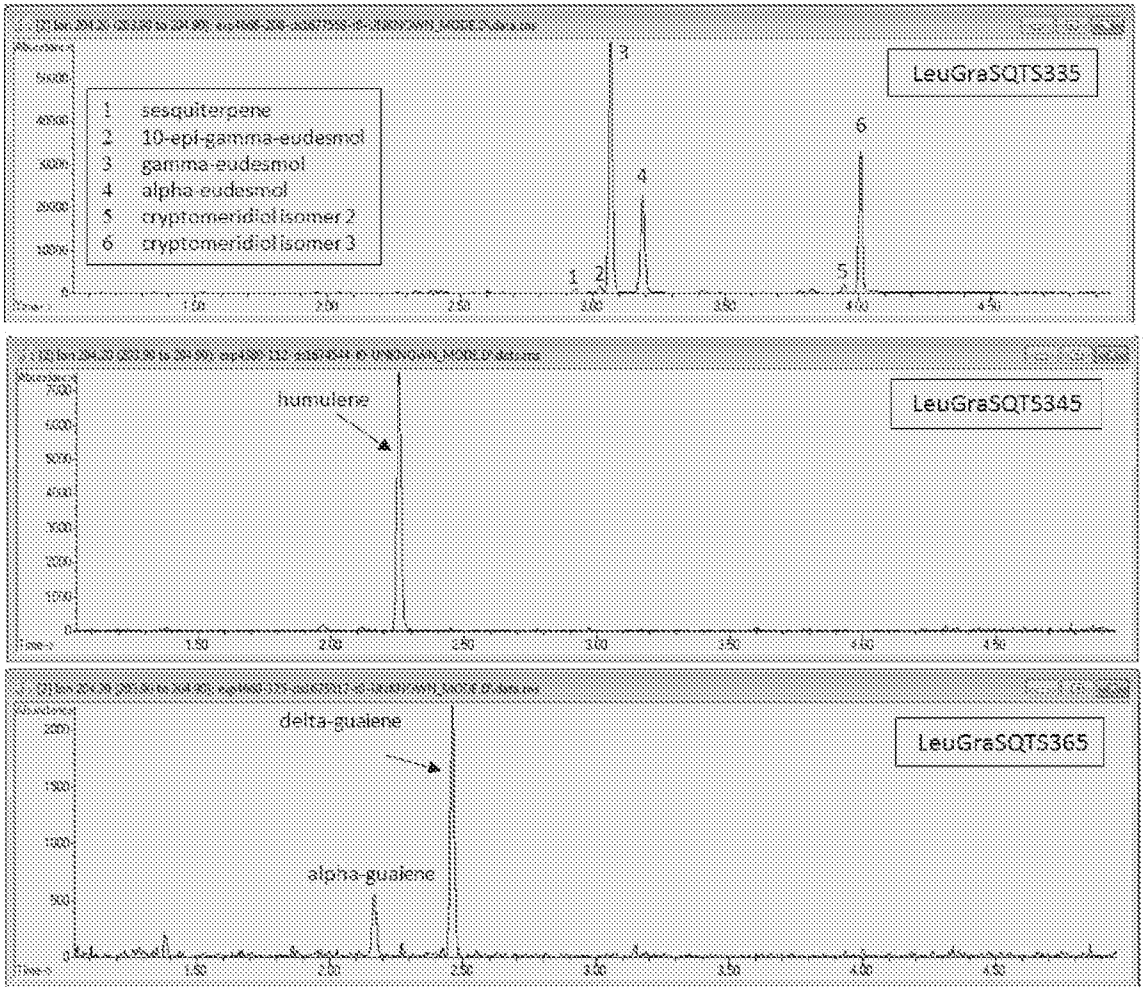

1 sesquiterpene
2 10-epi-gamma-eudesmol
3 gamma-eudesmol
4 alpha-eudesmol
5 cryptomeridiol isomer 2
6 cryptomeridiol isomer 3
LeuGraSQTS335
LeuGraSQTS345
humulene
LeuGraSQTS365
delta-guaiene
alpha-guaiene

FIG. 10

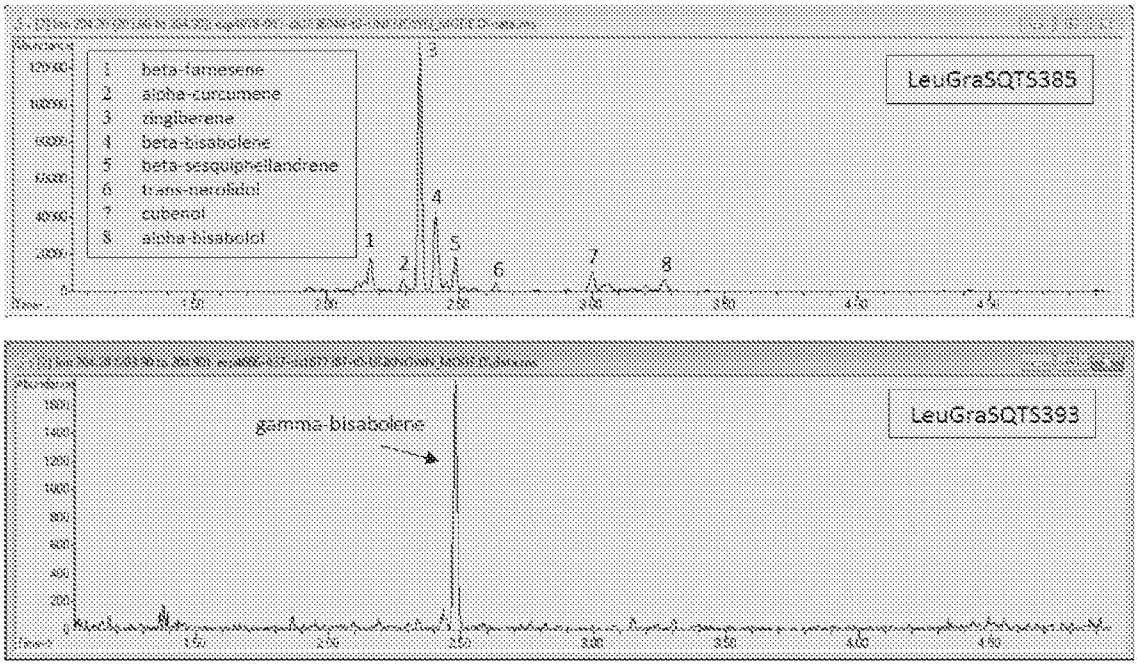

1 beta-farnesene
2 alpha-zurcumene
3 zingiberene
4 beta-bisabolene
5 beta-sesquiphellandrene
6 trans-nerolidol
7 cubenol
8 alpha-bisabolol
LeuGraSQTS385
gamma-bisabolene
LeuGraSQTS393

FIG. 11

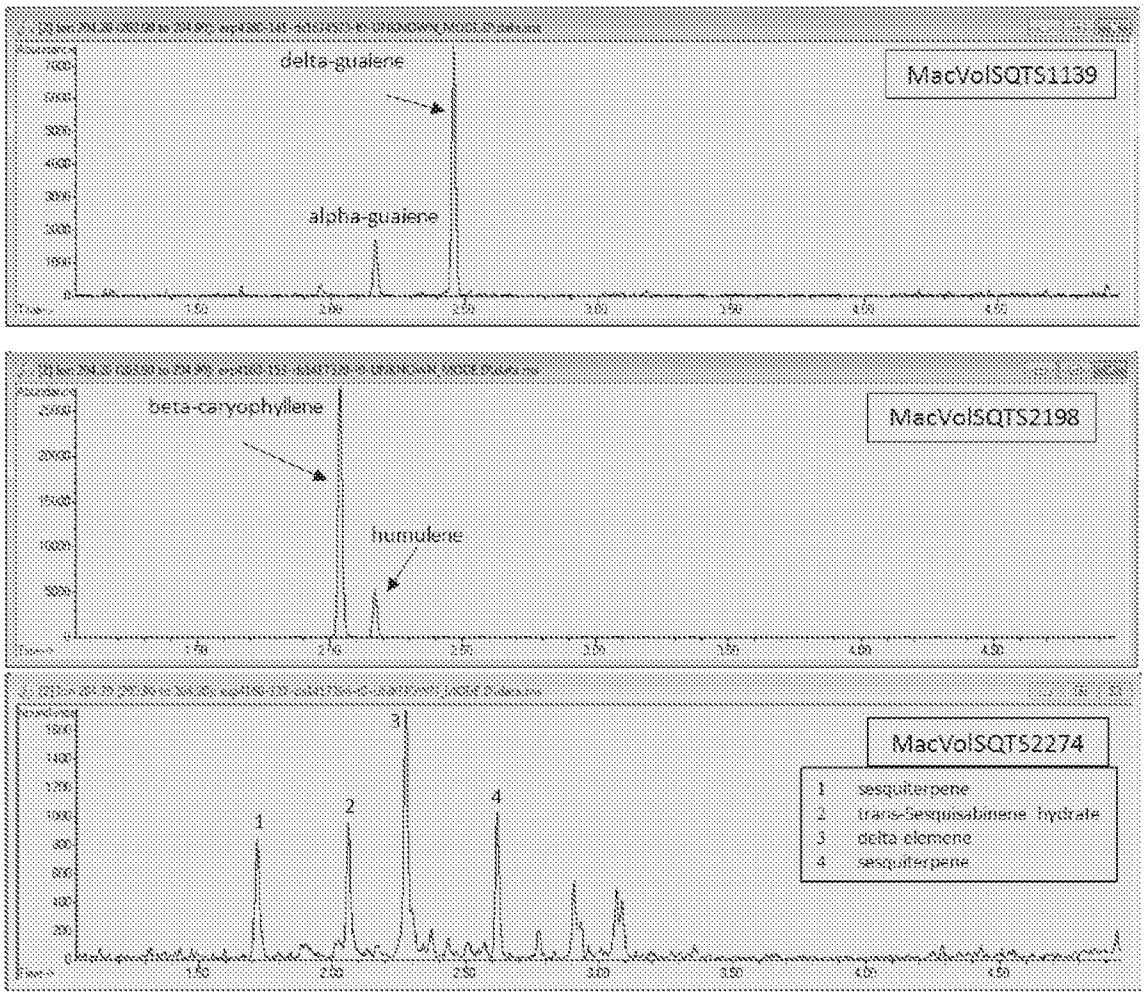

delta-guaiene
alpha-guaiene
MacVolSQTS1139
beta-caryophyllene
humulene
MacVolSQTS2198
1
2
3
4
MacVolSQTS2274
1 sesquiterpene
2 trans-Sesquisabinene hydrate
3 delta elemene
4 sesquiterpene

FIG. 12

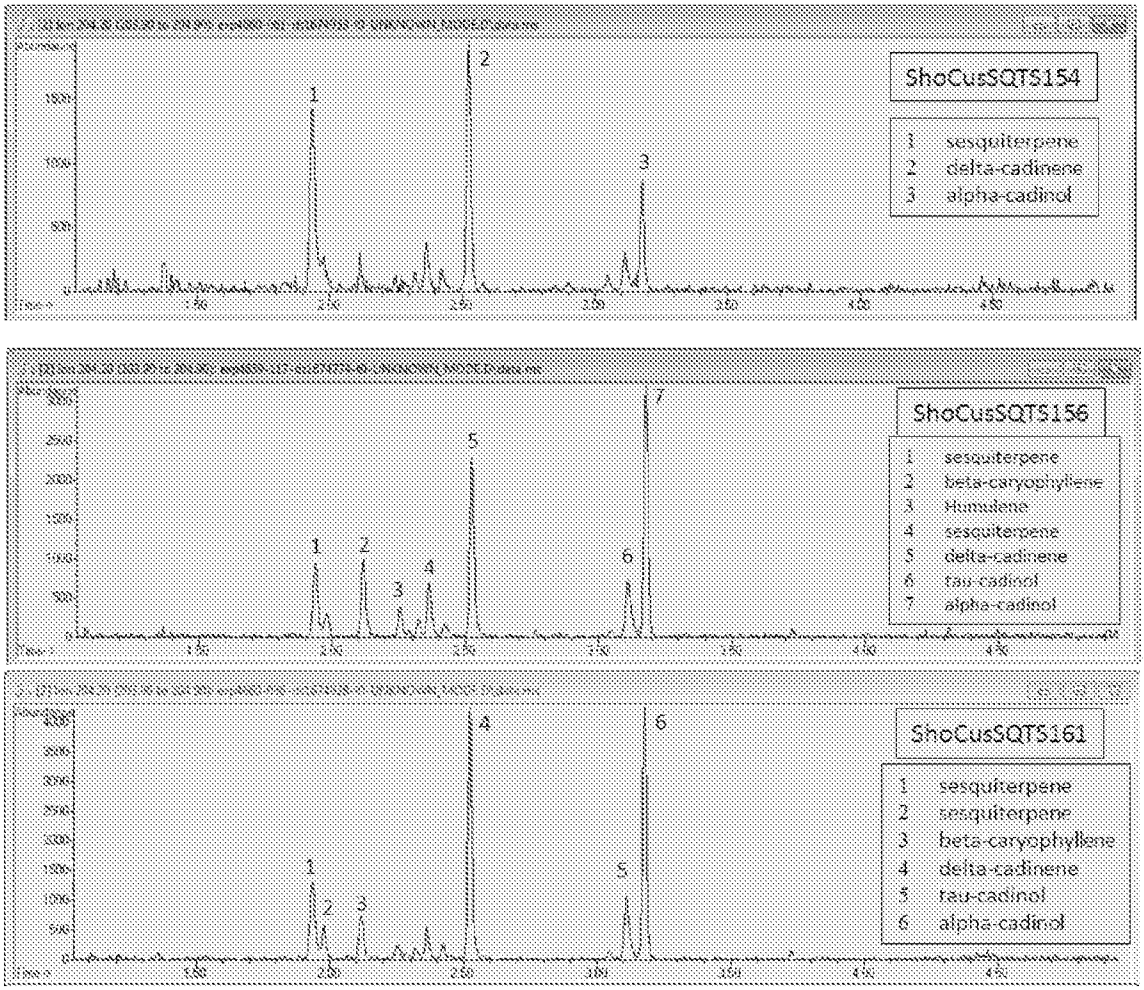

ShoCusSQTS154
1 sesquiterpene
2 delta-cadinene
3 alpha-cadinol
ShoCusSQTS156
1 sesquiterpene
2 beta-caryophyllene
3 Humulene
4 sesquiterpene
5 delta-cadinene
6 tau-cadinol
7 alpha-cadinol
ShoCusSQTS161
1 sesquiterpene
2 sesquiterpene
3 beta-caryophyllene
4 delta-cadinene
5 tau-cadinol
6 alpha-cadinol

FIG. 13

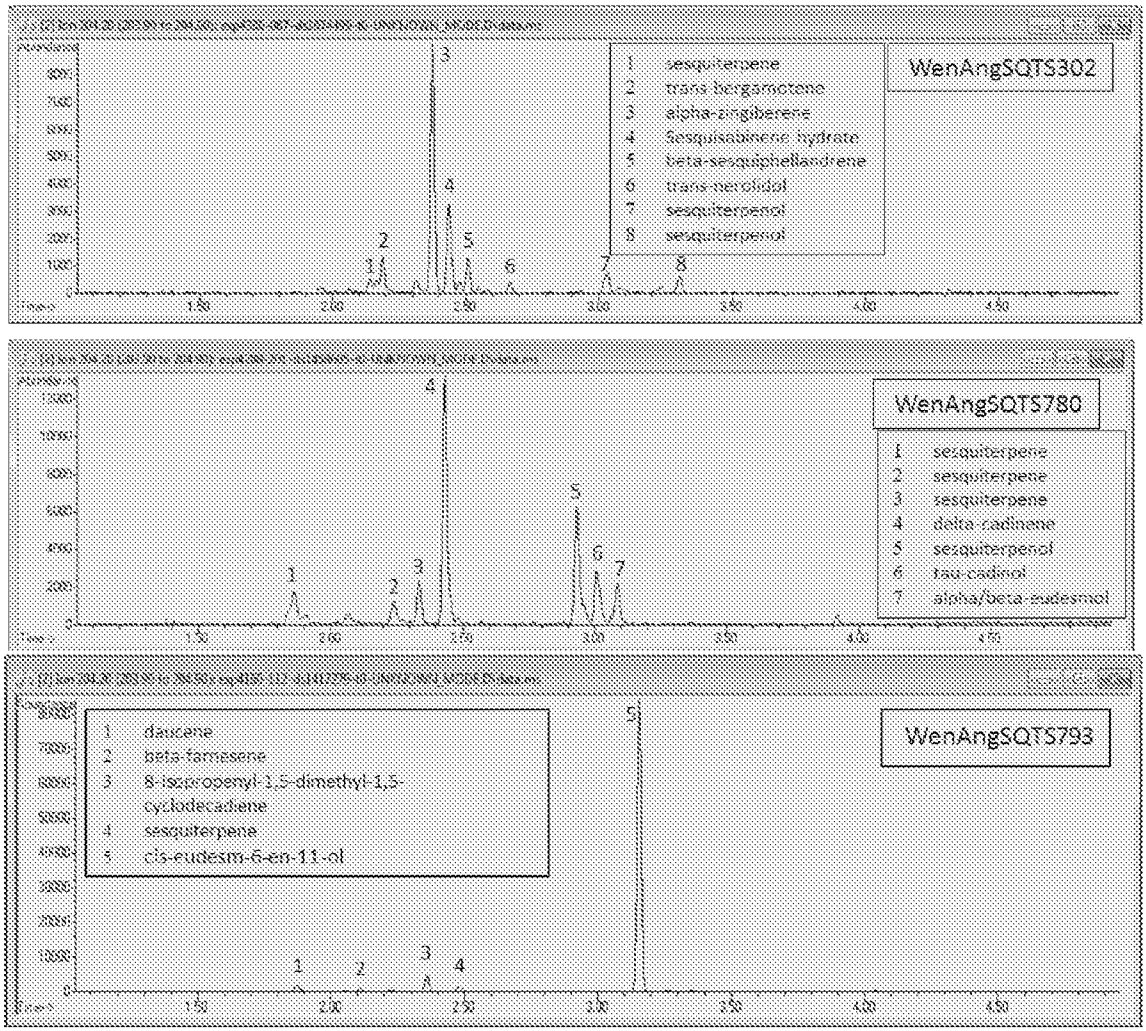

WenAngSQTS302
1 sesquiterpene
2 trans-bergamotene
3 alpha-zingiberene
4 Sesquisabinene hydrate
5 beta-sesquiphellandrene
6 trans-nerolidol
7 sesquiterpenol
8 sesquiterpenol
WenAngSQTS780
1 sesquiterpene
2 sesquiterpene
3 sesquiterpene
4 delta-cadinene
5 sesquiterpenol
6 tau-cadinol
7 alpha/beta-eudesmol
WenAngSQTS793
1 daucene
2 beta-farnesene
3 8-isopropenyl-1,5-dimethyl-1,5-cyclodecadiene
4 sesquiterpene
5 cis-eudesm-6-en-11-ol

FIG. 14

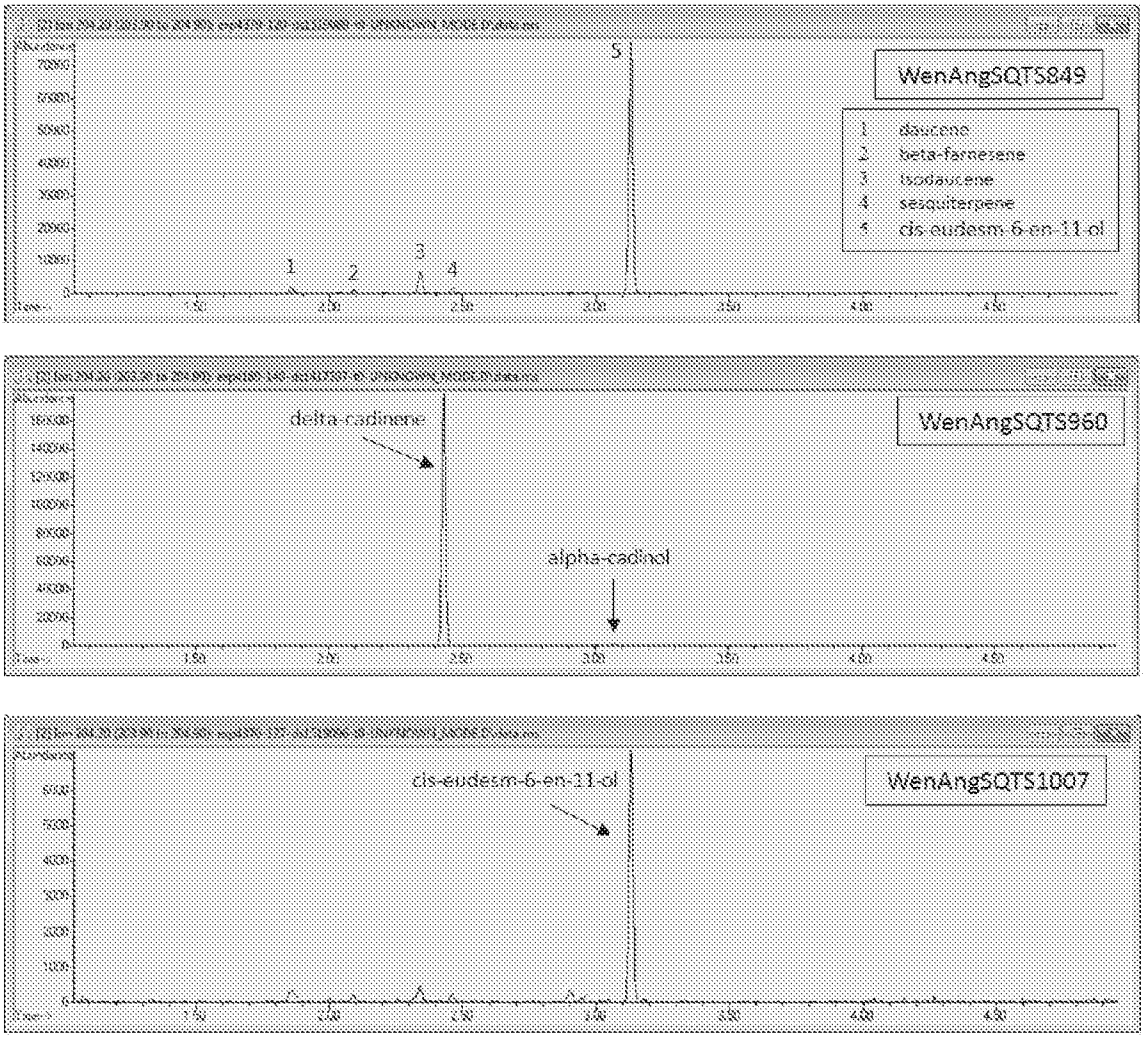

WenAngSQTS849
1 daucene
2 beta-farnesene
3 isodaucene
4 sesquiterpene
5 cis-eudesm-6-en-11-ol
1
2
3
4
5
WenAngSQTS960
delta-cadinene
alpha-cadinol
WenAngSQTS1007
cis-eudesm-6-en-11-ol

FIG. 15

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US2019/018122

Box No. II Observations where certain claims were found unsearchable (Continuation of item 2 of first sheet)

This international search report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

2. ☐ Claims Nos.:
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:

3. ☐ Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box No. III Observations where unity of invention is lacking (Continuation of item 3 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

see additional sheet

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.

2. ☐ As all searchable claims could be searched without effort justifying an additional fees, this Authority did not invite payment of additional fees.

3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:

4. ☒ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

1-20, 24-41(all partially)

Remark on Protest

☐ The additional search fees were accompanied by the applicant's protest and, where applicable, the payment of a protest fee.

☐ The additional search fees were accompanied by the applicant's protest but the applicable protest fee was not paid within the time limit specified in the invitation.

☐ No protest accompanied the payment of additional search fees.

INTERNATIONAL SEARCH REPORT

International application No
PCT/US2019/018122

A. CLASSIFICATION OF SUBJECT MATTER
INV. C12P5/00 C12P1/02 C12P1/04 C11D3/50
ADD.

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
C12P C11D

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
EPO-Internal, Sequence Search, EMBASE, FSTA, WPI Data

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	US 2012/246767 A1 (AMICK JEAN DAVIN [US] ET AL) 27 September 2012 (2012-09-27)	1,3-14, 18-20, 24-27, 29-41
Y	par.6, 7, 276-278, 495-518 par.286, 290, 525; sequence 403 & DATABASE Geneseq [Online] 21 June 2012 (2012-06-21), "Gossypium arboreum (+)-delta-cadinene synthase, SEQ ID 403.", retrieved from EBI accession no. GSP:AZV95121 Database accession no. AZV95121 the whole document	2,15-17, 28

-/--

[X] Further documents are listed in the continuation of Box C.

[X] See patent family annex.

* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
18 April 2019	03/07/2019

Name and mailing address of the ISA/	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Fax: (+31-70) 340-3016	Bonello, Steve

2

Form PCT/ISA/210 (second sheet) (April 2005)

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	DATABASE Geneseq [Online] 15 June 2007 (2007-06-15), "Cotton cadinene synthase.", XP002790773, retrieved from EBI accession no. GSP:ADY58798 Database accession no. ADY58798	1,3,7, 13, 18-20, 26,27, 29,30
Y	the whole document	2,15-17

X	DATABASE Geneseq [Online] 28 January 2016 (2016-01-28), "Gossypium raimondii agronomic improvement-related protein, SEQ:6690.", XP002790778, retrieved from EBI accession no. GSP:BCJ75688 Database accession no. BCJ75688	1,3,7, 13, 18-20, 26,27, 29,30
Y	the whole document	2,15-17

X	US 2008/268500 A1 (SCHALK MICHEL [FR]) 30 October 2008 (2008-10-30)	26,27, 29,30
Y	par.15-18	2,15-17

Y	US 2012/196315 A1 (ZHAO LISHAN [US] ET AL) 2 August 2012 (2012-08-02) par.9,163	2,15-17

X	WO 2017/050971 A1 (GIVAUDAN SA [CH]) 30 March 2017 (2017-03-30)	26,27
Y	p.1 l.9-14	28

A	WO 2014/206412 A1 (KØBENHAVNS UNI [DK]; FIRMENICH & CIE [CH]) 31 December 2014 (2014-12-31) p.3 l.25, p.5 l.5-20, Fig.1, 2, p.24 l.25	29,30

A	DATABASE JPO Proteins [Online] 23 December 2015 (2015-12-23), "JP 2015165798-A/747: MODIFIED VALENCENE SYNTHASE POLYPEPTIDES, ENCODING NUCLEIC ACID MOLECULES AND USES THEREOF.", XP002790775, retrieved from EBI accession no. JPOP:DM615049 Database accession no. DM615049 the whole document	1

2

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No
PCT/US2019/018122

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
US 2012246767 A1	27-09-2012	AU 2011320127 A1	02-05-2013
		BR 112013010244 A2	05-07-2016
		CA 2815829 A1	03-05-2012
		CN 103282492 A	04-09-2013
		EP 2633042 A1	04-09-2013
		JP 2013544084 A	12-12-2013
		JP 2015165798 A	24-09-2015
		US 2012246767 A1	27-09-2012
		WO 2012058636 A1	03-05-2012
US 2008268500 A1	30-10-2008	CN 101194013 A	04-06-2008
		EP 1893753 A2	05-03-2008
		EP 2311940 A1	20-04-2011
		JP 5430929 B2	05-03-2014
		JP 2008546384 A	25-12-2008
		US 2008268500 A1	30-10-2008
		WO 2006134523 A2	21-12-2006
US 2012196315 A1	02-08-2012	AU 2012212292 A1	15-08-2013
		CA 2826554 A1	09-08-2012
		CN 103608454 A	26-02-2014
		DK 2670846 T3	07-12-2015
		EP 2670846 A1	11-12-2013
		JP 5580488 B2	27-08-2014
		JP 2014505481 A	06-03-2014
		KR 20130110226 A	08-10-2013
		MX 346518 B	23-03-2017
		SG 192195 A1	30-09-2013
		US 2012196315 A1	02-08-2012
		US 2012270260 A1	25-10-2012
		WO 2012106405 A1	09-08-2012
		ZA 201305797 B	30-04-2014
WO 2017050971 A1	30-03-2017	CN 108138081 A	08-06-2018
		EP 3353272 A1	01-08-2018
		JP 2018529817 A	11-10-2018
		US 2018187123 A1	05-07-2018
		WO 2017050971 A1	30-03-2017
WO 2014206412 A1	31-12-2014	US 2016222401 A1	04-08-2016
		WO 2014206412 A1	31-12-2014

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

This International Searching Authority found multiple (groups of) inventions in this international application, as follows:

1. claims: 1-20, 24-41(all partially)

Directed towards a chimeric terpene synthase comprising an amino acid sequence at least 90% identical to an amino acid sequence consisting of SEQ ID No.1.

2-16. claims: 1-20, 24-41(all partially)

Directed towards a chimeric terpene synthase comprising an amino acid sequence at least 90% identical to an amino acid sequence consisting of SEQ ID No's 2-16, respectively.

---

17. claims: 1-41(partially)

Directed towards a chimeric terpene synthase comprising an amino acid sequence at least 90% identical to an amino acid sequence consisting of SEQ ID No. 17.

18-21. claims: 1-20, 24-41(all partially)

Directed towards a chimeric terpene synthase comprising an amino acid sequence at least 90% identical to an amino acid sequence consisting of SEQ ID No's 18-21, respectively.

---

22. claims: 1-41(partially)

Directed towards a chimeric terpene synthase comprising an amino acid sequence at least 90% identical to an amino acid sequence consisting of SEQ ID No. 22.

23-28. claims: 1-20, 24-41(all partially)

Directed towards a chimeric terpene synthase comprising an amino acid sequence at least 90% identical to an amino acid sequence consisting of SEQ ID No's 23-28, respectively.

---

29. claims: 1-41(partially)

Directed towards a chimeric terpene synthase comprising an amino acid sequence at least 90% identical to an amino acid sequence consisting of SEQ ID No. 29.

30-52. claims: 1-20, 24-41(all partially)

Directed towards a chimeric terpene synthase comprising an

amino acid sequence at least 90% identical to an amino acid sequence consisting of SEQ ID No's 30-52, respectively.
