Downloaded from orbit.dtu.dk on: Apr 09, 2024

DTU Library

Methods for production of strictosidine aglycone and monoterpenoid indole alkaloids

Jensen, Michael Krogh; Keasling, Jay D; Zang, Jie; Hansen, Lea Gram

Publication date:
2020

Document Version
Publisher's PDF, also known as Version of record

Link back to DTU Orbit

Citation (APA):
Jensen, M. K., Keasling, J. D., Zang, J., & Hansen, L. G. (2020). Methods for production of strictosidine aglycone and monoterpenoid indole alkaloids. (Patent No. *WO2020229516*).

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
19 November 2020 (19.11.2020)

WIPO | PCT

(10) International Publication Number
WO 2020/229516 A1

(51) International Patent Classification:
C12P 19/44 (2006.01) *C12P 17/18* (2006.01)
C12N 9/42 (2006.01)

(21) International Application Number:
PCT/EP2020/063283

(22) International Filing Date:
13 May 2020 (13.05.2020)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
62/846,820 13 May 2019 (13.05.2019) US
19175969.5 22 May 2019 (22.05.2019) EP

(71) Applicant: **DANMARKS TEKNISKE UNIVERSITET** [DK/DK]; Anker Engelunds Vej 101 A, 2800 Kgs. Lyngby (DK).

(72) Inventors: **JENSEN, Michael Krogh**; Hellebækgade 15, st., 2200 Copenhagen N (DK). **KEASLING, Jay D.**; 828 Contra Costa Ave., Berkeley, CA 94707 (US). **ZANG, Jie**; Højbovej 7A, 3450 Birkerød (DK). **HANSEN, Lea Gram**; Elmelundevej 23, 2.tv, 2700 Brønshøj (DK).

(74) Agent: **HØIBERG P/S**; Adelgade 12, 1304 Copenhagen K (DK).

(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, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, WS, 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).

(54) Title: METHODS FOR PRODUCTION OF STRICTOSIDINE AGLYCONE AND MONOTERPENOID INDOLE ALKALOIDS

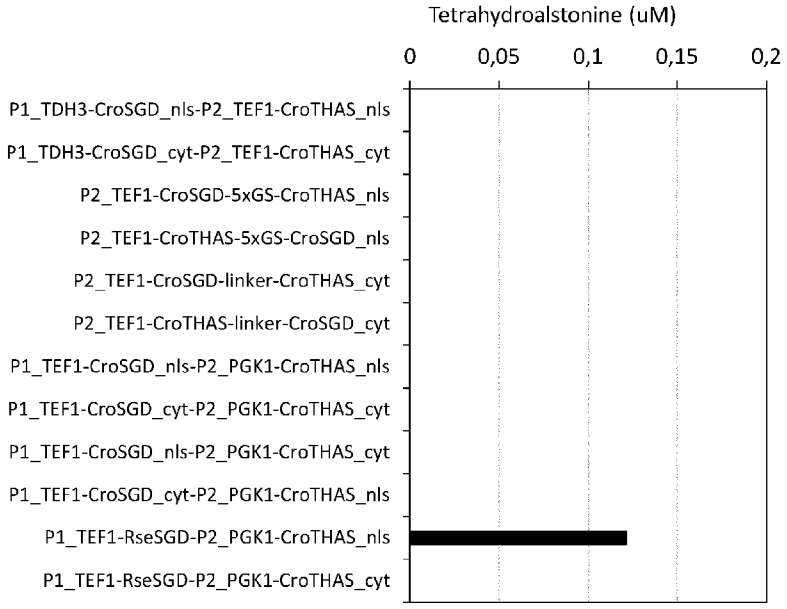

Fig. 1

(57) Abstract: Herein are provided microbial factories, in particular yeast factories, for production of strictosidine aglycone and optionally other plant-derived compounds. Also provided are methods for producing strictosidine aglycone in a microorganism, as well as useful nucleic acids, vectors and host cells.

WO 2020/229516 A1

[Continued on next page]

Declarations under Rule 4.17:

— *of inventorship (Rule 4.17(iv))*

Published:

— *with international search report (Art. 21(3))*

— *with sequence listing part of description (Rule 5.2(a))*

Methods for production of strictosidine aglycone and monoterpenoid indole alkaloids

Technical field

The present invention relates to microbial factories, such as microorganism factories in particular yeast factories and bacterial factories, for production of strictosidine aglycone and optionally other plant-derived compounds. Also provided are methods for producing strictosidine aglycone in a microorganism, as well as useful nucleic acids, vectors and host cells.

Background

Plants produce some of the most potent human therapeutics and have been used for millennia to treat illnesses. Despite the large repertoire of plant-derived pharmaceuticals, most of these products do not make it to the market because they are found in minute quantities in plants, they are difficult to extract, and there is limited knowledge about their biosynthetic pathways.

Furthermore, sourcing plant-derived pharmaceuticals based on plant-based extraction threatens to cause species extinction. New regulatory laws seek to create conditions to promote biodiversity conservation and sustainable use of genetic resources, which in the short term are expected to further affect the supply chains of many valuable plant natural products.

Moreover, many plant species are not readily genetically manipulated, and synthetic chemistry holds little promise for bulk production of complex plant-derived therapeutics. Together, supporting a need for refactored biosynthesis of new and existing pharmaceuticals, in genetically tractable and sustainable production hosts.

The monoterpenoid indole alkaloids (MIAs) are plant secondary metabolites that show a remarkable structural diversity and pharmaceutically valuable biological activities, such as anti-cancer and anti-psychosis properties. The productions of these alkaloids occurs through highly complicated pathways.

The common precursors for the different MIAs are strictosidine, and its deglycosylated form, strictosidine aglycone. Strictosidine is formed by the coupling of secologanin to

tryptamine in a reaction catalysed by the enzyme strictosidine synthase. Strictosidine alglycone is natively produced from hydrolyzing strictosidine by strictosidine-beta-glucosidase (SGD). Over 2,000 MIAs can be produced from strictosidine aglycone.

To enable a sustainable supply of therapeutic MIAs, researchers have for decades attempted to elucidate the biosynthetic pathways from MIA producing plants, including both the platform biosynthetic route to the common MIA precursor strictosidine and the anti-cancer drug vinblastine. Moreover, the platform biosynthetic route from geraniol to strictosidine, and the seven-step biosynthetic pathway from tabersonine to vindoline, the immediate precursor of vinblastine has also been refactored in yeast cell factories.

Current methods for production of strictosidine aglycone are mostly based on chemical synthesis or plant extraction. Such methods are not cost-effective and also have a significant impact on the environment. Therefore, methods for cost-effective and environmental-friendly production of strictosidine aglycone are required.

Summary

The invention concerns a microorganism capable of producing strictosidine aglycone and methods for strictosidine aglycone and monoterpenoid indole alkaloids (MIAs) production in a microorganism.

In one aspect is provided a microorganism capable of producing strictosidine aglycone, said microorganism expresses

a strictosidine-beta-glucosidase (SGD), capable of converting strictosidine to strictosidine aglycone,

wherein said SGD is a heterologous SGD selected from RseSGD (SEQ ID NO: 24), GseSGD (SEQ ID NO: 25), SapSGD (SEQ ID NO: 26), RveSGD (SEQ ID NO: 27), VmiSGD1 (SEQ ID NO: 47), AhuSGD (SEQ ID NO: 48), HimSGD2 (SEQ ID NO: 49), SinSGD (SEQ ID NO: 50), TelSGD (SEQ ID NO: 51), VunSGD (SEQ ID NO: 52), NsiSGD1 (SEQ ID NO: 53), LprSGD (SEQ ID NO: 54), AchSGD1 (SEQ ID NO: 55), HsuSGD (SEQ ID NO: 56), MroSGD (SEQ ID NO: 57), RseSGD2 (SEQ ID NO: 58), PgrSGD (SEQ ID NO: 59), OpuSGD (SEQ ID NO: 60), HpiSGD (SEQ ID NO: 61), HanSGD1 (SEQ ID NO: 62), AchSGD2 (SEQ ID NO: 63), HimSGD1 (SEQ ID NO: 64), IpeSGD (SEQ ID NO: 65), LsaSGD1 (SEQ ID NO: 66), or CarSGD (SEQ ID NO: 67) or

variants thereof having at least 70%, such as at least 80%, such as at least 90%, such as at least 91%, such as at least 92%, such as at least 93%, such as at least 94%, such as at least 95%, such as at least 96%, such as at least 97%, such as at least 98%, such as at least 99%, such as 100% identity thereto,

and/or;

wherein said SGD is a mosaic SGD, wherein said mosaic SGD comprises an amino acid sequence having the general formula

$$D_1\text{-}D_2\text{-}D_3\text{-}D_4$$

wherein D_1 is a first amino acid sequence from a first SGD,
wherein D_2 is a second amino acid sequence from a second SGD,
wherein D_3 is a third amino acid sequence comprising or consisting of amino acids of SEQ ID NO:91 or a variant thereof having at least 90% identity to SEQ ID NO: 91,
wherein D_4 is a fourth amino acid sequence from a fourth SGD or an amino acid sequence consisting of amino acids of SEQ ID NO:92 or a variant thereof having at least 90% identity to SEQ ID NO: 92,
wherein said first SGD, second SGD and fourth SGD can be the same or different, with the proviso that said first SGD, second SGD and fourth SGD are not all RseSGD.

Also provided herein are methods for producing strictosidine aglycone in a microorganism, comprising the steps of:

a) providing a microorganism, said cell expressing:
   a strictosidine-beta-glucosidase (SGD), capable of converting strictosidine to strictosidine aglycone;
b) incubating said microorganism in a medium comprising strictosidine or a substrate which can be converted to strictosidine by said microorganism;
c) optionally, recovering the strictosidine aglycone;
d) optionally, further converting the strictosidine aglycone to monoterpenoid indole alkaloids,

wherein said SGD is a heterologous SGD selected from RseSGD (SEQ ID NO: 24), GseSGD (SEQ ID NO: 25), SapSGD (SEQ ID NO: 26), RveSGD (SEQ ID NO: 27), VmiSGD1 (SEQ ID NO: 47), AhuSGD (SEQ ID NO: 48), HimSGD2 (SEQ ID NO: 49), SinSGD (SEQ ID NO: 50), TelSGD (SEQ ID NO: 51), VunSGD (SEQ ID NO: 52), NsiSGD1 (SEQ ID NO: 53), LprSGD (SEQ ID NO: 54), AchSGD1 (SEQ ID NO: 55),

HsuSGD (SEQ ID NO: 56), MroSGD (SEQ ID NO: 57), RseSGD2 (SEQ ID NO: 58), PgrSGD (SEQ ID NO: 59), OpuSGD (SEQ ID NO: 60), HpiSGD (SEQ ID NO: 61), HanSGD1 (SEQ ID NO: 62), AchSGD2 (SEQ ID NO: 63), HimSGD1 (SEQ ID NO: 64), IpeSGD (SEQ ID NO: 65), LsaSGD1 (SEQ ID NO: 66), or CarSGD (SEQ ID NO: 67) or variants thereof having at least 70%, such as at least 80%, such as at least 90%, such as at least 91%, such as at least 92%, such as at least 93%, such as at least 94%, such as at least 95%, such as at least 96%, such as at least 97%, such as at least 98%, such as at least 99%, such as 100% identity thereto,

and/or;

wherein said SGD is a mosaic SGD, wherein said mosaic SGD comprises an amino acid sequence having the general formula

$$D_1\text{-}D_2\text{-}D_3\text{-}D_4$$

wherein D_1 is a first amino acid sequence from a first SGD,
wherein D_2 is a second amino acid sequence from a second SGD,
wherein D_3 is a third amino acid sequence comprising or consisting of amino acids of SEQ ID NO:91 or a variant thereof having at least 90% identity to SEQ ID NO: 91,
wherein D_4 is a fourth amino acid sequence from a fourth SGD or an amino acid sequence consisting of amino acids of SEQ ID NO:92 or a variant thereof having at least 90% identity to SEQ ID NO: 92,
wherein said first SGD, second SGD and fourth SGD can be the same or different, with the proviso that said first SGD, second SGD and fourth SGD are not all RseSGD.

Also provided herein are nucleic acid constructs comprising a sequence identical to or having at least 90% identity, such as at least 91%, such as at least 92%, such as at least 93%, such as at least 94%, such as at least 95%, such as at least 96%, such as at least 97%, such as at least 98%, such as at least 99%, such as 100% identity to SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 3, SEQ ID NO: 4, SEQ ID NO:68, SEQ ID NO:69, SEQ ID NO:70, SEQ ID NO: 71, SEQ ID NO:72, SEQ ID NO: 73, SEQ ID NO:74, SEQ ID NO: 75, SEQ ID NO: 76, SEQ ID NO: 77, SEQ ID NO: 78, SEQ ID NO:79, SEQ ID NO:80, SEQ ID NO:81, SEQ ID NO:82, SEQ ID NO:83, SEQ ID NO:84, SEQ ID NO:85, SEQ ID NO:86, SEQ ID NO:87, SEQ ID NO:88, SEQ ID NO:100, SEQ ID NO:101, SEQ ID NO:102, SEQ ID NO:103, SEQ ID NO:104, SEQ ID NO:105, SEQ ID NO:106 and/or SEQ ID NO:107.

Also provided are vectors comprising the above nucleic acids, as well as host cells comprising said vectors and/or said nucleic acids.

Also provided is a kit of parts comprising a microorganism as described herein, and/or nucleic acid constructs as described herein, and/ or a vector as described herein, and instructions for use.

Also provided is the use of above nucleic acids, vectors or host cells for the production of strictosidine aglycone.

Also provided herein are methods for producing monoterpenoid indole alkaloids (MIAs) in a microorganism, said method comprising the steps of:

a) providing a microorganism capable of converting strictosidine aglycone to tabersonine and/or catharanthine, said cell expressing:
   optionally, a strictosidine synthase (STR);
   a strictosidine-beta-glucosidase (SGD);
   a NADPH--cytochrome P450 reductase (CPR);
   a Cytochrome b5 (CYB5);
   a Geissoschizine synthase (GS);
   a Geissoschizine oxidase (GO);
   a Redox1;
   a Redox2;
   a Stemmadenine O-acetyltransferase (SAT);
   a O-acetylstemmadenine oxidase (PAS);
   a Dehydroprecondylocarpine acetate synthase (DPAS);
   a Tabersonine synthase (TS); and/or
   a Catharanthine synthase (CS);
b) incubating said microorganism in a medium comprising strictosidine or a substrate which can be converted to strictosidine by said microorganism;
c) optionally, recovering the MIAs;
d) optionally, processing the MIAs into a pharmaceutical compound,

wherein said SGD is a heterologous SGD selected from RseSGD (SEQ ID NO: 24), GseSGD (SEQ ID NO: 25), SapSGD (SEQ ID NO: 26), RveSGD (SEQ ID NO: 27), VmiSGD1 (SEQ ID NO: 47), AhuSGD (SEQ ID NO: 48), HimSGD2 (SEQ ID NO: 49),

SinSGD (SEQ ID NO: 50), TelSGD (SEQ ID NO: 51), VunSGD (SEQ ID NO: 52), NsiSGD1 (SEQ ID NO: 53), LprSGD (SEQ ID NO: 54), AchSGD1 (SEQ ID NO: 55), HsuSGD (SEQ ID NO: 56), MroSGD (SEQ ID NO: 57), RseSGD2 (SEQ ID NO: 58), PgrSGD (SEQ ID NO: 59), OpuSGD (SEQ ID NO: 60), HpiSGD (SEQ ID NO: 61), HanSGD1 (SEQ ID NO: 62), AchSGD2 (SEQ ID NO: 63), HimSGD1 (SEQ ID NO: 64), IpeSGD (SEQ ID NO: 65), LsaSGD1 (SEQ ID NO: 66), or CarSGD (SEQ ID NO: 67) or variants thereof having at least 70%, such as at least 80%, such as at least 90%, such as at least 91%, such as at least 92%, such as at least 93%, such as at least 94%, such as at least 95%, such as at least 96%, such as at least 97%, such as at least 98%, such as at least 99%, such as 100% identity thereto,

and/or;

wherein said SGD is a mosaic SGD, wherein said mosaic SGD comprises an amino acid sequence having the general formula

$$D_1\text{-}D_2\text{-}D_3\text{-}D_4$$

wherein D_1 is a first amino acid sequence from a first SGD,
wherein D_2 is a second amino acid sequence from a second SGD,
wherein D_3 is a third amino acid sequence consisting of amino acids of SEQ ID NO:91 or a variant thereof having at least 90% identity to SEQ ID NO: 91,
wherein D_4 is a fourth amino acid sequence from a fourth SGD or an amino acid sequence consisting of amino acids of SEQ ID NO:92 or a variant thereof having at least 90% identity to SEQ ID NO: 92,
wherein said first SGD, second SGD and fourth SGD can be the same or different, with the proviso that said first SGD, second SGD and fourth SGD are not all RseSGD.
Also provided herein are strictosidine aglycone, tetrahydroalstonine, heteroyohimbine, rabersonine and/or catharanthine obtained by the method as described herein.

Also provided herein are methods for treating a disorder such as a cancer, arrhythmia, malaria, psychotic diseases, hypertension, depression, Alzheimer's disease, addiction and/or neuronal diseases, comprising administration of a therapeutic sufficient amount of an MIA or a pharmaceutical compound obtained by the as described herein.

Description of Drawings

Figure 1: High-resolution analytical results of tetrahydroalstonine (THA) obtained from LC-MS analysis of yeast cells (*Saccharomyces cerevisiae*) expressing SGD derived from *Catharanthus roseus* (CroSGD) alone and in various tagged and CroSGD-fusion versions, as well as SGD from *Rauvolfia serpentina* (RseSGD).

Figure 2: Sequence identity among SGD derived from *Catharanthus roseus (Cro*SGD*), Rauvolfia serpentina (Rse*SGD*), Rauvolfia verticillata (Rve*SGD*), Gelsemium sempervirens (Gse*SGD*), Camptotheca acuminate (Cac*SGD*), Scedosporium apiospermum (Sap*SGD*), Uncaria tomentosa (Uto*SGD*) and Glycine soja* (*GsoSGD).* The eight protein sequences were aligned with the t-Coffee web server.

Figure 3: Biosynthesis of the heteroyohimbine tetrahydroalstonine measured on LC-MS. The production of tetrahydroalstonine (THA) was measured in yeast strains expressing either GsoSGD, CacSGD, CroSGD, UtoSGD, GseSGD, SapSGD, RveSGD or RseSGD The yeast strain GsoSGD was used as a negative control. The p-value represents comparison between the negative control (GsoSGD) and CacSGD, CroSGD or UtoSGD, respectively.

Figure 4: GFP-tagged CroSGD and RseSGD localization in yeast. A) A yeast cell expressing GFP-CroSGD. B) A yeast cell expressing GFP-RseSGD. The arrows mark the localization of SGD in the yeast cells.

Figure 5: The biosynthesis of the heteroyohimbine alstonine in yeast cell factories, expressing RseSGD, CroTHAS and GseSBE, is shown in triplicates in figure 5. Alastonine was measured by Orbitrap Fusion™ Tribrid™ MS.

Figure 6: The yeast strain MIA-DC was feed with 0.1 mM of secologanine and 1 mM of tryptamine and the production of tabersonine and catharanthine were measured by LC-MS. A) Catharanthine production, B) Tabersonine production, C) Catharanthine standard, and D) Tabersonine standard.

Figure 7: The yeast strain MIA-DC was feed with 0.1 mM of secologanine and 1 mM of tryptamine and the concentration levels of tabersonine and catharanthine in MIA-DC and MIA-DA (control) were measured by LC-MS.

Figure 8: Biosynthesis of the heteroyohimbine tetrahydroalstonine measured on LC-MS. The production of tetrahydroalstonine (THA) was measured in yeast strains expressing either CroSGD, VmiSGD1, AhuSGD, HimSGD2, SinSGD, TelSGD, VunSGD, NsiSGD1, LprSGD, AchSGD1, HsuSGD, MroSGD, RseSGD2, PgrSGD, OpuSGD, HpiSGD, HanSGD1, AchSGD2, HimSGD1, IpeSGD, LsaSGD1, CarSGD, OeuSGD, AchSGD3, CmaSGD, MmySGD, VmiSGD3, IniSGD, or NsiSGD2. The *p*-value represents a comparison between the negative control (CroSGD) and OeuSGD, AchSGD3, CmaSGD, MmySGD, VmiSGD3, IniSGD, and NsiSGD2.

Figure 9: Biosynthesis of the heteroyohimbine tetrahydroalstonine measured on LC-MS. The production of tetrahydroalstonine (THA) was measured in yeast strains expressing one of the mosaic SGDs: RRCC-SGD, RCCC-SGD, CCCC-SGD, CRCC-SGD, CRCR-SGD, RRCR-SGD, CCCR-SGD, RCCR-SGD, CRRC-SGD, RRRC-SGD, RCRC-SGD, CCRC-RGD, RCRR-SGD, CRRR-SGD, RRRR-SGD, and CCRR-SGD. CCCC-SGD and RRRR-SGD are identical to the two wild type sequences CroSGD and RseSGD. The *p*-value represents comparisons between the negative control (CCCC-SGD/CroSGD) and all SGDs containing CroSGD domain 3: RRCC-SGD, RCCC-SGD, CRCC-SGD, CRCR-SGD, RRCR-SGD, CCCR-SGD and RCCR-SGD. The color indicates the identity of domain 3 and 4: Light grey – RseSGD domain 3 & 4, medium grey – RseSGD domain 3 & CroSGD domain 4, dark grey – CroSGD domain 3 & CroSGD/RseSGD domain 4.

Figure 10: Biosynthesis of the heteroyohimbine tetrahydroalstonine measured on LC-MS. The production of tetrahydroalstonine (THA) was measured in yeast strains expressing one of the wild type SGDs (UtoSGD, GseSGD, CroSGD, or RveSGD) or one of the engineered SGDs (UURR-SGD, GGRR-SGD, CCRR-SGD, or VVRR-SGD).

Figure 11: Biosynthesis of the common MIA precursor strictosidine (A) and heteroyohimbine tetrahydroalstonine (B) in *E. coli* measures by LC-MS. The production of strictosidine and tetrahydroalstonine were measures in bacterial strains expressing either CroSGD or RseSGD. A strain with an empty expression vector was included as a negative control.

Figure 12: Multiple sequence alignment of SGDs proteins derived from *Catharanthus roseus* (*Cro*SGD), *Rauvolfia serpentina* (*Rse*SGD and *Rse*SGD2), *Rauvolfia verticillata* (*Rve*SGD), *Gelsemium sempervirens* (*Gse*SGD), *Camptotheca acuminate* (*Cac*SGD), *Scedosporium apiospermum* (*Sap*SGD), *Uncaria tomentosa* (*Uto*SGD), *Glycine soja* (*GsoSGD*), *Vinca minor* (*Vmi*SGD1 and *Vmi*SGD3), *Tabernaemontana elegans* (*Tel*SGD), *Amsonia hubrichtii* (*Ahu*SGD), *Ophiorrhiza pumila*, (*Opu*SGD), *Nyssa sinensis*, (*Nsi*SGD1 and *Nsi*SGD2), *Coffea arabica* (*Car*SGD), *Carapichea ipecacuanha* (*Ipe*SGD), *Handroanthus impetiginosus* (*Him*SGD2 and *Him*SGD1), *Sesamum indicum* (*Sin*SGD), *Olea europaea* (*Oeu*SGD), *Actinidia chinensis var. chinensis* (*Ach*SGD1, *Ach*SGD2 and *Ach*SGD3), *Helianthus annuus* (*Han*SGD), *Lactuca sativa* (*Lse*SGD), *Ipomoea nil* (*Ini*SGD), *Chelidonium majus* (*Cma*SGD), *Vigna unguiculata* (*Vun*SGD), *Heliocybe sulcate* (*Hsu*SGD), *Pyricularia grisea* (*Pgr*SGD), *Lomentospora prolificans* (*Lpr*SGD), *Hydnomerulius pinastri MD-312* (*Hpi*SGD), *Madurella mycetomatis* (*Mmy*SGD), and *Moniliophthora roreri MCA 2997* (*Mro*SGD). The protein sequences were aligned with the t-Coffee web server.

Figure 13: Pairwise sequence identities among the 36 SGD protein sequences aligned in figure 8. The pairwise sequence identities were calculated from the alignment with CLC Main Workbench 8.

Detailed description

The present disclosure relates to microorganisms and method for production of strictosidine aglycone and monoterpenoid indole alkaloids (MIA).

The microorganism may be any non-natural or natural microorganism. By non-natural is meant an engineered microorganism, which comprises one or more genes which are not native to the microorganism. In some aspects of the present invention the microorganism expresses a heterologous SGD, mosaic SGD or variants thereof.

Microorganisms are microscopic organisms that exist as unicellular, multicellular, or cell clusters. Microorganism may be divided into different types such as bacteria, archaea, yeasts, fungi, protozoa, algae, and viruses. Thus, in one embodiment, the microorganism is selected from the group consisting of bacteria, archaea, yeasts, fungi, protozoa, algae, and viruses. In another embodiment, the microorganism is selected from the group consisting of bacteria, archaea, yeasts, fungi, protozoa and algae. In another embodiment, the microorganism is selected from the group consisting of

bacteria, archaea, yeasts, fungi, and algae. In another embodiment, the microorganism is selected from the group consisting of bacteria, archaea yeasts and fungi. In another embodiment, the microorganism is selected from bacteria, yeasts and fungi. In another embodiment, the microorganism is selected from bacteria or yeasts. In a preferred embodiment, the microorganism is a bacteria or a yeast.

In some embodiments, the microorganism is a bacteria. In one embodiment, the genus of said bacteria is selected from *Escherichia*, *Corynebacterium, Pseudomonas, Bacillus, Lactococcus, Lactobacillus, Halomonas, Bifidobacterium* and *Enterococcus*. In preferred embodiments, the genus of said bacteria is *Escherichia*. In another embodiment, the microorganism may be selected from the group consisting of *Escherichia, Corynebacterium glutamicum, Pseudomonas putida, Bacillus subtilis, Lactococcus bacillus, Halomonas elongate, Bifidobacterium infantis* and *Enterococcus faecali.* In preferred embodiments, the micororganims is an *Escherichia.* In some embodiments the bacteria is selected from the group consisting of *Escherichia coli, Corynebacterium glutamicum, Pseudomonas putida, Bacillus subtilis, Lactococcus bacillus, Halomonas elongate, Bifidobacterium infantis* and *Enterococcus faecal*

In some embodiments, the microorganism is a yeast. In some embodiments, the microorganism is a cell from a GRAS (Generally Recognized As Safe) organism or a non-pathogenic organism or strain. In some embodiments, the genus of said yeast is selected from *Saccharomyces*, *Pichia*, *Yarrowia*, *Kluyveromyces*, *Candida, Rhodotorula, Rhodosporidium, Cryptococcus, Trichosporon* and *Lipomyces.* In preferred embodiments, the genus of said yeast is *Saccharomyces.*

The microorganism may be selected from the group consisting of *Saccharomyces cerevisiae, Pichia pastoris, Kluyveromyces marxianus, Cryptococcus albidus, Lipomyces lipofera, Lipomyces starkeyi, Rhodosporidium toruloides, Rhodotorula glutinis, Trichosporon pullulan* and *Yarrowia lipolytica*. In preferred embodiments, the microorganism is a *Saccharomyces cerevisiae* cell.

Microorganism

Herein is thus provided a microorganism capable of producing strictosidine aglycone, said microorganism expresses

a strictosidine-beta-glucosidase (SGD), capable of converting strictosidine to strictosidine aglycone,

wherein said SGD is a heterologous SGD selected from RseSGD (SEQ ID NO: 24), GseSGD (SEQ ID NO: 25), SapSGD (SEQ ID NO: 26), RveSGD (SEQ ID NO: 27), VmiSGD1 (SEQ ID NO: 47), AhuSGD (SEQ ID NO: 48), HimSGD2 (SEQ ID NO: 49), SinSGD (SEQ ID NO: 50), TelSGD (SEQ ID NO: 51), VunSGD (SEQ ID NO: 52), NsiSGD1 (SEQ ID NO: 53), LprSGD (SEQ ID NO: 54), AchSGD1 (SEQ ID NO: 55), HsuSGD (SEQ ID NO: 56), MroSGD (SEQ ID NO: 57), RseSGD2 (SEQ ID NO: 58), PgrSGD (SEQ ID NO: 59), OpuSGD (SEQ ID NO: 60), HpiSGD (SEQ ID NO: 61), HanSGD1 (SEQ ID NO: 62), AchSGD2 (SEQ ID NO: 63), HimSGD1 (SEQ ID NO: 64), IpeSGD (SEQ ID NO: 65), LsaSGD1 (SEQ ID NO: 66), or CarSGD (SEQ ID NO: 67) or variants thereof having at least 70%, such as at least 80%, such as at least 90%, such as at least 91%, such as at least 92%, such as at least 93%, such as at least 94%, such as at least 95%, such as at least 96%, such as at least 97%, such as at least 98%, such as at least 99%, such as 100% identity thereto,

and/or;

wherein said SGD is a mosaic SGD, wherein said mosaic SGD comprises an amino acid sequence having the general formula

$$D_1\text{-}D_2\text{-}D_3\text{-}D_4$$

wherein D_1 is a first amino acid sequence from a first SGD,
wherein D_2 is a second amino acid sequence from a second SGD,
wherein D_3 is a third amino acid sequence comprising or consisting of amino acids of SEQ ID NO:91 or a variant thereof having at least 90% identity to SEQ ID NO: 91,
wherein D_4 is a fourth amino acid sequence from a fourth SGD or an amino acid sequence consisting of amino acids of SEQ ID NO:92 or a variant thereof having at least 90% identity to SEQ ID NO: 92,

wherein said first SGD, second SGD and fourth SGD can be the same or different, with the proviso that said first SGD, second SGD and fourth SGD are not all RseSGD.

The microorganismsdisclosed herein are thus all capable of converting strictosidine to strictosidine aglycone, when strictosidine is provided to the microorganism. In some embodiments, strictosidine is provided to the microorganism, for example by feeding strictosidine to the microorganism in the medium. In other embodiments, the microorganism is capable of synthesising strictosidine, for example the microorganism is further engineered as described below.

In another embodiment said microorganism further expresses a strictosidine synthase (STR), capable of converting secologanin and tryptamine to strictosidine. Thus, microorganisms further expressing STR are capable of converting secologanin and tryptamine to strictosidine aglycone, when secologanin and tryptamine are provided to the microorganism. Secologanin and tryptamine may be provided e.g. in the medium. However, in some embodiments the microorganism is capable of synthesising secologanin and/or tryptamine, for example the microorganismis further engineered to synthesis secologanin and/or tryptamine.

Strictosidine-O-beta-D-glucosidase (SGD)

The first heterologous enzyme expressed in the microorganism is capable of converting strictosidine to strictosidine aglycone. The first heterologous enzyme is not natively expressed in the microorganism. It may be derived from a eukaryote or a prokaryote, as detailed below, preferably a eukaryotic cell such as a plant cell.

In some embodiments, the first heterologous enzyme is a strictosidine-O-beta-D-glucosidase, herein also termed SGD, and having an EC number EC 3.2.1.105. This enzyme catalyses the following reaction:
Strictosidine + H_2O <=> D-glucose + strictosidine aglycone.

Heterologous SGD or variants thereof

Thus the microorganism expressing the first heterologous enzyme is capable of converting strictosidine to strictosidine aglycone by the action of the first heterologous enzyme.

The conversion of strictosidine to strictosidine aglycone, may be measured directly by the amount of strictosidine aglycone as known in the art, or surrogate measure of the conversion of strictosidine to strictosidine aglycone may be measured as known in the

art. Because strictosidine aglycone is highgly reactive, indirect determination of strictosidine aglycone may be preferred. For example, colorimetric assays to follow strictosidine consumption as described in Geerlings et al., 2000, may be used. The disappearance of strictosidine may also be monitored by UV, as described in Guirimand et al., 2010, or the general β-glucosidase activity in the cells may be measured, e.g. by UV detection of a synthetic substrate such as 4-methylumbelliferyl-β-D-glucoside (Guirimand et al., 2010).

Thus, to determine whether a SGD is capable of converting strictosidine to strictosidine aglycone, the person skilled in the art could use any of said methods, or could use high-precision mass spectrometry to detect the accurate mass of strictosidine aglycone after cultivation of a strain expressing an SGD or an enzyme suspected of having SGD activity in a medium; the cell is either provided with strictosidine in the medium or it has been engineered and can synthesise strictosidine. The strictosidine aglycone can be detected directly in the medium or in a pellet, after centrifugation of the culture broth. Alternatively, the appearance of other products, downstream of strictosidine aglycone, for example tetrahydroalstonine, can be monitored; such products will only form in the presence of a functional SGD, strictosidine, and an enzyme capable of using strictosidine aglycone, as described in e.g. Stavrinides et al., 2015.

In some embodiments, the first heterologous enzyme is an SGD which is native to *Rauvolfia serpentina, Gelsemium sempervirens, Scedosporium apiospermum or Rauvolfia verticillata*, *Vinca minor, Tabernaemontana elegans, Amsonia hubrichtii, Ophiorrhiza pumila, Nyssa sinensis, Coffea arabica, Carapichea ipecacuanha, Handroanthus impetiginosus, Sesamum indicum, Actinidia chinensis var. chinensis, Helianthus annuus, Lactuca sativa, Ipomoea nil, Vigna unguiculata, Heliocybe sulcate, Pyricularia grisea, Lomentospora prolificans, Hydnomerulius pinastri MD-312,* and *Moniliophthora roreri MCA 2997* or a functional variant thereof.

In other words, in some embodiments the SGD is derived from *Rauvolfia serpentina, Gelsemium sempervirens, Scedosporium apiospermum, Rauvolfia verticillata*, *Vinca minor, Tabernaemontana elegans, Amsonia hubrichtii, Ophiorrhiza pumila, Nyssa sinensis, Coffea arabica, Carapichea ipecacuanha, Handroanthus impetiginosus, Sesamum indicum, Actinidia chinensis var. chinensis, Helianthus annuus, Lactuca sativa, Ipomoea nil, Vigna unguiculata, Heliocybe sulcate, Pyricularia grisea,*

Lomentospora prolificans, Hydnomerulius pinastri MD-312, and *Moniliophthora roreri MCA 2997* or a functional variant thereof. Functional variants of SGD are modified enzymes which retain the capability to convert strictosidine to strictosidine aglycone. In some embodiments, the SGD is RseSGD as set forth in SEQ ID NO: 24 or a functional variant thereof having at least 70%, such as at least 80%, such as at least 90%, such as at least 91%, such as at least 92%, such as at least 93%, such as at least 94%, such as at least 95%, such as at least 96%, such as at least 97%, such as at least 98%, such as at least 99%, such as 100% identity to SEQ ID NO: 24. In other embodiments, the SGD is GseSGD as set forth in SEQ ID NO: 25 or a functional variant thereof having at least 70%, such as at least 80%, such as at least 90%, such as at least 91%, such as at least 92%, such as at least 93%, such as at least 94%, such as at least 95%, such as at least 96%, such as at least 97%, such as at least 98%, such as at least 99%, such as 100% identity to SEQ ID NO: 25. In other embodiments, the SGD is SapSGD as set forth in SEQ ID NO: 26 or a functional variant thereof having at least 70%, such as at least 80%, such as at least 90%, such as at least 91%, such as at least 92%, such as at least 93%, such as at least 94%, such as at least 95%, such as at least 96%, such as at least 97%, such as at least 98%, such as at least 99%, such as 100% identity to SEQ ID NO: 26. In other embodiments, the SGD is RveSGD as set forth in SEQ ID NO: 27 or a functional variant thereof having at least 70%, such as at least 80%, such as at least 90%, such as at least 91%, such as at least 92%, such as at least 93%, such as at least 94%, such as at least 95%, such as at least 96%, such as at least 97%, such as at least 98%, such as at least 99%, such as 100% identity to SEQ ID NO: 27. In other embodiments, the SGD is VmiSGD1 as set forth in SEQ ID NO: 47 or a functional variant thereof having at least 70%, such as at least 80%, such as at least 90%, such as at least 91%, such as at least 92%, such as at least 93%, such as at least 94%, such as at least 95%, such as at least 96%, such as at least 97%, such as at least 98%, such as at least 99%, such as 100% identity to SEQ ID NO: 47. In other embodiments, the SGD is AhuSGD as set forth in SEQ ID NO: 48 or a functional variant thereof having at least 70%, such as at least 80%, such as at least 90%, such as at least 91%, such as at least 92%, such as at least 93%, such as at least 94%, such as at least 95%, such as at least 96%, such as at least 97%, such as at least 98%, such as at least 99%, such as 100% identity to SEQ ID NO: 48. In other embodiments, the SGD is HimSGD2 as set forth in SEQ ID NO: 49 or a functional variant thereof having at least 70%, such as at least 80%, such as at least 90%, such

as at least 91%, such as at least 92%, such as at least 93%, such as at least 94%, such as at least 95%, such as at least 96%, such as at least 97%, such as at least 98%, such as at least 99%, such as 100% identity to SEQ ID NO: 49. In other embodiments, the SGD is SinSGD as set forth in SEQ ID NO: 50 or a functional variant thereof having at least 70%, such as at least 80%, such as at least 90%, such as at least 91%, such as at least 92%, such as at least 93%, such as at least 94%, such as at least 95%, such as at least 96%, such as at least 97%, such as at least 98%, such as at least 99%, such as 100% identity to SEQ ID NO: 50. In other embodiments, the SGD is TelSGD as set forth in SEQ ID NO: 51 or a functional variant thereof having at least 70%, such as at least 80%, such as at least 90%, such as at least 91%, such as at least 92%, such as at least 93%, such as at least 94%, such as at least 95%, such as at least 96%, such as at least 97%, such as at least 98%, such as at least 99%, such as 100% identity to SEQ ID NO: 51. In other embodiments, the SGD is VunSGD as set forth in SEQ ID NO: 52 or a functional variant thereof having at least 70%, such as at least 80%, such as at least 90%, such as at least 91%, such as at least 92%, such as at least 93%, such as at least 94%, such as at least 95%, such as at least 96%, such as at least 97%, such as at least 98%, such as at least 99%, such as 100% identity to SEQ ID NO: 52. In other embodiments, the SGD is NsiSGD1 as set forth in SEQ ID NO: 53 or a functional variant thereof having at least 70%, such as at least 80%, such as at least 90%, such as at least 91%, such as at least 92%, such as at least 93%, such as at least 94%, such as at least 95%, such as at least 96%, such as at least 97%, such as at least 98%, such as at least 99%, such as 100% identity to SEQ ID NO: 53. In other embodiments, the SGD is LprSGD as set forth in SEQ ID NO: 54 or a functional variant thereof having at least 70%, such as at least 80%, such as at least 90%, such as at least 91%, such as at least 92%, such as at least 93%, such as at least 94%, such as at least 95%, such as at least 96%, such as at least 97%, such as at least 98%, such as at least 99%, such as 100% identity to SEQ ID NO: 54. In other embodiments, the SGD is AchSGD1 as set forth in SEQ ID NO: 55 or a functional variant thereof having at least 70%, such as at least 80%, such as at least 90%, such as at least 91%, such as at least 92%, such as at least 93%, such as at least 94%, such as at least 95%, such as at least 96%, such as at least 97%, such as at least 98%, such as at least 99%, such as 100% identity to SEQ ID NO: 55. In other embodiments, the SGD is HsuSGD as set forth in SEQ ID NO: 56 or a functional variant thereof having at least 70%, such as at least 80%, such as at least 90%, such as at least 91%, such as at least 92%, such as at least 93%, such as at least 94%,

such as at least 95%, such as at least 96%, such as at least 97%, such as at least 98%, such as at least 99%, such as 100% identity to SEQ ID NO: 56. In other embodiments, the SGD is MroSGD as set forth in SEQ ID NO: 57 or a functional variant thereof having at least 70%, such as at least 80%, such as at least 90%, such as at least 91%, such as at least 92%, such as at least 93%, such as at least 94%, such as at least 95%, such as at least 96%, such as at least 97%, such as at least 98%, such as at least 99%, such as 100% identity to SEQ ID NO: 57. In other embodiments, the SGD is RseSGD2 as set forth in SEQ ID NO: 58 or a functional variant thereof having at least 70%, such as at least 80%, such as at least 90%, such as at least 91%, such as at least 92%, such as at least 93%, such as at least 94%, such as at least 95%, such as at least 96%, such as at least 97%, such as at least 98%, such as at least 99%, such as 100% identity to SEQ ID NO: 58. In other embodiments, the SGD is PgrSGD as set forth in SEQ ID NO: 59 or a functional variant thereof having at least 70%, such as at least 80%, such as at least 90%, such as at least 91%, such as at least 92%, such as at least 93%, such as at least 94%, such as at least 95%, such as at least 96%, such as at least 97%, such as at least 98%, such as at least 99%, such as 100% identity to SEQ ID NO: 59. In other embodiments, the SGD is OpuSGD as set forth in SEQ ID NO: 60 or a functional variant thereof having at least 70%, such as at least 80%, such as at least 90%, such as at least 91%, such as at least 92%, such as at least 93%, such as at least 94%, such as at least 95%, such as at least 96%, such as at least 97%, such as at least 98%, such as at least 99%, such as 100% identity to SEQ ID NO: 60. In other embodiments, the SGD is HpiSGD as set forth in SEQ ID NO: 61 or a functional variant thereof having at least 70%, such as at least 80%, such as at least 90%, such as at least 91%, such as at least 92%, such as at least 93%, such as at least 94%, such as at least 95%, such as at least 96%, such as at least 97%, such as at least 98%, such as at least 99%, such as 100% identity to SEQ ID NO: 61. In other embodiments, the SGD is HanSGD1 as set forth in SEQ ID NO: 62 or a functional variant thereof having at least 70%, such as at least 80%, such as at least 90%, such as at least 91%, such as at least 92%, such as at least 93%, such as at least 94%, such as at least 95%, such as at least 96%, such as at least 97%, such as at least 98%, such as at least 99%, such as 100% identity to SEQ ID NO: 62. In other embodiments, the SGD is AchSGD2 as set forth in SEQ ID NO: 63 or a functional variant thereof having at least 70%, such as at least 80%, such as at least 90%, such as at least 91%, such as at least 92%, such as at least 93%, such as at least 94%, such as at least 95%, such as at least 96%, such as at least

97%, such as at least 98%, such as at least 99%, such as 100% identity to SEQ ID NO: 63. In other embodiments, the SGD is HimSGD as set forth in SEQ ID NO: 64 or a functional variant thereof having at least 70%, such as at least 80%, such as at least 90%, such as at least 91%, such as at least 92%, such as at least 93%, such as at least 94%, such as at least 95%, such as at least 96%, such as at least 97%, such as at least 98%, such as at least 99%, such as 100% identity to SEQ ID NO: 64. In other embodiments, the SGD is IpeSGD as set forth in SEQ ID NO: 65 or a functional variant thereof having at least 70%, such as at least 80%, such as at least 90%, such as at least 91%, such as at least 92%, such as at least 93%, such as at least 94%, such as at least 95%, such as at least 96%, such as at least 97%, such as at least 98%, such as at least 99%, such as 100% identity to SEQ ID NO: 65. In other embodiments, the SGD is LsaSGD as set forth in SEQ ID NO: 66 or a functional variant thereof having at least 70%, such as at least 80%, such as at least 90%, such as at least 91%, such as at least 92%, such as at least 93%, such as at least 94%, such as at least 95%, such as at least 96%, such as at least 97%, such as at least 98%, such as at least 99%, such as 100% identity to SEQ ID NO: 66. In other embodiments, the SGD is CarSGD as set forth in SEQ ID NO: 67 or a functional variant thereof having at least 70%, such as at least 80%, such as at least 90%, such as at least 91%, such as at least 92%, such as at least 93%, such as at least 94%, such as at least 95%, such as at least 96%, such as at least 97%, such as at least 98%, such as at least 99%, such as 100% identity to SEQ ID NO: 67.

Preferably, the SGD is RseSGD or a functional variant thereof.

In some embodiments, the SGD originates from a MIA producing plant species, wherein said SGD shares at least 65% sequence identity to RseSGD. Thus, in some embodiments, the SGD is selected from the group consisting of RseSGD, RveSGD, TelSGD, or VmiSGD or variants thereof having at least 70%, such as at least 80%, such as at least 90%, such as at least 91%, such as at least 92%, such as at least 93%, such as at least 94%, such as at least 95%, such as at least 96%, such as at least 97%, such as at least 98%, such as at least 99%, such as 100% identity to SEQ ID NO: 24, SEQ ID NO: 27, SEQ ID NO: 51 or SEQ ID NO: 47.

In some embodiments, the SGD originates from a MIA producing plant species, wherein said SGD shares at the most 65% sequence identity to RseSGD. Thus, in

some embodiments, the SGD is selected from the group consisting of GseSGD, NsiSGD, OpuSGD, AhuSGD, or RseSGD2 or variants thereof having at least 70%, such as at least 80%, such as at least 90%, such as at least 91%, such as at least 92%, such as at least 93%, such as at least 94%, such as at least 95%, such as at least 96%, such as at least 97%, such as at least 98%, such as at least 99%, such as 100% identity to SEQ ID NO: 25, SEQ ID NO: 53 SEQ ID NO: 60, SEQ ID NO: 48 or SEQ ID NO: 58.

A person skilled in the art would know how to determine sequence identity between two species by using known methods in the art.

In some embodiments, the SGD originates from a non-MIA producing plant species. Thus, in some embodiments, the SGD is selected from the group consisting of AchSGD1, AchSGD2, CarSGD, HanSGD, HimSGD1, HimSGD2, LsaSGD1, SinSGD, VunSGD or IpeSGD or variants thereof having at least 70%, such as at least 80%, such as at least 90%, such as at least 91%, such as at least 92%, such as at least 93%, such as at least 94%, such as at least 95%, such as at least 96%, such as at least 97%, such as at least 98%, such as at least 99%, such as 100% identity to SEQ ID NO: 55, SEQ ID NO: 63, SEQ ID NO: 67, SEQ ID NO: 62, SEQ ID NO: 64, SEQ ID NO: 49, SEQ ID NO: 66, SEQ ID NO: 50, SEQ ID NO: 52 or SEQ ID NO: 65.

In some embodiments, the SGD originates from a non-MIA producing fungi species. Thus, in some embodiments, the SGD is selected from the group consisting of HpiSGD, HsuSGD, LprSGD, MroSGD, PgrSGD, or SapSGD or variants thereof having at least 70%, such as at least 80%, such as at least 90%, such as at least 91%, such as at least 92%, such as at least 93%, such as at least 94%, such as at least 95%, such as at least 96%, such as at least 97%, such as at least 98%, such as at least 99%, such as 100% identity to SEQ ID NO: 61, SEQ ID NO: 56, SEQ ID NO: 54, SEQ ID NO: 57, SEQ ID NO: 59 or SEQ ID NO: 26.

In other embodiments, said microorganism, such as the yeast cell or the bacteria cell, is capable of producing at least 1 µM tetrahydroalstonine. Thus, in some embodiments, the SGD is selected from the group consisting of RseSGD, VmiSGD or AhuSGD, or variants thereof having at least 70%, such as at least 80%, such as at least 90%, such as at least 91%, such as at least 92%, such as at least 93%, such as at least 94%,

such as at least 95%, such as at least 96%, such as at least 97%, such as at least 98%, such as at least 99%, such as 100% identity to SEQ ID NO: 24, SEQ ID NO: 47 or SEQ ID NO: 48.

In other embodiments the SGD is selected from the group consisting of RseSGD, GseSGD, SapSGD or RveSGD, or variants thereof having at least 70%, such as at least 80%, such as at least 90%, such as at least 91%, such as at least 92%, such as at least 93%, such as at least 94%, such as at least 95%, such as at least 96%, such as at least 97%, such as at least 98%, such as at least 99%, such as 100% identity to SEQ ID NO: 24, SEQ ID NO: 25, SEQ ID NO: 26 or SEQ ID NO: 27.

In other embodiments the SGD is selected from the group consisting of RseSGD, GseSGD, SapSG, RveSGD, VmiSGD, AhuSGD or variants thereof having at least 70%, such as at least 80%, such as at least 90%, such as at least 91%, such as at least 92%, such as at least 93%, such as at least 94%, such as at least 95%, such as at least 96%, such as at least 97%, such as at least 98%, such as at least 99%, such as 100% identity to SEQ ID NO: 24, SEQ ID NO: 25, SEQ ID NO: 26, SEQ ID NO: 27, SEQ ID NO: 47 or SEQ ID NO: 48.

In other embodiments the SGD is selected from the group consisting of RseSGD, RveSGD, VmiSGD, AhuSGD, HimSGD, SinSGD or TelSGD, or variants thereof having at least 70%, such as at least 80%, such as at least 90%, such as at least 91%, such as at least 92%, such as at least 93%, such as at least 94%, such as at least 95%, such as at least 96%, such as at least 97%, such as at least 98%, such as at least 99%, such as 100% identity to SEQ ID NO: 24, SEQ ID NO: 27, SEQ ID NO: 47, SEQ ID NO: 48, SEQ ID NO: 49, SEQ ID NO: 50 or SEQ ID NO: 51.

In some embodiments, said SGD is selected from RseSGD (SEQ ID NO: 24), GseSGD (SEQ ID NO: 25), SapSGD (SEQ ID NO: 26), RveSGD (SEQ ID NO: 27), VmiSGD1 (SEQ ID NO: 47), AhuSGD (SEQ ID NO: 48), HimSGD2 (SEQ ID NO: 49), SinSGD (SEQ ID NO: 50), TelSGD (SEQ ID NO: 51), VunSGD (SEQ ID NO: 52), NsiSGD1 (SEQ ID NO: 53), LprSGD (SEQ ID NO: 54), AchSGD1 (SEQ ID NO: 55), HsuSGD (SEQ ID NO: 56), MroSGD (SEQ ID NO: 57), RseSGD2 (SEQ ID NO: 58), PgrSGD (SEQ ID NO: 59), OpuSGD (SEQ ID NO: 60), HpiSGD (SEQ ID NO: 61), HanSGD1 (SEQ ID NO: 62), AchSGD2 (SEQ ID NO: 63), HimSGD1 (SEQ ID NO: 64), IpeSGD

(SEQ ID NO: 65), or LsaSGD1 (SEQ ID NO: 66), or variants thereof having at least 70%, such as at least 80%, such as at least 90%, such as at least 91%, such as at least 92%, such as at least 93%, such as at least 94%, such as at least 95%, such as at least 96%, such as at least 97%, such as at least 98%, such as at least 99%, such as 100% identity thereto.

In some embodiments, said SGD is selected from RseSGD (SEQ ID NO: 24), GseSGD (SEQ ID NO: 25), SapSGD (SEQ ID NO: 26), RveSGD (SEQ ID NO: 27), VmiSGD1 (SEQ ID NO: 47), AhuSGD (SEQ ID NO: 48), HimSGD2 (SEQ ID NO: 49), SinSGD (SEQ ID NO: 50), TelSGD (SEQ ID NO: 51), VunSGD (SEQ ID NO: 52), NsiSGD1 (SEQ ID NO: 53), LprSGD (SEQ ID NO: 54), AchSGD1 (SEQ ID NO: 55), HsuSGD (SEQ ID NO: 56), MroSGD (SEQ ID NO: 57), RseSGD2 (SEQ ID NO: 58), PgrSGD (SEQ ID NO: 59), OpuSGD (SEQ ID NO: 60), HpiSGD (SEQ ID NO: 61), HanSGD1 (SEQ ID NO: 62), AchSGD2 (SEQ ID NO: 63), HimSGD1 (SEQ ID NO: 64), IpeSGD (SEQ ID NO: 65), or CarSGD (SEQ ID NO: 67) or variants thereof having at least 70%, such as at least 80%, such as at least 90%, such as at least 91%, such as at least 92%, such as at least 93%, such as at least 94%, such as at least 95%, such as at least 96%, such as at least 97%, such as at least 98%, such as at least 99%, such as 100% identity thereto.

In some embodiments, said SGD is selected from RseSGD (SEQ ID NO: 24), GseSGD (SEQ ID NO: 25), SapSGD (SEQ ID NO: 26), RveSGD (SEQ ID NO: 27), VmiSGD1 (SEQ ID NO: 47), AhuSGD (SEQ ID NO: 48), HimSGD2 (SEQ ID NO: 49), SinSGD (SEQ ID NO: 50), TelSGD (SEQ ID NO: 51), VunSGD (SEQ ID NO: 52), NsiSGD1 (SEQ ID NO: 53), LprSGD (SEQ ID NO: 54), AchSGD1 (SEQ ID NO: 55), HsuSGD (SEQ ID NO: 56), MroSGD (SEQ ID NO: 57), RseSGD2 (SEQ ID NO: 58), PgrSGD (SEQ ID NO: 59), OpuSGD (SEQ ID NO: 60), HpiSGD (SEQ ID NO: 61), HanSGD1 (SEQ ID NO: 62), AchSGD2 (SEQ ID NO: 63), HimSGD1 (SEQ ID NO: 64), LsaSGD1 (SEQ ID NO: 66), or CarSGD (SEQ ID NO: 67) or variants thereof having at least 70%, such as at least 80%, such as at least 90%, such as at least 91%, such as at least 92%, such as at least 93%, such as at least 94%, such as at least 95%, such as at least 96%, such as at least 97%, such as at least 98%, such as at least 99%, such as 100% identity thereto.

In some embodiments, said SGD is selected from RseSGD (SEQ ID NO: 24), GseSGD (SEQ ID NO: 25), SapSGD (SEQ ID NO: 26), RveSGD (SEQ ID NO: 27), VmiSGD1 (SEQ ID NO: 47), AhuSGD (SEQ ID NO: 48), HimSGD2 (SEQ ID NO: 49), SinSGD (SEQ ID NO: 50), TelSGD (SEQ ID NO: 51), VunSGD (SEQ ID NO: 52), NsiSGD1 (SEQ ID NO: 53), LprSGD (SEQ ID NO: 54), AchSGD1 (SEQ ID NO: 55), HsuSGD (SEQ ID NO: 56), MroSGD (SEQ ID NO: 57), RseSGD2 (SEQ ID NO: 58), PgrSGD (SEQ ID NO: 59), OpuSGD (SEQ ID NO: 60), HpiSGD (SEQ ID NO: 61), HanSGD1 (SEQ ID NO: 62), AchSGD2 (SEQ ID NO: 63), IpeSGD (SEQ ID NO: 65), LsaSGD1 (SEQ ID NO: 66), or CarSGD (SEQ ID NO: 67) or variants thereof having at least 70%, such as at least 80%, such as at least 90%, such as at least 91%, such as at least 92%, such as at least 93%, such as at least 94%, such as at least 95%, such as at least 96%, such as at least 97%, such as at least 98%, such as at least 99%, such as 100% identity thereto.

In some embodiments, said SGD is selected from RseSGD (SEQ ID NO: 24), GseSGD (SEQ ID NO: 25), SapSGD (SEQ ID NO: 26), RveSGD (SEQ ID NO: 27), VmiSGD1 (SEQ ID NO: 47), AhuSGD (SEQ ID NO: 48), HimSGD2 (SEQ ID NO: 49), SinSGD (SEQ ID NO: 50), TelSGD (SEQ ID NO: 51), VunSGD (SEQ ID NO: 52), NsiSGD1 (SEQ ID NO: 53), LprSGD (SEQ ID NO: 54), AchSGD1 (SEQ ID NO: 55), HsuSGD (SEQ ID NO: 56), MroSGD (SEQ ID NO: 57), RseSGD2 (SEQ ID NO: 58), PgrSGD (SEQ ID NO: 59), OpuSGD (SEQ ID NO: 60), HpiSGD (SEQ ID NO: 61), HanSGD1 (SEQ ID NO: 62), HimSGD1 (SEQ ID NO: 64), IpeSGD (SEQ ID NO: 65), LsaSGD1 (SEQ ID NO: 66), or CarSGD (SEQ ID NO: 67) or variants thereof having at least 70%, such as at least 80%, such as at least 90%, such as at least 91%, such as at least 92%, such as at least 93%, such as at least 94%, such as at least 95%, such as at least 96%, such as at least 97%, such as at least 98%, such as at least 99%, such as 100% identity thereto.

In some embodiments, said SGD is selected from RseSGD (SEQ ID NO: 24), GseSGD (SEQ ID NO: 25), SapSGD (SEQ ID NO: 26), RveSGD (SEQ ID NO: 27), VmiSGD1 (SEQ ID NO: 47), AhuSGD (SEQ ID NO: 48), HimSGD2 (SEQ ID NO: 49), SinSGD (SEQ ID NO: 50), TelSGD (SEQ ID NO: 51), VunSGD (SEQ ID NO: 52), NsiSGD1 (SEQ ID NO: 53), LprSGD (SEQ ID NO: 54), AchSGD1 (SEQ ID NO: 55), HsuSGD (SEQ ID NO: 56), MroSGD (SEQ ID NO: 57), RseSGD2 (SEQ ID NO: 58), PgrSGD (SEQ ID NO: 59), OpuSGD (SEQ ID NO: 60), HpiSGD (SEQ ID NO: 61), HanSGD1

(SEQ ID NO: 62), HimSGD1 (SEQ ID NO: 64), IpeSGD (SEQ ID NO: 65), LsaSGD1 (SEQ ID NO: 66), or CarSGD (SEQ ID NO: 67) or variants thereof having at least 70%, such as at least 80%, such as at least 90%, such as at least 91%, such as at least 92%, such as at least 93%, such as at least 94%, such as at least 95%, such as at least 96%, such as at least 97%, such as at least 98%, such as at least 99%, such as 100% identity thereto.

In some embodiments, said SGD is selected from RseSGD (SEQ ID NO: 24), GseSGD (SEQ ID NO: 25), SapSGD (SEQ ID NO: 26), RveSGD (SEQ ID NO: 27), VmiSGD1 (SEQ ID NO: 47), AhuSGD (SEQ ID NO: 48), HimSGD2 (SEQ ID NO: 49), SinSGD (SEQ ID NO: 50), TelSGD (SEQ ID NO: 51), VunSGD (SEQ ID NO: 52), NsiSGD1 (SEQ ID NO: 53), LprSGD (SEQ ID NO: 54), AchSGD1 (SEQ ID NO: 55), HsuSGD (SEQ ID NO: 56), MroSGD (SEQ ID NO: 57), RseSGD2 (SEQ ID NO: 58), PgrSGD (SEQ ID NO: 59), OpuSGD (SEQ ID NO: 60), HpiSGD (SEQ ID NO: 61), AchSGD2 (SEQ ID NO: 63), HimSGD1 (SEQ ID NO: 64), IpeSGD (SEQ ID NO: 65), LsaSGD1 (SEQ ID NO: 66), or CarSGD (SEQ ID NO: 67) or variants thereof having at least 70%, such as at least 80%, such as at least 90%, such as at least 91%, such as at least 92%, such as at least 93%, such as at least 94%, such as at least 95%, such as at least 96%, such as at least 97%, such as at least 98%, such as at least 99%, such as 100% identity thereto.

In some embodiments, said SGD is selected from RseSGD (SEQ ID NO: 24), GseSGD (SEQ ID NO: 25), SapSGD (SEQ ID NO: 26), RveSGD (SEQ ID NO: 27), VmiSGD1 (SEQ ID NO: 47), AhuSGD (SEQ ID NO: 48), HimSGD2 (SEQ ID NO: 49), SinSGD (SEQ ID NO: 50), TelSGD (SEQ ID NO: 51), VunSGD (SEQ ID NO: 52), NsiSGD1 (SEQ ID NO: 53), LprSGD (SEQ ID NO: 54), AchSGD1 (SEQ ID NO: 55), HsuSGD (SEQ ID NO: 56), MroSGD (SEQ ID NO: 57), RseSGD2 (SEQ ID NO: 58), PgrSGD (SEQ ID NO: 59), OpuSGD (SEQ ID NO: 60), HpiSGD HanSGD1 (SEQ ID NO: 62), AchSGD2 (SEQ ID NO: 63), HimSGD1 (SEQ ID NO: 64), IpeSGD (SEQ ID NO: 65), LsaSGD1 (SEQ ID NO: 66), or CarSGD (SEQ ID NO: 67) or variants thereof having at least 70%, such as at least 80%, such as at least 90%, such as at least 91%, such as at least 92%, such as at least 93%, such as at least 94%, such as at least 95%, such as at least 96%, such as at least 97%, such as at least 98%, such as at least 99%, such as 100% identity thereto.

In some embodiments, said SGD is selected from RseSGD (SEQ ID NO: 24), GseSGD (SEQ ID NO: 25), SapSGD (SEQ ID NO: 26), RveSGD (SEQ ID NO: 27), VmiSGD1 (SEQ ID NO: 47), AhuSGD (SEQ ID NO: 48), HimSGD2 (SEQ ID NO: 49), SinSGD (SEQ ID NO: 50), TelSGD (SEQ ID NO: 51), VunSGD (SEQ ID NO: 52), NsiSGD1 (SEQ ID NO: 53), LprSGD (SEQ ID NO: 54), AchSGD1 (SEQ ID NO: 55), HsuSGD (SEQ ID NO: 56), MroSGD (SEQ ID NO: 57), RseSGD2 (SEQ ID NO: 58), PgrSGD (SEQ ID NO: 59), HpiSGD (SEQ ID NO: 61), HanSGD1 (SEQ ID NO: 62), AchSGD2 (SEQ ID NO: 63), HimSGD1 (SEQ ID NO: 64), IpeSGD (SEQ ID NO: 65), LsaSGD1 (SEQ ID NO: 66), or CarSGD (SEQ ID NO: 67) or variants thereof having at least 70%, such as at least 80%, such as at least 90%, such as at least 91%, such as at least 92%, such as at least 93%, such as at least 94%, such as at least 95%, such as at least 96%, such as at least 97%, such as at least 98%, such as at least 99%, such as 100% identity thereto.

In some embodiments, said SGD is selected from RseSGD (SEQ ID NO: 24), GseSGD (SEQ ID NO: 25), SapSGD (SEQ ID NO: 26), RveSGD (SEQ ID NO: 27), VmiSGD1 (SEQ ID NO: 47), AhuSGD (SEQ ID NO: 48), HimSGD2 (SEQ ID NO: 49), SinSGD (SEQ ID NO: 50), TelSGD (SEQ ID NO: 51), VunSGD (SEQ ID NO: 52), NsiSGD1 (SEQ ID NO: 53), LprSGD (SEQ ID NO: 54), AchSGD1 (SEQ ID NO: 55), HsuSGD (SEQ ID NO: 56), MroSGD (SEQ ID NO: 57), RseSGD2 (SEQ ID NO: 58), OpuSGD (SEQ ID NO: 60), HpiSGD (SEQ ID NO: 61), HanSGD1 (SEQ ID NO: 62), AchSGD2 (SEQ ID NO: 63), HimSGD1 (SEQ ID NO: 64), IpeSGD (SEQ ID NO: 65), LsaSGD1 (SEQ ID NO: 66), or CarSGD (SEQ ID NO: 67) or variants thereof having at least 70%, such as at least 80%, such as at least 90%, such as at least 91%, such as at least 92%, such as at least 93%, such as at least 94%, such as at least 95%, such as at least 96%, such as at least 97%, such as at least 98%, such as at least 99%, such as 100% identity thereto.

In some embodiments, said SGD is selected from RseSGD (SEQ ID NO: 24), GseSGD (SEQ ID NO: 25), SapSGD (SEQ ID NO: 26), RveSGD (SEQ ID NO: 27), VmiSGD1 (SEQ ID NO: 47), AhuSGD (SEQ ID NO: 48), HimSGD2 (SEQ ID NO: 49), SinSGD (SEQ ID NO: 50), TelSGD (SEQ ID NO: 51), VunSGD (SEQ ID NO: 52), NsiSGD1 (SEQ ID NO: 53), LprSGD (SEQ ID NO: 54), AchSGD1 (SEQ ID NO: 55), HsuSGD (SEQ ID NO: 56), MroSGD (SEQ ID NO: 57), PgrSGD (SEQ ID NO: 59), OpuSGD (SEQ ID NO: 60), HpiSGD (SEQ ID NO: 61), HanSGD1 (SEQ ID NO: 62), AchSGD2

(SEQ ID NO: 63), HimSGD1 (SEQ ID NO: 64), IpeSGD (SEQ ID NO: 65), LsaSGD1 (SEQ ID NO: 66), or CarSGD (SEQ ID NO: 67) or variants thereof having at least 70%, such as at least 80%, such as at least 90%, such as at least 91%, such as at least 92%, such as at least 93%, such as at least 94%, such as at least 95%, such as at least 96%, such as at least 97%, such as at least 98%, such as at least 99%, such as 100% identity thereto.

In some embodiments, said SGD is selected from RseSGD (SEQ ID NO: 24), GseSGD (SEQ ID NO: 25), SapSGD (SEQ ID NO: 26), RveSGD (SEQ ID NO: 27), VmiSGD1 (SEQ ID NO: 47), AhuSGD (SEQ ID NO: 48), HimSGD2 (SEQ ID NO: 49), SinSGD (SEQ ID NO: 50), TelSGD (SEQ ID NO: 51), VunSGD (SEQ ID NO: 52), NsiSGD1 (SEQ ID NO: 53), LprSGD (SEQ ID NO: 54), AchSGD1 (SEQ ID NO: 55), HsuSGD (SEQ ID NO: 56), RseSGD2 (SEQ ID NO: 58), PgrSGD (SEQ ID NO: 59), OpuSGD (SEQ ID NO: 60), HpiSGD (SEQ ID NO: 61), HanSGD1 (SEQ ID NO: 62), AchSGD2 (SEQ ID NO: 63), HimSGD1 (SEQ ID NO: 64), IpeSGD (SEQ ID NO: 65), LsaSGD1 (SEQ ID NO: 66), or CarSGD (SEQ ID NO: 67) or variants thereof having at least 70%, such as at least 80%, such as at least 90%, such as at least 91%, such as at least 92%, such as at least 93%, such as at least 94%, such as at least 95%, such as at least 96%, such as at least 97%, such as at least 98%, such as at least 99%, such as 100% identity thereto.

In some embodiments, said SGD is selected from RseSGD (SEQ ID NO: 24), GseSGD (SEQ ID NO: 25), SapSGD (SEQ ID NO: 26), RveSGD (SEQ ID NO: 27), VmiSGD1 (SEQ ID NO: 47), AhuSGD (SEQ ID NO: 48), HimSGD2 (SEQ ID NO: 49), SinSGD (SEQ ID NO: 50), TelSGD (SEQ ID NO: 51), VunSGD (SEQ ID NO: 52), NsiSGD1 (SEQ ID NO: 53), LprSGD (SEQ ID NO: 54), AchSGD1 (SEQ ID NO: 55), MroSGD (SEQ ID NO: 57), RseSGD2 (SEQ ID NO: 58), PgrSGD (SEQ ID NO: 59), OpuSGD (SEQ ID NO: 60), HpiSGD (SEQ ID NO: 61), HanSGD1 (SEQ ID NO: 62), AchSGD2 (SEQ ID NO: 63), HimSGD1 (SEQ ID NO: 64), IpeSGD (SEQ ID NO: 65), LsaSGD1 (SEQ ID NO: 66), or CarSGD (SEQ ID NO: 67) or variants thereof having at least 70%, such as at least 80%, such as at least 90%, such as at least 91%, such as at least 92%, such as at least 93%, such as at least 94%, such as at least 95%, such as at least 96%, such as at least 97%, such as at least 98%, such as at least 99%, such as 100% identity thereto.

In some embodiments, said SGD is selected from RseSGD (SEQ ID NO: 24), GseSGD (SEQ ID NO: 25), SapSGD (SEQ ID NO: 26), RveSGD (SEQ ID NO: 27), VmiSGD1 (SEQ ID NO: 47), AhuSGD (SEQ ID NO: 48), HimSGD2 (SEQ ID NO: 49), SinSGD (SEQ ID NO: 50), TelSGD (SEQ ID NO: 51), VunSGD (SEQ ID NO: 52), NsiSGD1 (SEQ ID NO: 53), LprSGD (SEQ ID NO: 54), HsuSGD (SEQ ID NO: 56), MroSGD (SEQ ID NO: 57), RseSGD2 (SEQ ID NO: 58), PgrSGD (SEQ ID NO: 59), OpuSGD (SEQ ID NO: 60), HpiSGD (SEQ ID NO: 61), HanSGD1 (SEQ ID NO: 62), AchSGD2 (SEQ ID NO: 63), HimSGD1 (SEQ ID NO: 64), IpeSGD (SEQ ID NO: 65), LsaSGD1 (SEQ ID NO: 66), or CarSGD (SEQ ID NO: 67) or variants thereof having at least 70%, such as at least 80%, such as at least 90%, such as at least 91%, such as at least 92%, such as at least 93%, such as at least 94%, such as at least 95%, such as at least 96%, such as at least 97%, such as at least 98%, such as at least 99%, such as 100% identity thereto.

In some embodiments, said SGD is selected from RseSGD (SEQ ID NO: 24), GseSGD (SEQ ID NO: 25), SapSGD (SEQ ID NO: 26), RveSGD (SEQ ID NO: 27), VmiSGD1 (SEQ ID NO: 47), AhuSGD (SEQ ID NO: 48), HimSGD2 (SEQ ID NO: 49), SinSGD (SEQ ID NO: 50), TelSGD (SEQ ID NO: 51), VunSGD (SEQ ID NO: 52), NsiSGD1 (SEQ ID NO: 53), AchSGD1 (SEQ ID NO: 55), HsuSGD (SEQ ID NO: 56), MroSGD (SEQ ID NO: 57), RseSGD2 (SEQ ID NO: 58), PgrSGD (SEQ ID NO: 59), OpuSGD (SEQ ID NO: 60), HpiSGD (SEQ ID NO: 61), HanSGD1 (SEQ ID NO: 62), AchSGD2 (SEQ ID NO: 63), HimSGD1 (SEQ ID NO: 64), IpeSGD (SEQ ID NO: 65), LsaSGD1 (SEQ ID NO: 66), or CarSGD (SEQ ID NO: 67) or variants thereof having at least 70%, such as at least 80%, such as at least 90%, such as at least 91%, such as at least 92%, such as at least 93%, such as at least 94%, such as at least 95%, such as at least 96%, such as at least 97%, such as at least 98%, such as at least 99%, such as 100% identity thereto.

In some embodiments, said SGD is selected from RseSGD (SEQ ID NO: 24), GseSGD (SEQ ID NO: 25), SapSGD (SEQ ID NO: 26), RveSGD (SEQ ID NO: 27), VmiSGD1 (SEQ ID NO: 47), AhuSGD (SEQ ID NO: 48), HimSGD2 (SEQ ID NO: 49), SinSGD (SEQ ID NO: 50), TelSGD (SEQ ID NO: 51), VunSGD (SEQ ID NO: 52), LprSGD (SEQ ID NO: 54), AchSGD1 (SEQ ID NO: 55), HsuSGD (SEQ ID NO: 56), MroSGD (SEQ ID NO: 57), RseSGD2 (SEQ ID NO: 58), PgrSGD (SEQ ID NO: 59), OpuSGD (SEQ ID NO: 60), HpiSGD (SEQ ID NO: 61), HanSGD1 (SEQ ID NO: 62), AchSGD2 (SEQ ID

NO: 63), HimSGD1 (SEQ ID NO: 64), IpeSGD (SEQ ID NO: 65), LsaSGD1 (SEQ ID NO: 66), or CarSGD (SEQ ID NO: 67) or variants thereof having at least 70%, such as at least 80%, such as at least 90%, such as at least 91%, such as at least 92%, such as at least 93%, such as at least 94%, such as at least 95%, such as at least 96%, such as at least 97%, such as at least 98%, such as at least 99%, such as 100% identity thereto.

In some embodiments, said SGD is selected from RseSGD (SEQ ID NO: 24), GseSGD (SEQ ID NO: 25), SapSGD (SEQ ID NO: 26), RveSGD (SEQ ID NO: 27), VmiSGD1 (SEQ ID NO: 47), AhuSGD (SEQ ID NO: 48), HimSGD2 (SEQ ID NO: 49), SinSGD (SEQ ID NO: 50), TelSGD (SEQ ID NO: 51), NsiSGD1 (SEQ ID NO: 53), LprSGD (SEQ ID NO: 54), AchSGD1 (SEQ ID NO: 55), HsuSGD (SEQ ID NO: 56), MroSGD (SEQ ID NO: 57), RseSGD2 (SEQ ID NO: 58), PgrSGD (SEQ ID NO: 59), OpuSGD (SEQ ID NO: 60), HpiSGD (SEQ ID NO: 61), HanSGD1 (SEQ ID NO: 62), AchSGD2 (SEQ ID NO: 63), HimSGD1 (SEQ ID NO: 64), IpeSGD (SEQ ID NO: 65), LsaSGD1 (SEQ ID NO: 66), or CarSGD (SEQ ID NO: 67) or variants thereof having at least 70%, such as at least 80%, such as at least 90%, such as at least 91%, such as at least 92%, such as at least 93%, such as at least 94%, such as at least 95%, such as at least 96%, such as at least 97%, such as at least 98%, such as at least 99%, such as 100% identity thereto.

In some embodiments, said SGD is selected from RseSGD (SEQ ID NO: 24), GseSGD (SEQ ID NO: 25), SapSGD (SEQ ID NO: 26), RveSGD (SEQ ID NO: 27), VmiSGD1 (SEQ ID NO: 47), AhuSGD (SEQ ID NO: 48), HimSGD2 (SEQ ID NO: 49), SinSGD (SEQ ID NO: 50), VunSGD (SEQ ID NO: 52), NsiSGD1 (SEQ ID NO: 53), LprSGD (SEQ ID NO: 54), AchSGD1 (SEQ ID NO: 55), HsuSGD (SEQ ID NO: 56), MroSGD (SEQ ID NO: 57), RseSGD2 (SEQ ID NO: 58), PgrSGD (SEQ ID NO: 59), OpuSGD (SEQ ID NO: 60), HpiSGD (SEQ ID NO: 61), HanSGD1 (SEQ ID NO: 62), AchSGD2 (SEQ ID NO: 63), HimSGD1 (SEQ ID NO: 64), IpeSGD (SEQ ID NO: 65), LsaSGD1 (SEQ ID NO: 66), or CarSGD (SEQ ID NO: 67) or variants thereof having at least 70%, such as at least 80%, such as at least 90%, such as at least 91%, such as at least 92%, such as at least 93%, such as at least 94%, such as at least 95%, such as at least 96%, such as at least 97%, such as at least 98%, such as at least 99%, such as 100% identity thereto.

In some embodiments, said SGD is selected from RseSGD (SEQ ID NO: 24), GseSGD (SEQ ID NO: 25), SapSGD (SEQ ID NO: 26), RveSGD (SEQ ID NO: 27), VmiSGD1 (SEQ ID NO: 47), AhuSGD (SEQ ID NO: 48), HimSGD2 (SEQ ID NO: 49), TelSGD (SEQ ID NO: 51), VunSGD (SEQ ID NO: 52), NsiSGD1 (SEQ ID NO: 53), LprSGD (SEQ ID NO: 54), AchSGD1 (SEQ ID NO: 55), HsuSGD (SEQ ID NO: 56), MroSGD (SEQ ID NO: 57), RseSGD2 (SEQ ID NO: 58), PgrSGD (SEQ ID NO: 59), OpuSGD (SEQ ID NO: 60), HpiSGD (SEQ ID NO: 61), HanSGD1 (SEQ ID NO: 62), AchSGD2 (SEQ ID NO: 63), HimSGD1 (SEQ ID NO: 64), IpeSGD (SEQ ID NO: 65), LsaSGD1 (SEQ ID NO: 66), or CarSGD (SEQ ID NO: 67) or variants thereof having at least 70%, such as at least 80%, such as at least 90%, such as at least 91%, such as at least 92%, such as at least 93%, such as at least 94%, such as at least 95%, such as at least 96%, such as at least 97%, such as at least 98%, such as at least 99%, such as 100% identity thereto.

In some embodiments, said SGD is selected from RseSGD (SEQ ID NO: 24), GseSGD (SEQ ID NO: 25), SapSGD (SEQ ID NO: 26), RveSGD (SEQ ID NO: 27), VmiSGD1 (SEQ ID NO: 47), AhuSGD (SEQ ID NO: 48), SinSGD (SEQ ID NO: 50), TelSGD (SEQ ID NO: 51), VunSGD (SEQ ID NO: 52), NsiSGD1 (SEQ ID NO: 53), LprSGD (SEQ ID NO: 54), AchSGD1 (SEQ ID NO: 55), HsuSGD (SEQ ID NO: 56), MroSGD (SEQ ID NO: 57), RseSGD2 (SEQ ID NO: 58), PgrSGD (SEQ ID NO: 59), OpuSGD (SEQ ID NO: 60), HpiSGD (SEQ ID NO: 61), HanSGD1 (SEQ ID NO: 62), AchSGD2 (SEQ ID NO: 63), HimSGD1 (SEQ ID NO: 64), IpeSGD (SEQ ID NO: 65), LsaSGD1 (SEQ ID NO: 66), or CarSGD (SEQ ID NO: 67) or variants thereof having at least 70%, such as at least 80%, such as at least 90%, such as at least 91%, such as at least 92%, such as at least 93%, such as at least 94%, such as at least 95%, such as at least 96%, such as at least 97%, such as at least 98%, such as at least 99%, such as 100% identity thereto.

In some embodiments, said SGD is selected from RseSGD (SEQ ID NO: 24), GseSGD (SEQ ID NO: 25), SapSGD (SEQ ID NO: 26), RveSGD (SEQ ID NO: 27), VmiSGD1 (SEQ ID NO: 47), HimSGD2 (SEQ ID NO: 49), SinSGD (SEQ ID NO: 50), TelSGD (SEQ ID NO: 51), VunSGD (SEQ ID NO: 52), NsiSGD1 (SEQ ID NO: 53), LprSGD (SEQ ID NO: 54), AchSGD1 (SEQ ID NO: 55), HsuSGD (SEQ ID NO: 56), MroSGD (SEQ ID NO: 57), RseSGD2 (SEQ ID NO: 58), PgrSGD (SEQ ID NO: 59), OpuSGD (SEQ ID NO: 60), HpiSGD (SEQ ID NO: 61), HanSGD1 (SEQ ID NO: 62), AchSGD2

(SEQ ID NO: 63), HimSGD1 (SEQ ID NO: 64), IpeSGD (SEQ ID NO: 65), LsaSGD1 (SEQ ID NO: 66), or CarSGD (SEQ ID NO: 67) or variants thereof having at least 70%, such as at least 80%, such as at least 90%, such as at least 91%, such as at least 92%, such as at least 93%, such as at least 94%, such as at least 95%, such as at least 96%, such as at least 97%, such as at least 98%, such as at least 99%, such as 100% identity thereto.

In some embodiments, said SGD is selected from RseSGD (SEQ ID NO: 24), GseSGD (SEQ ID NO: 25), SapSGD (SEQ ID NO: 26), RveSGD (SEQ ID NO: 27), AhuSGD (SEQ ID NO: 48), HimSGD2 (SEQ ID NO: 49), SinSGD (SEQ ID NO: 50), TelSGD (SEQ ID NO: 51), VunSGD (SEQ ID NO: 52), NsiSGD1 (SEQ ID NO: 53), LprSGD (SEQ ID NO: 54), AchSGD1 (SEQ ID NO: 55), HsuSGD (SEQ ID NO: 56), MroSGD (SEQ ID NO: 57), RseSGD2 (SEQ ID NO: 58), PgrSGD (SEQ ID NO: 59), OpuSGD (SEQ ID NO: 60), HpiSGD (SEQ ID NO: 61), HanSGD1 (SEQ ID NO: 62), AchSGD2 (SEQ ID NO: 63), HimSGD1 (SEQ ID NO: 64), IpeSGD (SEQ ID NO: 65), LsaSGD1 (SEQ ID NO: 66), or CarSGD (SEQ ID NO: 67) or variants thereof having at least 70%, such as at least 80%, such as at least 90%, such as at least 91%, such as at least 92%, such as at least 93%, such as at least 94%, such as at least 95%, such as at least 96%, such as at least 97%, such as at least 98%, such as at least 99%, such as 100% identity thereto.

In some embodiments, said SGD is selected from RseSGD (SEQ ID NO: 24), GseSGD (SEQ ID NO: 25), SapSGD (SEQ ID NO: 26), VmiSGD1 (SEQ ID NO: 47), AhuSGD (SEQ ID NO: 48), HimSGD2 (SEQ ID NO: 49), SinSGD (SEQ ID NO: 50), TelSGD (SEQ ID NO: 51), VunSGD (SEQ ID NO: 52), NsiSGD1 (SEQ ID NO: 53), LprSGD (SEQ ID NO: 54), AchSGD1 (SEQ ID NO: 55), HsuSGD (SEQ ID NO: 56), MroSGD (SEQ ID NO: 57), RseSGD2 (SEQ ID NO: 58), PgrSGD (SEQ ID NO: 59), OpuSGD (SEQ ID NO: 60), HpiSGD (SEQ ID NO: 61), HanSGD1 (SEQ ID NO: 62), AchSGD2 (SEQ ID NO: 63), HimSGD1 (SEQ ID NO: 64), IpeSGD (SEQ ID NO: 65), LsaSGD1 (SEQ ID NO: 66), or CarSGD (SEQ ID NO: 67) or variants thereof having at least 70%, such as at least 80%, such as at least 90%, such as at least 91%, such as at least 92%, such as at least 93%, such as at least 94%, such as at least 95%, such as at least 96%, such as at least 97%, such as at least 98%, such as at least 99%, such as 100% identity thereto.

In some embodiments, said SGD is selected from RseSGD (SEQ ID NO: 24), GseSGD (SEQ ID NO: 25), RveSGD (SEQ ID NO: 27), VmiSGD1 (SEQ ID NO: 47), AhuSGD (SEQ ID NO: 48), HimSGD2 (SEQ ID NO: 49), SinSGD (SEQ ID NO: 50), TelSGD (SEQ ID NO: 51), VunSGD (SEQ ID NO: 52), NsiSGD1 (SEQ ID NO: 53), LprSGD (SEQ ID NO: 54), AchSGD1 (SEQ ID NO: 55), HsuSGD (SEQ ID NO: 56), MroSGD (SEQ ID NO: 57), RseSGD2 (SEQ ID NO: 58), PgrSGD (SEQ ID NO: 59), OpuSGD (SEQ ID NO: 60), HpiSGD (SEQ ID NO: 61), HanSGD1 (SEQ ID NO: 62), AchSGD2 (SEQ ID NO: 63), HimSGD1 (SEQ ID NO: 64), IpeSGD (SEQ ID NO: 65), LsaSGD1 (SEQ ID NO: 66), or CarSGD (SEQ ID NO: 67) or variants thereof having at least 70%, such as at least 80%, such as at least 90%, such as at least 91%, such as at least 92%, such as at least 93%, such as at least 94%, such as at least 95%, such as at least 96%, such as at least 97%, such as at least 98%, such as at least 99%, such as 100% identity thereto.

In some embodiments, said SGD is selected from RseSGD (SEQ ID NO: 24), SapSGD (SEQ ID NO: 26), RveSGD (SEQ ID NO: 27), VmiSGD1 (SEQ ID NO: 47), AhuSGD (SEQ ID NO: 48), HimSGD2 (SEQ ID NO: 49), SinSGD (SEQ ID NO: 50), TelSGD (SEQ ID NO: 51), VunSGD (SEQ ID NO: 52), NsiSGD1 (SEQ ID NO: 53), LprSGD (SEQ ID NO: 54), AchSGD1 (SEQ ID NO: 55), HsuSGD (SEQ ID NO: 56), MroSGD (SEQ ID NO: 57), RseSGD2 (SEQ ID NO: 58), PgrSGD (SEQ ID NO: 59), OpuSGD (SEQ ID NO: 60), HpiSGD (SEQ ID NO: 61), HanSGD1 (SEQ ID NO: 62), AchSGD2 (SEQ ID NO: 63), HimSGD1 (SEQ ID NO: 64), IpeSGD (SEQ ID NO: 65), LsaSGD1 (SEQ ID NO: 66), or CarSGD (SEQ ID NO: 67) or variants thereof having at least 70%, such as at least 80%, such as at least 90%, such as at least 91%, such as at least 92%, such as at least 93%, such as at least 94%, such as at least 95%, such as at least 96%, such as at least 97%, such as at least 98%, such as at least 99%, such as 100% identity thereto.

In some embodiments, said SGD is selected from GseSGD (SEQ ID NO: 25), SapSGD (SEQ ID NO: 26), RveSGD (SEQ ID NO: 27), VmiSGD1 (SEQ ID NO: 47), AhuSGD (SEQ ID NO: 48), HimSGD2 (SEQ ID NO: 49), SinSGD (SEQ ID NO: 50), TelSGD (SEQ ID NO: 51), VunSGD (SEQ ID NO: 52), NsiSGD1 (SEQ ID NO: 53), LprSGD (SEQ ID NO: 54), AchSGD1 (SEQ ID NO: 55), HsuSGD (SEQ ID NO: 56), MroSGD (SEQ ID NO: 57), RseSGD2 (SEQ ID NO: 58), PgrSGD (SEQ ID NO: 59), OpuSGD (SEQ ID NO: 60), HpiSGD (SEQ ID NO: 61), HanSGD1 (SEQ ID NO: 62), AchSGD2

(SEQ ID NO: 63), HimSGD1 (SEQ ID NO: 64), IpeSGD (SEQ ID NO: 65), LsaSGD1 (SEQ ID NO: 66), or CarSGD (SEQ ID NO: 67) or variants thereof having at least 70%, such as at least 80%, such as at least 90%, such as at least 91%, such as at least 92%, such as at least 93%, such as at least 94%, such as at least 95%, such as at least 96%, such as at least 97%, such as at least 98%, such as at least 99%, such as 100% identity thereto.

Thus, in some embodiments the microorganism according to the present invention may express a SGD as described herein above. In other embodiments, the microorganism according to the present invention may express a mosaic SGD. The microorganism may be a yeast cell or a bacteria cell, as described herein.

Mosaic SGD or variants thereof

The inventors have engineered new and active mosaic SGDs capable of converting strictosidine into strictosidine aglycone. Said mosaic SGDs are useful in microorganism factories, such as yeast factories and bacteria factories, for production of strictosidine aglycone, tetrahydroalstonine and/or other MIA products.

Thus, the present invention also relates to a mosaic SGD, wherein said mosaic SGD comprises an amino acid sequence having the general formula

$$D_1\text{-}D_2\text{-}D_3\text{-}D_4$$

wherein D_1 is a first amino acid sequence from a first SGD,
wherein D_2 is a second amino acid sequence from a second SGD,
wherein D_3 is a third amino acid sequence comprising or consisting of amino acids of SEQ ID NO:91 or a variant thereof having at least 90% identity to SEQ ID NO: 91,
wherein D_4 is a fourth amino acid sequence from a fourth SGD or an amino acid sequence consisting of amino acids of SEQ ID NO:92 or a variant thereof having at least 90% identity to SEQ ID NO: 92,
wherein said first SGD, second SGD and fourth SGD can be the same or different, with the proviso that said first SGD, second SGD and fourth SGD are not all RseSGD.

The mosaic SGD thus comprises at least one domain of RseSGD, namely the third domain D_3, and at least one other domain as defined above which is not a domain of RseSGD.

The inventors found that a SGD can be divided into four domains:

- Domain 1 (D_1)
- Domain 2 (D_2)
- Domain 3 (D_3)
- Domain 4 (D_4)

Examples hereof are described in Examples 8 and 9 herein below.

Each of domain 1-4 consists of a consecutive sequence of amino acids. Domain 1 is the most N-terminal amino acid sequence in the SGD. The first amino acid residue in domain 1 is typically methionine, as this is the first amino acid which is translated from a start codon, however it may occur that the first domain actually starts with another residue in embodiments where part of the domain would be cleaved off, thereby removing the methionine. Being the first domain in SGD, domain 1 is followed by domain 2, which is followed by domain 3, which is followed by domain 4. Domain 4 is the most C-terminal amino acid sequence in the SGD. The last amino acid residue in domain 4 is the last amino acid residue in the consecutive sequence of the SGD.

The positions of the amino acids in each domain 1-4 of a SGD may be defined by aligning the SGD amino acid sequence to the amino acid sequence RseSGD of SEQ ID NO:24, hereby using RseSGD as a reference sequence. Thus, is it to be understood that following alignment between a SGD amino acid sequence and the reference amino acid sequence of SEQ ID NO:24, an amino acid corresponds to position X of SEQ ID NO:24 if it aligns to the same position.

For example, the domains can be defined as follows. Starting from an SGD which is not RseSGD, and which hereinafter is termed XxxSGD, a pairwise alignment of the two amino acid sequences of RseSGD and XxxSGD is performed to determine the boundaries of the domains in XxxSGC.

Domain 1 in XxxSGD can thus be defined as follows. Domain 1 of RseSGD (as set forth in SEQ ID NO: 89) is used to align XxxSGD. The first domain is then defined as the region of XxxSGD starting with the amino acid that aligns with the first residue of SEQ ID NO: 89 and finishing with the amino acid that aligns with the last residue of SEQ ID NO: 89. In embodiments where this amino acid is not a methionine, the

introduction of a methionine immediately upstream of this first domain may be necessary in order to ensure proper translation of the protein, as is known in the art.

The same procedure can be repeated for domains 2 and 3, as needed. Domain 2 in XxxSGD can thus be defined as follows. Domain 2 of RseSGD (as set forth in SEQ ID NO: 90) is used to align XxxSGD. The second domain is then defined as the region of XxxSGD starting with the amino acid that aligns with the first residue of SEQ ID NO: 90 and finishing with the amino acid that aligns with the last residue of SEQ ID NO: 90. Domain 3 in XxxSGD can thus be defined as follows. Domain 3 of RseSGD (as set forth in SEQ ID NO: 91) is used to align XxxSGD. The third domain is then defined as the region of XxxSGD starting with the amino acid that aligns with the first residue of SEQ ID NO: 91 and finishing with the amino acid that aligns with the last residue of SEQ ID NO: 91. The third domain of the mosaic SGD is domain D_3 of RseSGD as set forth in SEQ ID NO: 91, but it may still be useful to determine the position of domain 3 in XxxSGD, particularly in order to determine the position of domain 4 in XxxSGD.

Domain 4 in XxxSGD preferably corresponds to the region starting with the first amino acid immediately downstream of domain 3 of the same XxxSGD and finishing with the last amino acid of XxxSGD. In other words, if domain 3 of XxxSGD ends with residue number *n*, then domain 4 starts with residue *n* + *1*, where *n* is an integer.

The term “domain 1” as used herein refers to one or more sequential groups of amino acids corresponding to amino acids from position 1 to 115 of SEQ ID NO:24.

The term “domain 2” as used herein refers to one or more sequential groups of amino acids corresponding to amino acids from position 116 to 266 of SEQ ID NO:24.

The term “domain 3” as used herein refers to one or more sequential groups of amino acids corresponding to amino acids from position 267 to 456 of SEQ ID NO:24.

The term “domain 4” as used herein refers to one or more sequential groups of amino acids corresponding to amino acids from position 457 to 532 of SEQ ID NO:24.

The four domains of the mosaic SGD may be linked by, or separated by, small sequences, for example amino acid linkers, as is known in the art. It will thus be

understood that the mosaic SGD may comprise additional amino acids which can be added to each of the four domains, as is known in the art.

In some embodiments, the mosaic SGD may be further modified, for example by the introduction of additional domains which may increase the stability or longevity or half-life of the protein, or localidation domains targeting the mosaic SGD to specific cellular localisations. Relevant additional domains are known in the art.

A non-functional SGD as used herein referes to a SGD which is not capable of converting strictosidine to strictosidine aglycone, whereas in contrast, a functional SGD is capable of converting strictosidine to strictosidine aglycone. By introducing some domains of RseSGD into a non-functional SGD however, it may be possible to restore function of a non-functional SGD, as shown in the examples, thus obtaining a functional mosaic SGD.

In some embodiments, D_1 is a first amino acid sequence from a first SGD. Said first SGD may be any SGD, such as a functional or a non-functional SGD. It is preferred that said first SGD has at least 70 %, such as at least 75%, such as at least 80%, such as at least 85%, such as at least 90%, such as at least 95% identity to RseSGD of SEQ ID NO: 24.

In some embodiments, D_2 is a second amino acid sequence from a second SGD. Said second SGD may be any SGD, such as a functional or a non-functional SGD. It is preferred that said second SGD has at least 70 %, such as at least 75%, such as at least 80%, such as at least 85%, such as at least 90%, such as at least 95% identity to RseSGD of SEQ ID NO: 24.

Interestingly, the inventors found that domain 3 (D_3) of RseSGD consisting of an amino acid sequence of SEQ ID NO:91 is capable of rescuing the inability of a non-functional SGDs of converting strictosidine to strictosidine aglycone (see Figures 9 and 10). Thus in preferred embodiments, the mosaic SGD comprises 4 domains, of which at least one comprises or consists of domain 3 of RseSGD; this domain is set forth in SEQ ID NO: 91.

Thus, in some embodiments of the present invention, the mosaic SGD comprises a D_3, wherein said D_3 is a third amino acid sequence consisting of amino acids of SEQ ID NO:91 or a variant thereof having at least 70 %, such as at least 75%, such as at least 80%, such as at least 85%, such as at least 90% identity to SEQ ID NO: 91. In other words, said D_3 is an amio acid sequence of domain 3 of RseSGD.

In some embodiments, D_4 is a fourth amino acid sequence from a fourth SGD or an amino acid sequence consisting of amino acids of SEQ ID NO:92 or a variant thereof having at least 70 %, such as at least 75%, such as at least 80%, such as at least 85%, such as at least 90% identity to SEQ ID NO: 92. Said fourth SGD may be any SGD, such as a functional or a non-functional SGD. It is preferred that said fourth SGD has at least 70 %, such as at least 75%, such as at least 80%, such as at least 85%, such as at least 90%, such as at least 95% identity to RseSGD of SEQ ID NO: 24.

In a preferred embodiment, said mosaic SGD comprises a D_4, wherein said D_4 is a fourth amino acid sequence consisting of amino acids of SEQ ID NO:92 or a variant thereof.

Said first SGD, second SGD and fourth SGD can be the same or different, with the proviso that said first SGD, second SGD and fourth SGD are not all RseSGD. In other words, said mosaic SGD may not be an RseSGD of SEQ ID NO: 24. Thus, said first first SGD, second SGD and fourth SGD, may be of the same species or different species, however said first first SGD, second SGD and fourth SGD may not all be native to *Rauvolfia serpentina.*

The third domain of the mosaic SGD comprises or consists of the third domain of RseSGD as detailed above, and at least one of the first domain, the second domain and the fourth domain is from a second organism which is not *Rauvolfia serpentina*, for example at least one of D_1, D_2 or D_4 is from an SGD native to an organism selected from *Gelsemium sempervirens, Scedosporium apiospermum or Rauvolfia verticillata*, *Vinca minor, Tabernaemontana elegans, Amsonia hubrichtii, Ophiorrhiza pumila, Nyssa sinensis, Coffea arabica, Carapichea ipecacuanha, Handroanthus impetiginosus, Sesamum indicum, Actinidia chinensis var. chinensis, Helianthus annuus, Lactuca sativa, Ipomoea nil, Vigna unguiculata, Heliocybe sulcate, Pyricularia grisea, Lomentospora prolificans, Hydnomerulius pinastri MD-312,* and *Moniliophthora*

roreri MCA 2997 or a variant thereof – as explained above, the variant here does not need to be functional to begin with, as its activity may be rescued by the D_3 domain of RseSGD.

In some embodiments, each of D_1, D_2 and D_4 are from different SGDs, and are derived from different organisms independently selected from the group consisting of *Scedosporium apiospermum, Rauvolfia verticillata*, *Vinca minor, Tabernaemontana elegans, Amsonia hubrichtii, Ophiorrhiza pumila, Nyssa sinensis, Coffea arabica, Carapichea ipecacuanha, Handroanthus impetiginosus, Sesamum indicum, Actinidia chinensis var. chinensis, Helianthus annuus, Lactuca sativa, Ipomoea nil, Vigna unguiculata, Heliocybe sulcate, Pyricularia grisea, Lomentospora prolificans, Hydnomerulius pinastri MD-312,* and *Moniliophthora roreri MCA 299*. In such embodiments, one of D_1, D_2 and D_4 may be D_1, D_2 or D_4 from RseSGD as set forth in SEQ ID NO: 89, SEQ ID NO: 90 or SEQ ID NO: 92, respectively, or variants thereof having at least 70% identity or homology thereto.

In some embodiments, two of D_1, D_2 and D_4 are from the same SGD, and are derived from one organism and the remaining domain is from another SGD. Relevant organisms and SGDs have been described above in the section " Strictosidine-O-beta-D-glucosidase". For example, D_1 and D_2 are from one SGD from a first organism, and D_4 is from another SGD from another organism; or D_1 and D_4 are from one SGD from a first organism, and D_2 is from another SGD from another organism; or D_2 and D_4 are from one SGD from a first organism, and D_1 is from another SGD from another organism, which may be *Rauvolfia serpentina*. The first organism and the other organism may be different organisms which are independently selected from the group consisting of *Scedosporium apiospermum, Rauvolfia verticillata*, *Vinca minor, Tabernaemontana elegans, Amsonia hubrichtii, Ophiorrhiza pumila, Nyssa sinensis, Coffea arabica, Carapichea ipecacuanha, Handroanthus impetiginosus, Sesamum indicum, Actinidia chinensis var. chinensis, Helianthus annuus, Lactuca sativa, Ipomoea nil, Vigna unguiculata, Heliocybe sulcate, Pyricularia grisea, Lomentospora prolificans, Hydnomerulius pinastri MD-312,* and *Moniliophthora roreri MCA 299*.

In some embodiments, all of D_1, D_2 and D_4 are from the same SGD of the same organism, which is not *Rauvolfia serpentina*. D_1, D_2 and D_4 may be of an SGD native to an organism selected from the group consisting of *Scedosporium apiospermum,*

Rauvolfia verticillata, *Vinca minor, Tabernaemontana elegans, Amsonia hubrichtii, Ophiorrhiza pumila, Nyssa sinensis, Coffea arabica, Carapichea ipecacuanha, Handroanthus impetiginosus, Sesamum indicum, Actinidia chinensis var. chinensis, Helianthus annuus, Lactuca sativa, Ipomoea nil, Vigna unguiculata, Heliocybe sulcate, Pyricularia grisea, Lomentospora prolificans, Hydnomerulius pinastri MD-312,* and *Moniliophthora roreri MCA 299*.

Thus in some embodiments, the first, second and fourth SGD are all from the same SGD, which is not RseSGD. In other embodiments, the first and second SGD are from the same SGD and the fourth SGD is from another SGD; at least one said two SGDs is not RseSGD. In other embodiments, the first and third SGD are from the same SGD and the fourth SGD is from another SGD; at least one said two SGDs is not RseSGD. In other embodiments, the fourth and second SGD are from the same SGD and the fourth SGD is from another SGD; at least one said two SGDs is not RseSGD. In some embodiments, the first, second and fourth SGD are all from different SGDs, one of which may be RseSGD.

In one embodiment, the mosaic SGD comprises or consists of an amino acid sequence of SEQ ID NO: 93, SEQ ID NO: 94, SEQ ID NO: 95, SEQ ID NO: 96, SEQ ID NO: 97, SEQ ID NO: 98, SEQ ID NO: 99 or SEQ ID NO: 108, or variants thereof having at least 90% identity or homology thereto, such as at least 91%, such as at least 92%, such as at least 93%, such as at least 94%, such as at least 95%, such as at least 96%, such as at least 97%, such as at least 98%, such as at least 99% identity or homology thereto.

The SGD may be expressed in the microorganism by introducing a nucleic acid sequence as detailed further below, which encodes a SGD. In particular, the nucleic acid sequence is identical to or has at least 90% identity to SEQ ID NO: 1, such as at least 91%, such as at least 92%, such as at least 93%, such as at least 94%, such as at least 95%, such as at least 96%, such as at least 97%, such as at least 98%, such as at least 99%, such as 100% identity to SEQ ID NO: 1. Thus, the microorganism of the invention or the microorganism used in the methods of the invention preferably comprises at least a nucleic acid sequence identical to or having at least 90% identity to SEQ ID NO: 1.

In other embodiments, the nucleic acid sequence is identical to or has at least 90% identity to SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:68, SEQ ID NO:69, SEQ ID NO:70, SEQ ID NO: 71, SEQ ID NO:72, SEQ ID NO: 73, SEQ ID NO:74, SEQ ID NO: 75, SEQ ID NO: 76, SEQ ID NO: 77, SEQ ID NO: 78, SEQ ID NO:79, SEQ ID NO:80, SEQ ID NO:81, SEQ ID NO:82, SEQ ID NO:83, SEQ ID NO:84, SEQ ID NO:85, SEQ ID NO:86, SEQ ID NO:87, SEQ ID NO:88, SEQ ID NO:100, SEQ ID NO:101, SEQ ID NO:102, SEQ ID NO:103, SEQ ID NO:104, SEQ ID NO:105, SEQ ID NO:106 or SEQ ID NO:107 such as at least 91%, such as at least 92%, such as at least 93%, such as at least 94%, such as at least 95%, such as at least 96%, such as at least 97%, such as at least 98%, such as at least 99%, such as 100% identity to SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:68, SEQ ID NO:69, SEQ ID NO:70, SEQ ID NO: 71, SEQ ID NO:72, SEQ ID NO: 73, SEQ ID NO:74, SEQ ID NO: 75, SEQ ID NO: 76, SEQ ID NO: 77, SEQ ID NO: 78, SEQ ID NO:79, SEQ ID NO:80, SEQ ID NO:81, SEQ ID NO:82, SEQ ID NO:83, SEQ ID NO:84, SEQ ID NO:85, SEQ ID NO:86, SEQ ID NO:87, SEQ ID NO:88 SEQ ID NO:100, SEQ ID NO:101, SEQ ID NO:102, SEQ ID NO:103, SEQ ID NO:104, SEQ ID NO:105, SEQ ID NO:106 or SEQ ID NO:107.

As is known in the art, in the event that the first domain of XxxSGD used in the mosaic SGD is not a methionine, the skilled person will readily be able to introduce a start codon in the nucleic acid sequence encoding the mosaic SGD in order to ensure proper translation of the mosaic SGD. The skilled person will also know how to introduce short nucleic acid sequences corresponding to linkers separating the different domains in the mosaic SGD.

The microorganism according to the present invention, expressing a heterologous SGD or variant thereof, and/or a mosaic SGD or variant thereof, is capable of converting strictosidine to strictosidine aglycone.

The conversion of strictosidine to strictosidine aglycone, may be measured directly by the amount of strictosidine aglycone as known in the art, or surrogate measure of the conversion of strictosidine to strictosidine aglycone may be measured as known in the art. Because strictosidine aglycone is highgly reactive, indirect determination of strictosidine aglycone may be preferred. For example, colorimetric assays to follow

strictosidine consumption as described in Geerlings et al., 2000, may be used. The disappearance of strictosidine may also be monitored by UV, as described in Guirimand et al., 2010, or the general β-glucosidase activity in the cells may be measured, e.g. by UV detection of a synthetic substrate such as 4-methylumbelliferyl-β-D-glucoside (Guirimand et al., 2010).

Thus, to determine whether a SGD is capable of converting strictosidine to strictosidine aglycone, the person skilled in the art could use any of said methods, or could use high-precision mass spectrometry to detect the accurate mass of strictosidine aglycone after cultivation of a strain expressing an SGD or an enzyme suspected of having SGD activity in a medium; the cell is either provided with strictosidine in the medium or it has been engineered and can synthesise strictosidine. The strictosidine aglycone can be detected directly in the medium or in a pellet, after centrifugation of the culture broth. Alternatively, the appearance of other products, downstream of strictosidine aglycone, for example tetrahydroalstonine, can be monitored; such products will only form in the presence of a functional SGD, strictosidine, and an enzyme capable of using strictosidine aglycone, as described in e.g. Stavrinides et al., 2015.

Strictosidine synthase (STR)

Strictosidine may be provided to the microorganism, for example as part of the medium the cell is incubated in. In some embodiments, however, the microorganism is engineered and is capable of synthesising strictosidine from secologanin and tryptamine.

Thus in some embodiments the microorganism expresses a heterologous strictosidine synthase having an EC number EC 4.3.3.2. Such enzymes catalyse a Pictet-Spengler reaction between the aldehyde group of secologanin and the amino group of tryptamine to yield strictosidine.

Thus microorganisms expressing a heterologous STR are capable of converting secologanin and tryptamine to strictosidine.

In some embodiments, the STR is the STR native to *Catharanthus roseus* or a functional variant thereof which retains the ability to convert secologanin and tryptamine to strictosidine. Thus in some embodiments, the STR is CroSTR as set forth

in SEQ ID NO: 30 or a variant thereof having at least 90%, such as at least 91%, such as at least 92%, such as at least 93%, such as at least 94%, such as at least 95%, such as at least 96%, such as at least 97%, such as at least 98%, such as at least 99%, such as 100% identity to SEQ ID NO: 30.

Thus, in some embodiments, the microorganism expresses RseSGD as set forth in SEQ ID NO: 24 and CroSTR as set forth in SEQ ID NO: 30, or functional variants thereof having at least 90%, such as at least 91%, such as at least 92%, such as at least 93%, such as at least 94%, such as at least 95%, such as at least 96%, such as at least 97%, such as at least 98%, such as at least 99%, such as 100% identity thereto. In some embodiments, the microorganism expresses GseSGD as set forth in SEQ ID NO: 25 and CroSTR as set forth in SEQ ID NO: 30, or functional variants thereof having at least 90%, such as at least 91%, such as at least 92%, such as at least 93%, such as at least 94%, such as at least 95%, such as at least 96%, such as at least 97%, such as at least 98%, such as at least 99%, such as 100% identity thereto. In some embodiments, the microorganism expresses SapSGD as set forth in SEQ ID NO: 26 and CroSTR as set forth in SEQ ID NO: 30, or functional variants thereof having at least 90%, such as at least 91%, such as at least 92%, such as at least 93%, such as at least 94%, such as at least 95%, such as at least 96%, such as at least 97%, such as at least 98%, such as at least 99%, such as 100% identity thereto. In some embodiments, the microorganism expresses RveSGD as set forth in SEQ ID NO: 27 and CroSTR as set forth in SEQ ID NO: 30, or functional variants thereof having at least 90%, such as at least 91%, such as at least 92%, such as at least 93%, such as at least 94%, such as at least 95%, such as at least 96%, such as at least 97%, such as at least 98%, such as at least 99%, such as 100% identity thereto.

The STR may be expressed in the microorganism by introducing a nucleic acid sequence as detailed further below, which encodes an STR. In particular, the nucleic acid sequence is identical to or has at least 90% identity to SEQ ID NO: 7, such as at least 91%, such as at least 92%, such as at least 93%, such as at least 94%, such as at least 95%, such as at least 96%, such as at least 97%, such as at least 98%, such as at least 99%, such as 100% identity to SEQ ID NO: 7.

Tetrahydroalstonine synthase, heteroyohimbine synthase

In addition to the above, the microorganism may be further engineered so that it can produce tetrahydroalstonine.

In some embodiments, the microorganism expresses an SGD and optionally an STR, and further expresses a heterologous tetrahydroalstonine synthase (THAS), which is not natively present in the cell. Tetrahydroalstonine synthase has an EC number EC 1.-.-.- and catalyses conversion of strictosidine aglycone to tetrahydroalstonine. The microorganism when expressing a THAS is thus able to convert strictosidine aglycone to tetrahydroalstonine, thus producing tetrahydroalstonine.

In some embodiments, the microorganism expresses an SGD and optionally an STR, and further expresses a heteroyohimbine synthase (HYS), which is not natively present in the cell. Heteroyohimbine synthase has an EC number EC 1.-.-.- and catalyses conversion of strictosidine aglycone to tetrahydroalstonine, ajmalicine, or mayumbine. The microorganism when expressing an HYS is thus able to convert strictosidine aglycone to tetrahydroalstonine, ajmalicine, or mayumbine, thus producing tetrahydroalstonine.

In some embodiments, the microorganism expresses a SGD and optionally an STR and further expresses a THAS and an HYS.

In preferred embodiments, the THAS is the THAS native to *Catharanthus roseus* or a functional variant thereof which retains the ability to convert strictosidine aglycone to tetrahydroalstonine. Thus in some embodiments, the THAS is CroTHAS as set forth in SEQ ID NO: 28 or a functional variant thereof having at least 90%, such as at least 91%, such as at least 92%, such as at least 93%, such as at least 94%, such as at least 95%, such as at least 96%, such as at least 97%, such as at least 98%, such as at least 99%, such as 100% identity to SEQ ID NO: 28.

The THAS may be expressed in the microorganism by introducing a nucleic acid sequence as detailed further below, which encodes a THAS. In particular, the nucleic acid sequence is identical to or has at least 90% identity to SEQ ID NO: 5, such as at least 91%, such as at least 92%, such as at least 93%, such as at least 94%, such as

at least 95%, such as at least 96%, such as at least 97%, such as at least 98%, such as at least 99%, such as 100% identity to SEQ ID NO: 5.

In other preferred embodiments, the HYS is the HYS native to *Catharanthus roseus* or a functional variant thereof which retains the ability to convert strictosidine aglycone to tetrahydroalstonine, ajmalicine, or mayumbine. Thus in some embodiments, the HYS is CroHYS as set forth in SEQ ID NO: 46 or variant thereof having at least 90%, such as at least 91%, such as at least 92%, such as at least 93%, such as at least 94%, such as at least 95%, such as at least 96%, such as at least 97%, such as at least 98%, such as at least 99%, such as 100% identity to SEQ ID NO: 46.

The HYS may be expressed in the microorganism by introducing a nucleic acid sequence as detailed further below, which encodes an HYS. In particular, the nucleic acid sequence is identical to or has at least 90% to SEQ ID NO: 23, such as at least 91%, such as at least 92%, such as at least 93%, such as at least 94%, such as at least 95%, such as at least 96%, such as at least 97%, such as at least 98%, such as at least 99%, such as 100% identity to SEQ ID NO: 23.

In some embodiments, the microorganism expresses CroHYS and/or CroTHAS or functional variants thereof having at least 90%, such as at least 91%, such as at least 92%, such as at least 93%, such as at least 94%, such as at least 95%, such as at least 96%, such as at least 97%, such as at least 98%, such as at least 99%, such as 100% identity to SEQ ID NO: 46 and/or SEQ ID NO: 28.

The microorganism expressing THAS and/or HYS further expresses an SGD as described herein, in particular RseSGD as set forth in SEQ ID NO: 24, GseSGD as set forth in SEQ ID NO: 25, SapSGD as set forth in SEQ ID NO: 26, or RveSGD as set forth in SEQ ID NO: 27, or functional variants thereof having at least 90% identity thereto.

The cell may also further express an STR as described herein, in particular CroSTR as set forth in SEQ ID NO: 30, or a functional variant thereof having at least 90% identity thereto. In some embodiments, the microorganism thus also expresses RseSGD as set forth in SEQ ID NO: 24 and CroSTR as set forth in SEQ ID NO: 30; GseSGD as set forth in SEQ ID NO: 25 and CroSTR as set forth in SEQ ID NO: 30; SapSGD as set

forth in SEQ ID NO: 26 and CroSTR as set forth in SEQ ID NO: 30; or RveSGD as set forth in SEQ ID NO: 27 and CroSTR as set forth in SEQ ID NO: 30, or functional variants thereof having at least 90% identity thereto.

Sarpargan bridge enzyme (SBE)

In addition to the above, the microorganism may be further engineered so that it can produce a heteroyohimbine, in particular alstonine and serpentine. Heteroyohimbines are a prevalent subclass of the monoterpene indole alkaloids, which are found in many plant species, primarily from the *Apocynaceae* and *Rubiaceae* families. Examples of heteroyohimbines include the α1-adrenergic receptor antagonist ajmalicine, and the benzodiazepine receptor ligand mayumbine (19-epi-ajmalicine). Oxidized β-carboline heteroyohimbines also exhibit potent pharmacological activity: serpentine has shown topoisomerase inhibition activity and alstonine has been shown to interact with 5-HT2A/C receptors and may act as an anti-psychotic agent. In addition, heteroyohimbines are biosynthetic precursors of many oxindole alkaloids, which also display a wide range of biological activities.

In some embodiments, the microorganism expresses an SGD and optionally an STR, and further expresses a heterologous sarpargan bridge enzyme (SBE), which is not natively present in the cell. This enzyme has an EC number EC 1.14.14.- and catalyses conversion of tetrahydroalstonine and ajmalicine to the corresponding alstonine and serpentine, respectively, or converts by cyclization the strictosidine-derived geissoschizine to the sarpagan alkaloid polyneuridine aldehyde. The microorganism when expressing an SBE is thus able to convert tetrahydroalstonine to alstonine and serpentine. In embodiments where the cell is capable of producing ajmalicine, the microorganism when expressing an SBE is able to convert tetrahydroalstonine and ajmalicine to alstonine and serpentine.

In preferred embodiments, the SBE is the SBE native to *Gelsemium sempervirens* or a functional variant thereof which retains the ability to convert tetrahydroalstonine and ajmalicine to alstonine and serpentine. Thus in some embodiments, the SBE is GseSBE as set forth in SEQ ID NO: 29 or a functional variant thereof having at least 90%, such as at least 91%, such as at least 92%, such as at least 93%, such as at least 94%, such as at least 95%, such as at least 96%, such as at least 97%, such as at least 98%, such as at least 99%, such as 100% identity to SEQ ID NO: 29.

The SBE may be expressed in the microorganism by introducing a nucleic acid sequence as detailed further below, which encodes an SBE. In particular, the nucleic acid sequence is identical to or has at least 90% identity to SEQ ID NO: 6, such as at least 91%, such as at least 92%, such as at least 93%, such as at least 94%, such as at least 95%, such as at least 96%, such as at least 97%, such as at least 98%, such as at least 99%, such as 100% identity to SEQ ID NO: 6.

The microorganism also expresses a SGD as described herein, in particular RseSGD as set forth in SEQ ID NO: 24, GseSGD as set forth in SEQ ID NO: 25, SapSGD as set forth in SEQ ID NO: 26, or RveSGD as set forth in SEQ ID NO: 27, or functional variants thereof having at least 90% identity thereto.

The cell may also further express an STR as described herein, in particular CroSTR as set forth in SEQ ID NO: 30, or a functional variant thereof having at least 90% identity thereto. In some embodiments, the microorganism thus also expresses RseSGD as set forth in SEQ ID NO: 24 and CroSTR as set forth in SEQ ID NO: 30; GseSGD as set forth in SEQ ID NO: 25 and CroSTR as set forth in SEQ ID NO: 30; SapSGD as set forth in SEQ ID NO: 26 and CroSTR as set forth in SEQ ID NO: 30; or RveSGD as set forth in SEQ ID NO: 27 and CroSTR as set forth in SEQ ID NO: 30, or functional variants thereof having at least 90% identity thereto.

The microorganism may also express a THAS and/or an HYS as described herein, in particular the microorganism expresses CroHYS and/or CroTHAS or functional variants thereof having at least 90%, such as at least 91%, such as at least 92%, such as at least 93%, such as at least 94%, such as at least 95%, such as at least 96%, such as at least 97%, such as at least 98%, such as at least 99%, such as 100% identity to SEQ ID NO: 46 and SEQ ID NO: 28.

<u>NADPH-cytochrome P450 reductase, Cytochrome b5 and Geissoschizine synthase</u>
The microorganism may be further engineered so that it can produce 19*E*-geissoschizine.

In some embodiments, the microorganism expresses an SGD and optionally an STR, and further expresses a heterologous NADPH--cytochrome P450 reductase (CPR), a

heterologous Cytochrome b5 (CYB5) and a heterologous Geissoschizine synthase (GS) which are not natively present in the microorganism. NADPH--cytochrome P450 reductase has an EC number EC 1.6.2.4 and is required for electron transfer from NADP to cytochrome P450. Cytochrome b5 has an EC number EC 1.6.2.2 and is a membrane bound hemoprotein which function as an electron carrier. Geissoschizine synthase has an EC number EC 1.3.1.36 and catalyzes the reduction of strictosidine aglycone to 19*E*-geissoschizine. The microorganism when expressing CPR, CYB5 and GS is thus able to convert strictosidine aglycone to 19*E*-geissoschizine, thus producing 19*E*-geissoschizine.

In some embodiments, the microorganism expresses an SGD and optionally an STR and further expresses CPR, CYB5 and GS.

In preferred embodiments, the CPR is the CPR native to *Catharanthus roseus* or a functional variant thereof which retains the ability to transfer electrons from NADP to cytochrome P450. Thus in some embodiments, the CPR is CroCPR as set forth in SEQ ID NO: 31 or a variant thereof having at least 90%, such as at least 91%, such as at least 92%, such as at least 93%, such as at least 94%, such as at least 95%, such as at least 96%, such as at least 97%, such as at least 98%, such as at least 99%, such as 100% identity to SEQ ID NO: 31.

The CPR may be expressed in the microorganism by introducing a nucleic acid sequence as detailed further below, which encodes a CPR. In particular, the nucleic acid sequence is identical to or has at least 90% identity to SEQ ID NO: 8, such as at least 91%, such as at least 92%, such as at least 93%, such as at least 94%, such as at least 95%, such as at least 96%, such as at least 97%, such as at least 98%, such as at least 99%, such as 100% identity to SEQ ID NO: 8.

In preferred embodiments, the CYB5 is the CYB5 native to *Catharanthus roseus* or a functional variant thereof which retains the ability to function as an electron carrier. Thus in some embodiments, the CYB5 is CroCYB5 as set forth in SEQ ID NO: 32 or a variant thereof having at least 90%, such as at least 91%, such as at least 92%, such as at least 93%, such as at least 94%, such as at least 95%, such as at least 96%, such as at least 97%, such as at least 98%, such as at least 99%, such as 100% identity to SEQ ID NO: 32.

The CYB5 may be expressed in the microorganism by introducing a nucleic acid sequence as detailed further below, which encodes a CYB5. In particular, the nucleic acid sequence is identical to or has at least 90% identity to SEQ ID NO: 9, such as at least 91%, such as at least 92%, such as at least 93%, such as at least 94%, such as at least 95%, such as at least 96%, such as at least 97%, such as at least 98%, such as at least 99%, such as 100% identity to SEQ ID NO: 9.

In preferred embodiments, the GS is the GS native to *Catharanthus roseus* or a functional variant thereof which retains the ability to catalyze the reduction of strictosidine aglycone to 19*E*-geissoschizine. Thus in some embodiments, the GS is CroGS as set forth in SEQ ID NO: 33 or a variant thereof having at least 90%, such as at least 91%, such as at least 92%, such as at least 93%, such as at least 94%, such as at least 95%, such as at least 96%, such as at least 97%, such as at least 98%, such as at least 99%, such as 100% identity to SEQ ID NO: 33.

The GS may be expressed in the microorganism by introducing a nucleic acid sequence as detailed further below, which encodes a GS. In particular, the nucleic acid sequence is identical to or has at least 90% identity to SEQ ID NO: 10, such as at least 91%, such as at least 92%, such as at least 93%, such as at least 94%, such as at least 95%, such as at least 96%, such as at least 97%, such as at least 98%, such as at least 99%, such as 100% identity to SEQ ID NO: 10.

The microorganism further expresses an SGD as described herein, in particular RseSGD as set forth in SEQ ID NO: 24, GseSGD as set forth in SEQ ID NO: 25, SapSGD as set forth in SEQ ID NO: 26, or RveSGD as set forth in SEQ ID NO: 27, or functional variants thereof having at least 90% identity thereto.

The cell may also further express an STR as described herein, in particular CroSTR as set forth in SEQ ID NO: 30, or a functional variant thereof having at least 90% identity thereto. In some embodiments, the microorganism thus also expresses RseSGD as set forth in SEQ ID NO: 24 and CroSTR as set forth in SEQ ID NO: 30; GseSGD as set forth in SEQ ID NO: 25 and CroSTR as set forth in SEQ ID NO: 30; SapSGD as set forth in SEQ ID NO: 26 and CroSTR as set forth in SEQ ID NO: 30; or RveSGD as set

forth in SEQ ID NO: 27 and CroSTR as set forth in SEQ ID NO: 30, or functional variants thereof having at least 90% identity thereto.

Geissoschizine oxidase, Redox1 and Redox2

The microorganism may be further engineered so that it can produce stemmadenine. The microorganism may be as described herein above. In some embodiments, the microorganism is a yeast cell. In other embodiments the microorganism is a bacterial cell.

In some embodiments, the microorganism expresses an SGD and optionally an STR, CPR, CYB5 and GS and further expresses a Geissoschizine oxidase (GO), a Redox1 and a Redox2, which are not natively present in the cell. Geissoschizine oxidase has an EC number EC 1.14.14.- and catalyzes the oxidation of 19E-geissoschizine to produce a short-lived MIA unstable intermediate which can be oxidized either by Redox1 and Redox2 to produce stemmadenine and 16S/R-deshydroxymethylstemmadenine (16S/R-DHS) or by spontaneous conversion to akuammicine. Redox1 has a EC number EC 1.14.14.- and catalyses the first of two oxidation steps that the converts the unstable product resulting from oxidation of 19E-geissoschizine by geissoschizine oxidase (GO) to stemmadenine. Redox2 has an EC number EC 1.7.1.- and catalyses the second of two oxidation steps that the converts the unstable product resulting from oxidation of 19E-geissoschizine by geissoschizine oxidase (GO) to stemmadenine. The microorganism when expressing GO, Redox1 and Redox2 is thus able to convert 19*E*-geissoschizine to stemmadenine, thus producing 19*E*- stemmadenine.

In some embodiments, the microorganism expresses an SGD and optionally an STR, CPR, CYB5 and GS and further expresses GO, Redox1 and Redox2.

In preferred embodiments, the GO is the GO native to *Catharanthus roseus* or a functional variant thereof which retains the ability to catalyze the oxidation of 19E-geissoschizine to produce a short-lived MIA unstable intermediate which can be oxidized either by Redox1 and Redox2 to produce stemmadenine. Thus in some embodiments, the GO is CroGO as set forth in SEQ ID NO: 34 or a variant thereof having at least 90%, such as at least 91%, such as at least 92%, such as at least 93%, such as at least 94%, such as at least 95%, such as at least 96%, such as at least

97%, such as at least 98%, such as at least 99%, such as 100% identity to SEQ ID NO: 34.

The GO may be expressed in the microorganism by introducing a nucleic acid sequence as detailed further below, which encodes a GO. In particular, the nucleic acid sequence is identical to or has at least 90% identity to SEQ ID NO: 11, such as at least 91%, such as at least 92%, such as at least 93%, such as at least 94%, such as at least 95%, such as at least 96%, such as at least 97%, such as at least 98%, such as at least 99%, such as 100% identity to SEQ ID NO: 11.

In preferred embodiments, the Redox1 is the Redox1 native to *Catharanthus roseus* or a functional variant thereof which retains the ability to catalyse the first of two oxidation steps that the converts the unstable product resulting from oxidation of 19E-geissoschizine by geissoschizine oxidase (GO) to stemmadenine. Thus in some embodiments, the Redox1 is CroRedox1 as set forth in SEQ ID NO: 35 or a variant thereof having at least 90%, such as at least 91%, such as at least 92%, such as at least 93%, such as at least 94%, such as at least 95%, such as at least 96%, such as at least 97%, such as at least 98%, such as at least 99%, such as 100% identity to SEQ ID NO: 35.

The Redox1 may be expressed in the microorganism by introducing a nucleic acid sequence as detailed further below, which encodes a Redox1. In particular, the nucleic acid sequence is identical to or has at least 90% identity to SEQ ID NO: 12, such as at least 91%, such as at least 92%, such as at least 93%, such as at least 94%, such as at least 95%, such as at least 96%, such as at least 97%, such as at least 98%, such as at least 99%, such as 100% identity to SEQ ID NO: 12.

In preferred embodiments, the Redox2 is the Redox2 native to *Catharanthus roseus* or a functional variant thereof which retains the ability to catalyse the second of two oxidation steps that the converts the unstable product resulting from oxidation of 19E-geissoschizine by geissoschizine oxidase (GO) to stemmadenine. Thus in some embodiments, the Redox2 is CroRedox2 as set forth in SEQ ID NO: 36 or a variant thereof having at least 90%, such as at least 91%, such as at least 92%, such as at least 93%, such as at least 94%, such as at least 95%, such as at least 96%, such as

at least 97%, such as at least 98%, such as at least 99%, such as 100% identity to SEQ ID NO: 36.

The Redox2 may be expressed in the microorganism by introducing a nucleic acid sequence as detailed further below, which encodes a Redox2. In particular, the nucleic acid sequence is identical to or has at least 90% identity to SEQ ID NO: 13, such as at least 91%, such as at least 92%, such as at least 93%, such as at least 94%, such as at least 95%, such as at least 96%, such as at least 97%, such as at least 98%, such as at least 99%, such as 100% identity to SEQ ID NO: 13.

The microorganism further expresses an SGD as described herein, in particular RseSGD as set forth in SEQ ID NO: 24, GseSGD as set forth in SEQ ID NO: 25, SapSGD as set forth in SEQ ID NO: 26, or RveSGD as set forth in SEQ ID NO: 27, or functional variants thereof having at least 90% identity thereto.

The cell may also further express an STR as described herein, in particular CroSTR as set forth in SEQ ID NO: 30, or a functional variant thereof having at least 90% identity thereto. In some embodiments, the microorganism thus also expresses RseSGD as set forth in SEQ ID NO: 24 and CroSTR as set forth in SEQ ID NO: 30; GseSGD as set forth in SEQ ID NO: 25 and CroSTR as set forth in SEQ ID NO: 30; SapSGD as set forth in SEQ ID NO: 26 and CroSTR as set forth in SEQ ID NO: 30; or RveSGD as set forth in SEQ ID NO: 27 and CroSTR as set forth in SEQ ID NO: 30, or functional variants thereof having at least 90% identity thereto.

Stemmadenine O-acetyltransferase

The microorganism may be further engineered so that it can produce O-acetylstemmadenine.

In some embodiments, the microorganism expresses an SGD and optionally an STR, CPR, CYB5, GS, GO, Redox1 and Redox2, and further expresses Stemmadenine O-acetyltransferase which is not natively present in the cell. Stemmadenine O-acetyltransferase has an EC number EC 1.7.1.- and catalyzes the acetylation of stemmadenine to O-acetylstemmadenine. The microorganism when expressing SAT is thus able to convert stemmadenine to O-acetylstemmadenine, thus producing O-acetylstemmadenine.

In some embodiments, the microorganism expresses an SGD and optionally an STR, CPR, CYB5, GS GO, Redox1 and Redox2 and further expresses SAT.

In preferred embodiments, the SAT is the SAT native to *Catharanthus roseus* or a functional variant thereof which retains the ability to convert stemmadenine to O-acetylstemmadenine. Thus in some embodiments, the SAT is CroSAT as set forth in SEQ ID NO: 37 or a variant thereof having at least 90%, such as at least 91%, such as at least 92%, such as at least 93%, such as at least 94%, such as at least 95%, such as at least 96%, such as at least 97%, such as at least 98%, such as at least 99%, such as 100% identity identityto SEQ ID NO: 37.

The SAT may be expressed in the microorganism by introducing a nucleic acid sequence as detailed further below, which encodes a SAT. In particular, the nucleic acid sequence is identical to or has at least 90% identity to SEQ ID NO: 14, such as at least 91%, such as at least 92%, such as at least 93%, such as at least 94%, such as at least 95%, such as at least 96%, such as at least 97%, such as at least 98%, such as at least 99%, such as 100% identity to SEQ ID NO: 14.

The microorganism further expresses an SGD as described herein, in particular RseSGD as set forth in SEQ ID NO: 24, GseSGD as set forth in SEQ ID NO: 25, SapSGD as set forth in SEQ ID NO: 26, or RveSGD as set forth in SEQ ID NO: 27, or functional variants thereof having at least 90% identity thereto.

The cell may also further express an STR as described herein, in particular CroSTR as set forth in SEQ ID NO: 30, or a functional variant thereof having at least 90% identity thereto. In some embodiments, the microorganism thus also expresses RseSGD as set forth in SEQ ID NO: 24 and CroSTR as set forth in SEQ ID NO: 30; GseSGD as set forth in SEQ ID NO: 25 and CroSTR as set forth in SEQ ID NO: 30; SapSGD as set forth in SEQ ID NO: 26 and CroSTR as set forth in SEQ ID NO: 30; or RveSGD as set forth in SEQ ID NO: 27 and CroSTR as set forth in SEQ ID NO: 30, or functional variants thereof having at least 90% identity thereto.

O-acetylstemmadenine oxidase

The microorganism may be further engineered so that it can produce dihydroprecondylocarpine acetate.

In some embodiments, the microorganism expresses an SGD and optionally an STR, CPR, CYB5, GS, GO, Redox1, Redox2 and SAT, and further expresses O-acetylstemmadenine oxidase (PAS) which is not natively present in the cell. O-acetylstemmadenine oxidase has an EC number EC 1.21.3.- and converts O-acetylstemmadenine to precondylocarpine acetate. The microorganism when expressing PAS is thus able to convert O-acetylstemmadenine to precondylocarpine acetate, thus producing precondylocarpine acetate.

In some embodiments, the microorganism expresses an SGD and optionally an STR, CPR, CYB5, GS GO, Redox1, Redox2, and SAT and further expresses PAS.

In preferred embodiments, the PAS is the PAS native to *Catharanthus roseus* or a functional variant thereof which retains the ability to convert O-acetylstemmadenine to precondylocarpine acetate. Thus in some embodiments, the PAS is CroPAS as set forth in SEQ ID NO: 38 or a variant thereof having at least 90%, such as at least 91%, such as at least 92%, such as at least 93%, such as at least 94%, such as at least 95%, such as at least 96%, such as at least 97%, such as at least 98%, such as at least 99%, such as 100% identity to SEQ ID NO: 38.

The PAS may be expressed in the microorganism by introducing a nucleic acid sequence as detailed further below, which encodes a PAS. In particular, the nucleic acid sequence is identical to or has at least 90% identity to SEQ ID NO: 15, such as at least 91%, such as at least 92%, such as at least 93%, such as at least 94%, such as at least 95%, such as at least 96%, such as at least 97%, such as at least 98%, such as at least 99%, such as 100% identity to SEQ ID NO: 15.

The microorganism further expresses an SGD as described herein, in particular RseSGD as set forth in SEQ ID NO: 24, GseSGD as set forth in SEQ ID NO: 25, SapSGD as set forth in SEQ ID NO: 26, or RveSGD as set forth in SEQ ID NO: 27, or functional variants thereof having at least 90% identity thereto.

The cell may also further express an STR as described herein, in particular CroSTR as set forth in SEQ ID NO: 30, or a functional variant thereof having at least 90% identity thereto. In some embodiments, the microorganism thus also expresses RseSGD as set forth in SEQ ID NO: 24 and CroSTR as set forth in SEQ ID NO: 30; GseSGD as set forth in SEQ ID NO: 25 and CroSTR as set forth in SEQ ID NO: 30; SapSGD as set forth in SEQ ID NO: 26 and CroSTR as set forth in SEQ ID NO: 30; or RveSGD as set forth in SEQ ID NO: 27 and CroSTR as set forth in SEQ ID NO: 30, or functional variants thereof having at least 90% identity thereto.

Dehydroprecondylocarpine acetate synthase

The microorganism may be further engineered so that it can produce dihydroprecondylocarpine acetate.

In some embodiments, the microorganism expresses an SGD and optionally an STR, CPR, CYB5, GS, GO, Redox1, Redox2, SAT and PAS, and further expresses dihydroprecondylocarpine acetate synthase (DPAS) which is not natively present in the cell. Dihydroprecondylocarpine acetate synthase has an EC number EC 1.1.1.- and converts precondylocarpine acetate to dihydroprecondylocarpine acetate. The microorganism when expressing DPAS is thus able to convert precondylocarpine acetate to dihydroprecondylocarpine acetate, thus producing dihydroprecondylocarpine acetate.

In some embodiments, the microorganism expresses an SGD and optionally an STR, CPR, CYB5, GS GO, Redox1, Redox2, SAT and PAS and further expresses DPAS.

In preferred embodiments, the DPAS is the DPAS native to *Catharanthus roseus* or a functional variant thereof which retains the ability to convert precondylocarpine acetate to dihydroprecondylocarpine acetate. Thus in some embodiments, the DPAS is CroDPAS as set forth in SEQ ID NO: 39 or a variant thereof having at least 90%, such as at least 91%, such as at least 92%, such as at least 93%, such as at least 94%, such as at least 95%, such as at least 96%, such as at least 97%, such as at least 98%, such as at least 99%, such as 100% identity to SEQ ID NO: 39.

The DPAS may be expressed in the microorganism by introducing a nucleic acid sequence as detailed further below, which encodes a DPAS. In particular, the nucleic

acid sequence is identical to or has at least 90% identity to SEQ ID NO: 16, such as at least 91%, such as at least 92%, such as at least 93%, such as at least 94%, such as at least 95%, such as at least 96%, such as at least 97%, such as at least 98%, such as at least 99%, such as 100% identity to SEQ ID NO: 16.

The microorganism further expresses an SGD as described herein, in particular RseSGD as set forth in SEQ ID NO: 24, GseSGD as set forth in SEQ ID NO: 25, SapSGD as set forth in SEQ ID NO: 26, or RveSGD as set forth in SEQ ID NO: 27, or functional variants thereof having at least 90% identity thereto.

The cell may also further express an STR as described herein, in particular CroSTR as set forth in SEQ ID NO: 30, or a functional variant thereof having at least 90% identity thereto. In some embodiments, the microorganism thus also expresses RseSGD as set forth in SEQ ID NO: 24 and CroSTR as set forth in SEQ ID NO: 30; GseSGD as set forth in SEQ ID NO: 25 and CroSTR as set forth in SEQ ID NO: 30; SapSGD as set forth in SEQ ID NO: 26 and CroSTR as set forth in SEQ ID NO: 30; or RveSGD as set forth in SEQ ID NO: 27 and CroSTR as set forth in SEQ ID NO: 30, or functional variants thereof having at least 90% identity thereto.

Tabersonine synthase

The microorganism may be further engineered so that it can produce tabersonine.

In some embodiments, the microorganism expresses an SGD and optionally an STR, CPR, CYB5, GS, GO, Redox1, Redox2, SAT, PAS and DPAS, and further expresses Tabersonine synthase (TS) which is not natively present in the cell. Tabersonine synthase has an EC number EC 4.-.-.- and converts dihydroprecondylocarpine acetate to tabersonine. The microorganism when expressing TS is thus able to convert dihydroprecondylocarpine acetate to tabersonine, thus producing tabersonine.

In some embodiments, the microorganism expresses an SGD and optionally an STR, CPR, CYB5, GS GO, Redox1, Redox2, SAT, PAS and DPAS, and further expresses TS.

In preferred embodiments, the TS is the TS native to *Catharanthus roseus* or a functional variant thereof which retains the ability to convert dihydroprecondylocarpine

acetate to tabersonine. Thus in some embodiments, the TS is CroTS as set forth in SEQ ID NO: 40 or a variant thereof having at least 90%, such as at least 91%, such as at least 92%, such as at least 93%, such as at least 94%, such as at least 95%, such as at least 96%, such as at least 97%, such as at least 98%, such as at least 99%, such as 100% identity to SEQ ID NO: 40.

The TS may be expressed in the microorganism by introducing a nucleic acid sequence as detailed further below, which encodes a TS. In particular, the nucleic acid sequence is identical to or has at least 90% identity to SEQ ID NO: 17, such as at least 91%, such as at least 92%, such as at least 93%, such as at least 94%, such as at least 95%, such as at least 96%, such as at least 97%, such as at least 98%, such as at least 99%, such as 100% identity to SEQ ID NO: 17.

The microorganism further expresses an SGD as described herein, in particular RseSGD as set forth in SEQ ID NO: 24, GseSGD as set forth in SEQ ID NO: 25, SapSGD as set forth in SEQ ID NO: 26, or RveSGD as set forth in SEQ ID NO: 27, or functional variants thereof having at least 90% identity thereto.

The cell may also further express an STD as described herein, in particular CroSTR as set forth in SEQ ID NO: 30, or a functional variant thereof having at least 90% identity thereto. In some embodiments, the microorganism thus also expresses RseSGD as set forth in SEQ ID NO: 24 and CroSTR as set forth in SEQ ID NO: 30; GseSGD as set forth in SEQ ID NO: 25 and CroSTR as set forth in SEQ ID NO: 30; SapSGD as set forth in SEQ ID NO: 26 and CroSTR as set forth in SEQ ID NO: 30; or RveSGD as set forth in SEQ ID NO: 27 and CroSTR as set forth in SEQ ID NO: 30, or functional variants thereof having at least 90% identity thereto.

Catharanthine synthase

The microorganism may be further engineered so that it can produce catharanthine.

In some embodiments, the microorganism expresses an SGD and optionally an STR, CPR, CYB5, GS, GO, Redox1, Redox2, SAT, PAS and DPAS, and further expresses Catharanthine synthase (CS) which is not natively present in the cell. Catharanthine synthase has an EC number EC 4.-.-.- and converts dihydroprecondylocarpine acetate

to catharanthine. The microorganism when expressing CS is thus able to convert dihydroprecondylocarpine acetate to catharanthine, thus producing catharanthine.

In some embodiments, the microorganism expresses an SGD and optionally an STR, CPR, CYB5, GS GO, Redox1, Redox2, SAT, PAS and DPAS, and further expresses CS. Optionally the microorganism also expresses TS.

In preferred embodiments, the CS is the CS native to *Catharanthus roseus* or a functional variant thereof which retains the ability to convert dihydroprecondylocarpine acetate to catharanthine. Thus in some embodiments, the CS is CroCS as set forth in SEQ ID NO: 41 or a variant thereof having at least 90%, such as at least 91%, such as at least 92%, such as at least 93%, such as at least 94%, such as at least 95%, such as at least 96%, such as at least 97%, such as at least 98%, such as at least 99%, such as 100% identity to SEQ ID NO: 41.

The CS may be expressed in the microorganism by introducing a nucleic acid sequence as detailed further below, which encodes a CS. In particular, the nucleic acid sequence is identical to or has at least 90% identity to SEQ ID NO: 18, such as at least 91%, such as at least 92%, such as at least 93%, such as at least 94%, such as at least 95%, such as at least 96%, such as at least 97%, such as at least 98%, such as at least 99%, such as 100% identity to SEQ ID NO: 18.

The microorganism further expresses an SGD as described herein, in particular RseSGD as set forth in SEQ ID NO: 24, GseSGD as set forth in SEQ ID NO: 25, SapSGD as set forth in SEQ ID NO: 26, or RveSGD as set forth in SEQ ID NO: 27, or functional variants thereof having at least 90% identity thereto.

The cell may also further express an STR as described herein, in particular CroSTR as set forth in SEQ ID NO: 30, or a functional variant thereof having at least 90% identity thereto. In some embodiments, the microorganism thus also expresses RseSGD as set forth in SEQ ID NO: 24 and CroSTR as set forth in SEQ ID NO: 30; GseSGD as set forth in SEQ ID NO: 25 and CroSTR as set forth in SEQ ID NO: 30; SapSGD as set forth in SEQ ID NO: 26 and CroSTR as set forth in SEQ ID NO: 30; or RveSGD as set forth in SEQ ID NO: 27 and CroSTR as set forth in SEQ ID NO: 30, or functional variants thereof having at least 90% identity thereto.

Methods for producing strictosidine aglycone and monoterpenoid indole alkaloids

The microorganisms described herein are useful as platform for producing plant compounds, in particular strictosidine aglycone and monoterpenoid indole alkaloids (MIAs).

Herein is provided a method of producing strictosidine aglycone in a microorganism, said method comprising the steps of:

a) providing a microorganism, said cell expressing:
   a strictosidine-beta-glucosidase (SGD), capable of converting strictosidine to strictosidine aglycone;
b) incubating said microorganism in a medium comprising strictosidine or a substrate which can be converted to strictosidine by said microorganism;
c) optionally, recovering the strictosidine aglycone;
d) optionally, further converting the strictosidine aglycone to monoterpenoid indole alkaloids.

The microorganism may be as described herein above. Thus, the microorganism may be any microorganism.

Thus, in one embodiment, the microorganism is selected from the group consisting of bacteria, archaea, yeasts, fungi, protozoa, algae, and viruses. In another embodiment, the microorganism is selected from the group consisting of bacteria, archaea, yeasts, fungi, protozoa and algae. In another embodiment, the microorganism is selected from the group consisting of bacteria, archaea, yeasts, fungi, and algae. In another embodiment, the microorganism is selected from the group consisting of bacteria, archaea yeasts and fungi. In another embodiment, the microorganism is selected from bacteria, yeasts and fungi. In another embodiment, the microorganism is selected from bacteria or yeasts. In a preferred embodiment, the microorganism is a bacteria or a yeast.

In some embodiments, the microorganism is a bacteria. In one embodiment, the genus of said bacteria is selected from *Escherichia*, *Corynebacterium, Pseudomonas,*

Bacillus, Lactococcus, Lactobacillus, Halomonas, Bifidobacterium and *Enterococcus*. In preferred embodiments, the genus of said bacteria is *Escherichia*. In another embodiment, the microorganism may be selected from the group consisting of *Escherichia, Corynebacterium glutamicum, Pseudomonas putida, Bacillus subtilis, Lactococcus bacillus, Halomonas elongate, Bifidobacterium infantis* and *Enterococcus faecali.* In preferred embodiments, the micororganims is an *Escherichia.*

In some embodiments, the microorganism is a yeast. In some embodiments, the microorganism is a cell from a GRAS (Generally Recognized As Safe) organism or a non-pathogenic organism or strain. In some embodiments, the genus of said yeast is selected from *Saccharomyces*, *Pichia*, *Yarrowia*, *Kluyveromyces*, *Candida, Rhodotorula, Rhodosporidium, Cryptococcus, Trichosporon* and *Lipomyces.* In preferred embodiments, the genus of said yeast is *Saccharomyces.*

The microorganism may be selected from the group consisting of *Saccharomyces cerevisiae, Pichia pastoris, Kluyveromyces marxianus, Cryptococcus albidus, Lipomyces lipofera, Lipomyces starkeyi, Rhodosporidium toruloides, Rhodotorula glutinis, Trichosporon pullulan* and *Yarrowia lipolytica*. In preferred embodiments, the microorganism is a *Saccharomyces cerevisiae* cell.

The strictosidine aglycone produced in the cell may in some embodiments of the methods be further converted into monoterpenoid indole alkaloids. The term “further conversion” herein simply means that the produced strictosidine aglycone is transformed or converted into another compound which is a monoterpenoid indole alkaloid. The conversion may happen in vivo, i.e. within the cell, which may be capable of catalysing further conversion of the strictosidine aglycone into other compounds. The methods however may also comprise the steps of recovering the strictosidine aglycone from the microorganism or from the medium by methods known in the art, and thereafter converting the strictosidine aglycone into monoterpenoid indole alkaloids, i.e. the further conversion may be an ex vivo conversion.

Preferably, the microorganism expresses an SGD as described herein; the SGD may be a heterologous SGD or a mosaic SGD as described herein above. In preferred embodiments, the SGD is selected from RseSGD (SEQ ID NO: 24), GseSGD (SEQ ID NO: 25), SapSGD (SEQ ID NO: 26), RveSGD (SEQ ID NO: 27), VmiSGD1 (SEQ ID

NO: 47), AhuSGD (SEQ ID NO: 48), HimSGD2 (SEQ ID NO: 49), SinSGD (SEQ ID NO: 50), TelSGD (SEQ ID NO: 51), VunSGD (SEQ ID NO: 52), NsiSGD1 (SEQ ID NO: 53), LprSGD (SEQ ID NO: 54), AchSGD1 (SEQ ID NO: 55), HsuSGD (SEQ ID NO: 56), MroSGD (SEQ ID NO: 57), RseSGD2 (SEQ ID NO: 58), PgrSGD (SEQ ID NO: 59), OpuSGD (SEQ ID NO: 60), HpiSGD (SEQ ID NO: 61), HanSGD1 (SEQ ID NO: 62), AchSGD2 (SEQ ID NO: 63), HimSGD1 (SEQ ID NO: 64), IpeSGD (SEQ ID NO: 65), LsaSGD1 (SEQ ID NO: 66), or CarSGD (SEQ ID NO: 67) and functional variants thereof having at least 90%, such as at least 91%, such as at least 92%, such as at least 93%, such as at least 94%, such as at least 95%, such as at least 96%, such as at least 97%, such as at least 98%, such as at least 99%, such as 100% identity hereto.

The microorganism may be any of the microorganisms described herein. Thus, the microorganism in some embodiments expresses an SGD as described in the section "Strictosidine-O-beta-glucosidase (SGD)" and is capable of converting strictosidine to strictosidine aglycone. In some embodiments the SGD is a heterologous SGD as described in the section "Heterologous SGD or variants thereof". In some embodiments, the SGD is a mosaic SGD as described in the section "Mosaic SGD or variants thereof". The mosaic SGD is as described above and comprises an amino acid sequence having the general formula

$$D_1\text{-}D_2\text{-}D_3\text{-}D_4$$

wherein D_1 is a first amino acid sequence from a first SGD,
wherein D_2 is a second amino acid sequence from a second SGD,
wherein D_3 is a third amino acid sequence comprising or consisting of amino acids of SEQ ID NO:91 or a variant thereof having at least 90% identity to SEQ ID NO: 91,
wherein D_4 is a fourth amino acid sequence from a fourth SGD or an amino acid sequence consisting of amino acids of SEQ ID NO:92 or a variant thereof having at least 90% identity to SEQ ID NO: 92,
wherein said first SGD, second SGD and fourth SGD can be the same or different, with the proviso that said first SGD, second SGD and fourth SGD are not all RseSGD.

The microorganism may also express an STR as described in the section "Strictosidine synthase (STR)" and may thus be capable of synthesising strictosidine from secologanin and tryptamine. Preferably, secologanin and tryptamine are provided to

the cell, e.g. in the medium; in such embodiments, the medium need not comprise strictosidine. In other embodiments, particularly where the microorganism cannot synthesise strictosidine, strictosidine is provided to the microorganism as part of the medium.

The microorganism may be further engineered to produce tetrahydroalstonine as described in the section "Tetrahydroalstonine synthases, heteroyohimbine synthase". For example, the microorganism may express a heterologous THAS and/or a heterologous HYS.

The microorganism may be further engineered to produce a heteroyohimbine, in particular alstonine and serpentine, as described in the section "Sarpargan bridge enzyme (SBE)". For example, the microorganism may express a heterologous sarpargan bridge enzyme (SBE).

The microorganism may be further engineered to produce tabersonine and/or caranthine as described herein. In particular, the microorganism may be further engineered to synthesise 19*E*-geissoschizine as described in the section "NADPH--cytochrome P450 reductase, Cytochrome b5 and Geissoschizine synthase". For example, the microorganism may express a heterologous NADPH--cytochrome P450 reductase (CPR), a heterologous Cytochrome b5 (CYB5) and a heterologous Geissoschizine synthase (GS). The microorganism may be further engineered so that it can synthesise stemmadenine, as described in the section "Geissoschizine oxidase, Redox1 and Redox2". For example, the microorganism may express a GO, a Redox1 and a Redox2. The microorganism may be further engineered so that it can synthesise O-acetylstemmadenine as described in section "Stemmadenine O-acetyltransferase". For example, the microorganism may express SAT. The microorganism may be further engineered so that it can synthesise dihydroprecondylocarpine acetate as described in section "O-acetylstemmadenine oxidase". For example, the microorganism may express a PAS. The microorganism may be further engineered so that it can produce dihydroprecondylocarpine acetate, as described in the section "Dehydroprecondylocarpine acetate synthase". For example, the microorganism may express a DPAS. The microorganism may be further engineered so that it can produce tabersonine, as described in the section "Tabersonine synthase". For example, the microorganism expresses TS. The microorganism may be further engineered so that it

can produce catharanthine, as described in the section “Catharanthine synthase”. For example, the microorganism may express a CS.

Thus, the microorganism may be as described above, and may produce one or more of:

- strictosidine
- strictosidine aglycone
- tetrahydroalstonine
- alstonine
- tabersonine
- catharanthine

The necessary substrates for each product may be provided to the cell as part of the medium used to grow the cells. Alternatively, the substrates for each of the above products may be synthesised by the cell itself. In all cases, the microorganism is capable of synthesising strictosidine aglycone.

Each of the above products may be recovered from the medium by methods known in the art if desirable. Accordingly, the method may comprise the step of recovering one or more of:

- strictosidine
- strictosidine aglycone
- tetrahydroalstonine
- alstonine
- tabersonine
- catharanthine

In some embodiments, the medium comprises a substrate which is strictosidine. The microorganism can convert said strictosidine to strictosidine aglycone as described in detail herein above.

In some embodiments, the medium comprises strictosidine, at a concentration of at least 0.05 mM, such as at least 0.1 mM, such as at least 0.5 mM, such as at least 1 mM.

In other embodiments, the medium comprises tryptamine and secologanin, preferably at a concentration of at least 0.05 mM, such as at least 0.1 mM, such as at least 0.5 mM, such as at least 1 mM.

The present invention also related to a method of producing indole alkaloids (MIAs) in a microorganism.

Thus, herein is provided a method of producing monoterpenoid indole alkaloids (MIAs) in a microorganism, said method comprising the steps of:

i) providing a microorganism capable of converting strictosidine to tabersonine and/or catharanthine, said cell expressing:
   a strictosidine-beta-glucosidase (SGD);
   a NADPH--cytochrome P450 reductase (CPR);
   a Cytochrome b5 (CYB5);
   a Geissoschizine synthase (GS);
   a Geissoschizine oxidase (GO);
   a Redox1;
   a Redox2;
   a Stemmadenine O-acetyltransferase (SAT);
   a O-acetylstemmadenine oxidase (PAS);
   a Dehydroprecondylocarpine acetate synthase (DPAS);
   a Tabersonine synthase (TS); and/or
   a Catharanthine synthase (CS);
ii) incubating said microorganism in a medium comprising strictosidine or a substrate which can be converted to strictosidine by said microorganism;
iii) optionally, recovering the MIAs;
iv) optionally, processing the MIAs into a pharmaceutical compound,

wherein said SGD is a heterologous SGD selected from RseSGD (SEQ ID NO: 24), GseSGD (SEQ ID NO: 25), SapSGD (SEQ ID NO: 26), RveSGD (SEQ ID NO: 27), VmiSGD1 (SEQ ID NO: 47), AhuSGD (SEQ ID NO: 48), HimSGD2 (SEQ ID NO: 49), SinSGD (SEQ ID NO: 50), TelSGD (SEQ ID NO: 51), VunSGD (SEQ ID NO: 52), NsiSGD1 (SEQ ID NO: 53), LprSGD (SEQ ID NO: 54), AchSGD1 (SEQ ID NO: 55), HsuSGD (SEQ ID NO: 56), MroSGD (SEQ ID NO: 57), RseSGD2 (SEQ ID NO: 58), PgrSGD (SEQ ID NO: 59), OpuSGD (SEQ ID

NO: 60), HpiSGD (SEQ ID NO: 61), HanSGD1 (SEQ ID NO: 62), AchSGD2 (SEQ ID NO: 63), HimSGD1 (SEQ ID NO: 64), IpeSGD (SEQ ID NO: 65), LsaSGD1 (SEQ ID NO: 66), or CarSGD (SEQ ID NO: 67) or variants thereof having at least 70%, such as at least 80%, such as at least 90%, such as at least 91%, such as at least 92%, such as at least 93%, such as at least 94%, such as at least 95%, such as at least 96%, such as at least 97%, such as at least 98%, such as at least 99%, such as 100% identity thereto,

and/or;

wherein said SGD is a mosaic SGD, wherein said mosaic SGD comprises an amino acid sequence having the general formula

$$D_1\text{-}D_2\text{-}D_3\text{-}D_4$$

wherein D_1 is a first amino acid sequence from a first SGD,
wherein D_2 is a second amino acid sequence from a second SGD,
wherein D_3 is a third amino acid sequence comprising or consisting of amino acids of SEQ ID NO:91 or a variant thereof having at least 90% identity to SEQ ID NO: 91,
wherein D_4 is a fourth amino acid sequence from a fourth SGD or an amino acid sequence consisting of amino acids of SEQ ID NO:92 or a variant thereof having at least 90% identity to SEQ ID NO: 92,
wherein said first SGD, second SGD and fourth SGD can be the same or different, with the proviso that said first SGD, second SGD and fourth SGD are not all RseSGD.

The microorganism may optionally further express a strictosidine synthase (STR).

The microorganism capable of producing monoterpenoid indole alkaloids (MIAs) may be any microorgsnims as described herein under section “Deteiled description”.

Titers

The microorganisms and methods disclosed herein can be used to produce different plant-derived compounds at high titers. Strictosidine aglycone may thus be obtained with a total titer of at least 0.1 µM, such as at least 0.5 µM, such as at least 1 µM, such as at least 2 µM, such as at least 3 µM, such as at least 4 µM, such as at least 5 µM,

such as at least 6 µM, such as at least 7 µM L, such as at least 8 µM, such as at least 9 µM, such as at least 10 µM, such as at least 11 µM, such as at least 12 µM, such as at least 13 µM, such as at least 14 µM, such as at least 15 µM, such as at least 20 µM, such as at least 25 µM, such as at least 30 µM, such as at least 35 µM, such as at least 40 µM, such as at least 50 µM, or more, wherein the total titer is the sum of the intracellular strictosidine aglycone titer and the extracellular strictosidine aglycone. Indeed, the produced strictosidine aglycone may be secreted from the cell – extracellular strictosidine aglycone – or it may be retained in the cell – intracellular strictosidine aglycone.

The microorganism may be capable of producing extracellular strictosidine aglycone with a titer of at least 0.1 µM, such as at least 0.5 µM, such as at least 1 µM, such as at least 2 µM, such as at least 3 µM, such as at least 4 µM, such as at least 5 µM, such as at least 6 µM, such as at least 7 µM L, such as at least 8 µM, such as at least 9 µM, such as at least 10 µM, such as at least 11 µM, such as at least 12 µM, such as at least 13 µM, such as at least 14 µM, such as at least 15 µM, such as at least 20 µM, such as at least 25 µM, such as at least 30 µM, such as at least 35 µM, such as at least 40 µM, such as at least 50 µM, or more.

The microorganism may be capable of producing intracellular strictosidine aglycone with a titer of at least 0.1 µM, such as at least 0.5 µM, such as at least 1 µM, such as at least 2 µM, such as at least 3 µM, such as at least 4 µM, such as at least 5 µM, such as at least 6 µM, such as at least 7 µM L, such as at least 8 µM, such as at least 9 µM, such as at least 10 µM, such as at least 11 µM, such as at least 12 µM, such as at least 13 µM, such as at least 14 µM, such as at least 15 µM, such as at least 20 µM, such as at least 25 µM, such as at least 30 µM, such as at least 35 µM, such as at least 40 µM, such as at least 50 µM, or more.

Methods for determining the strictosidine aglycone titer are known in the art. For example, the cells can be lysed and the titers determined by Orbitrap Fusion Tribid MS (see example 5) to determine the intracellular or secreted strictosidine aglycone titers. The titers can also be determined by Orbitrap Fusion Tribid MS in supernatant fractions from which the cells have been removed.

The microorganism may be capable of producing tetrahydroalstonine with a titre of at least 1 μM, such as at least 2 μM, such as at least 4 μM, such as at least 6 μM, such as at least 8 μM such as at least 10 μM or more.

The microorganism may be capable of producing alstonine with a titre of at least 0.1 μM, such as at least 0.5 μM, such as at least 1 μM, such as at least 2 μM, such as at least 3 μM, such as at least 4 μM, such as at least 5 μM, such as at least 6 μM, such as at least 7 μM L, such as at least 8 μM, such as at least 9 μM, such as at least 10 μM, such as at least 11 μM, such as at least 12 μM, such as at least 13 μM, such as at least 14 μM, such as at least 15 μM, such as at least 20 μM or more.

The microorganism may be capable of producing tabersonine with a titre of at least 0.01 μM, such as at least 0.02 μM, such as at least 0.5 μM, such as at least 1 μM, such as at least 2 μM, such as at least 3 μM, such as at least 4 μM, such as at least 5 μM, such as at least 6 μM, such as at least 7 μM L, such as at least 8 μM, such as at least 9 μM, such as at least 10 μM, such as at least 11 μM, such as at least 12 μM, such as at least 13 μM, such as at least 14 μM, such as at least 15 μM, such as at least 20 μM or more.

The microorganism may be capable of producing catharanthine with a titre of at least 0.01 μM, such as at least 0.02 μM, such as at least 0.5 μM, such as at least 1 μM, such as at least 2 μM, such as at least 3 μM, such as at least 4 μM, such as at least 5 μM, such as at least 6 μM, such as at least 7 μM L, such as at least 8 μM, such as at least 9 μM, such as at least 10 μM, such as at least 11 μM, such as at least 12 μM, such as at least 13 μM, such as at least 14 μM, such as at least 15 μM, such as at least 20 μM or more.

Nucleic acids, vectors and host cells

Also disclosed herein are useful nucleic acid constructs for constructing a microorganism as described above, or useful in general in the methods described herein. Such nucleic acid constructs encode the heterologous enzymes useful for constructing the microorganisms of the invention.

It will be understood that the term “nucleic acid constructs” may refer to one nucleic acid molecule, or to a plurality of nucleic acid molecules, comprising the relevant

nucleic acid sequences. The nucleic acid construct may thus be one nucleic acid molecule, which may encode several enzymes, or it may be several nucleic acid molecules, each comprising one sequence encoding an enzyme. The relevant nucleic acid sequences may thus be comprised on one vector, or on several vectors. They may also be integrated in the genome, on one chromosome or even together in one location, or they may be integrated on different chromosomes. It is also possible to have some sequences on one or more vectors, and some integrated in the genome.

Also provided herein are nucleic acid constructs comprising a nucleic acid sequence identical to or having at least 90%, such as at least 91%, such as at least 92%, such as at least 93%, such as at least 94%, such as at least 95%, such as at least 96%, such as at least 97%, such as at least 98%, such as at least 99%, such as 100% identity to SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 3 or SEQ ID NO: 4, SEQ ID NO:68, SEQ ID NO:69, SEQ ID NO:70, SEQ ID NO: 71, SEQ ID NO:72, SEQ ID NO: 73, SEQ ID NO:74, SEQ ID NO: 75, SEQ ID NO: 76, SEQ ID NO: 77, SEQ ID NO: 78, SEQ ID NO:79, SEQ ID NO:80, SEQ ID NO:81, SEQ ID NO:82, SEQ ID NO:83, SEQ ID NO:84, SEQ ID NO:85, SEQ ID NO:86, SEQ ID NO:87, SEQ ID NO:88, SEQ ID NO:100, SEQ ID NO:101, SEQ ID NO:102, SEQ ID NO:103, SEQ ID NO:104, SEQ ID NO:105, SEQ ID NO:106 or SEQ ID NO:107. Thus, the microorganism of the invention or the microorganism used in the methods of the invention preferably comprises at least a nucleic acid sequence identical to or having at least 90% identity to SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 3 or SEQ ID NO: 4, SEQ ID NO:68, SEQ ID NO:69, SEQ ID NO:70, SEQ ID NO: 71, SEQ ID NO:72, SEQ ID NO: 73, SEQ ID NO:74, SEQ ID NO: 75, SEQ ID NO: 76, SEQ ID NO: 77, SEQ ID NO: 78, SEQ ID NO:79, SEQ ID NO:80, SEQ ID NO:81, SEQ ID NO:82, SEQ ID NO:83, SEQ ID NO:84, SEQ ID NO:85, SEQ ID NO:86, SEQ ID NO:87, SEQ ID NO:88, SEQ ID NO:100, SEQ ID NO:101, SEQ ID NO:102, SEQ ID NO:103, SEQ ID NO:104, SEQ ID NO:105, SEQ ID NO:106 or SEQ ID NO:107. Preferably the nucleic acid is identical to or has at least 90% identity to SEQ ID NO: 1.

As is known in the art, in the event that the first domain of XxxSGD used in the mosaic SGD is not a methionine, the skilled person will readily be able to introduce a start codon in the nucleic acid sequence encoding the mosaic SGD in order to ensure proper translation of the mosaic SGD. The skilled person will also know how to

introduce short nucleic acid sequences corresponding to linkers separating the different domains in the mosaic SGD.

The nucleic acid construct may further comprise a nucleic acid sequence identical to or having at 90%, such as at least 91%, such as at least 92%, such as at least 93%, such as at least 94%, such as at least 95%, such as at least 96%, such as at least 97%, such as at least 98%, such as at least 99%, such as 100% identity to SEQ ID NO: 7.

The nucleic acid construct may further comprise a sequence identical to or having at least 90%, such as at least 91%, such as at least 92%, such as at least 93%, such as at least 94%, such as at least 95%, such as at least 96%, such as at least 97%, such as at least 98%, such as at least 99%, such as 100% identity to SEQ ID NO: 5 and/or SEQ ID NO: 23.

The nucleic acid construct may further comprise a nucleic acid sequence identical to or having at least 90% identity, such as at least 91%, such as at least 92%, such as at least 93%, such as at least 94%, such as at least 95%, such as at least 96%, such as at least 97%, such as at least 98%, such as at least 99%, such as 100% identity to SEQ ID NO: 6.

The nucleic acid construct may further comprise a nucleic acid sequence identical to or having at least 90% identity, such as at least 91%, such as at least 92%, such as at least 93%, such as at least 94%, such as at least 95%, such as at least 96%, such as at least 97%, such as at least 98%, such as at least 99%, such as 100% identity to SEQ ID NO: 8, SEQ ID NO: 9, SEQ ID NO: 10, SEQ ID NO: 11, SEQ ID NO: 12, SEQ ID NO: 13, SEQ ID NO: 14, SEQ ID NO: 15, SEQ ID NO: 16, SEQ ID NO: 17 and/or SEQ ID NO: 18.

All nucleic acid sequences may have been codon-optimised for expression in the microorganism, as is known in the art.

It may be of interest to take advantage of inducible promoters. Thus in some embodiments, the nucleic acid constructs comprises one or more of the above nucleic acid sequences under the control of an inducible promoter. This allows more control of when the enzyme encoded by the sequence is actually expressed, and can be

advantageous for example if production of one of the plant compounds negatively affects cell growth. The skilled person will have no difficulty in identifying suitable inducible promoters.

In some embodiments, the nucleic acid construct is one or more vectors, for examples an integrative or a replicative vector. Suitable vectors are known in the art and readily available to the skilled person.

Also provided herein is a vector comprising one of more of the nucleic acid sequences above, in particular SEQ ID NO: 1 or a sequence having at least 90% identity thereto. The vector may further comprise any of SEQ ID NO: 7, SEQ ID NO: 5, SEQ ID NO: 23, SEQ ID NO: 6, SEQ ID NO: 8, SEQ ID NO: 9, SEQ ID NO: 10, SEQ ID NO: 11, SEQ ID NO: 12, SEQ ID NO: 13, SEQ ID NO: 14, SEQ ID NO: 15, SEQ ID NO: 16, SEQ ID NO: 17 and/or SEQ ID NO: 18 or a sequence having at least 90% identity thereto.

Also provided herein is a host cell comprising one or more nucleic acid sequence or vector as defined herein above, in particular SEQ ID NO: 1 or a sequence having at least 90% identity thereto, or a vector comprising SEQ ID NO: 1 or a sequence having at least 90% identity thereto, and one or more of SEQ ID NO: 7, SEQ ID NO: 5, SEQ ID NO: 23, SEQ ID NO: 6, SEQ ID NO: 8, SEQ ID NO: 9, SEQ ID NO: 10, SEQ ID NO: 11, SEQ ID NO: 12, SEQ ID NO: 13, SEQ ID NO: 14, SEQ ID NO: 15, SEQ ID NO: 16, SEQ IDNO: 17 and/or SEQ ID NO: 18 or a sequence having at least 90% identity thereto.

The host cell may be any host cell, such as a primary cell or a cell from a cell line. In preferred embodiments, the host cell is from a mammalian or human cell line. The host cell may be a prokaryote or a eukaryote. In a preferred embodiment, the cell is a eukaryote.

A host cell according to the present invention may be comprised within a host organism, such as an animal.

Also provided herein is the use of the nucleic acid constructs, the microorganisms, the vectors or the host cells described herein for producing strictosidine aglycone and/or tetrahydroalstonine, alstonine, tabersonine and/or catharanthine in a microorganism. In

some embodiments, the nucleic acid constructs, the microorganisms, the vectors or the host cells described herein are used in a method for producing strictosidine aglycone and/or tetrahydroalstonine, alstonine, tabersonine and/or catharanthine in a microorganism as described herein.

Pharmaceutical compounds

The plant compounds obtainable by the present methods may be useful for manufacturing pharmaceutical compounds. Thus, the methods may further comprise a step of producing a pharmaceutical compound from any of the compounds, in particular monoterpenoid indole alkaloids, produced by the microorganism of the present invention.

Thus is also provided a method of treating a disorder such as a cancer, arrhythmia, malaria, psychotic diseases, hypertension, depression, Alzheimer's disease, addiction and/or neuronal diseases, comprising administration of a therapeutic sufficient amount of an MIA or a pharmaceutical compound obtained by the methods described herein.

Sequences

Table 1

Sequence ID NO:	**Description**	**Details**
1	*DNA* *RseSGD* from *Rauvolfia serpentina*	Strictosidine-O-beta-D-glucosidase EC 3.2.1.105 Hydrolyses strictosidine to strictosidine aglycone
2	*DNA* *GseSGD* from *Gelsemium sempervirens*	strictosidine glucosidase EC 3.2.1.- Putative function: Hydrolyses O-glycosyl compounds
3	*DNA* *SapSGD* from *Scedosporium apiospermum*	3-alpha-(S)-strictosidine beta-glucosidase EC 3.2.1.105 Putative function: Hydrolyses strictosidine to strictosidine aglycone
4	*DNA*	Strictosidine-beta-D-glucosidase

	RveSGD from *Rauvolfia verticillata*	EC 3.2.1.105 Putative function: Hydrolyses strictosidine to strictosidine aglycone
5	*DNA* *CroTHAS* from *Chatharanthus roseus*	Tetrahydroalstonine synthase EC.1.-.-.- Converts strictosidine aglycone to tetrahydroalstonine
6	*DNA* *GseSBE* from *Gelsemium sempervirens*	Sarpagan bridge enzyme (CYP71AY5) EC 1.14.14.- Converts by aromatization the tetrahydroalstonine and ajmalicine to the corresponding alstonine and serpentine, respectively or converts by cyclization the strictosidine-derived geissoschizine to the sarpagan alkaloid polyneuridine aldehyde
7	*DNA* *CroSTR* from *Catharanthus roseus*	Strictosidine synthase EC 4.3.3.2 Converts secologanin and tryptamine to strictosidine by stereospecific condensation.
8	*DNA* *CroCPR* from *Catharanthus roseus*	NADPH-cytochrome P450 reductase EC 1.6.2.4 This enzyme is required for electron transfer from NADP to cytochrome P450
9	*DNA* *CroCYB5* from *Catharanthus roseus*	Cytochrome b5 EC 1.6.2.2 Membrane bound hemoprotein which function as an electron carrier
10	*DNA* *CroGS* from *Catharanthus roseus*	Geissoschizine synthase (CrADH14) EC 1.3.1.36 Catalyzes the reduction of strictosidine aglycone to 19*E*-geissoschizine

11	*DNA* *CroGO* from *Catharanthus roseus*	Geissoschizine oxidase (CYP71AY2) EC 1.14.14.- Catalyzes the oxidation of 19E-geissoschizine to produce a short-lived MIA unstable intermediate which can be oxidized either by Redox1 and Redox2 to produce stemmadenine and 16S/R-deshydroxymethylstemmadenine (16S/R-DHS) or by spontaneous conversion to akuammicine
12	*DNA* *CroRedox1* from *Catharanthus roseus*	Redox 1 EC 1.14.14.- Catalyzes the first of two oxidation steps that the converts the unstable product resulting from oxidation of 19E-geissoschizine by geissoschizine oxidase (GO) to stemmadenine biosynthesis
13	*DNA* *CroRedox2* from *Catharanthus roseus*	Redox 2 EC 1.7.1.- Catalyzes the second of two oxidation steps that the converts the unstable product resulting from oxidation of 19E-geissoschizine by geissoschizine oxidase (GO) to stemmadenine biosynthesis
14	*DNA* *CroSAT* from *Catharanthus roseus*	Stemmadenine O-acetyltransferase EC 1.7.1.- Catalyzes the acetylation of stemmadenine to O-acetylstemmadenine
15	*DNA* *CroPAS* from *Catharanthus roseus*	O-acetylstemmadenine oxidase (precondylocarpine acetate synthase) EC 1.21.3.-

		Converts O-acetylstemmadenine to dihydroprecondylocarpine acetate
16	*DNA* *CroDPAS* from *Catharanthus roseus*	Dehydroprecondylocarpine acetate synthase EC 1.1.1.- Converts precondylocarpine acetate to dihydroprecondylocarpine acetate
17	*DNA* *CroTS* from *Catharanthus roseus*	tabersonine synthase (Hydrolyase 2) EC 4.-.-.- Catalyzes the conversion of dihydroprecondylocarpine acetate to tabersonine
18	*DNA* *CroCS* from *Catharanthus roseus*	Catharanthine synthase (Hydrolase 1) EC 4.-.-.- Catalyzes the conversion of dihydroprecondylocarpine acetate to catharanthine
19	*DNA* *UtoSGD* from *Uncaria tomentosa*	Putative strictosidine beta-D-glucosidase EC 3.2.1.105 Putative function: Hydrolyses strictosidine to strictosidine aglycone
20	*DNA* *CroSGD* from *Catharanthus roseus*	Strictosidine-O-beta-D-glucosidase EC 3.2.1.105 Hydrolyses strictosidine to strictosidine aglycone
21	*DNA* *CacSGD* from *Camptotheca acuminata*	Putative strictosidine beta-D-glucosidase EC 3.2.1.105 Putative function: Hydrolyses strictosidine to strictosidine aglycone
22	*DNA* *GsoSGD* from *Glycine soja*	Uncharacterized protein EC 3.2.-.-

		Putative function: Hydrolyses O-glycosyl compounds
23	*DNA* *CroHYS*	Heteroyohimbine synthase EC.1.-.-.- Converts strictosidine aglycone to tetrahydroalstonine, ajmalicine, or mayumbine
24	Protein *RseSGD* from *Rauvolfia serpentina*	Strictosidine-O-beta-D-glucosidase EC 3.2.1.105 Q8GU20 Hydrolyses strictosidine to strictosidine aglycone
25	Protein *GseSGD* from *Gelsemium sempervirens*	strictosidine glucosidase EC 3.2.1.- AXK92564.1 Putative function: Hydrolyses O-glycosyl compounds
26	Protein *SapSGD* from *Scedosporium apiospermum*	3-alpha-(S)-strictosidine beta-glucosidase EC 3.2.1.105 A0A084GBX6 Putative function: Hydrolyses strictosidine to strictosidine aglycone
27	Protein *RveSGD* from *Rauvolfia verticillata*	Strictosidine-beta-D-glucosidase EC 3.2.1.105 M9NGS2 Putative function: Hydrolyses strictosidine to strictosidine aglycone
28	Protein *CroTHAS* from *Chatharanthus roseus*	Tetrahydroalstonine synthase EC.1.-.-.- A0A0F6SD02 Converts strictosidine aglycone to tetrahydroalstonine
29	Protein	Sarpagan bridge enzyme (CYP71AY5)

	GseSBE from *Gelsemium sempervirens*	EC 1.14.14.- P0DO14 Converts by aromatization the tetrahydroalstonine and ajmalicine to the corresponding alstonine and serpentine, respectively or converts by cyclization the strictosidine-derived geissoschizine to the sarpagan alkaloid polyneuridine aldehyde
30	Protein *CroSTR* from *Catharanthus roseus*	Strictosidine synthase EC 4.3.3.2 P18417 Converts secologanin and tryptamine to strictosidine by stereospecific condensation.
31	Protein *CroCPR* from *Catharanthus roseus*	NADPH-cytochrome P450 reductase EC 1.6.2.4 Q05001 This enzyme is required for electron transfer from NADP to cytochrome P450
32	Protein *CroCYB5* from *Catharanthus roseus*	Cytochrome b5 EC 1.6.2.2 A0A0C5DKP2 Membrane bound hemoprotein which function as an electron carrier
33	Protein *CroGS* from *Catharanthus roseus*	Geissoschizine synthase (CrADH14) EC 1.3.1.36 W8JWW7 Catalyzes the reduction of strictosidine aglycone to 19*E*-geissoschizine
34	Protein *CroGO* from *Catharanthus roseus*	Geissoschizine oxidase (CYP71AY2) EC 1.14.14.- I1TEM0

		Catalyzes the oxidation of 19E-geissoschizine to produce a short-lived MIA unstable intermediate which can be oxidized either by Redox1 and Redox2 to produce stemmadenine and 16S/R-deshydroxymethylstemmadenine (16S/R-DHS) or by spontaneous conversion to akuammicine
35	Protein *CroRedox1* from *Catharanthus roseus*	Redox 1 EC 1.14.14.- A0A2P1GIW4 Catalyzes the first of two oxidation steps that the converts the unstable product resulting from oxidation of 19E-geissoschizine by geissoschizine oxidase (GO) to stemmadenine biosynthesis
36	Protein *CroRedox2* from *Catharanthus roseus*	Redox 2 EC 1.7.1.- A0A2P1GIY9 Catalyzes the second of two oxidation steps that the converts the unstable product resulting from oxidation of 19E-geissoschizine by geissoschizine oxidase (GO) to stemmadenine biosynthesis
37	Protein *CroSAT* from *Catharanthus roseus*	Stemmadenine O-acetyltransferase EC 1.7.1.- A0A2P1GIW7 Catalyzes the acetylation of stemmadenine to O-acetylstemmadenine
38	Protein	O-acetylstemmadenine oxidase (precondylocarpine acetate synthase)

	CroPAS from *Catharanthus roseus*	EC 1.21.3.- MH213134.1 Converts O-acetylstemmadenine to dihydroprecondylocarpine acetate
39	Protein *CroDPAS* from *Catharanthus roseus*	Dehydroprecondylocarpine acetate synthase EC 1.1.1.- A0A1B1FHP3 Converts precondylocarpine acetate to dihydroprecondylocarpine acetate
40	Protein *CroTS* from *Catharanthus roseus*	tabersonine synthase (Hydrolyase 2) EC 4.-.-.- A0A2P1GIW3 Catalyzes the conversion of dihydroprecondylocarpine acetate to tabersonine
41	Protein *CroCS* from *Catharanthus roseus*	Catharanthine synthase (Hydrolase 1) EC 4.-.-.- A0A2P1GIW2 Catalyzes the conversion of dihydroprecondylocarpine acetate to catharanthine
42	Protein *UtoSGD* from *Uncaria tomentosa*	Putative strictosidine beta-D-glucosidase EC 3.2.1.105 I6ZQ42 Putative function: Hydrolyses strictosidine to strictosidine aglycone
43	Protein *CroSGD* from *Catharanthus roseus*	Strictosidine-O-beta-D-glucosidase EC 3.2.1.105 B8PRP4 Hydrolyses strictosidine to strictosidine aglycone

44	Protein *CacSGD* from *Camptotheca acuminata*	Putative strictosidine beta-D-glucosidase EC 3.2.1.105 G8E0P8 Putative function: Hydrolyses strictosidine to strictosidine aglycone
45	Protein *GsoSGD* from *Glycine soja*	Uncharacterized protein EC 3.2.-.- A0A0R0H2R3 Putative function: Hydrolyses O-glycosyl compounds
46	Protein *CroHYS* from *Catharanthus roseus*	Heteroyohimbine synthase EC.1.-.-.- A0A1B1FHP5 Converts strictosidine aglycone to tetrahydroalstonine, ajmalicine, or mayumbine
47	Protein *VmiSGD1* from *Vinca minor*	Uncharacterized protein EC 3.2.-.- Putative function: Hydrolyses O-glycosyl compounds
48	Protein *AhuSGD* from *Amsonia hubrichtii*	Uncharacterized protein EC 3.2.-.- Putative function: Hydrolyses O-glycosyl compounds
49	Protein *HimSGD2* from *Handroanthus impetiginosus*	Uncharacterized protein EC 3.2.-.- PIN06789.1 Putative function: Hydrolyses O-glycosyl compounds
50	Protein *SinSGD* from *Sesamum indicum*	Uncharacterized protein EC 3.2.-.- XP_011094151.1

		Putative function: Hydrolyses O-glycosyl compounds
51	Protein *TelSGD* from *Tabernaemontana elegans*	Uncharacterized protein EC 3.2.-.- Putative function: Hydrolyses O-glycosyl compounds
52	Protein *VunSGD* from *Vigna unguiculata*	Uncharacterized protein EC 3.2.-.- XP_027910736.1 Putative function: Hydrolyses O-glycosyl compounds
53	Protein *NsiSGD1* from *Nyssa sinensis*	Uncharacterized protein EC 3.2.-.- KAA8549635.1 Putative function: Hydrolyses O-glycosyl compounds
54	Protein *LprSGD* from *Lomentospora prolificans*	Uncharacterized protein EC 3.2.-.- PKS11920.1 Putative function: Hydrolyses O-glycosyl compounds
55	Protein *AchSGD1* from *Actinidia chinensis* var. *chinensis*	Uncharacterized protein EC 3.2.-.- PSS10019.1 Putative function: Hydrolyses O-glycosyl compounds
56	Protein *HsuSGD* from *Heliocybe sulcata*	Uncharacterized protein EC 3.2.-.- TFK52902.1 Putative function: Hydrolyses O-glycosyl compounds
57	Protein *MroSGD* from *Moniliophthora roreri* MCA 2997	Uncharacterized protein EC 3.2.-.- ESK96275.1

		Putative function: Hydrolyses O-glycosyl compounds
58	Protein *RseSGD2* from *Rauvolfia serpentina*	Raucaffricine-O-beta-D-glucosidase EC 3.2.1.125 AAF03675.1 Function: Hydrolyses the MIA raucaffricine
59	Protein *PgrSGD* from *Pyricularia grisea*	Uncharacterized protein EC 3.2.-.- AAX07701.1 Putative function: Hydrolyses O-glycosyl compounds
60	Protein *OpuSGD* from *Ophiorrhiza pumila*	Uncharacterized protein EC 3.2.-.- BAP90523.1 Putative function: Hydrolyses O-glycosyl compounds
61	Protein *HpiSGD* from *Hydnomerulius pinastri* MD-312	Uncharacterized protein EC 3.2.-.- KIJ63193.1 Putative function: Hydrolyses O-glycosyl compounds
62	Protein *HanSGD1* from *Helianthus annuus*	Uncharacterized protein EC 3.2.-.- XP_022015317.1 Putative function: Hydrolyses O-glycosyl compounds
63	Protein *AchSGD2* from *Actinidia chinensis* var. *chinensis*	Uncharacterized protein EC 3.2.-.- PSR88404.1 Putative function: Hydrolyses O-glycosyl compounds

64	Protein *HimSGD1* from *Handroanthus impetiginosus*	Uncharacterized protein EC 3.2.-.- PIN07435.1 Putative function: Hydrolyses O-glycosyl compounds
65	Protein *IpeSGD* from *Carapichea ipecacuanha*	beta-glucosidase EC 3.2.1.21 BAH02544.1 function: hydrolyses glucosidic Ipecac alkaloids
66	Protein *LsaSGD1* from *Lactuca sativa*	Uncharacterized protein EC 3.2.-.- XP_023770227.1 Putative function: Hydrolyses O-glycosyl compounds
67	Protein *CarSGD* from *Coffea arabica*	Uncharacterized protein EC 3.2.-.- XP_027073002.1 Putative function: Hydrolyses O-glycosyl compounds
68	*DNA VmiSGD1* from *Vinca minor*	Uncharacterized protein EC 3.2.-.- Putative function: Hydrolyses O-glycosyl compounds
69	*DNA AhuSGD* from *Amsonia hubrichtii*	Uncharacterized protein EC 3.2.-.- Putative function: Hydrolyses O-glycosyl compounds
70	*DNA HimSGD2* from *Handroanthus impetiginosus*	Uncharacterized protein EC 3.2.-.- Putative function: Hydrolyses O-glycosyl compounds
71	*DNA SinSGD* from *Sesamum indicum*	Uncharacterized protein EC 3.2.-.-

		Putative function: Hydrolyses O-glycosyl compounds
72	*DNA TelSGD* from *Tabernaemontana elegans*	Uncharacterized protein EC 3.2.-.- Putative function: Hydrolyses O-glycosyl compounds
73	*DNA VunSGD* from *Vigna unguiculata*	Uncharacterized protein EC 3.2.-.- Putative function: Hydrolyses O-glycosyl compounds
74	*DNA NsiSGD1* from *Nyssa sinensis*	Uncharacterized protein EC 3.2.-.- Putative function: Hydrolyses O-glycosyl compounds
75	*DNA LprSGD* from *Lomentospora prolificans*	Uncharacterized protein EC 3.2.-.- Putative function: Hydrolyses O-glycosyl compounds
76	*DNA AchSGD1* from *Actinidia chinensis* var. *chinensis*	Uncharacterized protein EC 3.2.-.- Putative function: Hydrolyses O-glycosyl compounds
77	*DNA HsuSGD* from *Heliocybe sulcata*	Uncharacterized protein EC 3.2.-.- Putative function: Hydrolyses O-glycosyl compounds
78	*DNA MroSGD* from *Moniliophthora roreri* MCA 2997	Uncharacterized protein EC 3.2.-.- Putative function: Hydrolyses O-glycosyl compounds
79	*DNA RseSGD2* from *Rauvolfia serpentina*	Raucaffricine-O-beta-D-glucosidase EC 3.2.1.125 Function: Hydrolyses the MIA raucaffricine

80	*DNA PgrSGD* from *Pyricularia grisea*	Uncharacterized protein EC 3.2.-.- Putative function: Hydrolyses O-glycosyl compounds
81	*DNA OpuSGD* from *Ophiorrhiza pumila*	Uncharacterized protein EC 3.2.-.- Putative function: Hydrolyses O-glycosyl compounds
82	*DNA HpiSGD* from *Hydnomerulius pinastri* MD-312	Uncharacterized protein EC 3.2.-.- Putative function: Hydrolyses O-glycosyl compounds
83	*DNA HanSGD1* from *Helianthus annuus*	Uncharacterized protein EC 3.2.-.- Putative function: Hydrolyses O-glycosyl compounds
84	*DNA AchSGD2* from *Actinidia chinensis* var. *chinensis*	Uncharacterized protein EC 3.2.-.- Putative function: Hydrolyses O-glycosyl compounds
85	*DNA HimSGD1* from *Handroanthus impetiginosus*	Uncharacterized protein EC 3.2.-.- Putative function: Hydrolyses O-glycosyl compounds
86	*DNA IpeSGD* from *Carapichea ipecacuanha*	Beta-glucosidase EC 3.2.1.21 Function: hydrolyses glucosidic Ipecac alkaloids
87	*DNA LsaSGD1* from *Lactuca sativa*	Uncharacterized protein EC 3.2.-.- Putative function: Hydrolyses O-glycosyl compounds
88	*DNA CarSGD* from *Coffea arabica*	Uncharacterized protein EC 3.2.-.-

		Putative function: Hydrolyses O-glycosyl compounds
89	Domain 1 of RseSGD from *Rauvolfia serpentina*	M1-R115 MDNTQAEPLVVAIVPKPNASTEHTNS HLIPVTRSKIVVHRRDFPQDFIFGAGG SAYQCEGAYNEGNRGPSIWDTFTQR SPAKISDGSNGNQAINCYHMYKEDIKI MKQTGLESYR
90	Domain 2 of RseSGD from *Rauvolfia serpentina*	F116-G266 FSISWSRVLPGGRLAAGVNKDGVKFY HDFIDELLANGIKPSVTLFHWDLPQAL EDEYGGFLSHRIVDDFCEYAEFCFWE FGDKIKYWTTFNEPHTFAVNGYALGE FAPGRGGKGDEGDPAIEPYVVTHNIL LAHKAAVEEYRNKFQKCQEG
91	Domain 3 of RseSGD from *Rauvolfia serpentina*	E267-G456 IGIVLNSMWMEPLSDVQADIDAQKRA LDFMLGWFLEPLTTGDYPKSMRELVK GRLPKFSADDSEKLKGCYDFIGMNYY TATYVTNAVKSNSEKLSYETDDQVTK TFERNQKPIGHALYGGWQHVVPWGL YKLLVYTKETYHVPVLYVTESGMVEE NKTKILLSEARRDAERTDYHQKHLAS VRDAIDDG
92	Domain 4 of RseSGD from *Rauvolfia serpentina*	V457-T532 VNVKGYFVWSFFDNFEWNLGYICRY GIIHVDYKSFERYPKESAIWYKNFIAG KSTT SPAKRRREEAQVELVKRQKT
93	Protein sequence of CCRR	Mosaic SGD
94	Protein sequence of CRRR	Mosaic SGD
95	Protein sequence of RCRR	Mosaic SGD

96	Protein sequence of RRRC	Mosaic SGD
97	Protein sequence of RCRC	Mosaic SGD
98	Protein sequence of CCRC	Mosaic SGD
99	Protein sequence of VVRR	Mosaic SGD
100	DNA of CCRR	Mosaic SGD
101	DNA of CRRR	Mosaic SGD
102	DNA of RCRR	Mosaic SGD
103	DNA of CRRC	Mosaic SGD
104	DNA of RRRC	Mosaic SGD
105	DNA of RCRC	Mosaic SGD
106	DNA of CCRC	Mosaic SGD
107	DNA of VVRR	Mosaic SGD
108	Protein sequence of CRRC	Mosaic SGD

Examples

Strains

Different strains were developed to validate the functionalization of RseSGD in the production of strictosidine aglycone and selected MIAs.

Table 2

Strain	Genotype	Substrate → Product
MIA-BJ	Cas9 @ XII-1, atf1Δ oye2Δ, oye3Δ ari1Δ adh6Δ , [CroG8H-CroCYB5] @X-3, [CroCPR-Cro8HGO] @XI-3, [CroIS-CroIO] @XII-2, [CroSTR-CroSLS] @X-4, [Cro7DLGT-Cro7DLH] @XI-1, [CroLAMT-CroADH2] @XII-4	Secologanin + tryptamine → strictosidine OR Geraniol + tryptamine → strictosidine

MIA-CA-1	Cas9 @ XII-1, atf1Δ oye2Δ, oye3Δ ari1Δ adh6Δ , [CroG8H-CroCYB5] @X-3, [CroCPR-Cro8HGO] @XI-3, [CroIS-CroIO] @XII-2, [CroSTR-CroSLS] @X-4, [Cro7DLGT-Cro7DLH] @XI-1, [CroLAMT-CroADH2] @XII-4 [**Cro**SGD-CroHYS]@XII-5	Secologanin + tryptamine → strictosidine* * or tetrahydroalstonine if the candidate SGD does function
MIA-CA-2	Cas9 @ XII-1, atf1Δ oye2Δ, oye3Δ ari1Δ adh6Δ , [CroG8H-CroCYB5] @X-3, [CroCPR-Cro8HGO] @XI-3, [CroIS-CroIO] @XII-2, [CroSTR-CroSLS] @X-4, [Cro7DLGT-Cro7DLH] @XI-1, [CroLAMT-CroADH2] @XII-4 [**Rse**SGD-CroHYS]@XII-5	Secolocanin + tryptamine → tetrahydroalstonine
MIA-CA-3	Cas9 @ XII-1, atf1Δ oye2Δ, oye3Δ ari1Δ adh6Δ , [CroG8H-CroCYB5] @X-3, [CroCPR-Cro8HGO] @XI-3, [CroIS-CroIO] @XII-2, [CroSTR-CroSLS] @X-4, [Cro7DLGT-Cro7DLH] @XI-1, [CroLAMT-CroADH2] @XII-4 [**Rve**SGD-CroHYS]@XII-5	Secolocanin + tryptamine → tetrahydroalstonine
MIA-CA-4	Cas9 @ XII-1, atf1Δ oye2Δ, oye3Δ ari1Δ adh6Δ , [CroG8H-CroCYB5] @X-3, [CroCPR-Cro8HGO] @XI-3, [CroIS-CroIO] @XII-2, [CroSTR-CroSLS] @X-4, [Cro7DLGT-Cro7DLH] @XI-1, [CroLAMT-CroADH2] @XII-4 [**Gse**SGD-CroHYS]@XII-5	Secolocanin + tryptamine → tetrahydroalstonine
MIA-CA-5	Cas9 @ XII-1, atf1Δ oye2Δ, oye3Δ ari1Δ adh6Δ , [CroG8H-CroCYB5] @X-3, [CroCPR-Cro8HGO] @XI-3, [CroIS-CroIO] @XII-2, [CroSTR-CroSLS] @X-4, [Cro7DLGT-Cro7DLH] @XI-1, [CroLAMT-	Secolocanin + tryptamine → strictosidine*

	CroADH2] @XII-4 [**Cac**SGD-CroHYS]@XII-5	* or tetrahydroalstonine if the candidate SGD does function
MIA-CA-6	Cas9 @ XII-1, atf1Δ oye2Δ, oye3Δ ari1Δ adh6Δ , [CroG8H-CroCYB5] @X-3, [CroCPR-Cro8HGO] @XI-3, [CroIS-CroIO] @XII-2, [CroSTR-CroSLS] @X-4, [Cro7DLGT-Cro7DLH] @XI-1, [CroLAMT-CroADH2] @XII-4 [**Sap**SGD-CroHYS]@XII-5	Secolocanine+ tryptamine → tetrahydroalstonine
MIA-CA-7	Cas9 @ XII-1, atf1Δ oye2Δ, oye3Δ ari1Δ adh6Δ , [CroG8H-CroCYB5] @X-3, [CroCPR-Cro8HGO] @XI-3, [CroIS-CroIO] @XII-2, [CroSTR-CroSLS] @X-4, [Cro7DLGT-Cro7DLH] @XI-1, [CroLAMT-CroADH2] @XII-4 [**Uto**SGD-CroHYS]@XII-5	Secolocanin + tryptamine → strictosidine* * or tetrahydroalstonine if the candidate SGD does function
MIA-CA-8	Cas9 @ XII-1, atf1Δ oye2Δ, oye3Δ ari1Δ adh6Δ , [CroG8H-CroCYB5] @X-3, [CroCPR-Cro8HGO] @XI-3, [CroIS-CroIO] @XII-2, [CroSTR-CroSLS] @X-4, [Cro7DLGT-Cro7DLH] @XI-1, [CroLAMT-CroADH2] @XII-4 [**Gso**SGD-CroHYS]@XII-5	Secolocanin + tryptamine → strictosidine* * or tetrahydroalstonine if the candidate SGD does function
MIA-BZ-1	Cas9 @ XII-1, atf1Δ oye2Δ, oye3Δ ari1Δ adh6Δ , [CroG8H-CroCYB5] @X-3, [CroCPR-Cro8HGO] @XI-3, [CroIS-CroIO] @XII-2, [CroSTR-CroSLS] @X-4, [Cro7DLGT-Cro7DLH] @XI-1, [CroLAMT-CroADH2] @XII-4 [**Cro**SGD]@XII-5	Secolocanin + tryptamine → strictosidine* * or strictosidine aglycone if the candidate SGD does function
MIA-BZ-2	Cas9 @ XII-1, atf1Δ oye2Δ, oye3Δ ari1Δ adh6Δ , [CroG8H-CroCYB5] @X-3,	Secolocanin + tryptamine → strictosidine aglycone

	[CroCPR-Cro8HGO] @XI-3, [CroIS-CroIO] @XII-2, [CroSTR-CroSLS] @X-4, [Cro7DLGT-Cro7DLH] @XI-1, [CroLAMT-CroADH2] @XII-4 [**Rse**SGD]@XII-5	
MIA-BZ-3	Cas9 @ XII-1, atf1Δ oye2Δ, oye3Δ ari1Δ adh6Δ , [CroG8H-CroCYB5] @X-3, [CroCPR-Cro8HGO] @XI-3, [CroIS-CroIO] @XII-2, [CroSTR-CroSLS] @X-4, [Cro7DLGT-Cro7DLH] @XI-1, [CroLAMT-CroADH2] @XII-4 [**Cro**SGD-CroTHAS]@XII-5	Secolocanin + tryptamine → strictosidine* * or tetrahydroalstonine if the candidate SGD does function
MIA-BZ-4	Cas9 @ XII-1, atf1Δ oye2Δ, oye3Δ ari1Δ adh6Δ , [CroG8H-CroCYB5] @X-3, [CroCPR-Cro8HGO] @XI-3, [CroIS-CroIO] @XII-2, [CroSTR-CroSLS] @X-4, [Cro7DLGT-Cro7DLH] @XI-1, [CroLAMT-CroADH2] @XII-4 [**Rse**SGD-CroTHAS]@XII-5	Secolocanin + tryptamine → tetrahydroalstonine
MIA-DA	Cas9@XII-1, atf1Δ oye2Δ, oye3Δ ari1Δ adh6Δ, [CroCPR-CroCYB5]@XI-3	No production
MIA-DC	Cas9@XII-1, atf1Δ oye2Δ, oye3Δ ari1Δ adh6Δ, [CroCPR-CroCYB5]@XI-3, [CroSTR-CroGS-RseSGD-CroGO-CroRedox1-CroRedox2]@XII-5, [CroSAT-CroPAS-CroDPAS-CroTS-CroCG]@XI-5	Secologanin + tryptamine → tabersonine + catharanthine
MIA-DE	Cas9@XII-1, atf1Δ oye2Δ, oye3Δ ari1Δ adh6Δ, [CroCPR-CroCYB5]@XI-3, [CroNMT-CroD4H-CroDAT-CroPER-CroT16H1]@X-4, [CroT16H2-Cro16OMT-CroT3O-CroT3R]@XII-4	tabersonine → Vindoline OR Tabersonine + catharanthine → vinblastine OR Vindoline + catharanthine → vinblastine

MIA-FA	Cas9 @ XII-1, atf1Δ oye2Δ, oye3Δ ari1Δ adh6Δ , [CroG8H-CroCYB5] @X-3, [CroCPR] @XI-3, [CroIO] @XII-2, [CroSTR-CroSLS] @X-4, [Cro7DLGT-Cro7DLH] @XI-1, [CroLAMT-CroADH2] @XII-4, [Vmi8HGO-A] @X-2, [NcMLP-NcISY] @XII-5, [CroHYS] @IV-1	Secologanin + tryptamine → strictosidine* OR Geraniol + tryptamine → strictosidine* **or tetrahydroalstonine if functional SGD is co-expressed*
MIA-FC-1	Cas9 @ XII-1, atf1Δ oye2Δ, oye3Δ ari1Δ adh6Δ , [CroG8H-CroCYB5] @X-3, [CroCPR] @XI-3, [CroIO] @XII-2, [CroSTR-CroSLS] @X-4, [Cro7DLGT-Cro7DLH] @XI-1, [CroLAMT-CroADH2] @XII-4, [Vmi8HGO-A] @X-2, [NcMLP-NcISY] @XII-5, [CroHYS] @IV-1 , [**Cro**SGD] @IV-2	Secolocanin + tryptamine → strictosidine* * *or tetrahydroalstonine if the candidate SGD does function*
MIA-FC-2	Cas9 @ XII-1, atf1Δ oye2Δ, oye3Δ ari1Δ adh6Δ , [CroG8H-CroCYB5] @X-3, [CroCPR] @XI-3, [CroIO] @XII-2, [CroSTR-CroSLS] @X-4, [Cro7DLGT-Cro7DLH] @XI-1, [CroLAMT-CroADH2] @XII-4, [Vmi8HGO-A] @X-2, [NcMLP-NcISY] @XII-5, [CroHYS] @IV-1 , [**Vmi**SGD1] @IV-2	Secologanin+ tryptamine → tetrahydroalstonine
MIA-FC-3	Cas9 @ XII-1, atf1Δ oye2Δ, oye3Δ ari1Δ adh6Δ , [CroG8H-CroCYB5] @X-3, [CroCPR] @XI-3, [CroIO] @XII-2, [CroSTR-CroSLS] @X-4, [Cro7DLGT-Cro7DLH] @XI-1, [CroLAMT-CroADH2] @XII-4, [Vmi8HGO-A] @X-2, [NcMLP-NcISY] @XII-5, [CroHYS] @IV-1 , [**Ahu**SGD] @IV-2	Secologanin+ tryptamine → tetrahydroalstonine
MIA-FC-4	Cas9 @ XII-1, atf1Δ oye2Δ, oye3Δ ari1Δ adh6Δ , [CroG8H-CroCYB5] @X-3,	Secologanin+ tryptamine → tetrahydroalstonine

	[CroCPR] @XI-3, [CroIO] @XII-2, [CroSTR-CroSLS] @X-4, [Cro7DLGT-Cro7DLH] @XI-1, [CroLAMT-CroADH2] @XII-4, [Vmi8HGO-A] @X-2, [NcMLP-NcISY] @XII-5, [CroHYS] @IV-1 , [**Him**SGD2] @IV-2	
MIA-FC-5	Cas9 @ XII-1, atf1Δ oye2Δ, oye3Δ ari1Δ adh6Δ , [CroG8H-CroCYB5] @X-3, [CroCPR] @XI-3, [CroIO] @XII-2, [CroSTR-CroSLS] @X-4, [Cro7DLGT-Cro7DLH] @XI-1, [CroLAMT-CroADH2] @XII-4, [Vmi8HGO-A] @X-2, [NcMLP-NcISY] @XII-5, [CroHYS] @IV-1 , [**Sin**SGD] @IV-2	Secologanin+ tryptamine → tetrahydroalstonine
MIA-FC-6	Cas9 @ XII-1, atf1Δ oye2Δ, oye3Δ ari1Δ adh6Δ , [CroG8H-CroCYB5] @X-3, [CroCPR] @XI-3, [CroIO] @XII-2, [CroSTR-CroSLS] @X-4, [Cro7DLGT-Cro7DLH] @XI-1, [CroLAMT-CroADH2] @XII-4, [Vmi8HGO-A] @X-2, [NcMLP-NcISY] @XII-5, [CroHYS] @IV-1 , [**Tel**SGD] @IV-2	Secologanin+ tryptamine → tetrahydroalstonine
MIA-FC-7	Cas9 @ XII-1, atf1Δ oye2Δ, oye3Δ ari1Δ adh6Δ , [CroG8H-CroCYB5] @X-3, [CroCPR] @XI-3, [CroIO] @XII-2, [CroSTR-CroSLS] @X-4, [Cro7DLGT-Cro7DLH] @XI-1, [CroLAMT-CroADH2] @XII-4, [Vmi8HGO-A] @X-2, [NcMLP-NcISY] @XII-5, [CroHYS] @IV-1 , [**Vun**SGD] @IV-2	Secologanin+ tryptamine → tetrahydroalstonine
MIA-FC-8	Cas9 @ XII-1, atf1Δ oye2Δ, oye3Δ ari1Δ adh6Δ , [CroG8H-CroCYB5] @X-3, [CroCPR] @XI-3, [CroIO] @XII-2, [CroSTR-CroSLS] @X-4, [Cro7DLGT-Cro7DLH] @XI-1, [CroLAMT-CroADH2]	Secologanin+ tryptamine → tetrahydroalstonine

	@XII-4, [Vmi8HGO-A] @X-2, [NcMLP-NclSY] @XII-5, [CroHYS] @IV-1 , [**Nsi**SGD1] @IV-2	
MIA-FC-9	Cas9 @ XII-1, atf1Δ oye2Δ, oye3Δ ari1Δ adh6Δ , [CroG8H-CroCYB5] @X-3, [CroCPR] @XI-3, [CrolO] @XII-2, [CroSTR-CroSLS] @X-4, [Cro7DLGT-Cro7DLH] @XI-1, [CroLAMT-CroADH2] @XII-4, [Vmi8HGO-A] @X-2, [NcMLP-NclSY] @XII-5, [CroHYS] @IV-1 , [**Lpr**SGD] @IV-2	Secologanin+ tryptamine → tetrahydroalstonine
MIA-FC-10	Cas9 @ XII-1, atf1Δ oye2Δ, oye3Δ ari1Δ adh6Δ , [CroG8H-CroCYB5] @X-3, [CroCPR] @XI-3, [CrolO] @XII-2, [CroSTR-CroSLS] @X-4, [Cro7DLGT-Cro7DLH] @XI-1, [CroLAMT-CroADH2] @XII-4, [Vmi8HGO-A] @X-2, [NcMLP-NclSY] @XII-5, [CroHYS] @IV-1 , [**Ach**SGD1] @IV-2	Secologanin+ tryptamine → tetrahydroalstonine
MIA-FC-11	Cas9 @ XII-1, atf1Δ oye2Δ, oye3Δ ari1Δ adh6Δ , [CroG8H-CroCYB5] @X-3, [CroCPR] @XI-3, [CrolO] @XII-2, [CroSTR-CroSLS] @X-4, [Cro7DLGT-Cro7DLH] @XI-1, [CroLAMT-CroADH2] @XII-4, [Vmi8HGO-A] @X-2, [NcMLP-NclSY] @XII-5, [CroHYS] @IV-1 , [**Hsu**SGD] @IV-2	Secologanin+ tryptamine → tetrahydroalstonine
MIA-FC-12	Cas9 @ XII-1, atf1Δ oye2Δ, oye3Δ ari1Δ adh6Δ , [CroG8H-CroCYB5] @X-3, [CroCPR] @XI-3, [CrolO] @XII-2, [CroSTR-CroSLS] @X-4, [Cro7DLGT-Cro7DLH] @XI-1, [CroLAMT-CroADH2] @XII-4, [Vmi8HGO-A] @X-2, [NcMLP-NclSY] @XII-5, [CroHYS] @IV-1 , [**Mro**SGD] @IV-2	Secologanin+ tryptamine → tetrahydroalstonine

MIA-FC-13	Cas9 @ XII-1, atf1Δ oye2Δ, oye3Δ ari1Δ adh6Δ , [CroG8H-CroCYB5] @X-3, [CroCPR] @XI-3, [CroIO] @XII-2, [CroSTR-CroSLS] @X-4, [Cro7DLGT-Cro7DLH] @XI-1, [CroLAMT-CroADH2] @XII-4, [Vmi8HGO-A] @X-2, [NcMLP-NcISY] @XII-5, [CroHYS] @IV-1 , [**Rse**SGD2] @IV-2	Secologanin+ tryptamine → tetrahydroalstonine
MIA-FC-14	Cas9 @ XII-1, atf1Δ oye2Δ, oye3Δ ari1Δ adh6Δ , [CroG8H-CroCYB5] @X-3, [CroCPR] @XI-3, [CroIO] @XII-2, [CroSTR-CroSLS] @X-4, [Cro7DLGT-Cro7DLH] @XI-1, [CroLAMT-CroADH2] @XII-4, [Vmi8HGO-A] @X-2, [NcMLP-NcISY] @XII-5, [CroHYS] @IV-1 , [**Pgr**SGD] @IV-2	Secologanin+ tryptamine → tetrahydroalstonine
MIA-FC-15	Cas9 @ XII-1, atf1Δ oye2Δ, oye3Δ ari1Δ adh6Δ , [CroG8H-CroCYB5] @X-3, [CroCPR] @XI-3, [CroIO] @XII-2, [CroSTR-CroSLS] @X-4, [Cro7DLGT-Cro7DLH] @XI-1, [CroLAMT-CroADH2] @XII-4, [Vmi8HGO-A] @X-2, [NcMLP-NcISY] @XII-5, [CroHYS] @IV-1 , [**Opu**SGD] @IV-2	Secologanin+ tryptamine → tetrahydroalstonine
MIA-FC-16	Cas9 @ XII-1, atf1Δ oye2Δ, oye3Δ ari1Δ adh6Δ , [CroG8H-CroCYB5] @X-3, [CroCPR] @XI-3, [CroIO] @XII-2, [CroSTR-CroSLS] @X-4, [Cro7DLGT-Cro7DLH] @XI-1, [CroLAMT-CroADH2] @XII-4, [Vmi8HGO-A] @X-2, [NcMLP-NcISY] @XII-5, [CroHYS] @IV-1 , [**Hpi**SGD] @IV-2	Secologanin+ tryptamine → tetrahydroalstonine
MIA-FC-17	Cas9 @ XII-1, atf1Δ oye2Δ, oye3Δ ari1Δ adh6Δ , [CroG8H-CroCYB5] @X-3, [CroCPR] @XI-3, [CroIO] @XII-2,	Secologanin+ tryptamine → tetrahydroalstonine

	[CroSTR-CroSLS] @X-4, [Cro7DLGT-Cro7DLH] @XI-1, [CroLAMT-CroADH2] @XII-4, [Vmi8HGO-A] @X-2, [NcMLP-NcISY] @XII-5, [CroHYS] @IV-1 , [**Han**SGD1] @IV-2	
MIA-FC-18	Cas9 @ XII-1, atf1Δ oye2Δ, oye3Δ ari1Δ adh6Δ , [CroG8H-CroCYB5] @X-3, [CroCPR] @XI-3, [CroIO] @XII-2, [CroSTR-CroSLS] @X-4, [Cro7DLGT-Cro7DLH] @XI-1, [CroLAMT-CroADH2] @XII-4, [Vmi8HGO-A] @X-2, [NcMLP-NcISY] @XII-5, [CroHYS] @IV-1 , [**Ach**SGD2] @IV-2	Secologanin+ tryptamine → tetrahydroalstonine
MIA-FC-19	Cas9 @ XII-1, atf1Δ oye2Δ, oye3Δ ari1Δ adh6Δ , [CroG8H-CroCYB5] @X-3, [CroCPR] @XI-3, [CroIO] @XII-2, [CroSTR-CroSLS] @X-4, [Cro7DLGT-Cro7DLH] @XI-1, [CroLAMT-CroADH2] @XII-4, [Vmi8HGO-A] @X-2, [NcMLP-NcISY] @XII-5, [CroHYS] @IV-1 , [**Him**SGD1] @IV-2	Secologanin+ tryptamine → tetrahydroalstonine
MIA-FC-20	Cas9 @ XII-1, atf1Δ oye2Δ, oye3Δ ari1Δ adh6Δ , [CroG8H-CroCYB5] @X-3, [CroCPR] @XI-3, [CroIO] @XII-2, [CroSTR-CroSLS] @X-4, [Cro7DLGT-Cro7DLH] @XI-1, [CroLAMT-CroADH2] @XII-4, [Vmi8HGO-A] @X-2, [NcMLP-NcISY] @XII-5, [CroHYS] @IV-1 , [**Ipe**SGD] @IV-2	Secologanin+ tryptamine → tetrahydroalstonine
MIA-FC-21	Cas9 @ XII-1, atf1Δ oye2Δ, oye3Δ ari1Δ adh6Δ , [CroG8H-CroCYB5] @X-3, [CroCPR] @XI-3, [CroIO] @XII-2, [CroSTR-CroSLS] @X-4, [Cro7DLGT-Cro7DLH] @XI-1, [CroLAMT-CroADH2] @XII-4, [Vmi8HGO-A] @X-2, [NcMLP-	Secologanin+ tryptamine → tetrahydroalstonine

	NcISY] @XII-5, [CroHYS] @IV-1 , [**Lsa**SGD1] @IV-2	
MIA-FC-22	Cas9 @ XII-1, atf1Δ oye2Δ, oye3Δ ari1Δ adh6Δ , [CroG8H-CroCYB5] @X-3, [CroCPR] @XI-3, [CroIO] @XII-2, [CroSTR-CroSLS] @X-4, [Cro7DLGT-Cro7DLH] @XI-1, [CroLAMT-CroADH2] @XII-4, [Vmi8HGO-A] @X-2, [NcMLP-NcISY] @XII-5, [CroHYS] @IV-1 , [**Car**SGD] @IV-2	Secologanin+ tryptamine → tetrahydroalstonine
MIA-FC-23	Cas9 @ XII-1, atf1Δ oye2Δ, oye3Δ ari1Δ adh6Δ , [CroG8H-CroCYB5] @X-3, [CroCPR] @XI-3, [CroIO] @XII-2, [CroSTR-CroSLS] @X-4, [Cro7DLGT-Cro7DLH] @XI-1, [CroLAMT-CroADH2] @XII-4, [Vmi8HGO-A] @X-2, [NcMLP-NcISY] @XII-5, [CroHYS] @IV-1 , [**Oeu**SGD2] @IV-2	Secolocanin + tryptamine → strictosidine* * or tetrahydroalstonine if the candidate SGD does function
MIA-FC-24	Cas9 @ XII-1, atf1Δ oye2Δ, oye3Δ ari1Δ adh6Δ , [CroG8H-CroCYB5] @X-3, [CroCPR] @XI-3, [CroIO] @XII-2, [CroSTR-CroSLS] @X-4, [Cro7DLGT-Cro7DLH] @XI-1, [CroLAMT-CroADH2] @XII-4, [Vmi8HGO-A] @X-2, [NcMLP-NcISY] @XII-5, [CroHYS] @IV-1 , [**Ach**SGD3] @IV-2	Secolocanin + tryptamine → strictosidine* * or tetrahydroalstonine if the candidate SGD does function
MIA-FC-25	Cas9 @ XII-1, atf1Δ oye2Δ, oye3Δ ari1Δ adh6Δ , [CroG8H-CroCYB5] @X-3, [CroCPR] @XI-3, [CroIO] @XII-2, [CroSTR-CroSLS] @X-4, [Cro7DLGT-Cro7DLH] @XI-1, [CroLAMT-CroADH2] @XII-4, [Vmi8HGO-A] @X-2, [NcMLP-NcISY] @XII-5, [CroHYS] @IV-1 , [**Cma**SGD] @IV-2	Secolocanin + tryptamine → strictosidine* ** or tetrahydroalstonine if the candidate SGD does function*

MIA-FC-26	Cas9 @ XII-1, atf1Δ oye2Δ, oye3Δ ari1Δ adh6Δ , [CroG8H-CroCYB5] @X-3, [CroCPR] @XI-3, [CroIO] @XII-2, [CroSTR-CroSLS] @X-4, [Cro7DLGT-Cro7DLH] @XI-1, [CroLAMT-CroADH2] @XII-4, [Vmi8HGO-A] @X-2, [NcMLP-NcISY] @XII-5, [CroHYS] @IV-1 , [**Mmy**SGD] @IV-2	Secolocanin + tryptamine → strictosidine* ** or tetrahydroalstonine if the candidate SGD does function*
MIA-FC-27	Cas9 @ XII-1, atf1Δ oye2Δ, oye3Δ ari1Δ adh6Δ , [CroG8H-CroCYB5] @X-3, [CroCPR] @XI-3, [CroIO] @XII-2, [CroSTR-CroSLS] @X-4, [Cro7DLGT-Cro7DLH] @XI-1, [CroLAMT-CroADH2] @XII-4, [Vmi8HGO-A] @X-2, [NcMLP-NcISY] @XII-5, [CroHYS] @IV-1 , [**Vmi**SGD3] @IV-2	Secolocanin + tryptamine → strictosidine* ** or tetrahydroalstonine if the candidate SGD does function*
MIA-FC-28	Cas9 @ XII-1, atf1Δ oye2Δ, oye3Δ ari1Δ adh6Δ , [CroG8H-CroCYB5] @X-3, [CroCPR] @XI-3, [CroIO] @XII-2, [CroSTR-CroSLS] @X-4, [Cro7DLGT-Cro7DLH] @XI-1, [CroLAMT-CroADH2] @XII-4, [Vmi8HGO-A] @X-2, [NcMLP-NcISY] @XII-5, [CroHYS] @IV-1 , [**Ini**SGD] @IV-2	Secolocanin + tryptamine → strictosidine* ** or tetrahydroalstonine if the candidate SGD does function*
MIA-FC-29	Cas9 @ XII-1, atf1Δ oye2Δ, oye3Δ ari1Δ adh6Δ , [CroG8H-CroCYB5] @X-3, [CroCPR] @XI-3, [CroIO] @XII-2, [CroSTR-CroSLS] @X-4, [Cro7DLGT-Cro7DLH] @XI-1, [CroLAMT-CroADH2] @XII-4, [Vmi8HGO-A] @X-2, [NcMLP-NcISY] @XII-5, [CroHYS] @IV-1 , [**Nsi**SGD2] @IV-2	Secolocanin + tryptamine → strictosidine* ** or tetrahydroalstonine if the candidate SGD does function*

Example 1

Construction of USER backbones

All USER vectors were constructed based on pCfB2315 (pRS413-HIS), linearized by restriction enzymes XhoI and SacI (Thermo-Fisher FastDigest™). All terminators were amplified from CEN.PK113-7D genome using primers flanked with XhoI and SacI restriction sites. A DNA cassette containing the ccdB counter-selection marker (Steyaert J. et al. 1993) was inserted into all USER vectors to ensure high cloning efficiency.

USER assembly of plasmids

All plasmids were constructed using the USER method (Jensen NB et al. 2013). Biobrick for plant genes were amplified from synthetic gBlocks (Integrated DNA Technologies and Twist Biosciences), codon optimized for expression in yeast host. Biobrick for promoters were amplified from yeast CEN.PK113-7D genome.

Construction of strains

All strains were constructed using the CRISPR-Cas9 method described in Jakočiūnas T. et al. 2015.

Example 2

Showing that CroSGD does not function in yeast

Geerlings et al. (Geerlings, A., 2000 and WO 00/42200) originally isolated a full-length cDNA clone from a Catharanthus roseus cDNA library giving rise to SGD activity in an in vitro assay.

To confirm if CroSGD could be validated and functionalized in yeast, CroSGD was expressed according to Geerlings et al. by using the strong glycolytic and constitutive active promoters TDH3 and TEF1, respectively.

The following yeast strains were produced, containing SGD and tetrahydroalstonine (THA) synthase both from *Catharantus roseus*, i.e. CroSGD and CroTHAS.

Strain MIA-BJ (EZ-Swap, full CroSTR) expressing:

- P1-TDH3-**Cro**SGD_nls-P2_TEF1-CroTHAS_nls
- P1-TDH3-**Cro**SGD_cyt-P2_TEF1-CroTHAS_cyt
- P2-TEF1-**Cro**SGD-5xGS-CroTHAS_nls
- P2-TEF1-CroTHAS-5xGS-**Cro**SGD_nls

- P2-TEF1-**Cro**SGD-5xGS-CroTHAS_cyt
- P2-TEF1-CroTHAS-5xGS-**Cro**SGD_cyt
- P1-TEF1-**Cro**SGD_nls-P2_PGK1-CroTHAS_nls
- P1-TEF1-**Cro**SGD_cyt-P2_PGK1-CroTHAS_cyt
- P1-TEF1-**Cro**SGD_nls-P2_PGK1-CroTHAS_cyt
- P1-TEF1-**Cro**SGD_cyt-P2_PGK1-CroTHAS_nls

The high-resolution analytical results obtained from LC-MS analysis expressing CroSGD alone and in various tagged and CroSGD-fusion versions contradicts the results presented by Geerlings et al. are not valid.

Figure 1 shows the LC-MS analysis of tetrahydroalstonine (THA). From figure 1 it can be seen that none of the strains expressing CroSGD could produce detectable amount of tetrahydroalstonine.

As a positive control, the following strains were created, strain MIA-BJ (EZ-Swap, full CroSTR) expressing:

- P1-TEF1-**Rse**SGD-P2_PGK1-CroTHAS_nls
- P1-TEF1-**Rse**SGD-P2_PGK1-CroTHAS_cyt

Surprisingly, and in contrast to the strains expressing CroSGD, the yeast stain expressing RseSGD (P1-TEF1-**Rse**SGD-P2_PGK1-CroTHAS_nls) was able to produce tetrahydroalstonine, thus showing that RseSGD is functional in yeast (Figure 1). Tetrahydroalstonine was detected in both samples from supernatant (filtered medium) and cell pellet.

Example 3

SGD homology search

To further investigate, and ultimately enable, functionalization of the critical SGD node in yeast, a homology-search for SGDs against the NCBI database and using the CroSGD protein sequence as a query was performed. From this search, eight different SGD homologs from *Catharanthus roseus* (*Cro*SGD), *Rauvolfia serpentina* (*Rse*SGD), *Rauvolfia verticillata* (*Rve*SGD), *Gelsemium sempervirens* (*Gse*SGD), *Camptotheca acuminate* (*Cac*SGD), *Scedosporium apiospermum* (*Sap*SGD), *Uncaria tomentosa* (*Uto*SGD) and *Glycine soja* (*GsoSGD)* were selected.

The eight protein sequences were aligned with the t-Coffee web server (Figure 2).

Among the eight SGDs selected for this test, two (*Catharanthus roseus* and *Rauvolfia serpentina*) are known to have SGD activity *in vitro*, four are putative SGD from MIA producing plants (*Rauvolfia verticillata, Gelsemium sempervirens, Camptotheca acuminate* and *Uncaria tomentosa*). *Scedosporium apiospermum* is a fungus known to produce other alkaloids. *Glycine soja*, which is unlikely to have SGD activity, was chosen as a negative control. See table 3 below.

Table 3.

Abbreviation	Function	Species	Family	MIA production in the origin organism
RseSGD	*In vitro* verified SGD	*Rauvolfia serpentina*	*Apocyanaceae*	Yes
RveSGD	Putative SGD	*Rauvolfia verticillate*	*Apocyanaceae*	Yes
CroSGD	*In vitro* verified SGD	*Catharanthus roseus*	*apocyanaceae*	Yes
GseSGD	Putative SGD	*Gelsemium sempervirens*	*Gelsemiacea*	Yes
UtoSGD	Putative SGD	*Uncaria tomentosa*	*Rubiaceae*	Yes
CacSGD	Putative SGD	*Camptotheca acuminata*	*Nyssaseae*	Yes
SapSGD	Putative SGD	*Scedosporium apiospermum*	*Microascaceae (fungi)*	No

GsoSGD	Putative GH1 beta-gucosidase	*Glycine soja*	*Phaseoleae*	No

Each one of the eight SGD together with the CroHYS (capable of converting strictosidine aglycone to tetrahydroalsoinine) gene were integrated into a MIA-BJ strain expressing CroG8H + CroCYB5 + CroCPR + Cro8HGO + CroIS + CroIO + CroSTR + CroSLS + Cro7DLGT + Cro7DLH + CroLAMT + CroADH2, resulting in strains MIA-CA-1 to MIA-CA-8

MIA-CA-1: MIA-BJ strain + CroSGD + CroHYS
MIA-CA-2: MIA-BJ strain + **Rse**SGD + CroHYS
MIA-CA-3: MIA-BJ strain + RveSGD + CroHYS
MIA-CA-4: MIA-BJ strain + GseSGD + CroHYS
MIA-CA-5: MIA-BJ strain + CacSGD + CroHYS
MIA-CA-6: MIA-BJ strain + SapSGD + CroHYS
MIA-CA-7: MIA-BJ strain + UtoSGD + CroHYS
MIA-CA-8: MIA-BJ strain + GsoSGD + CroHYS

First, all strains were grown (in triplicates) in 150 uL of YPD for overnight to saturation. Then, 10 ul preculture was transferred into 500 uL of synthetic complete (SC) medium with 2% glucose, supplemented with 0.1 mM of secologanin and 1 mM of tryptamine. After 6 days, 200 uL supernatant was filtered through a 0.2 µm filter membrane suitable for aquaeus solutions such as the AcroPrep™ Advance, 350 uL, 0.2 micron Supor® membrane for media/water. Next, 20 uL of 250 mg/L caffeine was added to each sample as internal standard before analysis on the LC-MS.

The sample caffeine mixtures were analysed on LC-MS to measure secologanin, strictosidine and tetrahydroalstonine concentrations.

Yeast strains expressing GseSGD, SapSGD, RveSGD and RseSGD were able to produce tetrahydroalstonine (Figure 3). Whereas, CacSGD, CroSGD and UtoSGD, as well as their control GsSGD were not able to produce tetrahydroalstonine. The p-value represents comparison between the negative control (GsoSGD) and each of CacSGD, CroSGD and UtoSGD.

The yeast strain expressing RseSGD was able to produce at least 10 µM tetrahydroalstonine.

Example 4

Cellular localisation and expression

In order to understand the functional discrepancy between CroSGD and RseSGD in yeast, the two enzymes were GFP-tagged and their subcellular localization was studied. A clear difference in both level of expression and localization was observed for CroSGD and RseSGD.

The yeast cells expressing GFP-linker-CroSGD showed weak expression of CroSGD, as well as a nuclear localization of the CroSGD, whereas the yeast cells expressing GFP-linker-RseSGD showed higher RseSGD expression and a supramolecular localization pattern (Figure 4) resembling CroSGD localization in planta.

Example 5

Production of strictosidine aglycone and heteroyohimbines

Strictosidine aglycone and tetrahydroalstonine

CroSGD or RseSGD alone or in combination with the CroTHAS were inserted into the MIA-BJ strain (CroG8H + CroCYB5 + CroCPR + Cro8HGO + CroIS + CroIO + CroSTR + CroSLS + Cro7DLGT + Cro7DLH + CroLAMT + CroADH2), resulting in strains MIA-BZ-1 to MIA-BZ-4:

- MIA-BZ-1: MIA-BJ strain + p*TEF1*->CroSGD-tADH1
- MIA-BZ-2: MIA-BJ strain + p*TEF1*->**Rse**SGD-tADH1
- MIA-BZ-3: MIA-BJ strain + t*CYC1*-CroTHAS<-p*PGK1*-p*TEF1*->CroSGD-t*ADH1*
- MIA-BZ-4: MIA-BJ strain + t*CYC1*-CroTHAS<-p*PGK1*-p*TEF1*->**Rse**SGD-t*ADH1*

The yeast strains MIA-BZ-1 to MIA-BZ-4 as well as their control (MIA-BJ strain), were tested in batch fermentation using 96-well deep plate as the following.

First, all strains were grown (in triplicates) in 150 uL of YPD for overnight to saturation. Then, 10 ul preculture was transferred into 500 uL of synthetic complete (SC) medium with 2% glucose, supplemented with 0.1 mM of secologanin and 1 mM of tryptamine.

After 6 days, 200 uL supernatant was filtered through a 0.2 µm filter membrane suitable for aquaeus solutions such as the AcroPrep™ Advance, 350 uL, 0.2 micron Supor® membrane for media/water. Next, 20 uL of 250 mg/L caffeine was added to each sample as an internal standard before analysis on the LC-MS.

Strictosidine aglycone was measured by Orbitrap Fusion™ Tribrid™ MS.

Analysis of strictosidine aglycone peaks on the Orbitrap Fusion™ Tribrid™ MS (positive mode, mass 351.1703 Da) is shown in table 4.

Table 4.

	Mass pos mode, 351.1703 Da Strictosidine aglycone production		
	4.08 min	4.40 min	4.52 min
MIA-BJ (EZ-Swap, full CroSTR)	N.D.	N.D.	N.D.
MIA-BJ + CroSGD	N.D.	N.D.	N.D.
MIA-BJ + **Rse**SGD	3.90E+06	7.31E+06	4.31E+06
MIA-BJ + CroSGD + CroTHAS	N.D.	N.D.	N.D.
MIA-BJ + **Rse**SGD + CroTHAS	1.56E+06	2.14E+06	1.18E+06

These results show that yeast strains expressing RseSGD are able to convert secologanin and tryptamine into strictosidine aglycone. Whereas the yeast strains expressing CroSGD, alone or in combination with CroTHAS, do not produce strictosidine aglycone. This shows that RseSGD is functional in yeast, while CroSGD is not functional in yeast.

Alstonine

To further explore if yeast could be used as a microbial platform for MIA biosynthesis RseSGD and CroTHAS were co-expressed with a sapargan bridge enzymes (SBE) from either *Gelsemium sempervirens* (GseSBE), *Catharantus roseus* (CroSBE) or *Rauvolfia serpentina* (RseSBE), thereby enabling production of a second heteroyohimbine, alstonine.

Strain MIA-BJ (EZ-Swap, full CroSTR) expressing:

- P1-TEF1-**Rse**SGD-P2_PGK1-CroTHAS_empty vector
- P1-TEF1-**Rse**SGD-P2_PGK1-CroTHAS_P1-FET1-CroSBE
- P1-TEF1-**Rse**SGD-P2_PGK1-CroTHAS_P1-FET1-RseSBE
- P1-TEF1-**Rse**SGD-P2_PGK1-CroTHAS_P1-FET1-**Gse**SBE

First, all strains were grown (in triplicates) in 150 uL of YPD for overnight to saturation. Then, 10 ul preculture was transferred into 500 uL of synthetic complete (SC) medium with 2% glucose, supplemented with 0.1 mM of secologanin and 1 mM of tryptamine. After 6 days, 200 uL supernatant was filtered through a 0.2 µm filter membrane suitable for aquaeus solutions such as the AcroPrep ™ Advance, 350 uL, 0.2 micron Supor ® membrane for media/water. Next, 20 uL of 250 mg/L caffeine was added to each sample as internal standard before analysis on the LC-MS.

The sample caffeine mixtures were analysed on LC-MS to measure secologanin, strictosidine and tetrahydroalstonine concentrations.

The biosynthesis of the heteroyohimbine alstonine in yeast cell factories is shown in triplicates in figure 5. Alastonine was measured by Orbitrap Fusion™ Tribrid™ MS.

The yeast cells expressing RseSGD, CroTHAS and GseSBE were capable of converting secologanin and tryptamine to strictosidine aglycone and further capable of converting strictosidine aglycone to tetrahydroalstonine and further capable of converting tetrahydroalstonine to alstonine. This example confirms that RseSGD is functional in yeast.

Example 6

Production of tabersonine and catharanthine

To further demonstrate functionalized RseSGD in yeast, the biosynthetic pathway steps from strictosidine aglycone to tabersonine and catharanthine (MIA-DC) were engineered.

Strain MIA-DC:
CroCPR + CroCYB5 + CroCPR + CroCYB5 + CroSTR + CroGS + RseSGD + CroGO + CroRedoc1 + CroRedox2 + CroSAT + CroPAS + CroCPAS + CroTS + CroCS

The MIA-DC and MIA-DA (control) strains were tested in batch fermentation using 96-

well deep plate as the following.

First, all strains were grown (in triplicates) in 150 uL YPD for overnight to saturation. Then, 10 ul preculture was transferred into 500 uL of synthetic complete (SC) medium with 2% glucose, supplemented with 0.1 mM of secologanin and 1 mM of tryptamine. After 6 days, 200 uL of supernatant was filtered through a 0.2 µm filter membrane suitable for aquaeus solutions such as the AcroPrep ™ Advance, 350 uL, 0.2 micron Supor ® membrane for media/water. Next, 20 uL of 250 mg/L caffeine was added to each sample as internal standard before analysis on the LC-MS.

The production of tabersonine and catharanthine were measured by LC-MS.

Yeast-based production of tabersonine and catharanthine were detected, based on precursor feeding of 0.1 mM of secologanine and 1 mM of tryptamine upstream the RseSGD in strain MIA-DC (Figure 6A-D and 7).

Example 7

Expanded SGD homology search

To further investigate, and ultimately enable, functionalization of the critical SGD node in yeast, a homology-search for SGDs against the NCBI database and the PhytoMetaSyn database was performed using the RseSGD and SapSGD protein sequences as queries. From this search, 28 different SGD homologs were selected from *Rauvolfia serpentina (Rse*SGD2*), Vinca minor* (*Vmi*SGD1 and *Vmi*SGD3*), Tabernaemontana elegans (Tel*SGD*), Amsonia hubrichtii (Ahu*SGD*), Ophiorrhiza pumila, (Opu*SGD*), Nyssa sinensis, (Nsi*SGD1 and *Nsi*SGD2*), Coffea arabica (Car*SGD*), Carapichea ipecacuanha (Ipe*SGD*), Handroanthus impetiginosus (Him*SGD2 and *Him*SGD1*), Sesamum indicum (Sin*SGD*), Olea europaea (Oeu*SGD*), Actinidia chinensis var. chinensis (Ach*SGD1, *Ach*SGD2 and *Ach*SGD3*), Helianthus annuus (Han*SGD*), Lactuca sativa (Lse*SGD*), Ipomoea nil (Ini*SGD*), Chelidonium majus (Cma*SGD*), Vigna unguiculata (Vun*SGD*), Heliocybe sulcate (Hsu*SGD*), Pyricularia grisea (Pgr*SGD*), Lomentospora prolificans (Lpr*SGD*), Hydnomerulius pinastri MD-312 (Hpi*SGD*), Madurella mycetomatis (Mmy*SGD*),* and *Moniliophthora roreri MCA 2997 (Mro*SGD*).*

The 28 protein sequences together with RseSGD, RveSGD, CroSGD, GseSGD, CacSGD, UtoSGD, GsoSGD, and SapSGD were aligned using the t-coffee server (Figure 12). Pairwise sequence identities were calculated from this alignment with CLC Main Workbench 8.0. (Figure 13)

Among the 28 selected sequences for this test two (RseSGD2 and IpeSGD) are known to have low SGD activity *in vitro*, seven are putative beta-glucosidases or hypothetical proteins from MIA producing plants (*Vinca minor, Tabernaemontana elegans, Amsonia hubrichtii, Ophiorrhiza pumila, Nyssa sinensis*), one (OeuSGD) is a oleuropein beta-glucosidase from *Olea europaea,* and 12 are putative beta-glucosidases with various putative activities from plants that do not produce MIAs but a range on different glycosylated natural products (*Coffea arabica, Handroanthus impetiginosus, Sesamum indicum, Actinidia chinensis var. chinensis, Helianthus annuus, Lactuca sativa, Ipomoea nil, Chelidonium majus,* and *Vigna unguiculata)*. Six of the selected sequences are putative beta-glucosidases and hypothetical proteins from fungi (*Heliocybe sulcate, Pyricularia grisea, Lomentospora prolificans, Hydnomerulius pinastri* MD-312, *Madurella mycetomatis,* and *Moniliophthora roreri* MCA 2997). Nothing has been reported on glycosylated natural products produced by any of these fungi.

Table 5

Abbreviation	Function	Species	Family	MIA production in the origin organism
RseSGD2	raucaffricine-O-beta-D-glucosidase	*Rauvolfia serpentina*	*Apocynaceae*	Yes
VmiSGD1	Putative beta-glucosidase	*Vinca minor*	*Apocynaceae*	Yes
VmiSGD3	Putative Beta-glucosidase	*Vinca minor*	*Apocynaceae*	Yes

TelSGD	Putative beta-glucosidase	*Tabernaemontana elegans*	*Apocynaceae*	Yes
AhuSGD	Putative beta-glucosidase	*Amsonia hubrichtii*	*Apocynaceae*	Yes
OpuSGD	Putative beta-glucosidase	*Ophiorrhiza pumila*	*Rubiaceae*	Yes
NsiSGD1	Hypothetical protein	*Nyssa sinensis*	*Nyssaceae*	Yes
NsiSGD2	Hypothetical protein	*Nyssa sinensis*	*Nyssaceae*	Yes
CarSGD	Putative raucaffricine-O-beta-D-glucosidase	*Coffea arabica*	*Rubiaceae*	No
IpeSGD	Beta-glucosidase	*Carapichea ipecacuanha*	*Rubiaceae*	No
HimSGD1	Putative beta-glucosidase	*Handroanthus impetiginosus*	*Bignoniaceae*	No
HimSGD2	Putative beta-glucosidase	*Handroanthus impetiginosus*	*Bignoniaceae*	No
SinSGD	Putative beta-glucosidase	*Sesamum indicum*	*Pedaliaceae*	No
OeuSGD	Oleuropein beta-glucosidase	*Olea europaea*	*Oleaceae*	No
AchSGD1	Putative beta-glucosidase	*Actinidia chinensis var. chinensis*	*Actinidiaceae*	No

AchSGD2	Putative beta-glucosidase	*Actinidia chinensis var. chinensis*	*Actinidiaceae*	No
AchSGD3	Putative beta-glucosidase	*Actinidia chinensis var. chinensis*	*Actinidiaceae*	No
HanSGD	Putative SGD	*Helianthus annuus*	*Asteraceae*	No
LsaSGD	Putative beta-glucosidase	*Lactuca sativa*	*Asteraceae*	No
IniSGD	Putative raucaffricine-O-beta-D-glucosidase	*Ipomoea nil*	*Convolvulaceae*	No
CmaSGD	Putative beta-glucosidase	*Chelidonium majus*	*Papaveraceae*	No
VunSGD	Putative cyanogenic beta-glucosidase	*Vigna unguiculata*	*Fabaceae*	No
HsuSGD	Putative beta-glucosidase	*Heliocybe sulcata*	*Gloeophyllaceae (fungi)*	No
PgrSGD	Putative lactase-phlorizin hydrolase	*Pyricularia grisea*	*Magnaporthaceae (fungi)*	No
LprSGD	Hypothetical protein	*Lomentospora prolificans*	*Microascaceae (fungi)*	No
HpiSGD	Putative GH1 family beta-glucosidase	*Hydnomerulius pinastri* MD-312	*(fungi)*	No

MmySGD	Putative Beta-glucosidase	*Madurella mycetomatis*	*(fungi)*	No
MroSGD	Putative beta-glucosidase	*Moniliophthora roreri* MCA 2997	*(fungi)*	No

Each one of the 28 SGD and CroSGD together with the CroHYS (capable of converting strictosidine aglycone to tetrahydroalsoinine) gene were integrated into a MIA-FA strain expressing CroG8H + Vmi8HGO-A + NcMLP + NclSY + CroCYB5 + CroCPR + CrolO + CroSTR + CroSLS + Cro7DLGT + Cro7DLH + CroLAMT + CroADH2 + CroHYS , resulting in strains MIA-FC-1 to MIA-FC-29. CroSGD was included as a negative control since it was already shown in example 2 to be unable to convert strictosidine to strictosidine aglycone in yeast.

MIA-FC-1: MIA-FA + CroSGD
MIA-FC-2: MIA-FA + VmiSGD1
MIA-FC-3: MIA-FA + AhuSGD
MIA-FC-4: MIA-FA + HimSGD2
MIA-FC-5: MIA-FA + SinSGD
MIA-FC-6: MIA-FA + TelSGD
MIA-FC-7: MIA-FA + VunSGD
MIA-FC-8: MIA-FA + NsiSGD1
MIA-FC-9: MIA-FA + LprSGD
MIA-FC-10: MIA-FA + AchSGD1
MIA-FC-11: MIA-FA + HsuSGD
MIA-FC-12: MIA-FA + MroSGD
MIA-FC-13: MIA-FA + RseSGD2
MIA-FC-14: MIA-FA + PgrSGD
MIA-FC-15: MIA-FA + OpuSGD
MIA-FC-16: MIA-FA + HpiSGD
MIA-FC-17: MIA-FA + HanSGD1
MIA-FC-18: MIA-FA + AchSGD2
MIA-FC-19: MIA-FA + HimSGD1
MIA-FC-20: MIA-FA + IpeSGD

MIA-FC-21: MIA-FA + LsaSGD1
MIA-FC-22: MIA-FA + CarSGD
MIA-FC-23: MIA-FA + OeuSGD
MIA-FC-24: MIA-FA + AchSGD3
MIA-FC-25: MIA-FA + CmaSGD
MIA-FC-26: MIA-FA + MmySGD
MIA-FC-27: MIA-FA + VmiSGD3
MIA-FC-28: MIA-FA + IniSGD
MIA-FC-29: MIA-FA + NsiSGD2

First, all strains were grown (in triplicates) in 150 uL of YPD overnight to saturation. Then, 10 ul preculture was transferred into 500 uL of synthetic complete (SC) medium with 2% glucose, supplemented with 0.1 mM of secologanin and 1 mM of tryptamine. After 6 days, 200 uL supernatant was filtered through a 0.2 µm filter membrane suitable for aquaeus solutions such as the AcroPrep™ Advance, 350 uL, 0.2 micron Supor® membrane for media/water. Next, 20 uL of 250 mg/L caffeine was added to each sample as internal standard before analysis on the LC-MS.

The sample caffeine mixtures were analysed on LC-MS to measure secologanin and tetrahydroalstonine concentrations.

Yeast strains expressing VmiSGD1, AhuSGD, HimSGD2, SinSGD, TelSGD, VunSGD, NsiSGD1, LprSGD, AchSGD1, HsuSGD, MroSGD, RseSGD2, PgrSGD, OpuSGD, HpiSGD, HanSGD1, AchSGD2, HimSGD1, IpeSGD, LsaSGD1, and CarSGD were able to produce tetrahydroalstonine and hereby also strictosidine aglycone (Figure 8) whereas yeast strains expressing OeuSGD, AchSGD3, CmaSGD, MmySGD, VmiSGD3, IniSGD, and NsiSGD2, as well as the negative control CroSGD were not able to produce tetrahydroalstonine. The p-value represents comparison between the negative control (CroSGD) and each of OeuSGD, AchSGD3, CmaSGD, MmySGD, VmiSGD3, IniSGD, and NsiSGD2. More homologs from MIA and non-MIA producing plants were tested, but none were able to produce tetrahydroalstonine.

Example 8

8. 1 Characterization of SGD domains

To investigate which sequence domains are critical for SGD functionalization in yeast the protein sequences of a functional SGD (**Rse**SGD) and a non-functional SGD (**Cro**SGD) were aligned and divided into four domains which were then reassembled in all 16 possible combinations. The domains of RseSGD are termed R and the domains of CroSGD are termend C in this Example. Two combinations (RRRR-SGD and CCCC-SGD) corresponds to the two wild type protein sequences (RseSGD and CroSGD). The four domains are 76 to 203 amino acids long with varying sequence identity (table 6).

Table 6

	Domain 1		**Domain 2**		**Domain 3**		**Domain 4**	
	start	stop	start	stop	start	stop	start	stop
RseSGD	M1	R115	F116	G266	E267	G456	V457	stop
	115		152		190		76	
CroSGD	M1	R123	F124	G274	E275	G477	V478	stop
	123		151		203		78	
Seq_ID	63.80%		79.60%		64.20%		77.60%	

Each of the 16 shuffled SGDs were cloned with USER fusion (Geu-Flores F et al. 2007) on a plasmid and transformed into a MIA-FA strain capable of expressingCroG8H + Vmi8HGO-A + NcMLP + NclSY + CroCYB5 + CroCPR + CroIO + CroSTR + CroSLS + Cro7DLGT + Cro7DLH + CroLAMT + CroADH2 + CroHYS, resulting in strains MIA-FD-1 to MIA-FD-16 (table 7). The MIA-FA strain is capable of synthesizing strictosidine when fed tryptamine and secologanin, or other precursors in the secologanin biosynthetic pathway from geraniol, and is also capable of converting strictosidine aclycone to tetrahydroalstonine if a functional SGD capable of converting strictosidine to strictosidine aglycone is coexpressed.

Table 7

Strain	Domain 1	Domain 2	Domain 3	Domain 4
MIA-FD-1: MIA-FA + pRS413U_pTEF1_CCCC-SGD	CroSGD	CroSGD	CroSGD	CroSGD
MIA-FD-2: MIA-FA + pRS413U_pTEF1_CRCC-SGD	CroSGD	RseSGD	CroSGD	CroSGD
MIA-FD-3: MIA-FA + pRS413U_pTEF1_CRCR-SGD	CroSGD	RseSGD	CroSGD	RseSGD
MIA-FD-4: MIA-FA + pRS413U_pTEF1_CCCR-SGD	CroSGD	CroSGD	CroSGD	RseSGD
MIA-FD-5: MIA-FA + pRS413U_pTEF1_CRRC-SGD	CroSGD	RseSGD	RseSGD	CroSGD
MIA-FD-6: MIA-FA + pRS413U_pTEF1_CCRC-SGD	CroSGD	CroSGD	RseSGD	RseSGD
MIA-FD-7: MIA-FA + pRS413U_pTEF1_CRRR-SGD	CroSGD	RseSGD	RseSGD	RseSGD
MIA-FD-8: MIA-FA + pRS413U_pTEF1_CCRR-SGD	CroSGD	CroSGD	RseSGD	RseSGD
MIA-FD-9: MIA-FA + pRS413U_pTEF1_RRCC-SGD	RseSGD	RseSGD	CroSGD	CroSGD
MIA-FD-10: MIA-FA + pRS413U_pTEF1_RCCC-SGD	RseSGD	CroSGD	CroSGD	CroSGD
MIA-FD-11: MIA-FA + pRS413U_pTEF1_RRCR-SGD	RseSGD	RseSGD	CroSGD	RseSGD
MIA-FD-12: MIA-FA + pRS413U_pTEF1_RCCR-SGD	RseSGD	CroSGD	CroSGD	RseSGD
MIA-FD-13: MIA-FA + pRS413U_pTEF1_RRRC-SGD	RseSGD	RseSGD	RseSGD	CroSGD
MIA-FD-14: MIA-FA + pRS413U_pTEF1_RCRC-SGD	RseSGD	CroSGD	RseSGD	CroSGD
MIA-FD-15: MIA-FA + pRS413U_pTEF1_RCRR-SGD	RseSGD	CroSGD	RseSGD	RseSGD
MIA-FD-16: MIA-FA + pRS413U_pTEF1_RRRR-SGD	RseSGD	RseSGD	RseSGD	RseSGD

First, all strains were grown (in triplicates) in 150 uL of synthetic complete without histidine (SC-HIS) overnight to saturation. Then, 10 ul preculture was transferred into 500 uL of SC-HIS medium with 2% glucose, supplemented with 0.1 mM of secologanin and 1 mM of tryptamine. After 6 days, 200 uL supernatant was filtered through a 0.2 µm filter membrane suitable for aquaeus solutions such as the AcroPrep™ Advance, 350 uL, 0.2 micron Supor® membrane for media/water. Next, 20 uL of 250 mg/L caffeine was added to each sample as internal standard before analysis on the LC-MS.

The sample caffeine mixtures were analysed on LC-MS to measure secologanin tetrahydroalstonine concentrations.

Results

Yeast strains expressing CRRC-SGD, RRRC-SGD, RCRC-SGD, CCRC-SGD, CRRR-SGD, CCRR-SGD, RCRR-SGD, and RRRR-SGD were able to produce tetrahydroalstonine (Figure 9). All functional SGD variants have RseSGD domain 3. All SGD variants with CroSGD domain 3 were not able to produce tetrahydroalstonine. The identity of domain 1 and 2 has low or no effect. Of the functional SGD variants, the four sequences with RseSGD domain 3 and domain 4 (CRRR-SGD, CCRR-SGD, RCRR-SGD, and RRRR-SGD) are able to produce the highest amount of tetrahydroalstonine. CCRR-SGD is the best variant capable of producing more tetrahydroalstonine than the wild type RseSGD (RRRR-SGD)

8.2 Production of tetrahydroalstonine in a yeast strain expressing CCRR_SGD

The best SGD variant (CCRR-SGD) were integrated in the MIA-FA strain MIA-FA capable of strain expressing CroG8H + Vmi8HGO-A + NcMLP + NclSY + CroCYB5 + CroCPR + CroIO + CroSTR + CroSLS + Cro7DLGT + Cro7DLH + CroLAMT + CroADH2 + CroHYS, resulting in the strain MIA-FE:

MIA-FE: MIA-FA + CCRR-SGD

First, MIA-FE was grown (in triplicates) in 150 uL of YPD overnight to saturation. Then, 10 ul preculture was transferred into 500 uL of synthetic complete (SC) medium with 2% glucose, supplemented with 0.1 mM of secologanin and 1 mM of tryptamine. After 6 days, 200 uL supernatant was filtered through t a 0.2 µm filter membrane suitable for aquaeus solutions such as he AcroPrep™ Advance, 350 uL, 0.2 micron Supor® membrane for media/water. Next, 20 uL of 250 mg/L caffeine was added to each sample as internal standard before analysis on the LC-MS.

The sample caffeine mixtures were analysed on LC-MS to measure tetrahydroalstonine concentrations.

Results

The yeast strain expressing CCRR-SGD was able to produce 13.30 µM (±1.29 µM) tetrahydroalstonine.

Example 9

Rescuing the function of other SGD homologs with RseSGD domain 3 and 4

Encouraged by the capability of RseSGD domain 3 and 4 to rescue the non-functional CroSGD in yeast three more SGD variants were cloned swapping domain 3 and 4 between RseSGD and UtoSGD (U), GseSGD (G), and RveSGD (V) respectively. Even though swapping domain 3 alone was able to make CroSGD functional swapping both domain 3 and domain 4 gave the largest improvement and therefor this swapping strategy was expanded to other SGD sequences.

The sequences of the four domains of UtoSGD, GseSGD and RveSGD were determined from a multiple sequence alignment (Figure 12). The first residue in domain 1 is always the start methionine and the last residue in domain 4 is always the last residue in the sequence. The remaining first and last residues are defined as the residues aligning with the first and last residues in the four RseSGD domains. Table 8 summarizes the four domains of RseSGD, CroSGD, UtoSGD, GseSGD, and RveSGD.

Table 8

	Domain 1		**Domain 2**		**Domain 3**		**Domain 4**		**Seq_ID to RseSGD**
	start	stop	start	stop	start	stop	start	stop	
RseSGD	M1	R115	F116	G266	E267	G456	V457	stop	
UtoSGD	M1	R88	F89	G277	K278	G459	V460	stop	40.70%
GseSGD	M1	R92	F93	G265	Q266	G456	V457	stop	53.90%
CroSGD	M1	R123	F124	G274	E275	G477	V478	stop	70.30%
RveSGD	M1	R115	F116	G265	E266	G459	V460	stop	89.90%

Three domain-swap SGD variants and the three wild type SGDs were cloned with USER fusion. The plasmids were transformed into a MIA-FA strain capable of expressing CroG8H + Vmi8HGO-A + NcMLP + NclSY + CroCYB5 + CroCPR + CroIO + CroSTR + CroSLS + Cro7DLGT + Cro7DLH + CroLAMT + CroADH2 + CroHYS, resulting in strains MIA-FD-17 to MIA-FD-22 (table 9). The MIA-FA strain is capable of

synthesizing strictosidine when fed tryptamine and secologanin, or other precursors in the secologanin biosynthetic pathway from geraniol, and is also capable of converting strictosidine aclycone to tetrahydroalstonine if a functional SGD capable of converting strictosidine to strictosidine aglycone is coexpressed

Table 9

MIA-FD-17: MIA-FA + pRS413U_pTEF1_UtoSGD-SGD	**UtoSGD**	**UtoSGD**	**UtoSGD**	**UtoSGD**
MIA-FD-18: MIA-FA + pRS413U_pTEF1_UURR-SGD	UtoSGD	UtoSGD	RseSGD	RseSGD
MIA-FD-19: MIA-FA + pRS413U_pTEF1_GseSGD-SGD	GseSGD	GseSGD	GseSGD	GseSGD
MIA-FD-20: MIA-FA + pRS413U_pTEF1_GGRR-SGD	GseSGD	GseSGD	RseSGD	RseSGD
MIA-FD-21: MIA-FA + pRS413U_pTEF1_RveSGD-SGD	RveSGD	RveSGD	RveSGD	RveSGD
MIA-FD-22: MIA-FA + pRS413U_pTEF1_VVRR-SGD	RveSGD	RveSGD	RseSGD	RseSGD

First, all six strains plus two control strains (MIA-FD-1 and 8) were grown (in triplicates) in 150 uL of synthetic complete without histidine (SC-HIS) overnight to saturation. Then, 10 ul preculture was transferred into 500 uL of SC-HIS medium with 2% glucose, supplemented with 0.1 mM of secologanin and 1 mM of tryptamine. After 6 days, 200 uL supernatant was filtered through a 0.2 µm filter membrane suitable for aquaeus solutions such as the AcroPrep™ Advance, 350 uL, 0.2 micron Supor® membrane for media/water. Next, 20 uL of 250 mg/L caffeine was added to each sample as internal standard before analysis on the LC-MS.

The sample caffeine mixtures were analysed on LC-MS to measure tetrahydroalstonine concentrations.

As already shown in example 9, swapping in RseSGD domain 3 and 4 rescued the function of the non-functional CroSGD (Figure 9). Wild type RveSGD is capable of producing tetrahydroalstonine. Swapping in RseSGD domain 3 and 4 improved the tetrahydroalstonine production about seven fold. GseSGD and UtoSGD have lower

sequence identity to RseSGD (53.9% and 40.7% respectively) than CroSGD and RveSGD (70.3 % and 89.9%). GseSGD can produce tetrahydroalstonine in low concentrations whereas UtoSGD is incapable of tetrahydroalstonine production. Swapping in RseSGD domain 3 and 4 into these two SGDs did not rescue the function of UtoSGD and abolished the low tetrahydroalstonine production of GseSGD.

Example 10

Minimum strictosidine aglycone production in yeast

Strictosidine aglycone is chemically unstable and was impossible to either purchase or purify to use as a standard for quantification. The minimum strictosidine aglycone produced by the tested SGD homologs was calculated from the measured tetrahydroalstonine produced by the yeast strains and the measured secologanin left in the media. It is possible that not all produced strictosidine aglycone is converted to tetrahydroalstonine, and therefore the true strictosidine aglycone titres might in some cases be higher than the estimated minimum production.

Strictosidine aglycone production in µM:

Since strictosidine aglycone is converted to tetrahydroalstonine in equimolar amounts, the minimum strictosidine aglycone titre equals the tetrahydroalstonine titre.

c(strictosidine aglycone) = c(tetrahydroalstonine)

Strictosidine aglycone yields:

The minimum strictosidine algycone yield can be estimated from the strictosidine aglycone titre and the theoretical strictosidine titre. It is assumed that all secologanin taken up by the yeast strain is converted to strictosidine.

Strictosidine_aglycone_% = c(strictosidine aglycone)/(c(secologanin supplemented in media) – c(secologanin left after cultivation))

Example 11

Production of THA in *Escherichia coli*

To test if RseSGD or CroSGD could be used for production of strictosidine aglycone and MIAs in prokaryotic microorganisms an expression system was established in the gram-negative bacterium *Escherichia coli* for *in vivo* conversion of secologanin and

tryptamine to strictosidine by CroSTR, conversion of strictosidine to strictosidine aglycone by RseSGD or CroSGD and conversion of strictosidine aglycone to tetrahydroalstonine by CroHYS. Two low-copy plasmids were cloned for co-expression of the three genes from a polycistronic mRNA under control of a medium strength constitutive promoter. The plasmids were based on pCfB3510(p15A_P2BCD2GFP). The two plasmids and an empty plasmid were transformed into the strain DH5-α giving the three strains MIA-ECO-1 to MIA-ECO-3.

MIA-ECO-1: DH5-α + p15A-AmpR-CroSTR-CroHYS-**Cro**SGD
MIA-ECO-2: DH5-α + p15A-AmpR-CroSTR-CroHYS-**Rse**SGD
MIA-ECO-3: DH5-α + p15A-AmpR

First, all three strains were grown (in triplicates) in 150 uL of Lysogeny broth (LB) medium with 100 µg/mL ampicillin overnight to saturation. Then, 10 ul preculture was transferred into 500 uL LB medium with 100 µg/mL ampicillin and supplemented with 0.1 mM of secologanin and 1 mM of tryptamine. After 48 hours, 200 uL supernatant was filtered through a 0.2 µm filter membrane suitable for aquaeus solutions such as the AcroPrep™ Advance, 350 uL, 0.2 micron Supor® membrane for media/water. Next, 20 uL of 250 mg/L caffeine was added to each sample as internal standard before analysis on the LC-MS.

The sample caffeine mixtures were analysed on LC-MS to measure secologanin, strictosidine, and tetrahydroalstonine concentrations.

Results

The *E. coli* strain MIA-ECO-2 expressing **Rse**SGD, CroSTR, and CroHYS was able to produce tetrahydroalstonine (Figure 11-B). . No strictosidine was detected in the media of the E.coli expressing RseSGD. MIA-ECO-1 expressing **Cro**SGD, CroSTR, and CroHYS produced strictosidine (Figure 11-A) but no tetrahydroalstonine, indicating that like in yeast **Rse**SGD is functional and CroSGD is non-functional.

References

Geerlings, A., Ibañez, M. M., Memelink, J., van Der Heijden, R. & Verpoorte, R. Molecular cloning and analysis of strictosidine beta-D-glucosidase, an enzyme in terpenoid indole alkaloid biosynthesis in Catharanthus roseus. J. Biol. Chem. 275, 3051–3056 (2000).

Fernando Geu-Flores, Hussam H. Nour-Eldin, Morten T. Nielsen and Barbara A. Halkier 2007. USER fusion: a rapid and efficient method for simultaneous fusion and cloning of multiple PCR products. Nucleic Acids Research, 2007, Vol. 35, No. 7 e55. doi:10.1093/nar/gkm106

Guirimand G., Courdavault V., Lanoue A., Mahroug S., Guihur A., Blanc N., Giglioli-Guivarc'h N., St-Pierre B., Burlat V. Strictosidine activation in Apocynaceae: towards a "nuclear time bomb"? BMC Plant Biology 2010, 10:182

Jakočiūnas T, Rajkumar AS, Zhang J, Arsovska D, Rodriguez A, Jendresen CB, Skjødt ML, Nielsen AT, Borodina I, Jensen MK, Keasling JD. CasEMBLR: Cas9-Facilitated Multiloci Genomic Integration of in Vivo Assembled DNA Parts in Saccharomyces cerevisiae.ACS Synth Biol. 2015 Nov 20;4(11):1226-34. doi: 0.1021/acssynbio.5b00007. Epub 2015 Mar 26.

Jensen NB, Strucko T, Kildegaard KR, David F, Maury J, Mortensen UH, Forster J, Nielsen J, Borodina I. EasyClone: method for iterative chromosomal integration of multiple genes in Saccharomyces cerevisiae. FEMS Yeast Res. 2014 Mar;14(2):238-48. doi: 10.1111/1567-1364.12118. Epub 2013 Nov 18.

Luijendick T.J.C., Stenvens, L.H., Verpoorte R. Reaction for the Localization of Strictosidine Glucosidase Activity on Polyacrylamide gels. Phytochemical analysis (1996). doi:3.0.CO;2-H">10.1002/(SICI)1099-1565(199601)7:1<16::AID-PCA280>3.0.CO;2-H.

Stavrinides A., Tatsis E.C., Foureau E., Caputi L., Kellner F., Courdavault V., O'Connor S.E. Unlocking the Diversity of Alkaloids in *Catharanthus roseus*: Nuclear Localization Suggests Metabolic Channeling in Secondary Metabolism. Chemistry & Biology 22, 336–341, March 19, 2015

Steyaert J, Van Melderen L, Bernard P, Thi MH, Loris R, Wyns L, Couturier M. J Mol Purification, circular dichroism analysis, crystallization and preliminary X-ray diffraction analysis of the F plasmid CcdB killer protein Biol. 1993 May 20;231(2):513-5.

WO 00/4220: Verpoorte, R., Van Der Heijden, R., Memelink, J. & Geerlings, A. Strictosidine glucosidase from catharanthus roseus and its use in alkaloid production. World Patent (2000).

Items

1. A microorganism capable of producing strictosidine aglycone, said microorganism expresses

a strictosidine-beta-glucosidase (SGD), capable of converting strictosidine to strictosidine aglycone,

wherein said SGD is a heterologous SGD selected from RseSGD (SEQ ID NO: 24), GseSGD (SEQ ID NO: 25), SapSGD (SEQ ID NO: 26), RveSGD (SEQ ID NO: 27), VmiSGD1 (SEQ ID NO: 47), AhuSGD (SEQ ID NO: 48), HimSGD2 (SEQ ID NO: 49), SinSGD (SEQ ID NO: 50), TelSGD (SEQ ID NO: 51), VunSGD (SEQ ID NO: 52), NsiSGD1 (SEQ ID NO: 53), LprSGD (SEQ ID NO: 54), AchSGD1 (SEQ ID NO: 55), HsuSGD (SEQ ID NO: 56), MroSGD (SEQ ID NO: 57), RseSGD2 (SEQ ID NO: 58), PgrSGD (SEQ ID NO: 59), OpuSGD (SEQ ID NO: 60), HpiSGD (SEQ ID NO: 61), HanSGD1 (SEQ ID NO: 62), AchSGD2 (SEQ ID NO: 63), HimSGD1 (SEQ ID NO: 64), IpeSGD (SEQ ID NO: 65), LsaSGD1 (SEQ ID NO: 66), or CarSGD (SEQ ID NO: 67) or variants thereof having at least 70%, such as at least 80%, such as at least 90%, such as at least 91%, such as at least 92%, such as at least 93%, such as at least 94%, such as at least 95%, such as at least 96%, such as at least 97%, such as at least 98%, such as at least 99%, such as 100% identity thereto,

and/or;

wherein said SGD is a mosaic SGD, wherein said mosaic SGD comprises an amino acid sequence having the general formula

$$D_1\text{-}D_2\text{-}D_3\text{-}D_4$$

wherein D_1 is a first amino acid sequence from a first SGD,
wherein D_2 is a second amino acid sequence from a second SGD,
wherein D_3 is a third amino acid sequence comprising or consisting of amino acids of SEQ ID NO:91 or a variant thereof having at least 90% identity to SEQ ID NO: 91,
wherein D_4 is a fourth amino acid sequence from a fourth SGD or an amino acid sequence consisting of amino acids of SEQ ID NO:92 or a variant thereof having at least 90% identity to SEQ ID NO: 92,

2. wherein said first SGD, second SGD and fourth SGD can be the same or different, with the proviso that said first SGD, second SGD and fourth SGD are not all RseSGD. The microorganism according to item 1, further expressing

a strictosidine synthase (STR), capable of converting secologanin and tryptamine to strictosidine, whereby the microorganism is capable of synthesising strictosidine,

wherein said STR is preferably CroSTR or variants thereof having at least 90%, such as at least 91%, such as at least 92%, such as at least 93%, such as at least 94%, such as at least 95%, such as at least 96%, such as at least 97%, such as at least 98%, such as at least 99%, such as 100% identity to SEQ ID NO: 30.

3. The microorganism according to any one of the preceding items, wherein D_1 comprises or consists of an amino acid sequence corresponding to amino acids M1 to R115 of SEQ ID NO:24.

4. The microorganism according to any one of the preceding items, wherein D_2 comprises or consists of an amino acid sequence corresponding to amino acids F116 to G266 of SEQ ID NO:24.

5. The microorganism according to any one of the preceding items, wherein D_4 comprises or consists of amino acids of SEQ ID NO:92 or a variant thereof having at least 90% identity to SEQ ID NO: 92.

6. The microorganism according to any one of the preceding items, wherein at least one of D_1, D_2 or D_4 is from an SGD which is native to a first organism selected from *Gelsemium sempervirens, Scedosporium apiospermum or Rauvolfia verticillata*, *Vinca minor, Tabernaemontana elegans, Amsonia hubrichtii, Ophiorrhiza pumila, Nyssa sinensis, Coffea arabica, Carapichea ipecacuanha, Handroanthus impetiginosus, Sesamum indicum, Actinidia chinensis var. chinensis, Helianthus annuus, Lactuca sativa, Ipomoea nil, Vigna unguiculata, Heliocybe sulcate, Pyricularia grisea, Lomentospora prolificans, Hydnomerulius pinastri MD-312,* and *Moniliophthora roreri MCA 2997.*

7. The microorgagnism according to any one of the preceding items, wherein the first SGD, the second SGD and the fourth SGD are identical or different.

8. The microorganism according to any one of the preceding items, wherein two of the first SGD, the second SGD and the fourth SGD are identical, or wherein the first SGD, the second SGD and the fourth SGD are different, or wherein the first SGD, the second SGD and the fourth SGD are identical.

9. The microorganism according to any one of the preceding items, wherein said mosaic SGD comprises or consists of an amino acid sequence of SEQ ID NO: 93, SEQ ID NO: 94, SEQ ID NO: 95, SEQ ID NO: 96, SEQ ID NO: 97, SEQ ID NO: 98, SEQ ID NO: 99, or SEQ ID NO: 108, or variants thereof having at least 90% identity or homology thereto, such as at least 91%, such as at least 92%, such as at least 93%, such as at least 94%, such as at least 95%, such as at least 96%, such as at least 97%, such as at least 98%, such as at least 99% identity or homology thereto.

10. The microorganism according to any one the preceding items, further expressing a tetrahydroalstonine synthase (THAS) and/or a heteroyohimbine synthase (HYS), capable of converting strictosidine aglycone to tetrahydroalstonine, whereby the microorganism is capable of synthesising tetrahydroalstonine,
wherein said THAS is preferably CroTHAS and/or HYS is CroHYS or variants thereof, having at least 90%, such as at least 91%, such as at least 92%, such as at least 93%, such as at least 94%, such as at least 95%, such as at least 96%, such as at least 97%, such as at least 98%, such as at least 99%, such as 100% identity to SEQ ID NO: 28 and/or SEQ ID NO: 46.

11. The microorganism according to any of the preceding items, further expressing a sarpargan bridge enzymes (SBE), capable of converting tetrahydroalstonine and ajmalicine to a heteroyohimbine selected from the group consisting of alstonine and serpentine, whereby the microorganism is capable of synthesising alstonine and serpentine,
wherein said SBE is preferably GseSBE or variants thereof having at least 90%, such as at least 91%, such as at least 92%, such as at least 93%, such as at

least 94%, such as at least 95%, such as at least 96%, such as at least 97%, such as at least 98%, such as at least 99%, such as 100% identity to SEQ ID NO: 29.

12. The microorganism according to any one of the preceding items, further expressing

a NADPH--cytochrome P450 reductase (CPR);
a Cytochrome b5 (CYB5);
a Geissoschizine synthase (GS);
a Geissoschizine oxidase (GO);
a Redox1;
a Redox2;
a Stemmadenine O-acetyltransferase (SAT);
a O-acetylstemmadenine oxidase (PAS);
a Dehydroprecondylocarpine acetate synthase (DPAS);
a Tabersonine synthase (TS); and/or
a Catharanthine synthase (CS),

whereby the microorganism is capable of synthesising tabersonine and/or catharanthine,

wherein preferably said CPR is CroCPR, said CYB5 is CroCYB5, said GS is CroSG, said GO is CroGO, said Redox1 is CroRedox1, said Redox2 is CroRedox2, said SAT is CroSAT, said PAS is CroPAS, said DPAS is CroDPAS, said TS is CroTS and/or said CS is CroCS or variants thereof having at least 90%, such as at least 91%, such as at least 92%, such as at least 93%, such as at least 94%, such as at least 95%, such as at least 96%, such as at least 97%, such as at least 98%, such as at least 99%, such as 100% identity to SEQ ID NO: 31, SEQ ID NO: 32, SEQ ID NO: 33, SEQ ID NO: 34, SEQ ID NO: 35, SEQ ID NO: 36, SEQ ID NO: 37, SEQ ID NO: 38, SEQ ID NO: 39, SEQ ID NO: 40 and/or SEQ ID NO: 41, respectively.

13. The microorganism according to any one of the preceding items, capable of producing strictosidine aglycone with a titre of at least 1 µM, such as at least 2 µM, such as at least 4 µM, such as at least 6 µM, such as at least 8 µM such as at least 10 µM or more.

14. The microorganism according to item 10, capable of producing tetrahydroalstonine with a titre of at least 1 µM, such as at least 2 µM, such as at least 4 µM, such as at least 6 µM, such as at least 8 µM such as at least 10 µM or more.

15. The microorganism according to item 11, capable of producing alstonine with a titre of at least 1 µM, such as at least 2 µM, such as at least 4 µM, such as at least 6 µM, such as at least 8 µM such as at least 10 µM or more.

16. The microorganism according to item 12, capable of producing tabersonine with a titre of at least 0.01 µM, such as at least 0.02 µM.

17. The microorganism according to item 12, capable of producing catharanthine with a titre of at least 0.01 µM, such as at least 0.02 µM.

18. The microorganism according to any of the preceding items, wherein the microorganism is selected from the group consisting of yeasts, bacteria, archaea, fungi, protozoa, algae, and viruses, preferably the microorganism is a yeast or a bacteria.

19. The microorganism according to any one of the preceding items, wherein the microorganism is a bacteria.

20. The microorganism according to item 19, wherein the genus of said bacteria is selected from the groups consisting of *Escherichia*, *Corynebacterium, Pseudomonas, Bacillus, Lactococcus, Lactobacillus, Halomonas, Bifidobacterium* and *Enterococcus*.

21. The microorganism according to any one of items 19 to 20, wherein the bacteria is selected from the group consisting of *Escherichia coli, Corynebacterium glutamicum, Pseudomonas putida, Bacillus subtilis, Lactococcus bacillus, Halomonas elongate, Bifidobacterium infantis* and *Enterococcus faecal.*

22. The microorganism according to any one of items 19 to 21, wherein the bacteria is *Escherichia coli.*

23. The microorganism according to any one of the preceding items, wherein the microorganism is a yeast.

24. The microorganism according to item 23, wherein the genus of said yeast cell is selected from the group consisting of *Saccharomyces*, *Pichia*, *Yarrowia*, *Kluyveromyces*, *Candida, Rhodotorula, Rhodosporidium, Cryptococcus, Trichosporon* and *Lipomyces*.

25. The microorganism according to any one of items 23 to 24, wherein the yeast is selected from the group consisting of *Saccharomyces cerevisiae, Pichia pastoris, Kluyveromyces marxianus, Cryptococcus albidus, Lipomyces lipofera, Lipomyces starkeyi, Rhodosporidium toruloides, Rhodotorula glutinis, Trichosporon pullulan* and *Yarrowia lipolytica*

26. The microorganism according to any one of items 23 to 25, wherein the yeast is *Saccharomyces cerevisiae.*

27. The microorganism according to any of the preceding items, wherein the microorganism comprises a nucleic acid encoding SGD, said nucleic acid having at least 90%, such as at least 91%, such as at least 92%, such as at least 93%, such as at least 94%, such as at least 95%, such as at least 96%, such as at least 97%, such as at least 98%, such as at least 99%, such as 100% identity to SEQ ID NO: 1.

28. A method of producing strictosidine aglycone in a microorganism, said method comprising the steps of:
    a) providing a microorganism, said cell expressing:
        a strictosidine-beta-glucosidase (SGD), capable of converting strictosidine to strictosidine aglycone;
    b) incubating said microorganism in a medium comprising strictosidine or a substrate which can be converted to strictosidine by said microorganism;
    c) optionally, recovering the strictosidine aglycone;
    d) optionally, further converting the strictosidine aglycone to monoterpenoid indole alkaloids,

wherein said SGD is a heterologous SGD selected from RseSGD (SEQ ID NO: 24), GseSGD (SEQ ID NO: 25), SapSGD (SEQ ID NO: 26), RveSGD (SEQ ID NO: 27), VmiSGD1 (SEQ ID NO: 47), AhuSGD (SEQ ID NO: 48), HimSGD2 (SEQ ID NO: 49), SinSGD (SEQ ID NO: 50), TelSGD (SEQ ID NO: 51), VunSGD (SEQ ID NO: 52), NsiSGD1 (SEQ ID NO: 53), LprSGD (SEQ ID NO: 54), AchSGD1 (SEQ ID NO: 55), HsuSGD (SEQ ID NO: 56), MroSGD (SEQ ID NO: 57), RseSGD2 (SEQ ID NO: 58), PgrSGD (SEQ ID NO: 59), OpuSGD (SEQ ID NO: 60), HpiSGD (SEQ ID NO: 61), HanSGD1 (SEQ ID NO: 62), AchSGD2 (SEQ ID NO: 63), HimSGD1 (SEQ ID NO: 64), IpeSGD (SEQ ID NO: 65), LsaSGD1 (SEQ ID NO: 66), or CarSGD (SEQ ID NO: 67) or variants thereof having at least 70%, such as at least 80%, such as at least 90%, such as at least 91%, such as at least 92%, such as at least 93%, such as at least 94%, such as at least 95%, such as at least 96%, such as at least 97%, such as at least 98%, such as at least 99%, such as 100% identity thereto,

and/or;

wherein said SGD is a mosaic SGD, wherein said mosaic SGD comprises an amino acid sequence having the general formula

$$D_1\text{-}D_2\text{-}D_3\text{-}D_4$$

wherein D_1 is a first amino acid sequence from a first SGD,
wherein D_2 is a second amino acid sequence from a second SGD,
wherein D_3 is a third amino acid sequence comprising or consisting of amino acids of SEQ ID NO:91 or a variant thereof having at least 90% identity to SEQ ID NO: 91,
wherein D_4 is a fourth amino acid sequence from a fourth SGD or an amino acid sequence consisting of amino acids of SEQ ID NO:92 or a variant thereof having at least 90% identity to SEQ ID NO: 92,
wherein said first SGD, second SGD and fourth SGD can be the same or different, with the proviso that said first SGD, second SGD and fourth SGD are not all RseSGD.

29. The microorganism according to item 28, wherein the SGD, the heterologous SGD and/or the mosaic SGD is as defined in any one of the preceding items.

30. The microorganism according to any one of items 28 to 29, wherein D_1 comprises or consists of an amino acid sequence corresponding to amino acids M1 to R115 of SEQ ID NO:24.

31. The microorganism according to any one of items 28 to 30, wherein D_2 comprises or consists of an amino acid sequence corresponding to amino acids F116 to G266 of SEQ ID NO:24.

32. The microorganism according to any one of items 28 to 31, wherein D_4 comprises or consists of amino acids of SEQ ID NO:92 or a variant thereof having at least 90% identity to SEQ ID NO: 92.

33. The microorganism according to any one of items 28 to 32, wherein at least one of D_1, D_2 or D_4 is from an SGD which is native to a first organism selected from *Gelsemium sempervirens, Scedosporium apiospermum or Rauvolfia verticillata*, *Vinca minor, Tabernaemontana elegans, Amsonia hubrichtii, Ophiorrhiza pumila, Nyssa sinensis, Coffea arabica, Carapichea ipecacuanha, Handroanthus impetiginosus, Sesamum indicum, Actinidia chinensis var. chinensis, Helianthus annuus, Lactuca sativa, Ipomoea nil, Vigna unguiculata, Heliocybe sulcate, Pyricularia grisea, Lomentospora prolificans, Hydnomerulius pinastri MD-312,* and *Moniliophthora roreri MCA 2997.*

34. The microorgagnism according to any one of items 28 to 33, wherein the first SGD, the second SGD and the fourth SGD are identical or different.

35. The microorganism according to any one of items 28 to 34, wherein two of the first SGD, the second SGD and the fourth SGD are identical, or wherein the first SGD, the second SGD and the fourth SGD are different, or wherein the first SGD, the second SGD and the fourth SGD are identical.

36. The microorganism according to items 28 to 35, wherein said mosaic SGD comprises or consists of an amino acid sequence of SEQ ID NO: 93, SEQ ID NO: 94, SEQ ID NO: 95, SEQ ID NO: 96, SEQ ID NO: 97, SEQ ID NO: 98, SEQ ID NO: 99, or SEQ ID NO: 108, or variants

thereof having at least 90% identity or homology thereto, such as at least 91%, such as at least 92%, such as at least 93%, such as at least 94%, such as at least 95%, such as at least 96%, such as at least 97%, such as at least 98%, such as at least 99% identity or homology thereto.

37. The method according to any one of items 28 to 36, wherein the substrate is secologanin and/or tryptamine, and wherein said microorganism further expresses:

a strictosidine synthase (STR), capable of converting secologanin and tryptamine to strictosidine;

wherein said STR is preferably CroSTR or variants thereof having at least 90%, such as at least 91%, such as at least 92%, such as at least 93%, such as at least 94%, such as at least 95%, such as at least 96%, such as at least 97%, such as at least 98%, such as at least 99%, such as 100% identity to SEQ ID NO: 30.

38. The method according to any one of items 28 to 37, wherein the method comprising step d) and wherein said microorganism further expresses:

a tetrahydroalstonine synthase (THAS) and/or or a heteroyohimbine synthase (HSY), capable of converting strictosidine aglycone to tetrahydroalstonine;

wherein preferably said THAS is identical to or has at least 90%, such as at least 91%, such as at least 92%, such as at least 93%, such as at least 94%, such as at least 95%, such as at least 96%, such as at least 97%, such as at least 98%, such as at least 99%, such as 100% identity to SEQ ID NO: 28 and/or HYS is identical to or has at least 90%, such as at least 91%, such as at least 92%, such as at least 93%, such as at least 94%, such as at least 95%, such as at least 96%, such as at least 97%, such as at least 98%, such as at least 99%, such as 100% identity to SEQ ID NO: 46.

39. The method according to items 28 to 38, wherein said method further comprises the step of recover tetrahydroalstonine.

40. The method according to any one of items 28 to 39, wherein the method comprising step d) and wherein said microorganism further expresses:

a sapargan bridge enzyme (SBE), capable of converting tetrahydroalstonine to alstonine;

wherein preferably said SBE is identical to or has at least 90%, such as at least 91%, such as at least 92%, such as at least 93%, such as at least 94%, such as at least 95%, such as at least 96%, such as at least 97%, such as at least 98%, such as at least 99%, such as 100% identity to SEQ ID NO: 29.

41. The method according to item 40, wherein said method further comprises the step of recovering alstonine.

42. The method according to any one of items 28 to 41, wherein the method comprises step d) and wherein said microorganism further expresses:

a NADPH--cytochrome P450 reductase (CPR);
a Cytochrome b5 (CYB5);
a Geissoschizine synthase (GS);
a Geissoschizine oxidase (GO);
a Redox1;
a Redox2;
a Stemmadenine O-acetyltransferase (SAT);
a O-acetylstemmadenine oxidase (PAS);
a Dehydroprecondylocarpine acetate synthase (DPAS);
a Tabersonine synthase (TS); and/or
a Catharanthine synthase (CS),

wherein preferably said CPR is CroCPR, said CYB5 is CroCYB5, said GS is CroSG, said GO is CroGO, said Redox1 is CroRedox1, said Redox2 is CroRedox2, said SAT is CroSAT, said PAS is CroPAS, said DPAS is CroDPAS, said TS is CroTS and/or said CS is CroCS or variants thereof having at least 90%, such as at least 91%, such as at least 92%, such as at least 93%, such as at least 94%, such as at least 95%, such as at least 96%, such as at least 97%, such as at least 98%, such as at least 99%, such as 100% identity to SEQ ID NO: 31, SEQ ID NO: 32, SEQ ID NO: 33, SEQ ID NO: 34, SEQ ID NO: 35, SEQ ID NO: 36, SEQ ID NO: 37, SEQ ID NO: 38, SEQ ID NO: 39, SEQ ID NO: 40 and/or SEQ ID NO: 41, respectively.

wherein the microorganism is capable of producing tabersonine and/or catharanthine, optionally wherein said method further comprises the step of recovering tabersonine and/or catharanthine.

43. The method according to any one of items 28 to 42, wherein the medium comprises at least strictosidine, preferably at a concentration of at least 0.05 mM, such as at least 0.1 mM, such as at least 0.5 mM, such as at least 1 mM.

44. The method according to any one of items 288 to 43, wherein the medium comprises at least tryptamine and secologanin, preferably at a concentration of at least 0.05 mM, such as at least 0.1 mM, such as at least 0.5 mM, such as at least 1 mM.

45. A nucleic acid construct comprising a sequence identical to or having at least 90% identity, such as at least 91%, such as at least 92%, such as at least 93%, such as at least 94%, such as at least 95%, such as at least 96%, such as at least 97%, such as at least 98%, such as at least 99%, such as 100% identity to SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 3,SEQ ID NO: 4, SEQ ID NO:68, SEQ ID NO:69, SEQ ID NO:70, SEQ ID NO: 71, SEQ ID NO:72, SEQ ID NO: 73, SEQ ID NO:74, SEQ ID NO: 75, SEQ ID NO: 76, SEQ ID NO: 77, SEQ ID NO: 78, SEQ ID NO:79, SEQ ID NO:80, SEQ ID NO:81, SEQ ID NO:82, SEQ ID NO:83, SEQ ID NO:84, SEQ ID NO:85, SEQ ID NO:86, SEQ ID NO:87, SEQ ID NO:88, SEQ ID NO:100, SEQ ID NO:101, SEQ ID NO:102, SEQ ID NO:103, SEQ ID NO:104, SEQ ID NO:105, SEQ ID NO:106 and/or SEQ ID NO:107.

46. The nucleic acid construct according to item 45, further comprising a sequence identical to or having at 90% identity, such as at least 91%, such as at least 92%, such as at least 93%, such as at least 94%, such as at least 95%, such as at least 96%, such as at least 97%, such as at least 98%, such as at least 99%, such as 100% identity to SEQ ID NO: 7.

47. The nucleic acid construct according to any of items 45 to 46, further comprising a sequence identical to or having at least 90% identity, such as at least 91%, such as at least 92%, such as at least 93%, such as at least 94%, such as at least 95%, such as at least 96%, such as at least 97%, such as at least 98%,

such as at least 99%, such as 100% identity to SEQ ID NO: 5 and/or SEQ ID NO: 23.

48. The nucleic acid construct according to any of items 45 to 47, further comprising a nucleic acid sequence identical to or having at least 90% identity, such as at least 91%, such as at least 92%, such as at least 93%, such as at least 94%, such as at least 95%, such as at least 96%, such as at least 97%, such as at least 98%, such as at least 99%, such as 100% identity to SEQ ID NO: 6.

49. The nucleic acid construct according to any one of items 45 to 48, further comprising a nucleic acid sequence identical to or having at least 90% identity, such as at least 91%, such as at least 92%, such as at least 93%, such as at least 94%, such as at least 95%, such as at least 96%, such as at least 97%, such as at least 98%, such as at least 99%, such as 100% identity to SEQ ID NO: 8, SEQ ID NO: 9, SEQ ID NO: 10, SEQ ID NO: 11, SEQ ID NO: 12, SEQ ID NO: 13, SEQ ID NO: 14, SEQ ID NO: 15, SEQ ID NO: 16, SEQ ID NO: 17 and/or SEQ ID NO: 18.

50. The nucleic acid construct according to any of items 45 to 49, wherein at least one of the one or more nucleic acid sequences are under the control of an inducible promoter.

51. The nucleic acid construct according to any of items 45 to 50, wherein the nucleic acid construct is a vector such as an integrative vector or a replicative vector.

52. A vector comprising a nucleic acid sequence as defined in any one of items 45 to 50.

53. A host cell comprising one or more nucleic acid sequence as defined in any of items 45 to 50, or the vector according to item 52.

54. A kit of parts comprising a microorganism according to any one of items 1 to 36, and/or nucleic acid constructs according to any one of items 45 to 50, and/ or a vector according to item 52, and instructions for use.

55. Use of the nucleic acid construct according to any one of items 45 to 50, of the microorganism according to any of items 1 to 36, the vector according to item 52, or the host cell according to item 53, for the production of strictosidine aglycone and/or tetrahydroalstonine, alstonine, tabersonine and/or catharanthine in a microorganism.

56. The use according to item 55 in the method according to items 37 to 44.

57. Strictosidine aglycone obtained by the method according to any of items 37 to 44.

58. Tetrahydroalstonine obtained by the method according to any of items 39 to 44.

59. Heteroyohimbine obtained by the method according to any of items 41 to 44.

60. Tabersonine and/or catharanthine obtained by the method according item 42 to 44.

61. A method of producing monoterpenoid indole alkaloids (MIAs) in a microorganism, said method comprising the steps of:

a) providing a microorganism capable of converting strictosidine to tabersonine and/or catharanthine, said cell expressing:

a strictosidine-beta-glucosidase (SGD);
a NADPH--cytochrome P450 reductase (CPR);
a Cytochrome b5 (CYB5);
a Geissoschizine synthase (GS);
a Geissoschizine oxidase (GO);
a Redox1;
a Redox2;
a Stemmadenine O-acetyltransferase (SAT);
a O-acetylstemmadenine oxidase (PAS);
a Dehydroprecondylocarpine acetate synthase (DPAS);
a Tabersonine synthase (TS); and/or
a Catharanthine synthase (CS);
optionally, a strictosidine synthase (STR);

b) incubating said microorganism in a medium comprising strictosidine or a substrate which can be converted to strictosidine by said microorganism;
c) optionally, recovering the MIAs;
d) optionally, processing the MIAs into a pharmaceutical compound,

wherein said SGD is a heterologous SGD selected from RseSGD (SEQ ID NO: 24), GseSGD (SEQ ID NO: 25), SapSGD (SEQ ID NO: 26), RveSGD (SEQ ID NO: 27), VmiSGD1 (SEQ ID NO: 47), AhuSGD (SEQ ID NO: 48), HimSGD2 (SEQ ID NO: 49), SinSGD (SEQ ID NO: 50), TelSGD (SEQ ID NO: 51), VunSGD (SEQ ID NO: 52), NsiSGD1 (SEQ ID NO: 53), LprSGD (SEQ ID NO: 54), AchSGD1 (SEQ ID NO: 55), HsuSGD (SEQ ID NO: 56), MroSGD (SEQ ID NO: 57), RseSGD2 (SEQ ID NO: 58), PgrSGD (SEQ ID NO: 59), OpuSGD (SEQ ID NO: 60), HpiSGD (SEQ ID NO: 61), HanSGD1 (SEQ ID NO: 62), AchSGD2 (SEQ ID NO: 63), HimSGD1 (SEQ ID NO: 64), IpeSGD (SEQ ID NO: 65), LsaSGD1 (SEQ ID NO: 66), or CarSGD (SEQ ID NO: 67) or variants thereof having at least 70%, such as at least 80%, such as at least 90%, such as at least 91%, such as at least 92%, such as at least 93%, such as at least 94%, such as at least 95%, such as at least 96%, such as at least 97%, such as at least 98%, such as at least 99%, such as 100% identity thereto,

and/or;

wherein said SGD is a mosaic SGD, wherein said mosaic SGD comprises an amino acid sequence having the general formula

$$D_1\text{-}D_2\text{-}D_3\text{-}D_4$$

wherein D_1 is a first amino acid sequence from a first SGD,
wherein D_2 is a second amino acid sequence from a second SGD,
wherein D_3 is a third amino acid sequence comprising or consisting of amino acids of SEQ ID NO:91 or a variant thereof having at least 90% identity to SEQ ID NO: 91,
wherein D_4 is a fourth amino acid sequence from a fourth SGD or an amino acid sequence consisting of amino acids of SEQ ID NO:92 or a variant thereof having at least 90% identity to SEQ ID NO: 92,

wherein said first SGD, second SGD and fourth SGD can be the same or different, with the proviso that said first SGD, second SGD and fourth SGD are not all RseSGD.

62. The method according to item 61, wherein the microorganism is as defined in any one of the preceding items.

63. A method of treating a disorder such as a cancer, arrhythmia, malaria, psychotic diseases, hypertension, depression, Alzheimer's disease, addiction and/or neuronal diseases, comprising administration of a therapeutic sufficient amount of an MIA or a pharmaceutical compound obtained by the method according to any of items 24 to 30, 47 or 61 to 62.

Claims

1. A microorganism capable of producing strictosidine aglycone, said microorganism expresses

a strictosidine-beta-glucosidase (SGD), capable of converting strictosidine to strictosidine aglycone,

wherein said SGD is a heterologous SGD selected from RseSGD (SEQ ID NO: 24), GseSGD (SEQ ID NO: 25), SapSGD (SEQ ID NO: 26), RveSGD (SEQ ID NO: 27), VmiSGD1 (SEQ ID NO: 47), AhuSGD (SEQ ID NO: 48), HimSGD2 (SEQ ID NO: 49), SinSGD (SEQ ID NO: 50), TelSGD (SEQ ID NO: 51), VunSGD (SEQ ID NO: 52), NsiSGD1 (SEQ ID NO: 53), LprSGD (SEQ ID NO: 54), AchSGD1 (SEQ ID NO: 55), HsuSGD (SEQ ID NO: 56), MroSGD (SEQ ID NO: 57), RseSGD2 (SEQ ID NO: 58), PgrSGD (SEQ ID NO: 59), OpuSGD (SEQ ID NO: 60), HpiSGD (SEQ ID NO: 61), HanSGD1 (SEQ ID NO: 62), AchSGD2 (SEQ ID NO: 63), HimSGD1 (SEQ ID NO: 64), IpeSGD (SEQ ID NO: 65), LsaSGD1 (SEQ ID NO: 66), or CarSGD (SEQ ID NO: 67) or variants thereof having at least 70%, such as at least 80%, such as at least 90%, such as at least 91%, such as at least 92%, such as at least 93%, such as at least 94%, such as at least 95%, such as at least 96%, such as at least 97%, such as at least 98%, such as at least 99%, such as 100% identity thereto,

and/or;

wherein said SGD is a mosaic SGD, wherein said mosaic SGD comprises an amino acid sequence having the general formula

$$D_1\text{-}D_2\text{-}D_3\text{-}D_4$$

wherein D_1 is a first amino acid sequence from a first SGD,
wherein D_2 is a second amino acid sequence from a second SGD,
wherein D_3 is a third amino acid sequence comprising or consisting of amino acids of SEQ ID NO:91 or a variant thereof having at least 90% identity to SEQ ID NO: 91,
wherein D_4 is a fourth amino acid sequence from a fourth SGD or an amino acid sequence consisting of amino acids of SEQ ID NO:92 or a variant thereof having at least 90% identity to SEQ ID NO: 92,

wherein said first SGD, second SGD and fourth SGD can be the same or different, with the proviso that said first SGD, second SGD and fourth SGD are not all RseSGD.

2. The microorganism according to claim 1, wherein the microorganism is selected from the group consisting of bacteria, archaea, yeast, fungi, protozoa, algae, and viruses, preferably the microorganism is a yeast or a bacteria, such as *Saccharomyces cerevisiae* or *Escherichia coli*.

3. The microorganism according to any one the preceding claims, further expressing

a strictosidine synthase (STR), capable of converting secologanin and tryptamine to strictosidine, whereby the microorganism is capable of synthesising strictosidine,

wherein said STR is preferably CroSTR or variants thereof having at least 90%, such as at least 91%, such as at least 92%, such as at least 93%, such as at least 94%, such as at least 95%, such as at least 96%, such as at least 97%, such as at least 98%, such as at least 99%, such as 100% identity to SEQ ID NO: 30.

4. The microorganism according to any one of the preceding claims, wherein D_1 comprises or consists of an amino acid sequence corresponding to amino acids M1 to R115 of SEQ ID NO:24.

5. The microorganism according to any one of the preceding claims, wherein D_2 comprises or consists of an amino acid sequence corresponding to amino acids F116 to G266 of SEQ ID NO:24.

6. The microorganism according to any one of the preceding claims, wherein D_4 comprises or consists of amino acids of SEQ ID NO:92 or a variant thereof having at least 90% identity to SEQ ID NO: 92.

7. The microorganism according to any one of the preceding claims, wherein at least one of D_1, D_2 or D_4 is from an SGD which is native to a first organism selected from *Gelsemium sempervirens, Scedosporium apiospermum or*

Rauvolfia verticillata, *Vinca minor, Tabernaemontana elegans, Amsonia hubrichtii, Ophiorrhiza pumila, Nyssa sinensis, Coffea arabica, Carapichea ipecacuanha, Handroanthus impetiginosus, Sesamum indicum, Actinidia chinensis var. chinensis, Helianthus annuus, Lactuca sativa, Ipomoea nil, Vigna unguiculata, Heliocybe sulcate, Pyricularia grisea, Lomentospora prolificans, Hydnomerulius pinastri MD-312,* and *Moniliophthora roreri MCA 2997.*

8. The microorgagnism according to any one of the preceding claims, wherein the first SGD, the second SGD and the fourth SGD are identical or different.

9. The microorganism according to any one of the preceding claims, wherein two of the first SGD, the second SGD and the fourth SGD are identical, or wherein the first SGD, the second SGD and the fourth SGD are different, or wherein the first SGD, the second SGD and the fourth SGD are identical.

10. The microorganism according to any one of the preceding claims, wherein said mosaic SGD comprises or consists of an amino acid sequence of SEQ ID NO: 93, SEQ ID NO: 94, SEQ ID NO: 95, SEQ ID NO: 96, SEQ ID NO: 97, SEQ ID NO: 98, SEQ ID NO: 99 or SEQ ID NO: 8, or variants thereof having at least 90% identity or homology thereto, such as at least 91%, such as at least 92%, such as at least 93%, such as at least 94%, such as at least 95%, such as at least 96%, such as at least 97%, such as at least 98%, such as at least 99% identity or homology thereto.

11. The microorganism according to any one the preceding claims, further expressing:

    i. a tetrahydroalstonine synthase (THAS) and/or a heteroyohimbine synthase (HYS), capable of converting strictosidine aglycone to tetrahydroalstonine, whereby the microorganism is capable of synthesising tetrahydroalstonine,

wherein said THAS is preferably CroTHAS and/or HYS is CroHYS or variants thereof, having at least 90%, such as at least 91%, such as at least 92%, such as at least 93%, such as at least 94%, such as at least 95%, such as at least 96%, such as at least 97%, such as at least 98%, such as at least 99%, such as 100% identity to SEQ ID NO: 28 and/or SEQ ID NO: 46,

and optionally further expressing

a sarpargan bridge enzymes (SBE), capable of converting tetrahydroalstonine and ajmalicine to a heteroyohimbine selected from the group consisting of alstonine and serpentine, whereby the microorganism is capable of synthesising alstonine and serpentine,

wherein said SBE is preferably GseSBE or variants thereof having at least 90%, such as at least 91%, such as at least 92%, such as at least 93%, such as at least 94%, such as at least 95%, such as at least 96%, such as at least 97%, such as at least 98%, such as at least 99%, such as 100% identity to SEQ ID NO: 29,

and/or

ii. further expressing

a NADPH--cytochrome P450 reductase (CPR);

a Cytochrome b5 (CYB5);

a Geissoschizine synthase (GS);

a Geissoschizine oxidase (GO);

a Redox1;

a Redox2;

a Stemmadenine O-acetyltransferase (SAT);

a O-acetylstemmadenine oxidase (PAS);

a Dehydroprecondylocarpine acetate synthase (DPAS);

a Tabersonine synthase (TS); and/or

a Catharanthine synthase (CS),

whereby the microorganism is capable of synthesising tabersonine and/or catharanthine,

wherein preferably said CPR is CroCPR, said CYB5 is CroCYB5, said GS is CroSG, said GO is CroGO, said Redox1 is CroRedox1, said Redox2 is CroRedox2, said SAT is CroSAT, said PAS is CroPAS, said DPAS is CroDPAS, said TS is CroTS and/or said CS is CroCS or variants thereof having at least 90%, such as at least 91%, such as at least 92%, such as at least 93%, such as at least 94%, such as at least 95%, such as at least 96%, such as at least 97%, such as at least 98%, such as at least 99%, such as 100% identity to SEQ ID NO: 31, SEQ ID NO: 32, SEQ ID NO: 33, SEQ ID NO: 34, SEQ ID NO: 35, SEQ ID NO: 36, SEQ ID NO: 37, SEQ ID NO: 38, SEQ ID NO: 39, SEQ ID NO: 40 and/or SEQ ID NO: 41, respectively.

12. The microorganism according to any one of the preceding claims, capable of producing strictosidine aglycone with a titre of at least 1 µM, such as at least 2 µM, such as at least 4 µM, such as at least 6 µM, such as at least 8 µM such as at least 10 µM or more.

13. The microorganism according to claim 11, capable of producing:

i. tetrahydroalstonine with a titre of at least 1 µM, such as at least 2 µM, such as at least 4 µM, such as at least 6 µM, such as at least 8 µM such as at least 10 µM or more, and optionally alstonine with a titre of at least 1 µM, such as at least 2 µM, such as at least 4 µM, such as at least 6 µM, such as at least 8 µM such as at least 10 µM or more, and/or

ii. tabersonine with a titre of at least 0.01 µM, such as at least 0.02 µM, and/or catharanthine with a titre of at least 0.01 µM, such as at least 0.02 µM.

14. A method of producing strictosidine aglycone in a microorganism, said method comprises the steps of:

a) providing a microorganism, said cell expressing:
a strictosidine-beta-glucosidase (SGD), capable of converting strictosidine to strictosidine aglycone;

b) incubating said microorganism in a medium comprising strictosidine or a substrate which can be converted to strictosidine by said microorganism;

c) optionally, recovering the strictosidine aglycone;

d) optionally, further converting the strictosidine aglycone to monoterpenoid indole alkaloids,

wherein said SGD is a heterologous SGD selected from RseSGD (SEQ ID NO: 24), GseSGD (SEQ ID NO: 25), SapSGD (SEQ ID NO: 26), RveSGD (SEQ ID NO: 27), VmiSGD1 (SEQ ID NO: 47), AhuSGD (SEQ ID NO: 48), HimSGD2 (SEQ ID NO: 49), SinSGD (SEQ ID NO: 50), TelSGD (SEQ ID NO: 51), VunSGD (SEQ ID NO: 52), NsiSGD1 (SEQ ID NO: 53), LprSGD (SEQ ID NO: 54), AchSGD1 (SEQ ID NO: 55), HsuSGD (SEQ ID NO: 56), MroSGD (SEQ ID NO: 57), RseSGD2 (SEQ ID NO: 58), PgrSGD (SEQ ID NO: 59), OpuSGD (SEQ ID NO: 60), HpiSGD (SEQ ID NO: 61), HanSGD1 (SEQ ID NO: 62), AchSGD2

(SEQ ID NO: 63), HimSGD1 (SEQ ID NO: 64), IpeSGD (SEQ ID NO: 65), LsaSGD1 (SEQ ID NO: 66), or CarSGD (SEQ ID NO: 67) or variants thereof having at least 70%, such as at least 80%, such as at least 90%, such as at least 91%, such as at least 92%, such as at least 93%, such as at least 94%, such as at least 95%, such as at least 96%, such as at least 97%, such as at least 98%, such as at least 99%, such as 100% identity thereto,

and/or;

wherein said SGD is a mosaic SGD, wherein said mosaic SGD comprises an amino acid sequence having the general formula

$$D_1\text{-}D_2\text{-}D_3\text{-}D_4$$

wherein D_1 is a first amino acid sequence from a first SGD,
wherein D_2 is a second amino acid sequence from a second SGD,
wherein D_3 is a third amino acid sequence comprising or consisting of amino acids of SEQ ID NO:91 or a variant thereof having at least 90% identity to SEQ ID NO: 91,
wherein D_4 is a fourth amino acid sequence from a fourth SGD or an amino acid sequence consisting of amino acids of SEQ ID NO:92 or a variant thereof having at least 90% identity to SEQ ID NO: 92,
wherein said first SGD, second SGD and fourth SGD can be the same or different, with the proviso that said first SGD, second SGD and fourth SGD are not all RseSGD.

15. The method according to claim 14, wherein the SGD, the heterologous SGD and/or the mosaic SGD is as defined in any one of claims 1 to 13.

16. The method according to any one of claims 14 to 15, wherein the substrate is secologanin and/or tryptamine, and wherein said microorganism further expresses:

a strictosidine synthase (STR), capable of converting secologanin and tryptamine to strictosidine;

wherein said STR is preferably CroSTR or variants thereof having at least 90%, such as at least 91%, such as at least 92%, such as at least 93%, such as at least 94%, such as at least 95%, such as at least 96%, such as at least 97%,

such as at least 98%, such as at least 99%, such as 100% identity to SEQ ID NO: 30.

17. The method according to any one of claims 14 to 16, wherein the method comprises step d) and wherein said microorganism further expresses:

i. a tetrahydroalstonine synthase (THAS) and/or or a heteroyohimbine synthase (HSY), capable of converting strictosidine aglycone to tetrahydroalstonine;

wherein preferably said THAS is identical to or has at least 90%, such as at least 91%, such as at least 92%, such as at least 93%, such as at least 94%, such as at least 95%, such as at least 96%, such as at least 97%, such as at least 98%, such as at least 99%, such as 100% identity to SEQ ID NO: 28 and/or HYS is identical to or has at least 90%, such as at least 91%, such as at least 92%, such as at least 93%, such as at least 94%, such as at least 95%, such as at least 96%, such as at least 97%, such as at least 98%, such as at least 99%, such as 100% identity to SEQ ID NO: 46, optionally wherein said method further comprises the step of recover tetrahydroalstonine,

and optionally wherein said microorganism further expresses:

a sapargan bridge enzyme (SBE), capable of converting tetrahydroalstonine to alstonine;

wherein preferably said SBE is identical to or has at least 90%, such as at least 91%, such as at least 92%, such as at least 93%, such as at least 94%, such as at least 95%, such as at least 96%, such as at least 97%, such as at least 98%, such as at least 99%, such as 100% identity to SEQ ID NO: 29, optionally wherein said method further comprises the step of recovering alstonine,

and/or

ii. wherein said microorganism further expresses:

a NADPH--cytochrome P450 reductase (CPR);

a Cytochrome b5 (CYB5);

a Geissoschizine synthase (GS);

a Geissoschizine oxidase (GO);

a Redox1;

a Redox2;

a Stemmadenine O-acetyltransferase (SAT);

a O-acetylstemmadenine oxidase (PAS);

a Dehydroprecondylocarpine acetate synthase (DPAS);
a Tabersonine synthase (TS); and/or
a Catharanthine synthase (CS),
wherein preferably said CPR is CroCPR, said CYB5 is CroCYB5, said GS is CroSG, said GO is CroGO, said Redox1 is CroRedox1, said Redox2 is CroRedox2, said SAT is CroSAT, said PAS is CroPAS, said DPAS is CroDPAS, said TS is CroTS and/or said CS is CroCS or variants thereof having at least 90%, such as at least 91%, such as at least 92%, such as at least 93%, such as at least 94%, such as at least 95%, such as at least 96%, such as at least 97%, such as at least 98%, such as at least 99%, such as 100% identity to SEQ ID NO: 31, SEQ ID NO: 32, SEQ ID NO: 33, SEQ ID NO: 34, SEQ ID NO: 35, SEQ ID NO: 36, SEQ ID NO: 37, SEQ ID NO: 38, SEQ ID NO: 39, SEQ ID NO: 40 and/or SEQ ID NO: 41, respectively,
wherein the microorganism is capable of producing tabersonine and/or catharanthine, optionally wherein said method further comprises the step of recovering tabersonine and/or catharanthine.

18. A nucleic acid construct comprising a sequence identical to or having at least 90% identity, such as at least 91%, such as at least 92%, such as at least 93%, such as at least 94%, such as at least 95%, such as at least 96%, such as at least 97%, such as at least 98%, such as at least 99%, such as 100% identity to SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 3, SEQ ID NO: 4, SEQ ID NO:68, SEQ ID NO:69, SEQ ID NO:70, SEQ ID NO: 71, SEQ ID NO:72, SEQ ID NO: 73, SEQ ID NO:74, SEQ ID NO: 75, SEQ ID NO: 76, SEQ ID NO: 77, SEQ ID NO: 78, SEQ ID NO:79, SEQ ID NO:80, SEQ ID NO:81, SEQ ID NO:82, SEQ ID NO:83, SEQ ID NO:84, SEQ ID NO:85, SEQ ID NO:86, SEQ ID NO:87, SEQ ID NO:88, SEQ ID NO:100, SEQ ID NO:101, SEQ ID NO:102, SEQ ID NO:103, SEQ ID NO:104, SEQ ID NO:105, SEQ ID NO:106 and/or SEQ ID NO:107, optionally, further comprising a sequence identical to or having at 90% identity, such as at least 91%, such as at least 92%, such as at least 93%, such as at least 94%, such as at least 95%, such as at least 96%, such as at least 97%, such as at least 98%, such as at least 99%, such as 100% identity to SEQ ID NO: 7.

19. The nucleic acid construct according to claim 18, further comprising a sequence identical to or having at least 90% identity, such as at least 91%, such as at least 92%, such as at least 93%, such as at least 94%, such as at least 95%, such as at least 96%, such as at least 97%, such as at least 98%, such as at least 99%, such as 100% identity to SEQ ID NO: 5 and/or SEQ ID NO: 23, and/or optionally further comprising a nucleic acid sequence identical to or having at least 90% identity, such as at least 91%, such as at least 92%, such as at least 93%, such as at least 94%, such as at least 95%, such as at least 96%, such as at least 97%, such as at least 98%, such as at least 99%, such as 100% identity to SEQ ID NO: 6, and/or further comprising a nucleic acid sequence identical to or having at least 90% identity, such as at least 91%, such as at least 92%, such as at least 93%, such as at least 94%, such as at least 95%, such as at least 96%, such as at least 97%, such as at least 98%, such as at least 99%, such as 100% identity to SEQ ID NO: 8, SEQ ID NO: 9, SEQ ID NO: 10, SEQ ID NO: 11, SEQ ID NO: 12, SEQ ID NO: 13, SEQ ID NO: 14, SEQ ID NO: 15, SEQ ID NO: 16, SEQ ID NO: 17 and/or SEQ ID NO: 18.

20. A vector comprising a nucleic acid sequence as defined in any one of claims 18 to 19.

21. A host cell comprising one or more nucleic acid sequence as defined in any one of claims 18 to 19, or the vector according to claim 20.

22. A kit of parts comprising a microorganism according to any one of claims 1 to 13, and/or nucleic acid constructs according to any one of claims 18 to 19, and/ or a vector according to claim 20, and instructions for use.

23. Use of the nucleic acid construct according to any one of claims 18 to 19, of the microorganism according to any of claims 1 to 13, the vector according to claim 20, or the host cell according to claim 21, for the production of strictosidine aglycone, tetrahydroalstonine, alstonine, tabersonine and/or catharanthine in a microorganism, preferably according to the method in claims 14 to 17.

24. A method of producing monoterpenoid indole alkaloids (MIAs) in a microorganism, said method comprising the steps of:

a) providing a microorganism capable of converting strictosidine to tabersonine and/or catharanthine, said cell expressing:
a strictosidine-beta-glucosidase (SGD);
a NADPH--cytochrome P450 reductase (CPR);
a Cytochrome b5 (CYB5);
a Geissoschizine synthase (GS);
a Geissoschizine oxidase (GO);
a Redox1;
a Redox2;
a Stemmadenine O-acetyltransferase (SAT);
a O-acetylstemmadenine oxidase (PAS);
a Dehydroprecondylocarpine acetate synthase (DPAS);
a Tabersonine synthase (TS); and/or
a Catharanthine synthase (CS);
b) incubating said microorganism in a medium comprising strictosidine or a substrate which can be converted to strictosidine by said microorganism;
c) optionally, recovering the MIAs;
d) optionally, processing the MIAs into a pharmaceutical compound,

wherein said SGD is a heterologous SGD selected from RseSGD (SEQ ID NO: 24), GseSGD (SEQ ID NO: 25), SapSGD (SEQ ID NO: 26), RveSGD (SEQ ID NO: 27), VmiSGD1 (SEQ ID NO: 47), AhuSGD (SEQ ID NO: 48), HimSGD2 (SEQ ID NO: 49), SinSGD (SEQ ID NO: 50), TelSGD (SEQ ID NO: 51), VunSGD (SEQ ID NO: 52), NsiSGD1 (SEQ ID NO: 53), LprSGD (SEQ ID NO: 54), AchSGD1 (SEQ ID NO: 55), HsuSGD (SEQ ID NO: 56), MroSGD (SEQ ID NO: 57), RseSGD2 (SEQ ID NO: 58), PgrSGD (SEQ ID NO: 59), OpuSGD (SEQ ID NO: 60), HpiSGD (SEQ ID NO: 61), HanSGD1 (SEQ ID NO: 62), AchSGD2 (SEQ ID NO: 63), HimSGD1 (SEQ ID NO: 64), IpeSGD (SEQ ID NO: 65), LsaSGD1 (SEQ ID NO: 66), or CarSGD (SEQ ID NO: 67) or variants thereof having at least 70%, such as at least 80%, such as at least 90%, such as at least 91%, such as at least 92%, such as at least 93%, such as at least 94%, such as at least 95%, such as at least 96%, such as at least 97%, such as at least 98%, such as at least 99%, such as 100% identity thereto,

and/or;

wherein said SGD is a mosaic SGD, wherein said mosaic SGD comprises an amino acid sequence having the general formula

D_1-D_2-D_3-D_4

wherein D_1 is a first amino acid sequence from a first SGD,
wherein D_2 is a second amino acid sequence from a second SGD,
wherein D_3 is a third amino acid sequence comprising or consisting of amino acids of SEQ ID NO:91 or a variant thereof having at least 90% identity to SEQ ID NO: 91,
wherein D_4 is a fourth amino acid sequence from a fourth SGD or an amino acid sequence consisting of amino acids of SEQ ID NO:92 or a variant thereof having at least 90% identity to SEQ ID NO: 92,
wherein said first SGD, second SGD and fourth SGD can be the same or different, with the proviso that said first SGD, second SGD and fourth SGD are not all RseSGD.

25. The method according to claim 24, wherein said microorganism further expresses strictosidine synthase (STR).

26. The method according to any one of claims 24-26, wherein said microorganism is as defined in any one of claims 1 to 14.

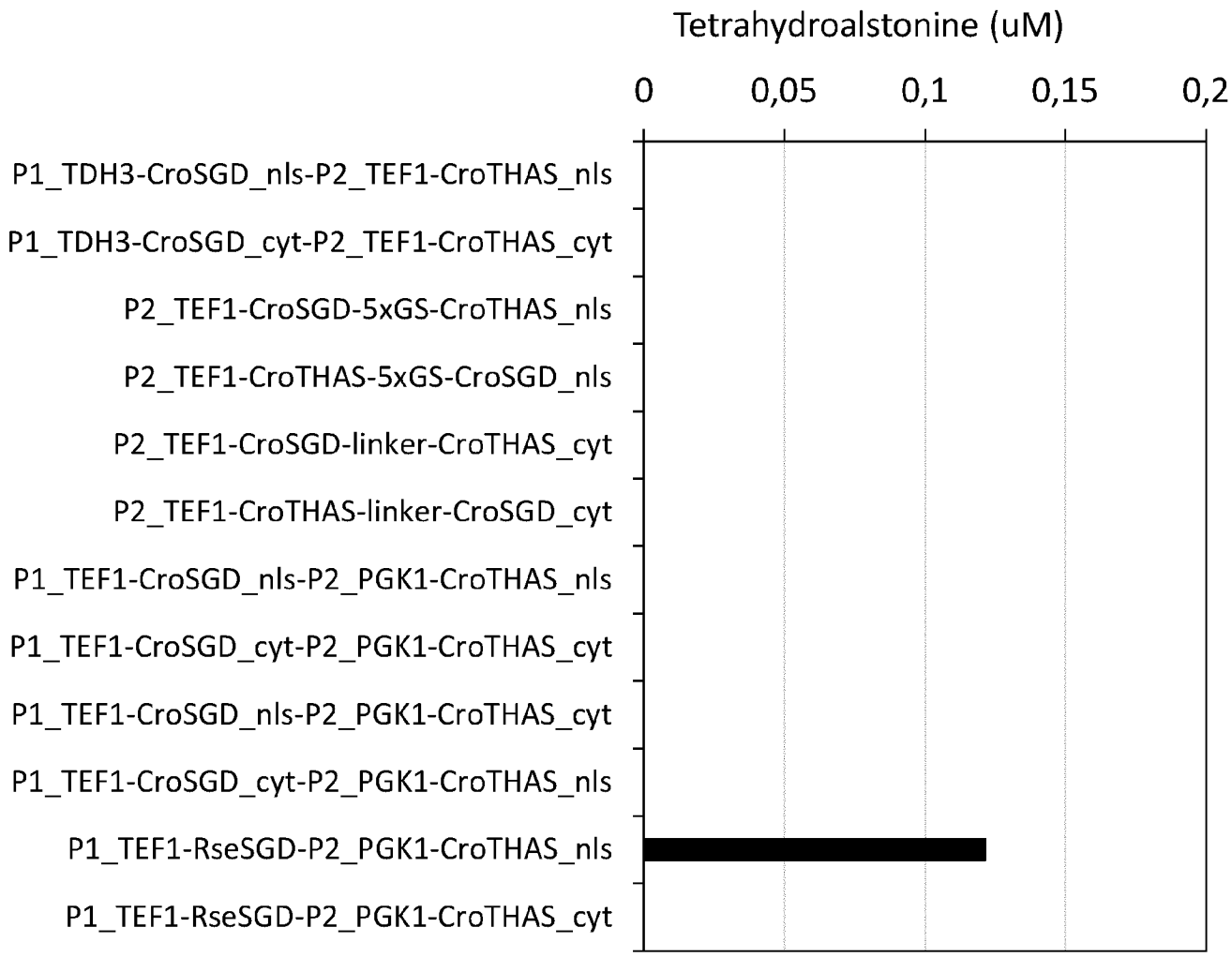
Tetrahydroalstonine (uM)
0
0,05
0,1
0,15
0,2
P1_TDH3-CroSGD_nls-P2_TEF1-CroTHAS_nls
P1_TDH3-CroSGD_cyt-P2_TEF1-CroTHAS_cyt
P2_TEF1-CroSGD-5xGS-CroTHAS_nls
P2_TEF1-CroTHAS-5xGS-CroSGD_nls
P2_TEF1-CroSGD-linker-CroTHAS_cyt
P2_TEF1-CroTHAS-linker-CroSGD_cyt
P1_TEF1-CroSGD_nls-P2_PGK1-CroTHAS_nls
P1_TEF1-CroSGD_cyt-P2_PGK1-CroTHAS_cyt
P1_TEF1-CroSGD_nls-P2_PGK1-CroTHAS_cyt
P1_TEF1-CroSGD_cyt-P2_PGK1-CroTHAS_nls
P1_TEF1-RseSGD-P2_PGK1-CroTHAS_nls
P1_TEF1-RseSGD-P2_PGK1-CroTHAS_cyt

Fig. 1

		1	2	3	4	5	6	7	8
1	RseSGD								
2	RveSGD	**89.94**							
3	CroSGD	**70.32**	**70.68**						
4	GseSGD	**53.86**	**53.65**	**48.79**					
5	CacSGD	**51.71**	**50.98**	**47.93**	**55.72**				
6	SapSGD	**34.63**	**33.94**	**34.22**	**37.24**	**35.89**			
7	UtoSGD	**40.68**	**40.17**	**37.85**	**44.85**	**43.63**	**32.45**		
8	GsoSGD	**48.98**	**46.89**	**45.66**	**51.79**	**51.53**	**36.74**	**41.55**	

Fig. 2

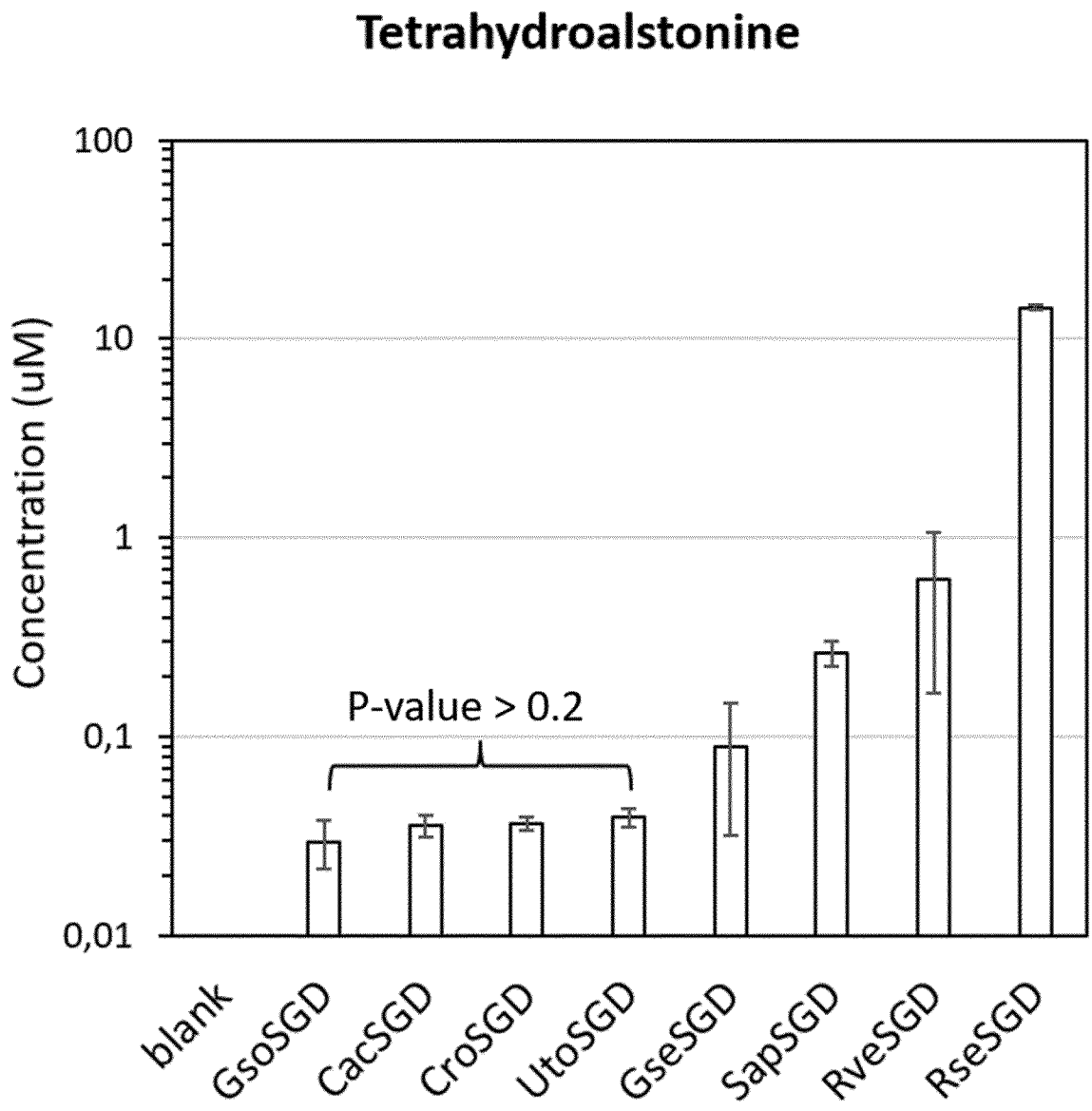
Tetrahydroalstonine
Concentration (uM)
100
10
1
0,1
0,01
P-value > 0.2
blank
GsoSGD
CacSGD
CroSGD
UtoSGD
GseSGD
SapSGD
RveSGD
RseSGD

Fig. 3

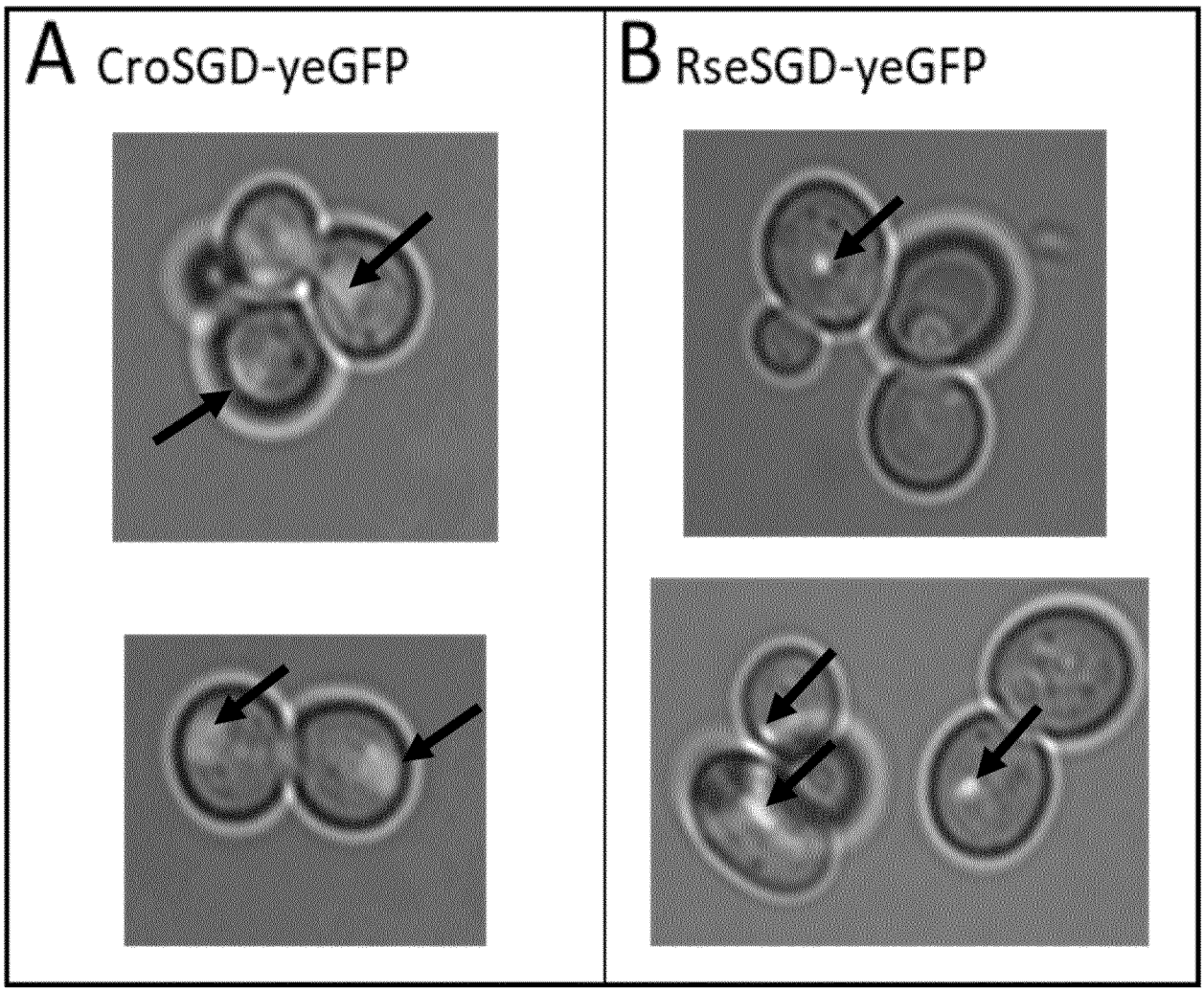

Fig. 4

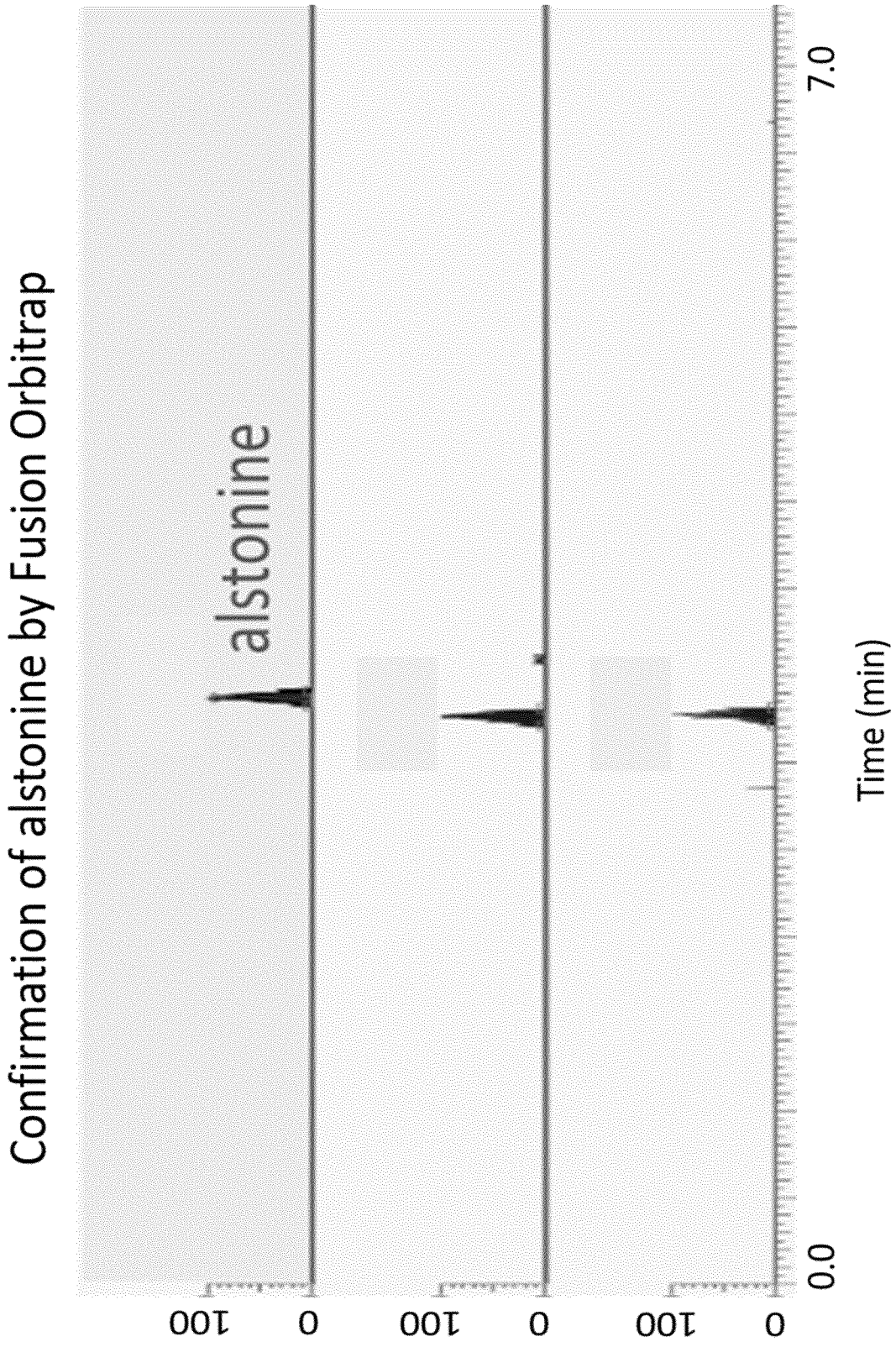
Confirmation of alstonine by Fusion Orbitrap
alstonine
100
0
100
0
100
0
0.0
7.0
Time (min)

Fig. 5

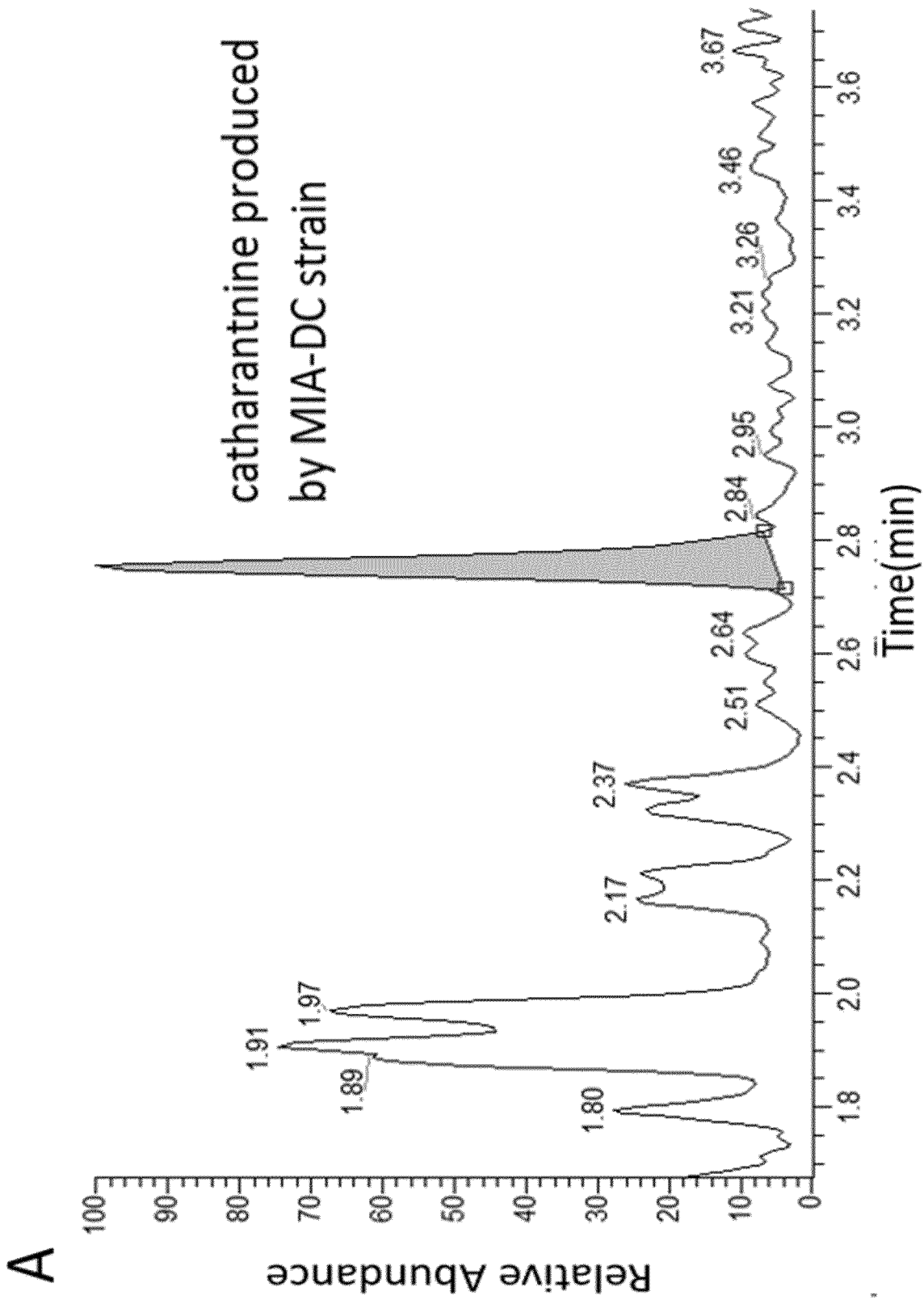
A
catharantnine produced
by MIA-DC strain
Relative Abundance
Time(min)
1.80
1.89
1.91
1.97
2.17
2.37
2.51
2.64
2.84
2.95
3.21
3.26
3.46
3.67

Fig. 6A

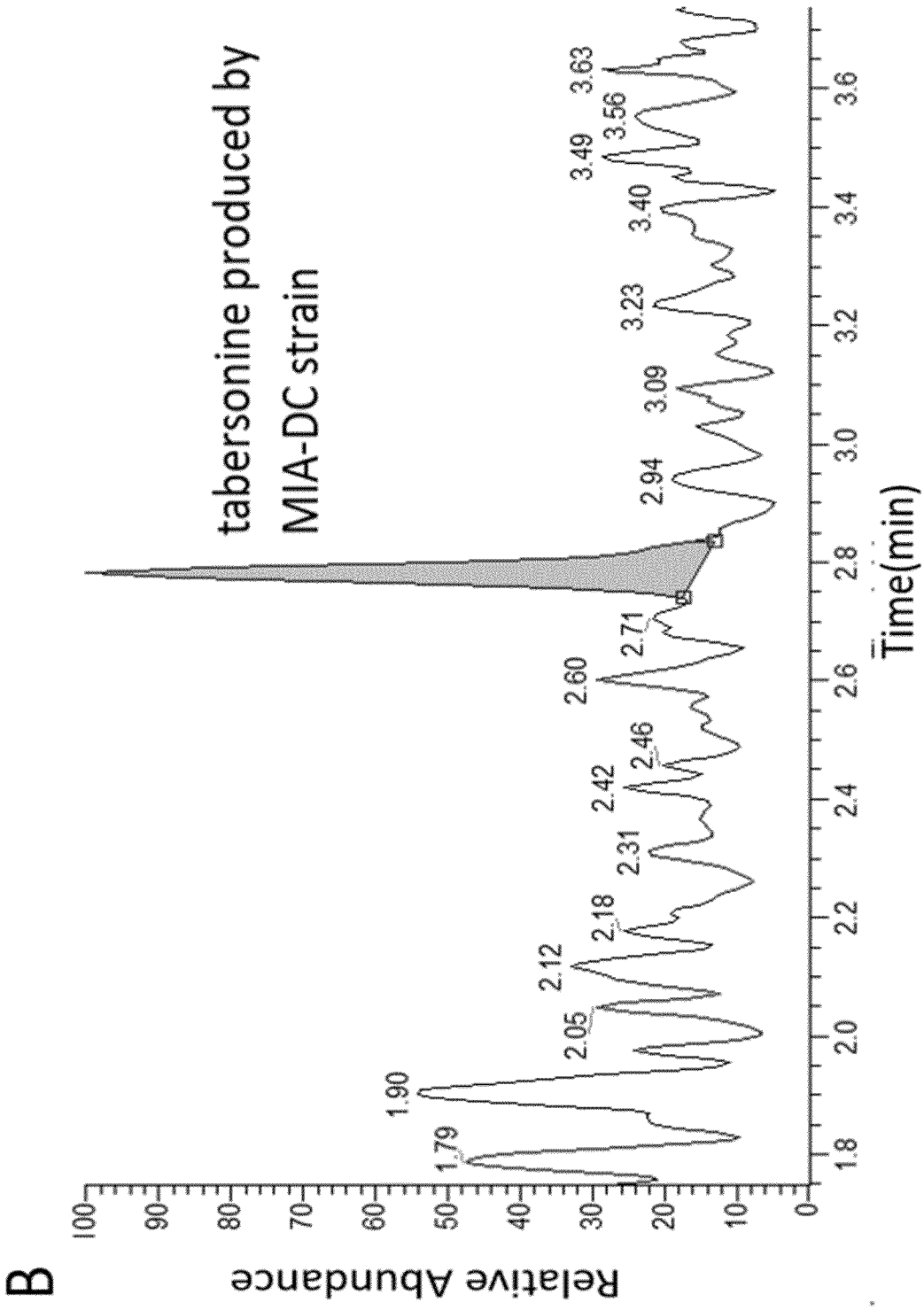
B
tabersonine produced by
MIA-DC strain
Relative Abundance
Time(min)
1.79
1.90
2.05
2.12
2.18
2.31
2.42
2.46
2.60
2.71
2.94
3.09
3.23
3.40
3.49
3.56
3.63

Fig. 6B

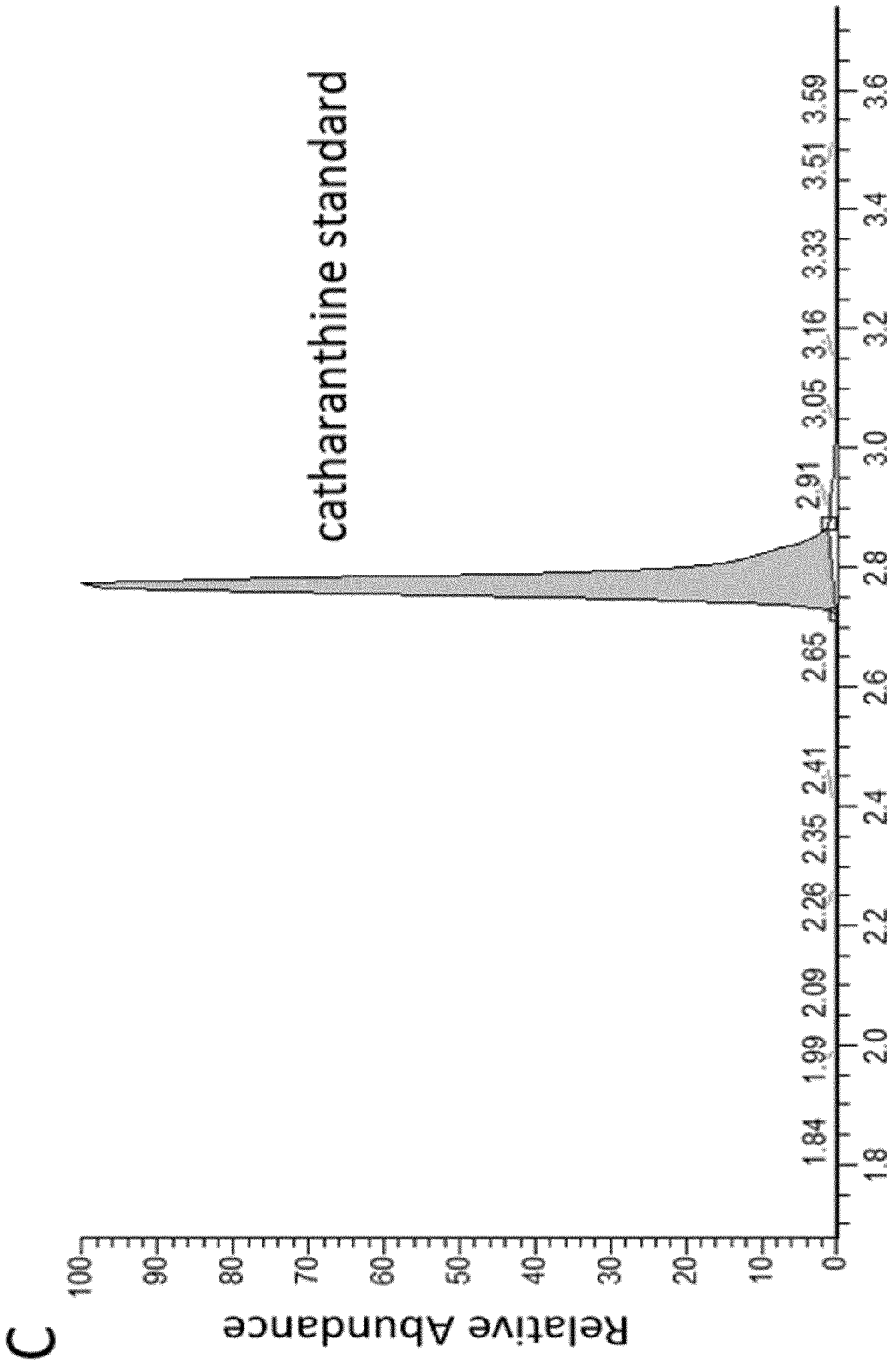
C
Relative Abundance
100
90
80
70
60
50
40
30
20
10
0
catharanthine standard
1.84
1.99
2.09
2.26
2.35
2.41
2.65
2.91
3.05
3.16
3.33
3.51
3.59
1.8
2.0
2.2
2.4
2.6
2.8
3.0
3.2
3.4
3.6

Fig. 6C

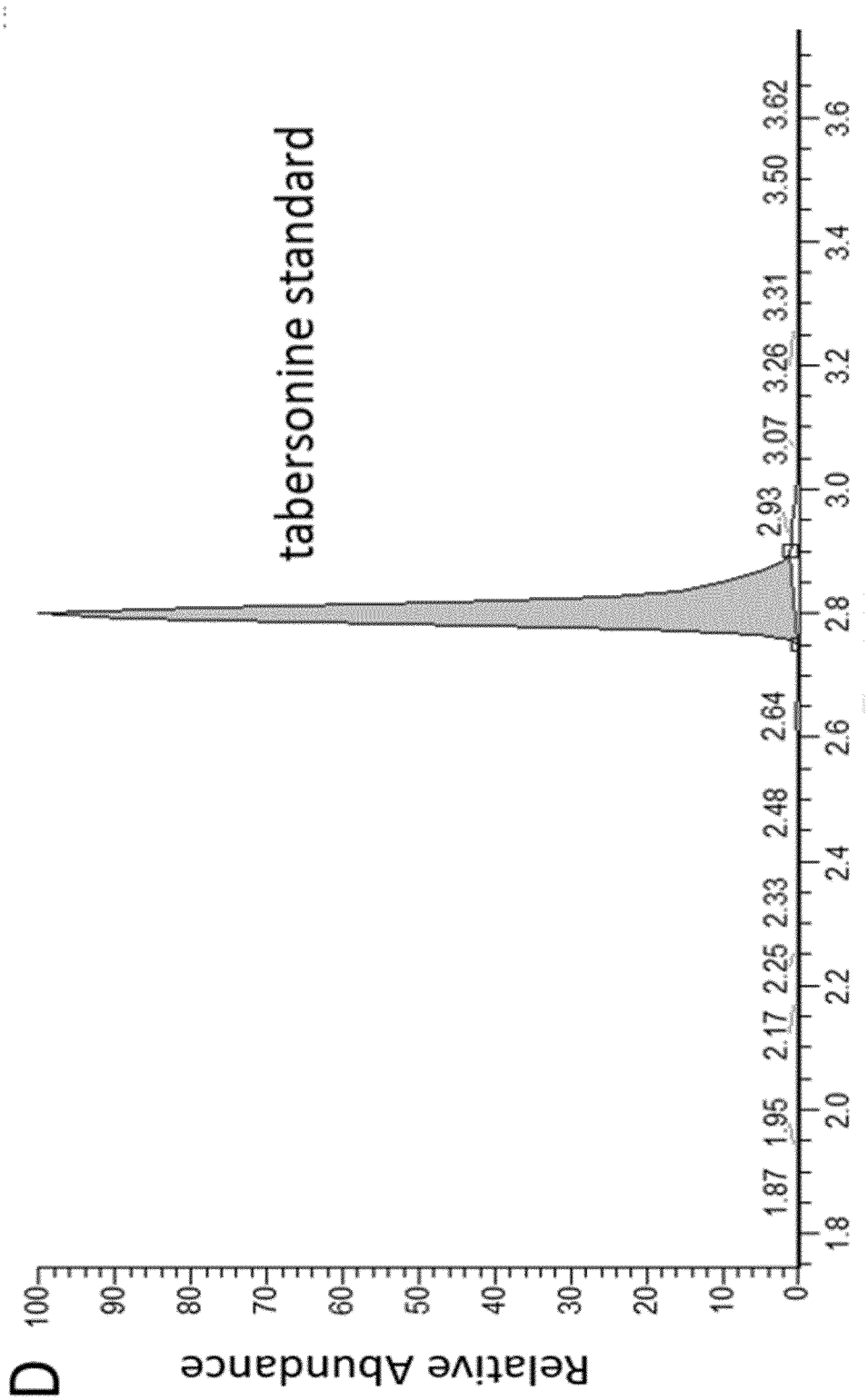
D
Relative Abundance
tabersonine standard
100
90
80
70
60
50
40
30
20
10
0
1.87
1.95
2.0
2.17
2.25
2.2
2.33
2.4
2.48
2.64
2.6
2.8
2.93
3.0
3.07
3.26
3.2
3.31
3.4
3.50
3.62
3.6
1.8

Fig. 6D

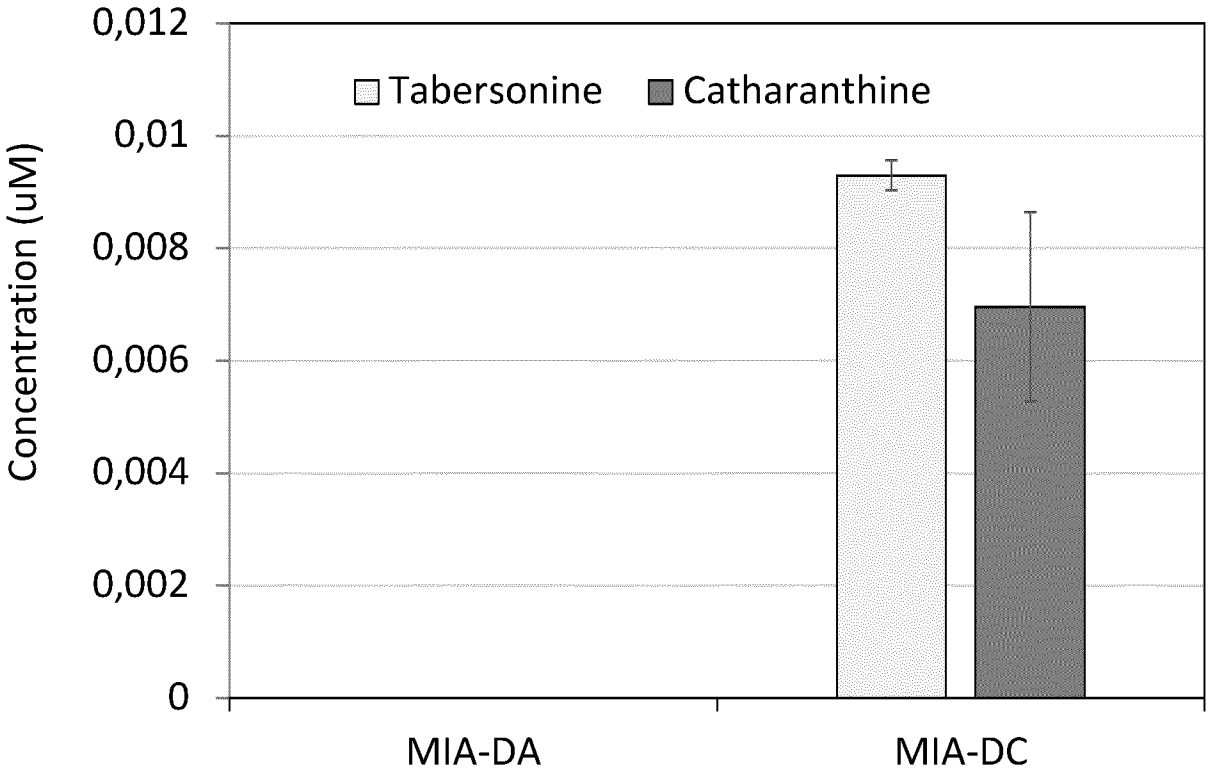
0,012
0,01
0,008
0,006
0,004
0,002
0
Concentration (uM)
Tabersonine
Catharanthine
MIA-DA
MIA-DC

Fig. 7

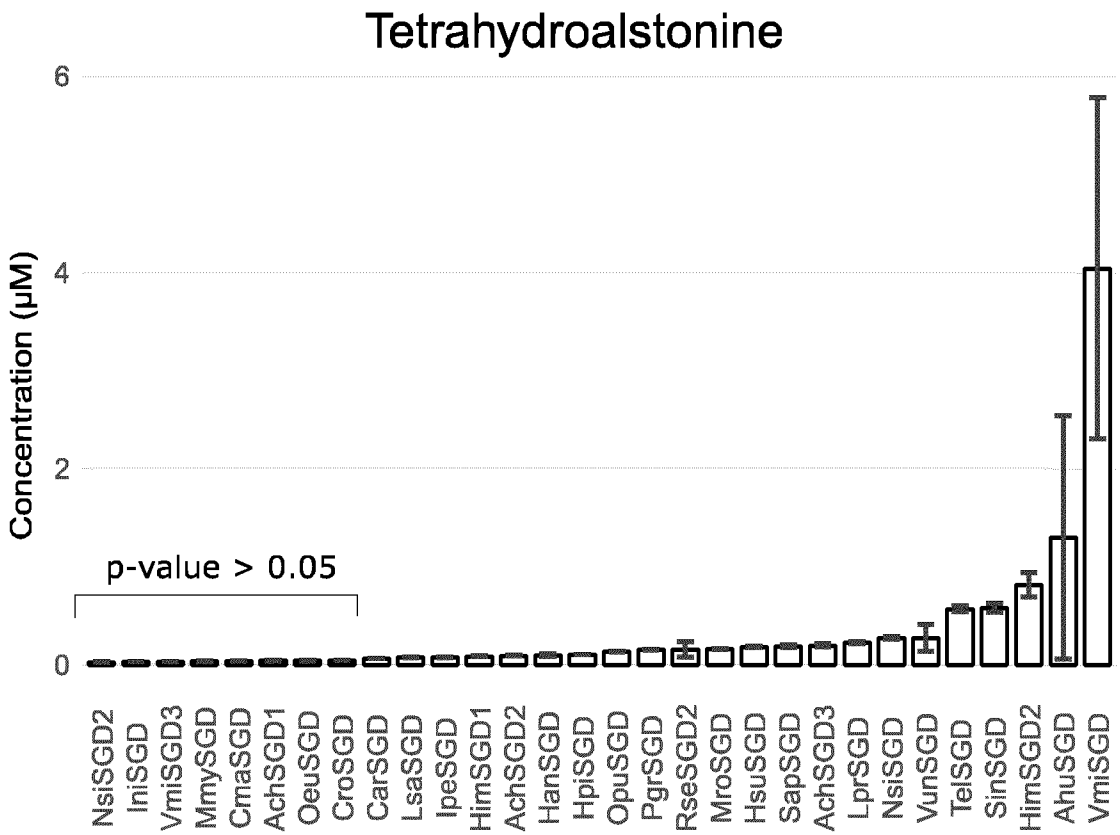
Tetrahydroalstonine
Concentration (µM)
6
4
2
0
p-value > 0.05
NsiSGD2
IniSGD
VmiSGD3
MmySGD
CmaSGD
AchSGD1
OeuSGD
CroSGD
CarSGD
LsaSGD
IpeSGD
HimSGD1
AchSGD2
HanSGD
HpiSGD
OpuSGD
PgrSGD
RseSGD2
MroSGD
HsuSGD
SapSGD
AchSGD3
LprSGD
NsiSGD
VunSGD
TelSGD
SinSGD
HimSGD2
AhuSGD
VmiSGD

Fig. 8

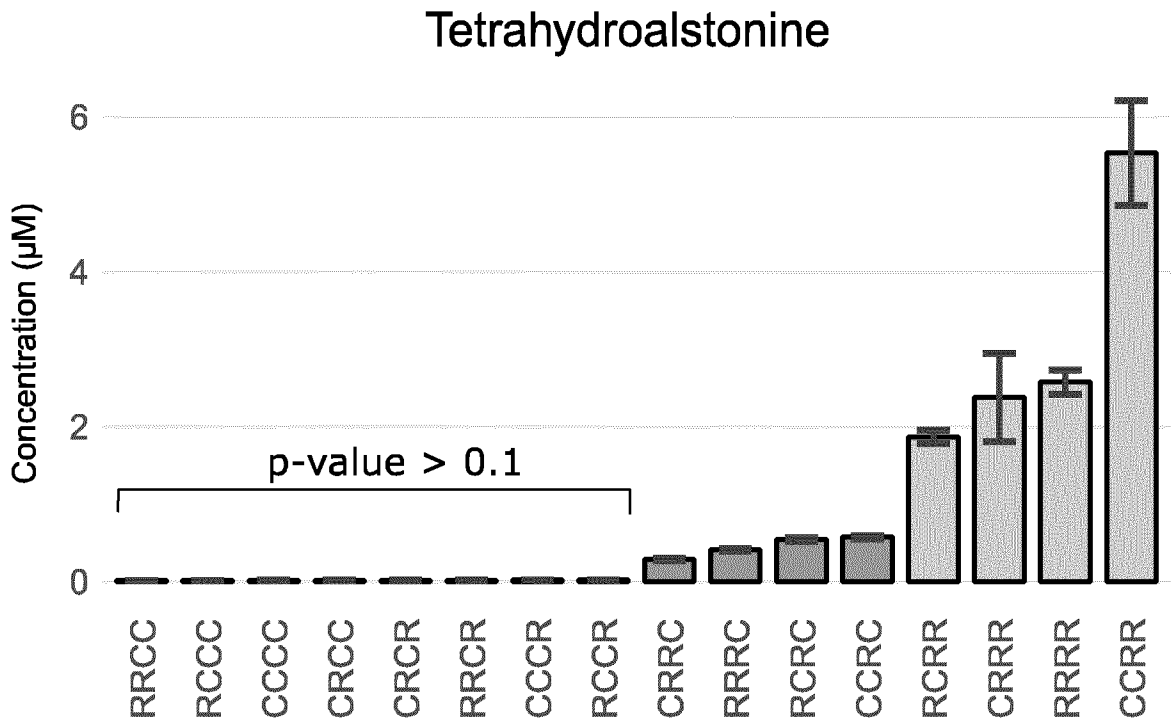
Tetrahydroalstonine
Concentration (µM)
6
4
2
0
p-value > 0.1
RRCC
RCCC
CCCC
CRCC
CRCR
RRCR
CCCR
RCCR
CRRC
RRRC
RCRC
CCRC
RCRR
CRRR
RRRR
CCRR

Fig. 9

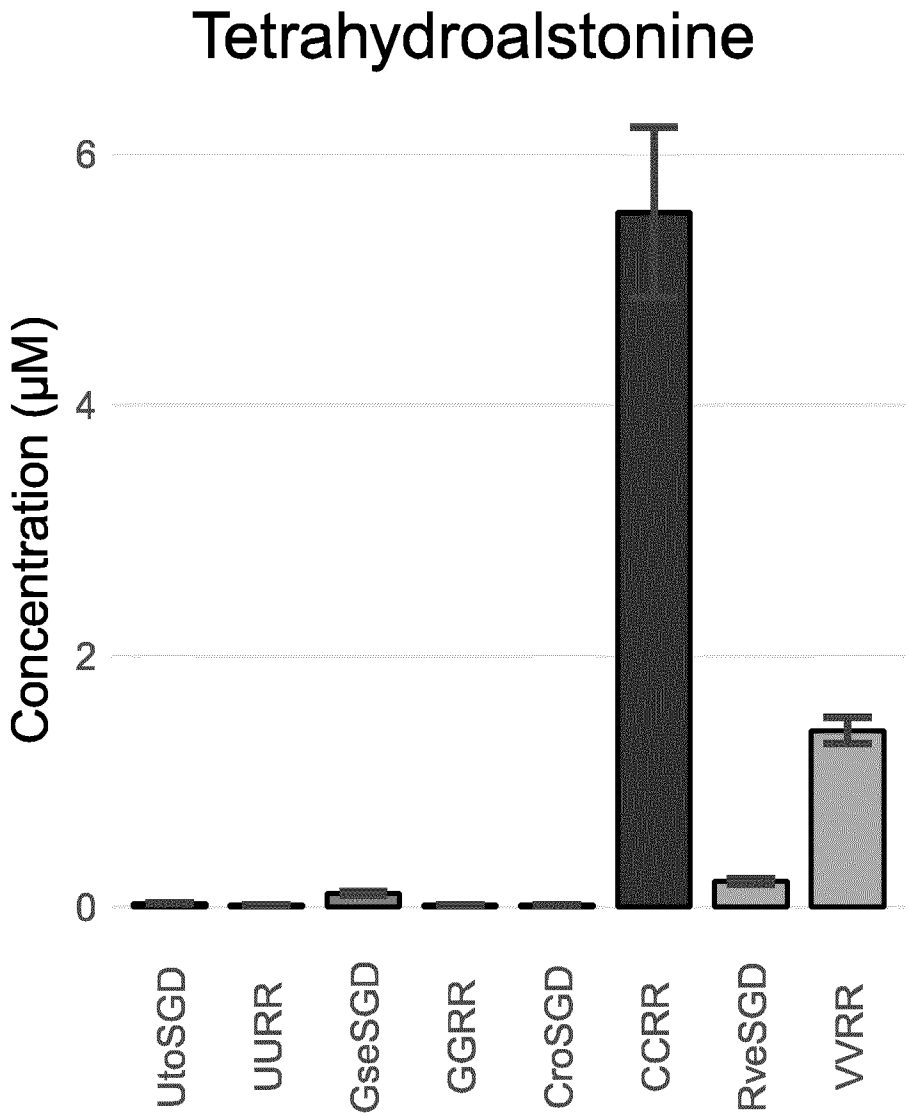
Tetrahydroalstonine
Concentration (µM)
6
4
2
0
UtoSGD
UURR
GseSGD
GGRR
CroSGD
CCRR
RveSGD
VVRR

Fig. 10

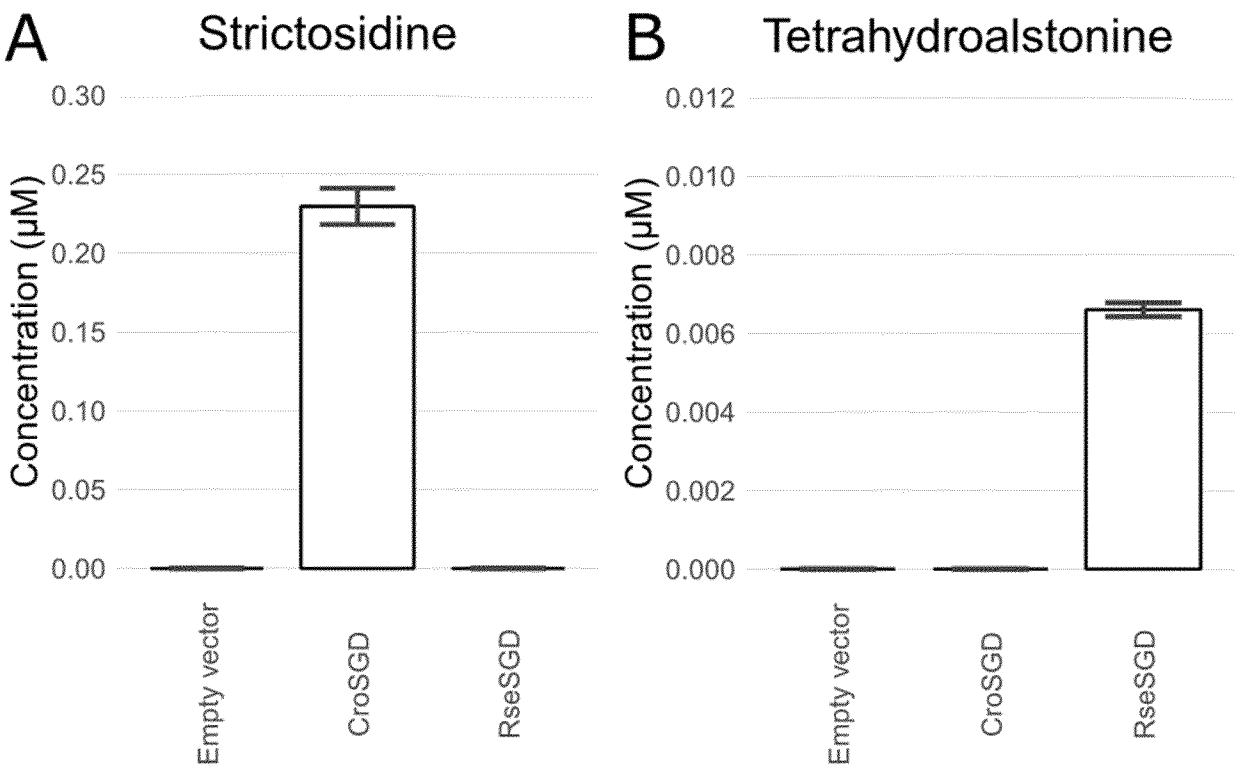
A
Strictosidine
Concentration (µM)
0.30
0.25
0.20
0.15
0.10
0.05
0.00
Empty vector
CroSGD
RseSGD
B
Tetrahydroalstonine
Concentration (µM)
0.012
0.010
0.008
0.006
0.004
0.002
0.000
Empty vector
CroSGD
RseSGD

Fig. 11

```
          1                   10                  20
AchSGD2   MRK.......GIVLAVVLV....VLRVQTCIAQI.........................
AchSGD3   MFS.......AYISAMRTE....VHLLLRVL..VV.VNTIVFSSEALSSNYGIG....SL
AchSGD1   MAI..N.....RALLILFCF.LA...ISNTEATS...KKYP.................P.
CacSGD    ....ME.A..QSIPLSVH.N...P...S.............................SI
CarSGD    MAAKSN.V.....................T.............................ND
CmaSGD    MEI.......SGSILFGFL....VVVGLVAS..RS.EAVHDLRTFN.............
CroSGD    MGS.......KDDQSLVVA....ISPAAEPNGNH...SVPIPFAYPSIPIQPRKHNKPI.
GseSGD    MAT.......PSSTIVP..DAT.....................................K.
GsoSGD    MAF.......KGYFVLGLI....ALVVVGTSKVTCEIEAD...KVSPII.........DF
HanSGD    MAT.......FDLTDQIAP....FPDE................................I
HimSGD2   MEN..G.....SGAVVAVGN....PQSAGSPNAV...PPDQ.DNS.............N.
HimSGD1   MNQ..D.....KMALQEYLA.TPTRIIR..............................
HpiSGD    MTE.........................................................
HsuSGD    MAQ.........................................................
IniSGD    ..MYGD.K....KMDNHL.T...GNGRK...............................FQ
IpeSGD    MSS.......VLPTPVLP....TPGRN.................................
LprSGD    .........................................................
LsaSGD    MET.......TTQNTGAKF....SLFQNL...............................V
MmySGD    .........................................................
MroSGD    .........................................................
NsiSGD1   ..MENS.S.....................D...............................LL
NsiSGD2   ....ME.S..EGNGVP...........T...............................YL
OeuSGD    MDI..Q.....SNVLTITSGST....PTDTSSNG...QAAKSTKE.............R.
OpuSGD    ......MEFLNPAFTR.............................................V
PgrSGD    .........................................................
AhuSGD    MAT.......IPKVI....DATNISRRPFPTDAS..........................K.
RseSGD    MAT.......QSSAVIDSNDAT.....................................R.
RseSGD    MDN.......TQAEPLVVA....IVPKPNASTEH...TNSHLIPVTRSKIV.........
RveSGD    MES.......NQGEPLVVA....IVPKPNASTEQ...KNSHLIPATRSKIV.........
SapSGD    .........................................................
SinSGD    MAN..N.....GPGA....QVARYVGA.................................K.
TelSGD    MET.......THSPLVVAI....APRPNAVADMK...NSNATRPASKV............
UtoSGD    MST..PATKF..................................................
VmiSGD1   MEI.......TNHVELVKP....NGFANNNNSHY...INSSNTRSKI............
VmiSGD3   MAD.......LVEFINPNS....SSENNPSDQPS...STIISK................
VunSGD    MAF.......YSTLFLGLFA...LL.LVRSSKV...TSHETVSVSPTI..........DI
```

Fig. 12 A

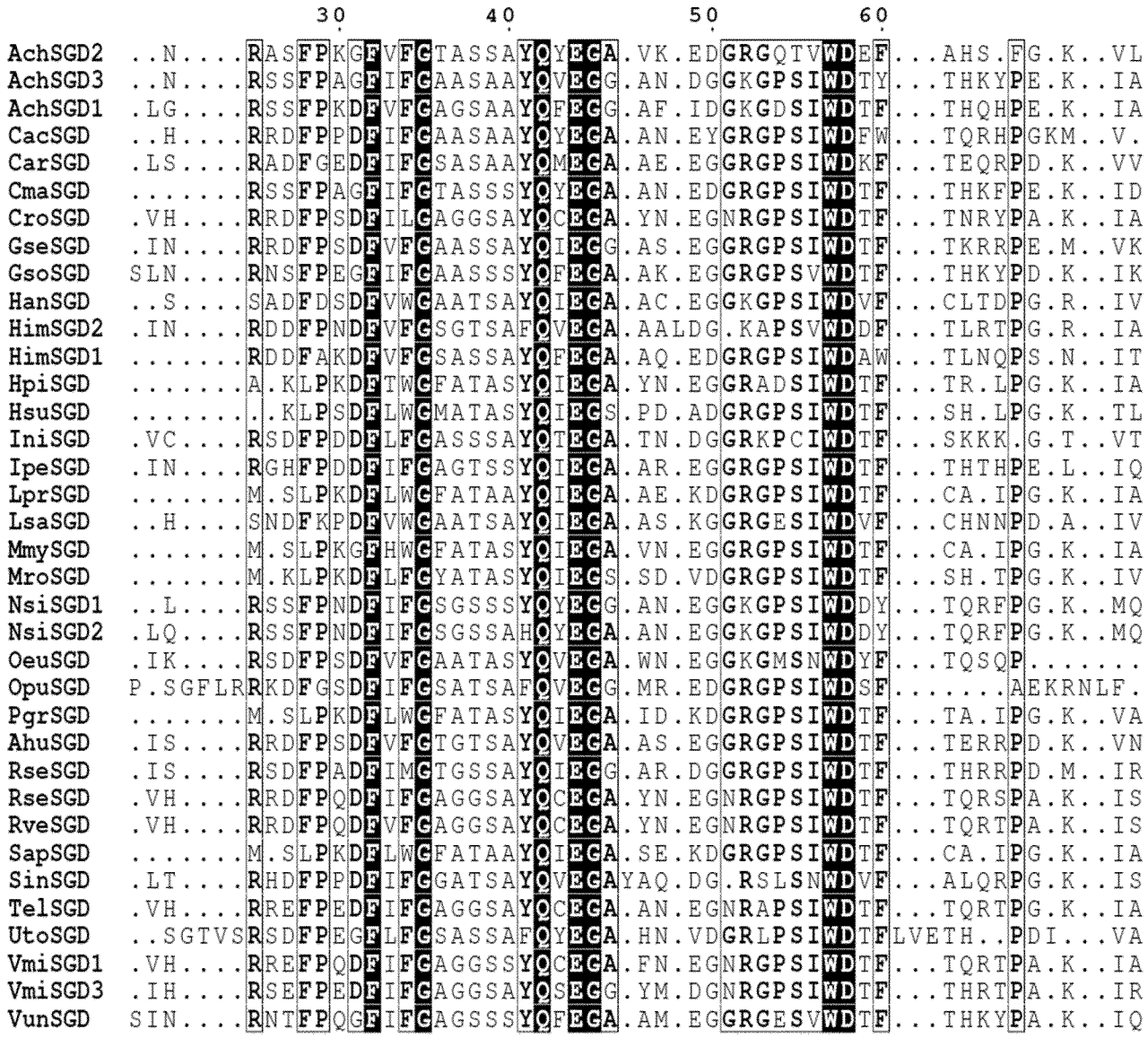
30
40
50
60
AchSGD2 ..N....RASFPKGFVFGTASSAYQYEGA.VK.EDGRGQTVWDEF...AHS.FG.K..VL
AchSGD3 ..N....RSSFPAGFIFGAASAAYQVEGG.AN.DGGKGPSIWDTY...THKYPE.K..IA
AchSGD1 .LG....RSSFPKDFVFGAGSAAYQFEGG.AF.IDGKGDSIWDTF...THQHPE.K..IA
CacSGD ..H....RRDFPPDFIFGAASAAYQYEGA.AN.EYGRGPSIWDFW...TQRHPGKM..V.
CarSGD .LS....RADFGEDFIFGSASAAYQMEGA.AE.EGGRGPSIWDKF...TEQRPD.K..VV
CmaSGD .......RSSFPAGFIFGTASSSYQYEGA.AN.EDGRGPSIWDTF...THKFPE.K..ID
CroSGD .VH....RRDFPSDFILGAGGSAYQCEGA.YN.EGNRGPSIWDTF...TNRYPA.K..IA
GseSGD .IN....RRDFPSDFVFGAASSAYQIEGG.AS.EGGRGPSIWDTF...TKRRPE.M..VK
GsoSGD SLN....RNSFPEGFIFGAASSSYQFEGA.AK.EGGRGPSVWDTF...THKYPD.K..IK
HanSGD ..S....SADFDSDFVWGAATSAYQIEGA.AC.EGGKGPSIWDVF...CLTDPG.R..IV
HimSGD2 .IN....RDDFPNDFVFGSGTSAFQVEGA.AALDG.KAPSVWDDF...TLRTPG.R..IA
HimSGD1 .......RDDFAKDFVFGSASSAYQFEGA.AQ.EDGRGPSIWDAW...TLNQPS.N..IT
HpiSGD .......A.KLPKDFTWGFATASYQIEGA.YN.EGGRADSIWDTF...TR.LPG.K..IA
HsuSGD .......KLPSDFLWGMATASYQIEGS.PD.ADGRGPSIWDTF...SH.LPG.K..TL
IniSGD .VC....RSDFPDDFLFGASSSAYQTEGA.TN.DGGRKPCIWDTF...SKKK.G.T..VT
IpeSGD .IN....RGHFPDDFIFGAGTSSYQIEGA.AR.EGGRGPSIWDTF...THTHPE.L..IQ
LprSGD .......M.SLPKDFLWGFATAAYQIEGA.AE.KDGRGPSIWDTF...CA.IPG.K..IA
LsaSGD ..H....SNDFKPDFVWGAATSAYQIEGA.AS.KGGRGESIWDVF...CHNNPD.A..IV
MmySGD .......M.SLPKGFHWGFATASYQIEGA.VN.EGGRGPSIWDTF...CA.IPG.K..IA
MroSGD .......M.KLPKDFLFGYATASYQIEGS.SD.VDGRGPSIWDTF...SH.TPG.K..IV
NsiSGD1 ..L....RSSFPNDFIFGSGSSSYQYEGG.AN.EGGKGPSIWDDY...TQRFPG.K..MQ
NsiSGD2 .LQ....RSSFPNDFIFGSGSSAHQYEGA.AN.EGGKGPSIWDDY...TQRFPG.K..MQ
OeuSGD .IK....RSDFPSDFVFGAATASYQVEGA.WN.EGGKGMSNWDYF...TQSQP.......
OpuSGD P.SGFLRRKDFGSDFIFGSATSAFQVEGG.MR.EDGRGPSIWDSF.......AEKRNLF.
PgrSGD .......M.SLPKDFLWGFATASYQIEGA.ID.KDGRGPSIWDTF...TA.IPG.K..VA
AhuSGD .IS....RRDFPSDFVFGTGTSAYQVEGA.AS.EGGRGPSIWDTF...TERRPD.K..VN
RseSGD .IS....RSDFPADFIMGTGSSAYQIEGG.AR.DGGRGPSIWDTF...THRRPD.M..IR
RseSGD .VH....RRDFPQDFIFGAGGSAYQCEGA.YN.EGNRGPSIWDTF...TQRSPA.K..IS
RveSGD .VH....RRDFPQDFVFGAGGSAYQCEGA.YN.EGNRGPSIWDTF...TQRTPA.K..IS
SapSGD .......M.SLPKDFLWGFATAAYQIEGA.SE.KDGRGPSIWDTF...CA.IPG.K..IA
SinSGD .LT....RHDFPPDFIFGGATSAYQVEGAYAQ.DG.RSLSNWDVF...ALQRPG.K..IS
TelSGD .VH....RREFPEDFIFGAGGSAYQCEGA.AN.EGNRAPSIWDTF...TQRTPG.K..IA
UtoSGD ..SGTVSRSDFPEGFLFGSASSAFQYEGA.HN.VDGRLPSIWDTFLVETH..PDI...VA
VmiSGD1 .VH....RREFPQDFIFGAGGSSYQCEGA.FN.EGNRGPSIWDTF...TQRTPA.K..IA
VmiSGD3 .IH....RSEFPEDFIFGAGGSAYQSEGG.YM.DGNRGPSIWDTF...THRTPA.K..IR
VunSGD SIN....RNTFPQGFIFGAGSSSYQFEGA.AM.EGGRGESVWDTF...THKYPA.K..IQ

Fig. 12 B

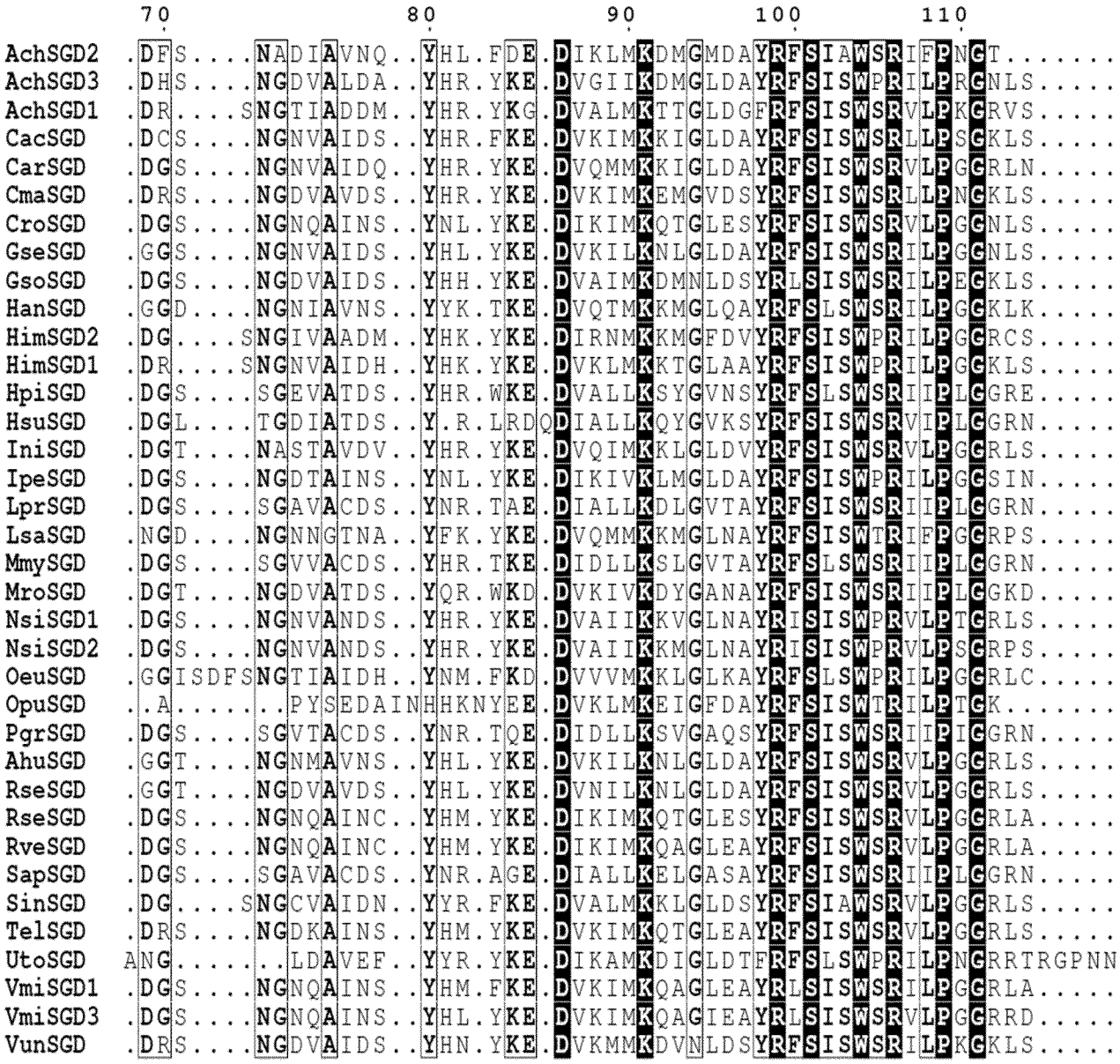
70
80
90
100
110
AchSGD2 .DFS....NADIAVNQ..YHL.FDE.DIKLMKDMGMDAYRFSIAWSRIFPNGT.....
AchSGD3 .DHS....NGDVALDA..YHR.YKE.DVGIIKDMGLDAYRFSISWPRILPRGNLS.....
AchSGD1 .DR....SNGTIADDM..YHR.YKG.DVALMKTTGLDGFRFSISWSRVLPKGRVS.....
CacSGD .DCS....NGNVAIDS..YHR.FKE.DVKIMKKIGLDAYRFSISWSRLLPSGKLS.....
CarSGD .DGS....NGNVAIDQ..YHR.YKE.DVQMMKKIGLDAYRFSISWSRVLPGGRLN.....
CmaSGD .DRS....NGDVAVDS..YHR.YKE.DVKIMKEMGVDSYRFSISWSRLLPNGKLS.....
CroSGD .DGS....NGNQAINS..YNL.YKE.DIKIMKQTGLESYRFSISWSRVLPGGNLS.....
GseSGD .GGS....NGNVAIDS..YHL.YKE.DVKILKNLGLDAYRFSISWSRILPGGNLS.....
GsoSGD .DGS....NGDVAIDS..YHH.YKE.DVAIMKDMNLDSYRLSISWSRILPEGKLS.....
HanSGD .GGD....NGNIAVNS..YYK.TKE.DVQTMKKMGLQAYRFSLSWSRILPGGKLK.....
HimSGD2 .DG....SNGIVAADM..YHK.YKE.DIRNMKKMGFDVYRFSISWPRILPGGRCS.....
HimSGD1 .DR....SNGNVAIDH..YHK.YKE.DVKLMKKTGLAAYRFSISWPRILPGGKLS.....
HpiSGD .DGS....SGEVATDS..YHR.WKE.DVALLKSYGVNSYRFSLSWSRIIPLGGRE.....
HsuSGD .DGL....TGDIATDS..Y.R.LRDQDIALLKQYGVKSYRFSISWSRVIPLGGRN.....
IniSGD .DGT....NASTAVDV..YHR.YKE.DVQIMKKLGLDVYRFSISWSRVLPGGRLS.....
IpeSGD .DGS....NGDTAINS..YNL.YKE.DIKIVKLMGLDAYRFSISWPRILPGGSIN.....
LprSGD .DGS....SGAVACDS..YNR.TAE.DIALLKDLGVTAYRFSISWSRIIPLGGRN.....
LsaSGD .NGD....NGNNGTNA..YFK.YKE.DVQMMKKMGLNAYRFSISWTRIFPGGRPS.....
MmySGD .DGS....SGVVACDS..YHR.TKE.DIDLLKSLGVTAYRFSLSWSRIIPLGGRN.....
MroSGD .DGT....NGDVATDS..YQR.WKD.DVKIVKDYGANAYRFSISWSRIIPLGGKD.....
NsiSGD1 .DGS....NGNVANDS..YHR.YKE.DVAIIKKVGLNAYRISISWPRVLPTGRLS.....
NsiSGD2 .DGS....NGNVANDS..YHR.YKE.DVAIIKKMGLNAYRISISWPRVLPSGRPS.....
OeuSGD .GGISDFSNGTIAIDH..YNM.FKD.DVVVMKKLGLKAYRFSLSWPRILPGGRLC.....
OpuSGD ..A.....PYSEDAINHHKNYEE.DVKLMKEIGFDAYRFSISWTRILPTGK.....
PgrSGD .DGS....SGVTACDS..YNR.TQE.DIDLLKSVGAQSYRFSISWSRIIPIGGRN.....
AhuSGD .GGT....NGNMAVNS..YHL.YKE.DVKILKNLGLDAYRFSISWSRVLPGGRLS.....
RseSGD .GGT....NGDVAVDS..YHL.YKE.DVNILKNLGLDAYRFSISWSRVLPGGRLS.....
RseSGD .DGS....NGNQAINC..YHM.YKE.DIKIMKQTGLESYRFSISWSRVLPGGRLA.....
RveSGD .DGS....NGNQAINC..YHM.YKE.DIKIMKQAGLEAYRFSISWSRVLPGGRLA.....
SapSGD .DGS....SGAVACDS..YNR.AGE.DIALLKELGASAYRFSISWSRIIPLGGRN.....
SinSGD .DG....SNGCVAIDN..YYR.FKE.DVALMKKLGLDSYRFSIAWSRVLPGGRLS.....
TelSGD .DRS....NGDKAINS..YHM.YKE.DVKIMKQTGLEAYRFSISWSRVLPGGRLS.....
UtoSGD ANG.....LDAVEF..YYR.YKE.DIKAMKDIGLDTFRFSLSWPRILPNGRRTRGPNN
VmiSGD1 .DGS....NGNQAINS..YHM.FKE.DVKIMKQAGLEAYRLSISWSRILPGGRLA.....
VmiSGD3 .DGS....NGNQAINS..YHL.YKE.DVKIMKQAGIEAYRLSISWSRVLPGGRRD.....
VunSGD .DRS....NGDVAIDS..YHN.YKE.DVKMMKDVNLDSYRFSISWSRILPKGKLS.....

Fig. 12 C

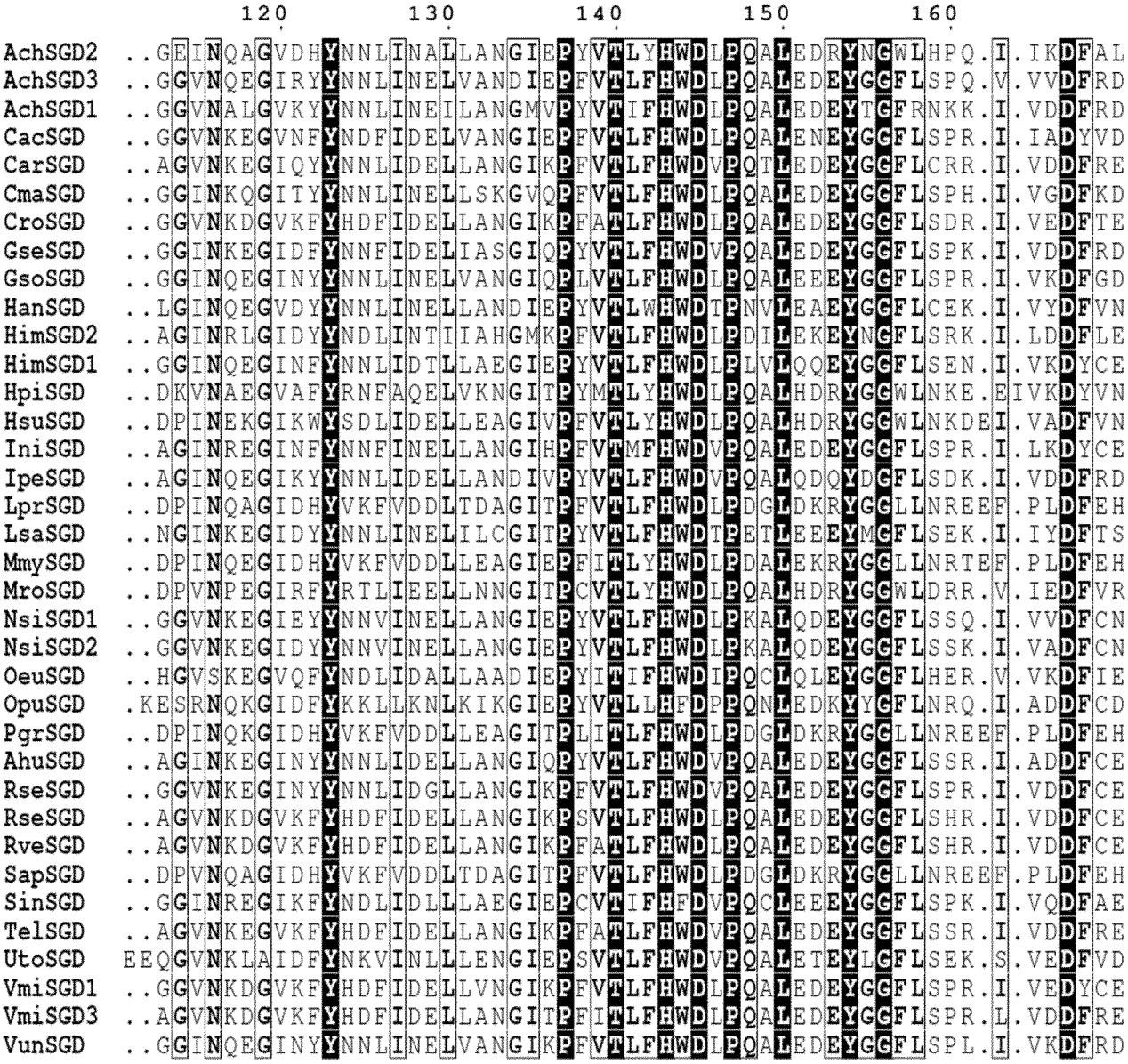
120
130
140
150
160
AchSGD2 ..GEINQAGVDHYNNLINALLANGIEPYVTLYHWDLPQALEDRYNGWLHPQ.I.IKDFAL
AchSGD3 ..GGVNQEGIRYYNNLINELVANDIEPFVTLFHWDLPQALEDEYGGFLSPQ.V.VVDFRD
AchSGD1 ..GGVNALGVKYYNNLINEILANGMVPYVTIFHWDLPQALEDEYTGFRNKK.I.VDDFRD
CacSGD ..GGVNKEGVNFYNDFIDELVANGIEPFVTLFHWDLPQALENEYGGFLSPR.I.IADYVD
CarSGD ..AGVNKEGIQYYNNLIDELLANGIKPFVTLFHWDVPQTLEDEYGGFLCRR.I.VDDFRE
CmaSGD ..GGINKQGITYYNNLINELLSKGVQPFVTLFHWDLPQALEDEYGGFLSPH.I.VGDFKD
CroSGD ..GGVNKDGVKFYHDFIDELLANGIKPFATLFHWDLPQALEDEYGGFLSDR.I.VEDFTE
GseSGD ..GGINKEGIDFYNNFIDELIASGIQPYVTLFHWDVPQALEDEYGGFLSPK.I.VDDFRD
GsoSGD ..GGINQEGINYYNNLINELVANGIQPLVTLFHWDLPQALEEEYGGFLSPR.I.VKDFGD
HanSGD ..LGINQEGVDYYNNLINELLANDIEPYVTLWHWDTPNVLEAEYGGFLCEK.I.VYDFVN
HimSGD2 ..AGINRLGIDYYNDLINTIIAHGMKPFVTLFHWDLPDILEKEYNGFLSRK.I.LDDFLE
HimSGD1 ..GGINQEGINFYNNLIDTLLAEGIEPYVTLFHWDLPLVLQQEYGGFLSEN.I.VKDYCE
HpiSGD ..DKVNAEGVAFYRNFAQELVKNGITPYMTLYHWDLPQALHDRYGGWLNKE.EIVKDYVN
HsuSGD ..DPINEKGIKWYSDLIDELLEAGIVPFVTLYHWDLPQALHDRYGGWLNKDEI.VADFVN
IniSGD ..AGINREGINFYNNFINELLANGIHPFVTMFHWDVPQALEDEYGGFLSPR.I.LKDYCE
IpeSGD ..AGINQEGIKYYNNLIDELLANDIVPYVTLFHWDVPQALQDQYDGFLSDK.I.VDDFRD
LprSGD ..DPINQAGIDHYVKFVDDLTDAGITPFVTLFHWDLPDGLDKRYGGLLNREEF.PLDFEH
LsaSGD ..NGINKEGIDYYNNLINELILCGITPYVTLFHWDTPETLEEEYMGFLSEK.I.IYDFTS
MmySGD ..DPINQEGIDHYVKFVDDLLEAGIEPFITLYHWDLPDALEKRYGGLLNRTEF.PLDFEH
MroSGD ..DPVNPEGIRFYRTLIEELLNNGITPCVTLYHWDLPQALHDRYGGWLDRR.V.IEDFVR
NsiSGD1 ..GGVNKEGIEYYNNVINELLANGIEPYVTLFHWDLPKALQDEYGGFLSSQ.I.VVDFCN
NsiSGD2 ..GGVNKEGIDYYNNVINELLANGIEPYVTLFHWDLPKALQDEYGGFLSSK.I.VADFCN
OeuSGD ..HGVSKEGVQFYNDLIDALLAADIEPYITIFHWDIPQCLQLEYGGFLHER.V.VKDFIE
OpuSGD .KESRNQKGIDFYKKLLKNLKIKGIEPYVTLLHFDPPQNLEDKYYGFLNRQ.I.ADDFCD
PgrSGD ..DPINQKGIDHYVKFVDDLLEAGITPLITLFHWDLPDGLDKRYGGLLNREEF.PLDFEH
AhuSGD ..AGINKEGINYYNNLIDELLANGIQPYVTLFHWDVPQALEDEYGGFLSSR.I.ADDFCE
RseSGD ..GGVNKEGINYYNNLIDGLLANGIKPFVTLFHWDVPQALEDEYGGFLSPR.I.VDDFCE
RseSGD ..AGVNKDGVKFYHDFIDELLANGIKPSVTLFHWDLPQALEDEYGGFLSHR.I.VDDFCE
RveSGD ..AGVNKDGVKFYHDFIDELLANGIKPFATLFHWDLPQALEDEYGGFLSHR.I.VDDFCE
SapSGD ..DPVNQAGIDHYVKFVDDLTDAGITPFVTLFHWDLPDGLDKRYGGLLNREEF.PLDFEH
SinSGD ..GGINREGIKFYNDLIDLLLAEGIEPCVTIFHFDVPQCLEEEYGGFLSPK.I.VQDFAE
TelSGD ..AGVNKEGVKFYHDFIDELLANGIKPFATLFHWDVPQALEDEYGGFLSSR.I.VDDFRE
UtoSGD EEQGVNKLAIDFYNKVINLLLENGIEPSVTLFHWDVPQALETEYLGFLSEK.S.VEDFVD
VmiSGD1 ..GGVNKDGVKFYHDFIDELLVNGIKPFVTLFHWDLPQALEDEYGGFLSPR.I.VEDYCE
VmiSGD3 ..AGVNKDGVKFYHDFIDELLANGITPFITLFHWDLPQALEDEYGGFLSPR.L.VDDFRE
VunSGD ..GGINQEGINYYNNLINELVANGIKPFVTLFHWDLPQALEDEYGGFLSPL.I.VKDFRD

Fig. 12 D

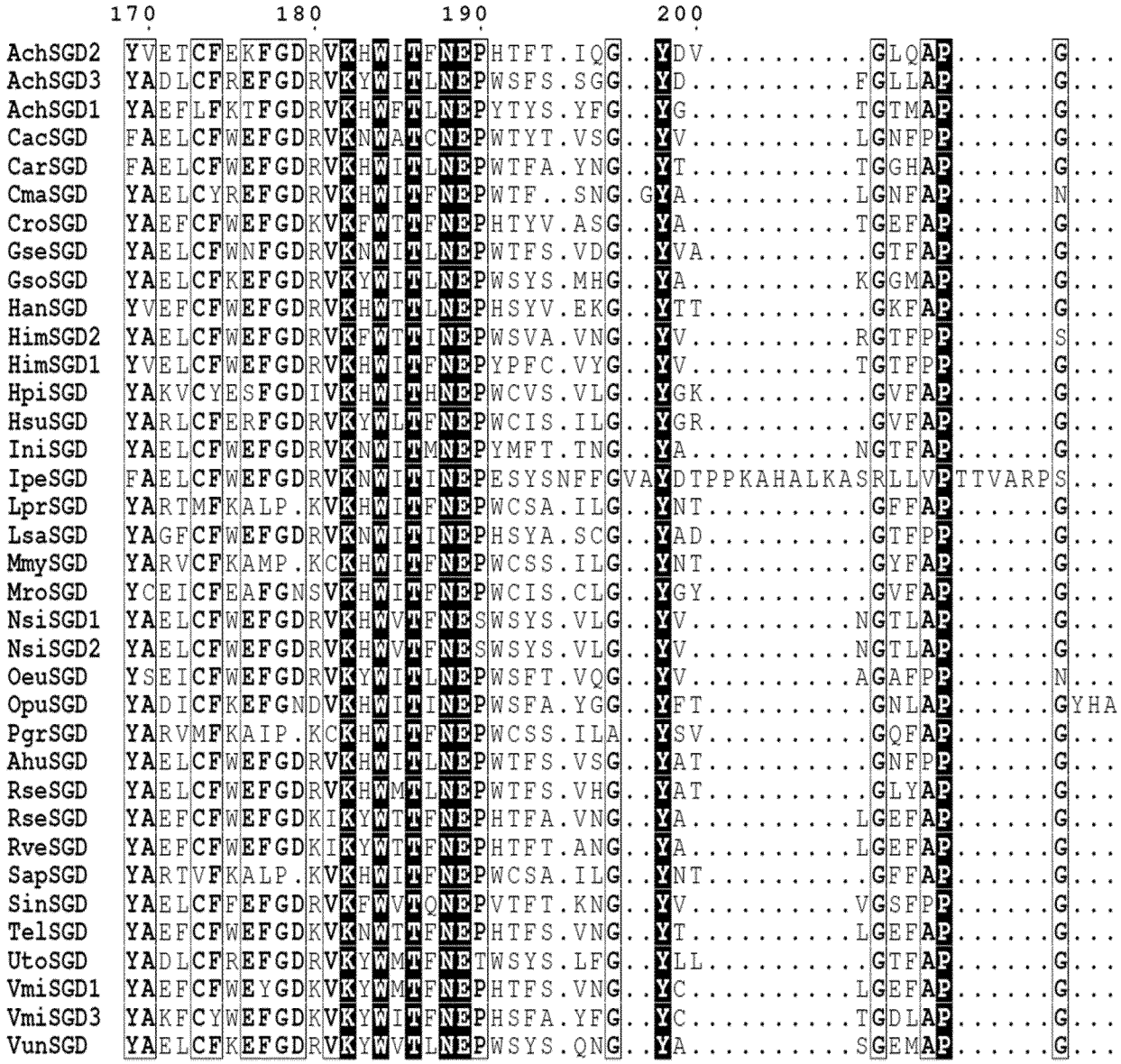
170
180
190
200
AchSGD2
AchSGD3
AchSGD1
CacSGD
CarSGD
CmaSGD
CroSGD
GseSGD
GsoSGD
HanSGD
HimSGD2
HimSGD1
HpiSGD
HsuSGD
IniSGD
IpeSGD
LprSGD
LsaSGD
MmySGD
MroSGD
NsiSGD1
NsiSGD2
OeuSGD
OpuSGD
PgrSGD
AhuSGD
RseSGD
RseSGD
RveSGD
SapSGD
SinSGD
TelSGD
UtoSGD
VmiSGD1
VmiSGD3
VunSGD

Fig. 12 E

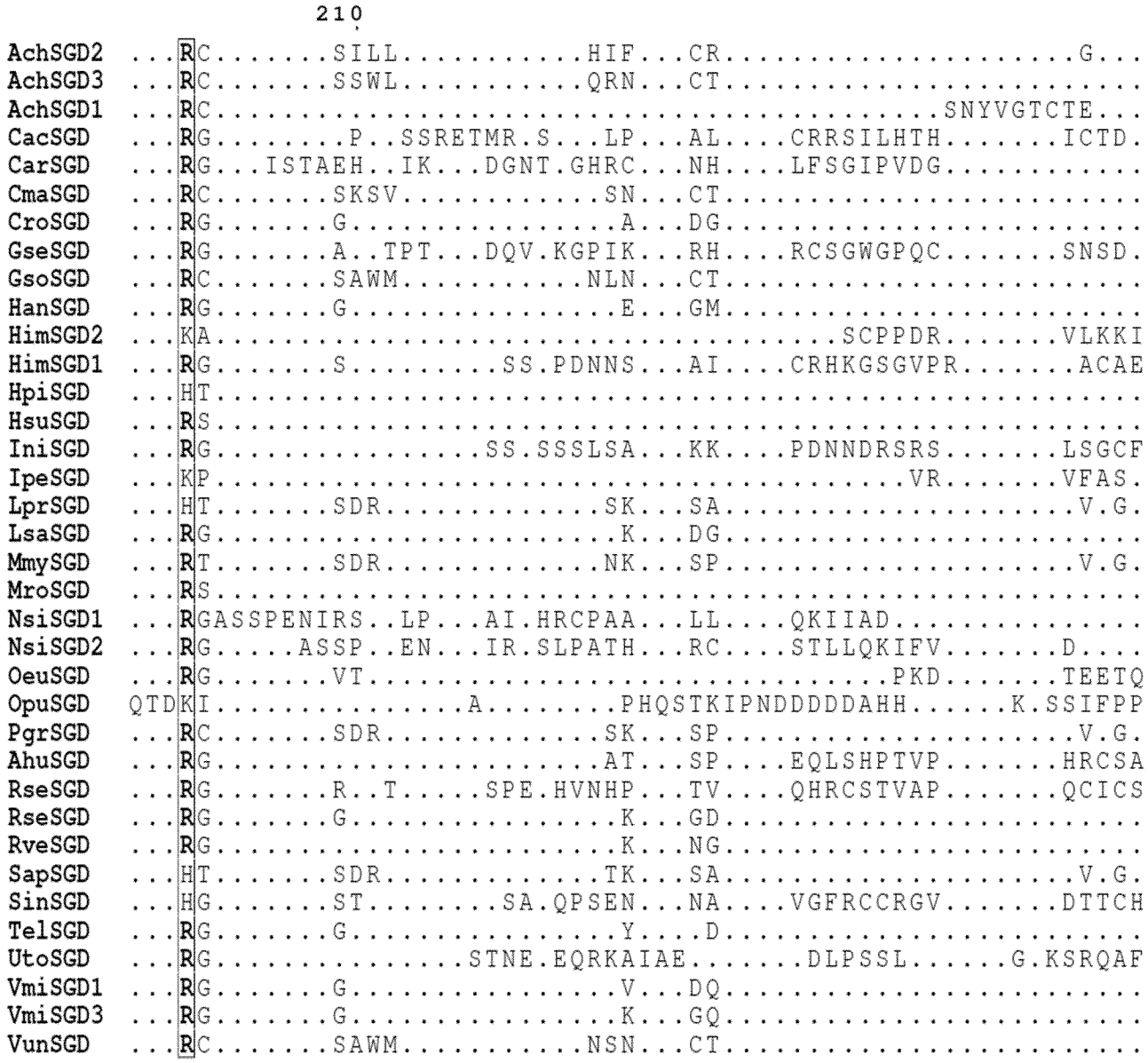
210
AchSGD2 ...RC.....SILL.....HIF...CR.....G...
AchSGD3 ...RC.....SSWL.....QRN...CT.....
AchSGD1 ...RC.....SNYVGTCTE...
CacSGD ...RG.....P..SSRETMR.S...LP...AL....CRRSILHTH.....ICTD.
CarSGD ...RG...ISTAEH..IK...DGNT.GHRC...NH....LFSGIPVDG.....
CmaSGD ...RC.....SKSV.....SN...CT.....
CroSGD ...RG.....G.....A...DG.....
GseSGD ...RG.....A..TPT...DQV.KGPIK...RH....RCSGWGPQC.....SNSD.
GsoSGD ...RC.....SAWM.....NLN...CT.....
HanSGD ...RG.....G.....E...GM.....
HimSGD2 ...KA.....SCPPDR.....VLKKI
HimSGD1 ...RG.....S.....SS.PDNNS...AI....CRHKGSGVPR.....ACAE
HpiSGD ...HT.....
HsuSGD ...RS.....
IniSGD ...RG.....SS.SSSLSA...KK....PDNNDRSRS.....LSGCF
IpeSGD ...KP.....VR.....VFAS.
LprSGD ...HT.....SDR.....SK...SA.....V.G.
LsaSGD ...RG.....K...DG.....
MmySGD ...RT.....SDR.....NK...SP.....V.G.
MroSGD ...RS.....
NsiSGD1 ...RGASSPENIRS..LP...AI.HRCPAA...LL....QKIIAD.....
NsiSGD2 ...RG.....ASSP..EN...IR.SLPATH...RC....STLLQKIFV.....D....
OeuSGD ...RG.....VT.....PKD.....TEETQ
OpuSGD QTDKI.....A.....PHQSTKIPNDDDDDAHH.....K.SSIFPP
PgrSGD ...RC.....SDR.....SK...SP.....V.G.
AhuSGD ...RG.....AT...SP....EQLSHPTVP.....HRCSA
RseSGD ...RG.....R..T.....SPE.HVNHP...TV....QHRCSTVAP.....QCICS
RseSGD ...RG.....G.....K...GD.....
RveSGD ...RG.....K...NG.....
SapSGD ...HT.....SDR.....TK...SA.....V.G.
SinSGD ...HG.....ST.....SA.QPSEN...NA....VGFRCCRGV.....DTTCH
TelSGD ...RG.....G.....Y....D.....
UtoSGD ...RG.....STNE.EQRKAIAE.....DLPSSL.....G.KSRQAF
VmiSGD1 ...RG.....G.....V...DQ.....
VmiSGD3 ...RG.....G.....K...GQ.....
VunSGD ...RC.....SAWM.....NSN...CT.....

Fig. 12 F

```
                                                220
AchSGD2  ........................GNSA............I.............
AchSGD3  ............G...........GDSG............I.............
AchSGD1  ........................GDSS............T.............
CacSGD   ........................GNPA............T.............
CarSGD   ..........................NPG............T.............
CmaSGD   ............G...........GNSG............T.............
CroSGD   ............K...........GNPG............K.............
GseSGD   ........................GNPG............T.............
GsoSGD   ............G...........GDSA............T.............
HanSGD   ............P...........GNPG............T.............
HimSGD2  PPHRSVQHSSATVPTTRQYSDIKYDKSDPA............K.............
HimSGD1  ........................GNPG............T.............
HpiSGD   ............................SD....RAKFHVGDSST.............
HsuSGD   ............................SD...RTRSPEGDSR.T.............
IniSGD   PWRLSLSSKVKISDN..........GDPG............L.............
IpeSGD   ............T...........ADPG............T............T
LprSGD   ........................DSA............R.............
LsaSGD   ............V...........GDPG............T.............
MmySGD   ........................DSA............R.............
MroSGD   ......................SNRNRSEAGDSTR.....................
NsiSGD1  ........................GDPG............I.............
NsiSGD2  ........................GDPG............R.............
OeuSGD   KHARLHRGGGKLLAAFKY.....GNPG............T.............
OpuSGD   SRFSLPPS................SSSA............SETPAIIPAKKLPYPD
PgrSGD   ........................DSS............R.............
AhuSGD   STMPCIRS.....T..........GNPG............T.............
RseSGD   T.......................GNPG............T.............
RseSGD   ............E...........GDPA............I.............
RveSGD   ............K...........GDPA............T.............
SapSGD   ........................DSA............R.............
SinSGD   G.......................GDAG............T.............
TelSGD   ............K...........GDPG............T.............
UtoSGD   AHSRTPRA................GDPS............TE............
VmiSGD1  ............K...........GDPG............I.............
VmiSGD3  ............E...........GDPG............R.............
VunSGD   ............G...........GDSS............T.............
```

Fig. 12 G

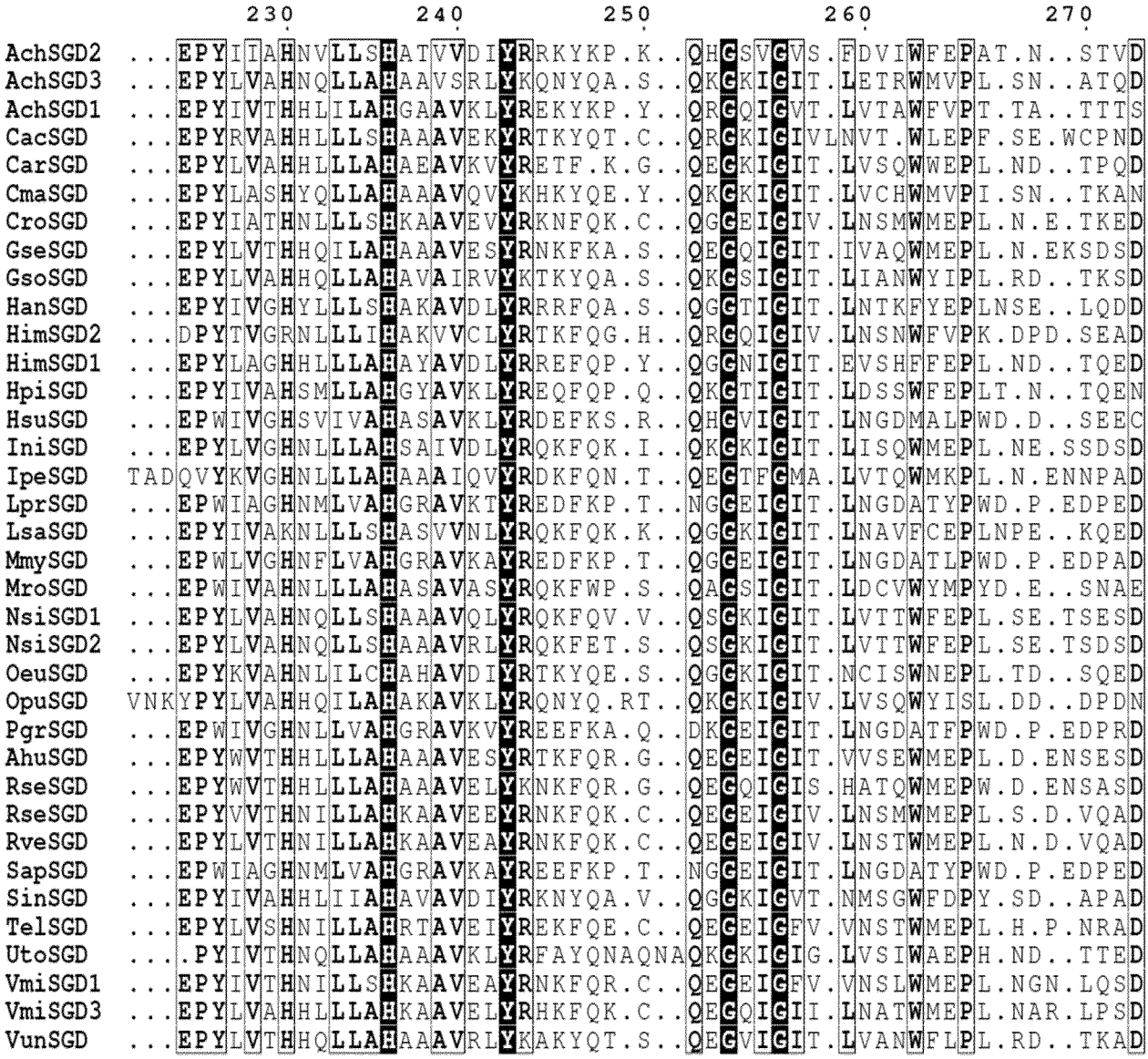
230
240
250
260
270
AchSGD2 ...EPYIIAHNVLLSHATVVDIYRRKYKP.K..QHGSVGVS.FDVIWFEPAT.N..STVD
AchSGD3 ...EPYLVAHNQLLAHAAVSRLYKQNYQA.S..QKGKIGIT.LETRWMVPL.SN..ATQD
AchSGD1 ...EPYIVTHHLILAHGAAVKLYREKYKP.Y..QRGQIGVT.LVTAWFVPT.TA..TTTS
CacSGD ...EPYRVAHHLLLSHAAAVEKYRTKYQT.C..QRGKIGIVLNVT.WLEPF.SE.WCPND
CarSGD ...EPYLVAHHLLLAHAEAVKVYRETF.K.G..QEGKIGIT.LVSQWWEPL.ND..TPQD
CmaSGD ...EPYLASHYQLLAHAAAVQVYKHKYQE.Y..QKGKIGIT.LVCHWMVPI.SN..TKAN
CroSGD ...EPYIATHNLLLSHKAAVEVYRKNFQK.C..QGGEIGIV.LNSMWMEPL.N.E.TKED
GseSGD ...EPYLVTHHQILAHAAAVESYRNKFKA.S..QEGQIGIT.IVAQWMEPL.N.EKSDSD
GsoSGD ...EPYLVAHHQLLAHAVAIRVYKTKYQA.S..QKGSIGIT.LIANWYIPL.RD..TKSD
HanSGD ...EPYIVGHYLLLSHAKAVDLYRRRFQA.S..QGGTIGIT.LNTKFYEPLNSE..LQDD
HimSGD2 ...DPYTVGRNLLLIHAKVVCLYRTKFQG.H..QRGQIGIV.LNSNWFVPK.DPD.SEAD
HimSGD1 ...EPYLAGHHLLLAHAYAVDLYREFQP.Y..QGGNIGIT.EVSHFFEPL.ND..TQED
HpiSGD ...EPYIVAHSMLLAHGYAVKLYREQFQP.Q..QKGTIGIT.LDSSWFEPLT.N..TQEN
HsuSGD ...EPWIVGHSVIVAHASAVKLYRDEFKS.R..QHGVIGIT.LNGDMALPWD.D..SEEC
IniSGD ...EPYLVGHNLLLAHSAIVDLYRQKFQK.I..QKGKIGIT.LISQWMEPL.NE.SSDSD
IpeSGD TADQVYKVGHNLLLAHAAAIQVYRDKFQN.T..QEGTFGMA.LVTQWMKPL.N.ENNPAD
LprSGD ...EPWIAGHNMLVAHGRAVKTYREDFKP.T..NGGEIGIT.LNGDATYPWD.P.EDPED
LsaSGD ...EPYIVAKNLLLSHASVVNLYRQKFQK.K..QGGKIGIT.LNAVFCEPLNPE..KQED
MmySGD ...EPWLVGHNFLVAHGRAVKAYREDFKP.T..QGGEIGIT.LNGDATLPWD.P.EDPAD
MroSGD ...EPWIVAHNLLLAHASAVASYRQKFWP.S..QAGSIGIT.LDCVWYMPYD.E..SNAE
NsiSGD1 ...EPYLVAHNQLLSHAAAVQLYRQKFQV.V..QSGKIGIT.LVTTWFEPL.SE.TSESD
NsiSGD2 ...EPYLVAHNQLLSHAAAVRLYRQKFET.S..QSGKIGIT.LVTTWFEPL.SE.TSDSD
OeuSGD ...EPYKVAHNLILCHAHAVDIYRTKYQE.S..QGGKIGIT.NCISWNEPL.TD..SQED
OpuSGD VNKYPYLVAHHQILAHAKAVKLYRQNYQ.RT..QKGKIGIV.LVSQWYISL.DD..DPDN
PgrSGD ...EPWIVGHNLLVAHGRAVKVYREEFKA.Q..DKGEIGIT.LNGDATFPWD.P.EDPRD
AhuSGD ...EPYWVTHHLLLAHAAAVESYRTKFQR.G..QEGEIGIT.VVSEWMEPL.D.ENSESD
RseSGD ...EPYWVTHHLLLAHAAAVELYKNKFQR.G..QEGQIGIS.HATQWMEPW.D.ENSASD
RseSGD ...EPYVVTHNILLAHKAAVEEYRNKFQK.C..QEGEIGIV.LNSMWMEPL.S.D.VQAD
RveSGD ...EPYLVTHNILLAHKAAVEAYRNKFQK.C..QEGEIGIV.LNSTWMEPL.N.D.VQAD
SapSGD ...EPWIAGHNMLVAHGRAVKAYREEFKP.T..NGGEIGIT.LNGDATYPWD.P.EDPED
SinSGD ...EPYIVAHHLIIAHAVAVDIYRKNYQA.V..QGGKIGVT.NMSGWFDPY.SD..APAD
TelSGD ...EPYLVSHNILLAHRTAVEIYREKFQE.C..QEGEIGFV.VNSTWMEPL.H.P.NRAD
UtoSGD ....PYIVTHNQLLAHAAAVKLYRFAYQNAQNAQKGKIGIG.LVSIWAEPH.ND..TTED
VmiSGD1 ...EPYIVTHNILLSHKAAVEAYRNKFQR.C..QEGEIGFV.VNSLWMEPL.NGN.LQSD
VmiSGD3 ...EPYLVAHHLLLAHKAAVELYRHKFQK.C..QEGQIGII.LNATWMEPL.NAR.LPSD
VunSGD ...EPYLVTHHQLLAHAAAVRLYKAKYQT.S..QEGVIGIT.LVANWFLPL.RD..TKAD

Fig. 12 H

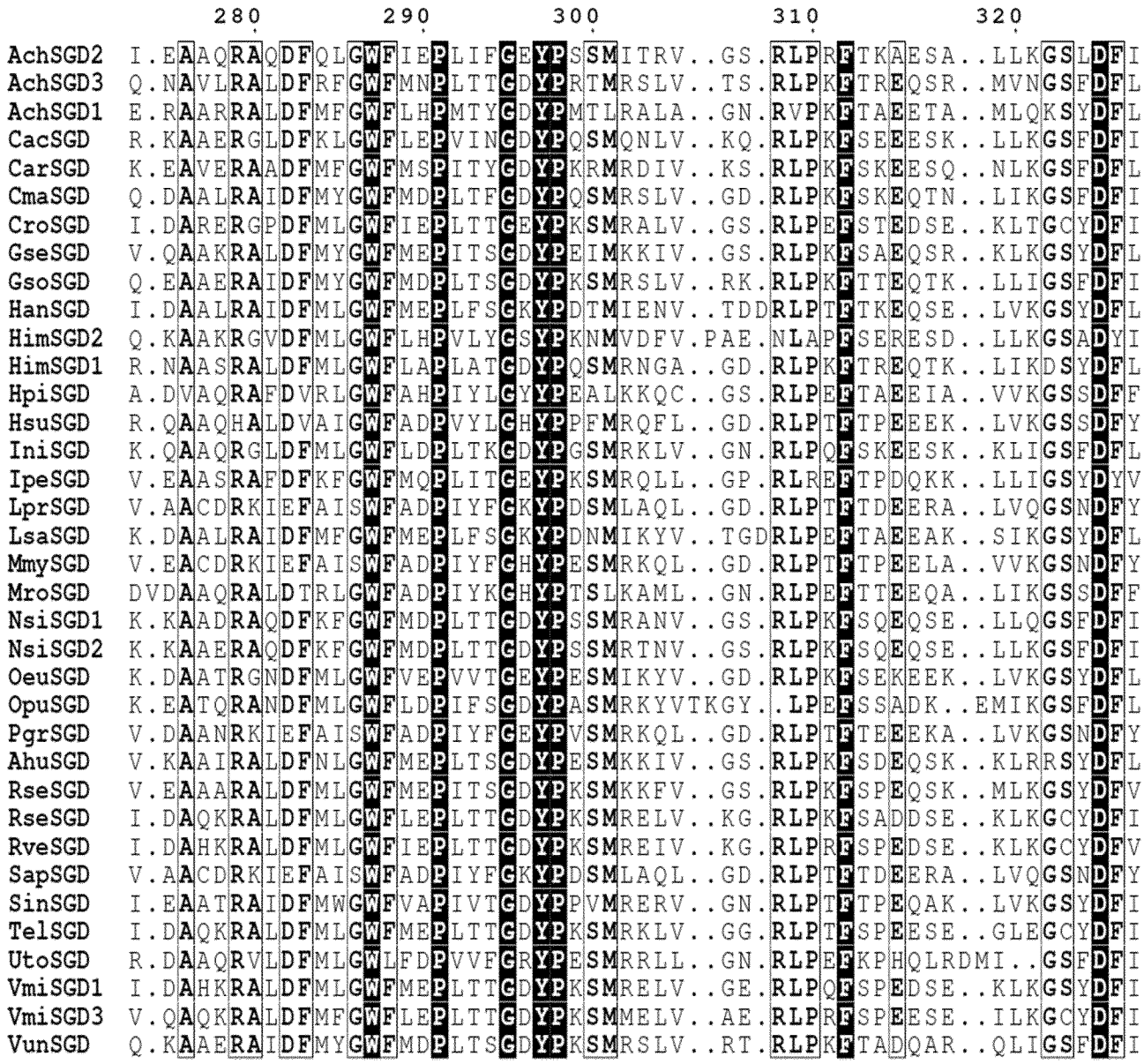
280
290
300
310
320
AchSGD2 I.EAAQRAQDFQLGWFIEPLIFGEYPSSMITRV..GS.RLPRFTKAESA..LLKGSLDFI
AchSGD3 Q.NAVLRALDFRFGWFMNPLTTGDYPRTMRSLV..TS.RLPKFTREQSR..MVNGSFDFL
AchSGD1 E.RAARRALDFMFGWFLHPMTYGDYPMTLRALA..GN.RVPKFTAEETA..MLQKSYDFL
CacSGD R.KAAERGLDFKLGWFLEPVINGDYPQSMQNLV..KQ.RLPKFSEEESK..LLKGSFDFI
CarSGD K.EAVERAADFMFGWFMSPITYGDYPKRMRDIV..KS.RLPKFSKEESQ..NLKGSFDFL
CmaSGD Q.DAALRAIDFMYGWFMDPLTFGDYPQSMRSLV..GD.RLPKFSKEQTN..LIKGSFDFI
CroSGD I.DARERGPDFMLGWFIEPLTTGEYPKSMRALV..GS.RLPEFSTEDSE..KLTGCYDFI
GseSGD V.QAAKRALDFMYGWFMEPITSGDYPEIMKKIV..GS.RLPKFSAEQSR..KLKGSYDFL
GsoSGD Q.EAAERAIDFMYGWFMDPLTSGDYPKSMRSLV..RK.RLPKFTTEQTK..LLIGSFDFI
HanSGD I.DAALRAIDFMLGWFMEPLFSGKYPDTMIENV..TDDRLPTFTKEQSE..LVKGSYDFL
HimSGD2 Q.KAAKRGVDFMLGWFLHPVLYGSYPKNMVDFV.PAE.NLAPFSERESD..LLKGSADYI
HimSGD1 R.NAASRALDFMLGWFLAPLATGDYPQSMRNGA..GD.RLPKFTREQTK..LIKDSYDFL
HpiSGD A.DVAQRAFDVRLGWFAHPIYLGYYPEALKKQC..GS.RLPEFTAEEIA..VVKGSSDFF
HsuSGD R.QAAQHALDVAIGWFADPVYLGHYPPFMRQFL..GD.RLPTFTPEEEK..LVKGSSDFY
IniSGD K.QAAQRGLDFMLGWFLDPLTKGDYPGSMRKLV..GN.RLPQFSKEESK..KLIGSFDFL
IpeSGD V.EAASRAFDFKFGWFMQPLITGEYPKSMRQLL..GP.RLREFTPDQKK..LLIGSYDYV
LprSGD V.AACDRKIEFAISWFADPIYFGKYPDSMLAQL..GD.RLPTFTDEERA..LVQGSNDFY
LsaSGD K.DAALRAIDFMFGWFMEPLFSGKYPDNMIKYV..TGDRLPEFTAEEAK..SIKGSYDFL
MmySGD V.EACDRKIEFAISWFADPIYFGHYPESMRKQL..GD.RLPTFTPEELA..VVKGSNDFY
MroSGD DVDAAQRALDTRLGWFADPIYKGHYPTSLKAML..GN.RLPEFTTEEQA..LIKGSSDFF
NsiSGD1 K.KAADRAQDFKFGWFMDPLTTGDYPSSMRANV..GS.RLPKFSQEQSE..LLQGSFDFI
NsiSGD2 K.KAAERAQDFKFGWFMDPLTTGDYPSSMRTNV..GS.RLPKFSQEQSE..LLKGSFDFI
OeuSGD K.DAATRGNDFMLGWFVEPVVTGEYPESMIKYV..GD.RLPKFSEKEEK..LVKGSYDFL
OpuSGD K.EATQRANDFMLGWFLDPIFSGDYPASMRKYVTKGY..LPEFSSADK..EMIKGSFDFL
PgrSGD V.DAANRKIEFAISWFADPIYFGEYPVSMRKQL..GD.RLPTFTEEEKA..LVKGSNDFY
AhuSGD V.KAAIRALDFNLGWFMEPLTSGDYPESMKKIV..GS.RLPKFSDEQSK..KLRRSYDFL
RseSGD V.EAAARALDFMLGWFMEPITSGDYPKSMKKFV..GS.RLPKFSPEQSK..MLKGSYDFV
RseSGD I.DAQKRALDFMLGWFLEPLTTGDYPKSMRELV..KG.RLPKFSADDSE..KLKGCYDFI
RveSGD I.DAHKRALDFMLGWFIEPLTTGDYPKSMREIV..KG.RLPRFSPEDSE..KLKGCYDFV
SapSGD V.AACDRKIEFAISWFADPIYFGKYPDSMLAQL..GD.RLPTFTDEERA..LVQGSNDFY
SinSGD I.EAATRAIDFMWGWFVAPIVTGDYPPVMRERV..GN.RLPTFTPEQAK..LVKGSYDFI
TelSGD I.DAQKRALDFMLGWFMEPLTTGDYPKSMRKLV..GG.RLPTFSPEESE..GLEGCYDFI
UtoSGD R.DAAQRVLDFMLGWLFDPVVFGRYPESMRRLL..GN.RLPEFKPHQLRDMI..GSFDFI
VmiSGD1 I.DAHKRALDFMLGWFMEPLTTGDYPKSMRELV..GE.RLPQFSPEDSE..KLKGSYDFI
VmiSGD3 V.QAQKRALDFMFGWFLEPLTTGDYPKSMMELV..AE.RLPRFSPEESE..ILKGCYDFI
VunSGD Q.KAAERAIDFMYGWFMDPLTSGDYPKSMRSLV..RT.RLPKFTADQAR..QLIGSFDFI

Fig. 12 I

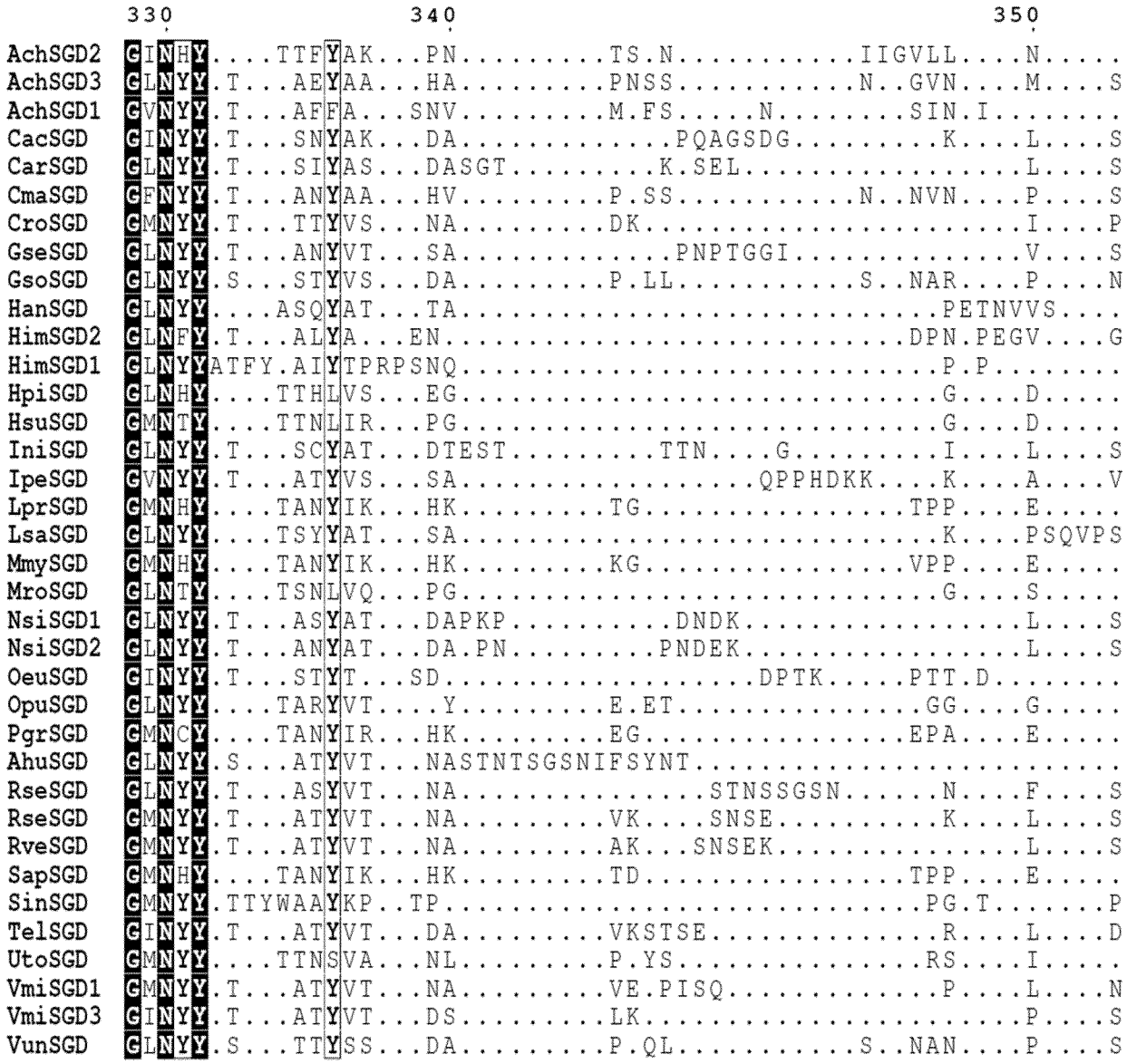
330
340
350
AchSGD2 GINHY....TTFYAK...PN.....TS.N.....IIGVLL....N.....
AchSGD3 GLNYY.T...AEYAA...HA.....PNSS.....N..GVN....M....S
AchSGD1 GVNYY.T...AFFA...SNV.....M.FS.....N.....SIN.I.....
CacSGD GINYY.T...SNYAK...DA.....PQAGSDG.....K....L....S
CarSGD GLNYY.T...SIYAS...DASGT.....K.SEL.....L....S
CmaSGD GFNYY.T...ANYAA...HV.....P.SS.....N..NVN....P....S
CroSGD GMNYY.T...TTYVS...NA.....DK.....I....P
GseSGD GLNYY.T...ANYVT...SA.....PNPTGGI.....V....S
GsoSGD GLNYY.S...STYVS...DA.....P.LL.....S..NAR....P....N
HanSGD GLNYY....ASQYAT...TA.....PETNVVS....
HimSGD2 GLNFY.T...ALYA...EN.....DPN.PEGV....G
HimSGD1 GLNYYATFY.AIYTPRPSNQ.....P.P.....
HpiSGD GLNHY....TTHLVS...EG.....G....D.....
HsuSGD GMNTY....TTNLIR...PG.....G....D.....
IniSGD GLNYY.T...SCYAT...DTEST.....TTN....G.....I....L....S
IpeSGD GVNYY.T...ATYVS...SA.....QPPHDKK....K....A....V
LprSGD GMNHY....TANYIK...HK.....TG.....TPP....E.....
LsaSGD GLNYY....TSYYAT...SA.....K....PSQVPS
MmySGD GMNHY....TANYIK...HK.....KG.....VPP....E.....
MroSGD GLNTY....TSNLVQ...PG.....G....S.....
NsiSGD1 GLNYY.T...ASYAT...DAPKP.....DNDK.....L....S
NsiSGD2 GLNYY.T...ANYAT...DA.PN.....PNDEK.....L....S
OeuSGD GINYY.T...STYT...SD.....DPTK.....PTT.D.....
OpuSGD GLNYY....TARYVT....Y.....E.ET.....GG....G.....
PgrSGD GMNCY....TANYIR...HK.....EG.....EPA....E.....
AhuSGD GLNYY.S...ATYVT...NASTNTSGSNIFSYNT.....
RseSGD GLNYY.T...ASYVT...NA.....STNSSGSN.....N....F....S
RseSGD GMNYY.T...ATYVT...NA.....VK....SNSE.....K....L....S
RveSGD GMNYY.T...ATYVT...NA.....AK...SNSEK.....L....S
SapSGD GMNHY....TANYIK...HK.....TD.....TPP....E.....
SinSGD GMNYY.TTYWAAYKP..TP.....PG.T.....P
TelSGD GINYY.T...ATYVT...DA.....VKSTSE.....R....L....D
UtoSGD GMNYY....TTNSVA...NL.....P.YS.....RS....I.....
VmiSGD1 GMNYY.T...ATYVT...NA.....VE.PISQ.....P....L....N
VmiSGD3 GINYY.T...ATYVT...DS.....LK.....P....S
VunSGD GLNYY.S...TTYSS...DA.....P.QL.....S..NAN....P....S

Fig. 12 J

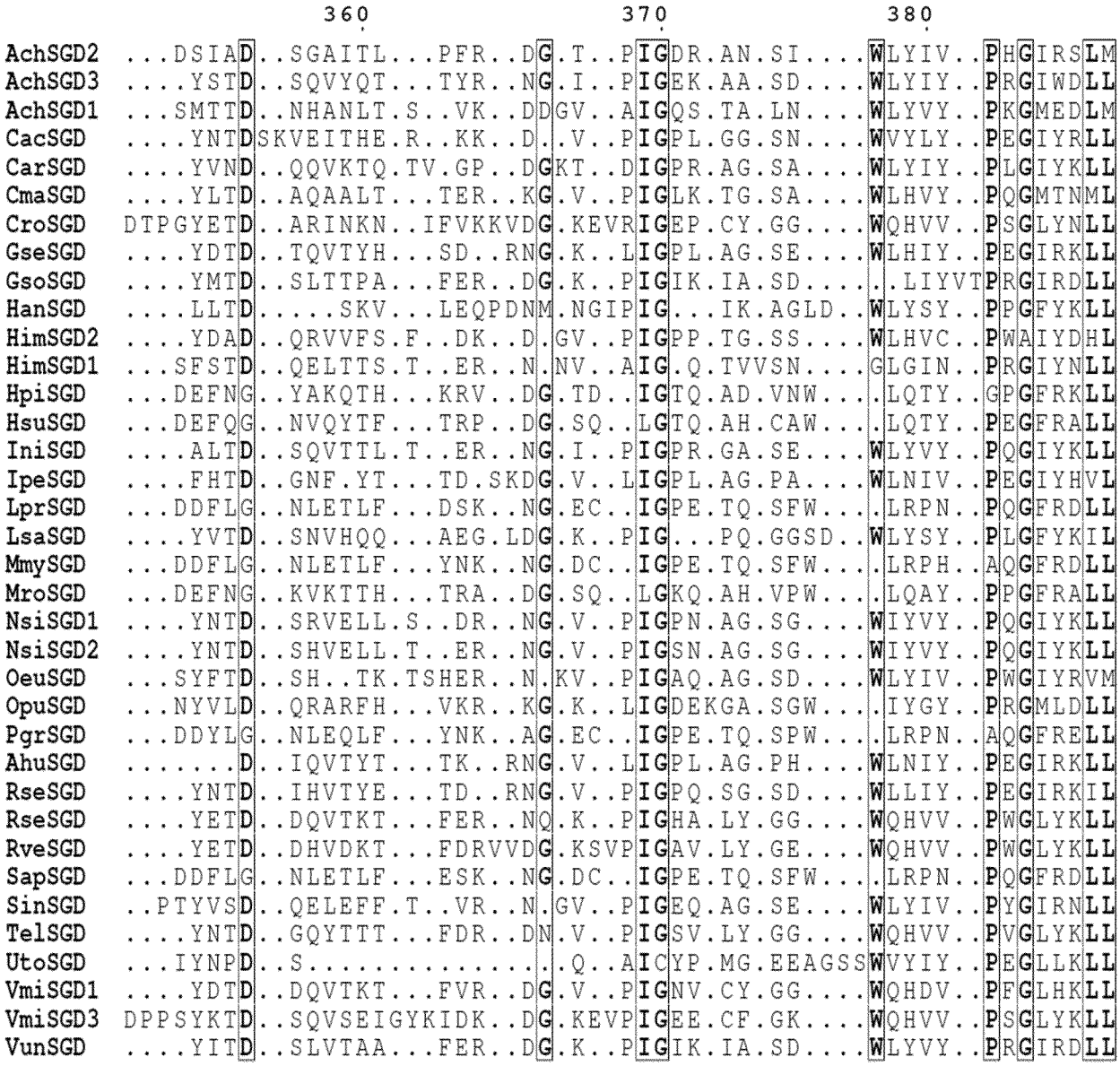
360
370
380
AchSGD2 ...DSIAD..SGAITL...PFR..DG.T..PIGDR.AN.SI....WLYIV..PHGIRSLM
AchSGD3 ....YSTD..SQVYQT...TYR..NG.I..PIGEK.AA.SD....WLYIY..PRGIWDLL
AchSGD1 ...SMTTD..NHANLT.S..VK..DDGV..AIGQS.TA.LN....WLYVY..PKGMEDLM
CacSGD ....YNTDSKVEITHE.R..KK..D..V..PIGPL.GG.SN....WVYLY..PEGIYRLL
CarSGD ....YVND..QQVKTQ.TV.GP..DGKT..DIGPR.AG.SA....WLYIY..PLGIYKLL
CmaSGD ....YLTD..AQAALT...TER..KG.V..PIGLK.TG.SA....WLHVY..PQGMTNML
CroSGD DTPGYETD..ARINKN..IFVKKVDG.KEVRIGEP.CY.GG....WQHVV..PSGLYNLL
GseSGD ....YDTD..TQVTYH...SD..RNG.K..LIGPL.AG.SE....WLHIY..PEGIRKLL
GsoSGD ....YMTD..SLTTPA...FER..DG.K..PIGIK.IA.SD.....LIYVTPRGIRDLL
HanSGD ....LLTD.....SKV...LEQPDNM.NGIPIG...IK.AGLD..WLYSY..PPGFYKLL
HimSGD2 ....YDAD..QRVVFS.F..DK..D.GV..PIGPP.TG.SS....WLHVC..PWAIYDHL
HimSGD1 ...SFSTD..QELTTS.T..ER..N.NV..AIG.Q.TVVSN....GLGIN..PRGIYNLL
HpiSGD ...DEFNG..YAKQTH...KRV..DG.TD..IGTQ.AD.VNW.....LQTY..GPGFRKLL
HsuSGD ...DEFQG..NVQYTF...TRP..DG.SQ..LGTQ.AH.CAW.....LQTY..PEGFRALL
IniSGD ....ALTD..SQVTTL.T..ER..NG.I..PIGPR.GA.SE....WLYVY..PQGIYKLL
IpeSGD ....FHTD..GNF.YT...TD.SKDG.V..LIGPL.AG.PA....WLNIV..PEGIYHVL
LprSGD ...DDFLG..NLETLF...DSK..NG.EC..IGPE.TQ.SFW.....LRPN..PQGFRDLL
LsaSGD ....YVTD..SNVHQQ...AEG.LDG.K..PIG...PQ.GGSD..WLYSY..PLGFYKIL
MmySGD ...DDFLG..NLETLF...YNK..NG.DC..IGPE.TQ.SFW.....LRPH..AQGFRDLL
MroSGD ...DEFNG..KVKTTH...TRA..DG.SQ..LGKQ.AH.VPW.....LQAY..PPGFRALL
NsiSGD1 ....YNTD..SRVELL.S..DR..NG.V..PIGPN.AG.SG....WIYVY..PQGIYKLL
NsiSGD2 ....YNTD..SHVELL.T..ER..NG.V..PIGSN.AG.SG....WIYVY..PQGIYKLL
OeuSGD ...SYFTD..SH..TK.TSHER..N.KV..PIGAQ.AG.SD....WLYIV..PWGIYRVM
OpuSGD ...NYVLD..QRARFH...VKR..KG.K..LIGDEKGA.SGW.....IYGY..PRGMLDLL
PgrSGD ...DDYLG..NLEQLF...YNK..AG.EC..IGPE.TQ.SPW.....LRPN..AQGFRELL
AhuSGD .......D..IQVTYT...TK..RNG.V..LIGPL.AG.PH....WLNIY..PEGIRKLL
RseSGD ....YNTD..IHVTYE...TD..RNG.V..PIGPQ.SG.SD....WLLIY..PEGIRKIL
RseSGD ....YETD..DQVTKT...FER..NQ.K..PIGHA.LY.GG....WQHVV..PWGLYKLL
RveSGD ....YETD..DHVDKT...FDRVVDG.KSVPIGAV.LY.GE....WQHVV..PWGLYKLL
SapSGD ...DDFLG..NLETLF...ESK..NG.DC..IGPE.TQ.SFW.....LRPN..PQGFRDLL
SinSGD ..PTYVSD..QELEFF.T..VR..N.GV..PIGEQ.AG.SE....WLYIV..PYGIRNLL
TelSGD ....YNTD..GQYTTT...FDR..DN.V..PIGSV.LY.GG....WQHVV..PVGLYKLL
UtoSGD ...IYNPD..S.............Q..AICYP.MG.EEAGSSWVYIY..PEGLLKLL
VmiSGD1 ....YDTD..DQVTKT...FVR..DG.V..PIGNV.CY.GG....WQHDV..PFGLHKLL
VmiSGD3 DPPSYKTD..SQVSEIGYKIDK..DG.KEVPIGEE.CF.GK....WQHVV..PSGLYKLL
VunSGD ....YITD..SLVTAA...FER..DG.K..PIGIK.IA.SD....WLYVY..PRGIRDLL

Fig. 12 K

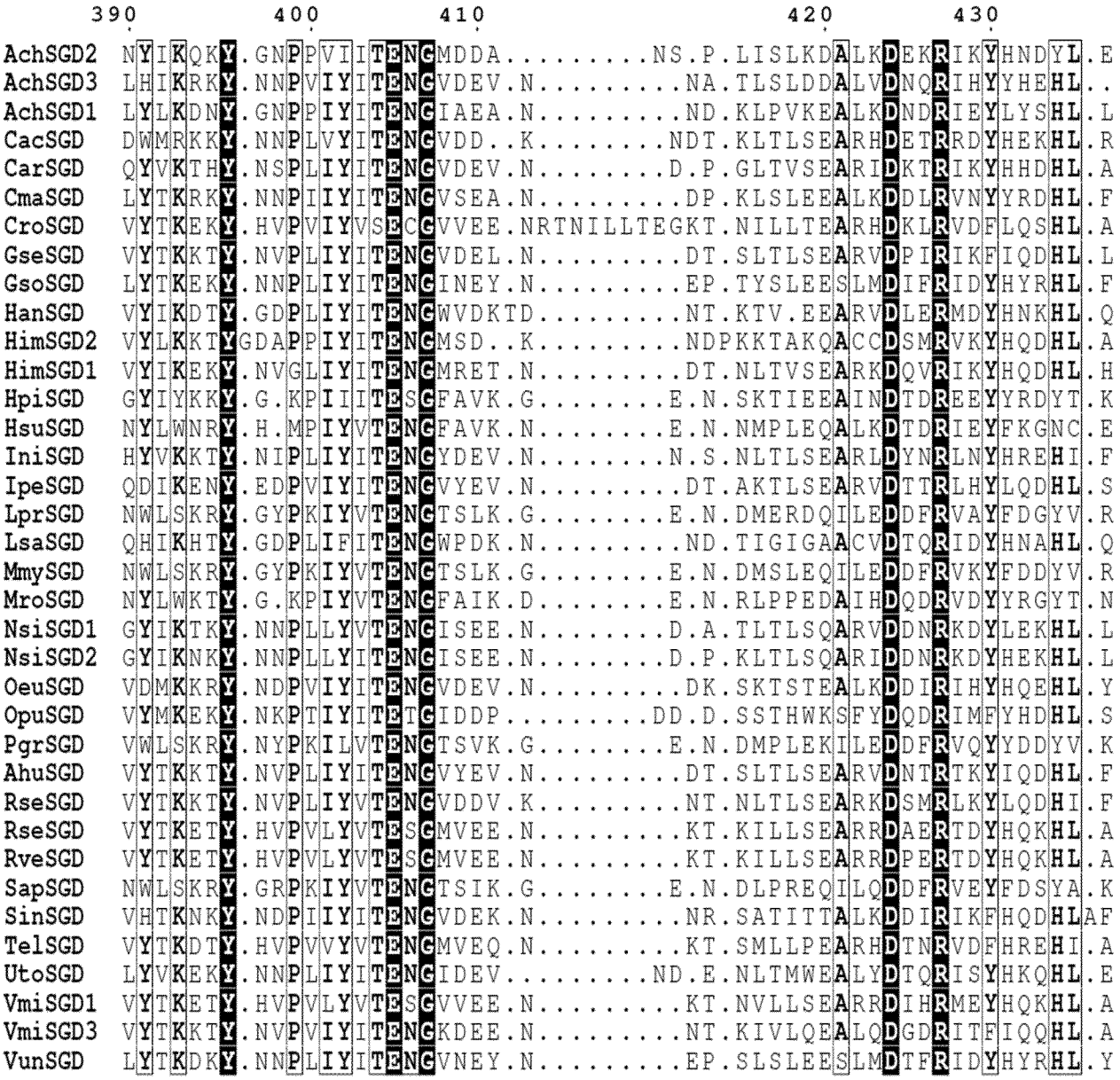
390
400
410
420
430
AchSGD2 NYIKQKY.GNPPVITTENGMDDA.....NS.P.LISLKDALKDEKRIKYHNDYL.E
AchSGD3 LHIKRKY.NNPVIYITTENGVDEV.N.....NA.TLSLDDALVDNQRIHYYHEHL..
AchSGD1 LYLKDNY.GNPPIYITTENGIAEA.N.....ND.KLPVKEALKDNDRIEYLYSHL.L
CacSGD DWMRKKY.NNPLVYITTENGVDD..K.....NDT.KLTLSEARHDETRRDYHEKHL.R
CarSGD QYVKTHY.NSPLIYITTENGVDEV.N.....D.P.GLTVSEARIDKTRIKYHHDHL.A
CmaSGD LYTKRKY.NNPIIYITTENGVSEA.N.....DP.KLSLEEALKDDLRVNNYYRDHL.F
CroSGD VYTKEKY.HVPVIYVSECGVVEE.NRTNILLTEGKT.NILLTEARHDKLRVDFLQSHL.A
GseSGD VYTKKTY.NVPLIYITTENGVDEL.N.....DT.SLTLSEARVDPIRIKFIQDHL.L
GsoSGD LYTKEKY.NNPLIYITTENGINEY.N.....EP.TYSLEESLMDIFRIDYHYRHL.F
HanSGD VYIKDTY.GDPLIYITTENGWVDKTD.....NT.KTV.EEARVDLERMDYHNKHL.Q
HimSGD2 VYLKKTYGDAPPIYITTENGMSD..K.....NDPKKTAKQACCDSMRVKYHQDHL.A
HimSGD1 VYIKEKY.NVGLIYITTENGMRET.N.....DT.NLTVSEARKDQVRIKYHQDHL.H
HpiSGD GYIYKKY.G.KPIIITTESGFAVK.G.....E.N.SKTIEEAINDTDREEYYRDYT.K
HsuSGD NYLWNRY.H.MPIYVTTENGFAVK.N.....E.N.NMPLEQALKDTDRIEYFKGNC.E
IniSGD HYVKKTY.NIPLIYITTENGYDEV.N.....N.S.NLTLSEARLDYNRLNYHREHI.F
IpeSGD QDIKENY.EDPVIYITTENGVYEV.N.....DT.AKTLSEARVDTTRLHYLQDHL.S
LprSGD NWLSKRY.GYPKIYVTTENGTSLK.G.....E.N.DMERDQILEDDFRVAYFDGYV.R
LsaSGD QHIKHTY.GDPLIFITTENGWPDK.N.....ND.TIGIGAACVDTQRIDYHNAHL.Q
MmySGD NWLSKRY.GYPKIYVTTENGTSLK.G.....E.N.DMSLEQILEDDFRVKYFDDYV.R
MroSGD NYLWKTY.G.KPIYVTTENGFAIK.D.....E.N.RLPPEDAIHDQDRVDYYRGYT.N
NsiSGD1 GYIKTKY.NNPLLYVTTENGISEE.N.....D.A.TLTLSQARVDDNRKDYLEKHL.L
NsiSGD2 GYIKNKY.NNPLLYITTENGISEE.N.....D.P.KLTLSQARIDDNRKDYHEKHL.L
OeuSGD VDMKKRY.NDPVIYITTENGVDEV.N.....DK.SKTSTEALKDDIRIHYHQEHL.Y
OpuSGD VYMKEKY.NKPTIYITTETGIDDP.....DD.D.SSTHWKSFYDQDRIMFYHDHL.S
PgrSGD VWLSKRY.NYPKILVTTENGTSVK.G.....E.N.DMPLEKILEDDFRVQYYDDYV.K
AhuSGD VYTKKTY.NVPLIYITTENGVYEV.N.....DT.SLTLSEARVDNTRTKYIQDHL.F
RseSGD VYTKKTY.NVPLIYVTTENGVDDV.K.....NT.NLTLSEARKDSMRLKYLQDHI.F
RseSGD VYTKETY.HVPVLYVTTESGMVEE.N.....KT.KILLSEARRDAERTDYHQKHL.A
RveSGD VYTKETY.HVPVLYVTTESGMVEE.N.....KT.KILLSEARRDPERTDYHQKHL.A
SapSGD NWLSKRY.GRPKIYVTTENGTSIK.G.....E.N.DLPREQILQDDFRVEYFDSYA.K
SinSGD VHTKNKY.NDPIIYITTENGVDEK.N.....NR.SATITTALKDDIRIKFHQDHLAF
TelSGD VYTKDTY.HVPVVYVTTENGMVEQ.N.....KT.SMLLPEARHDTNRVDFHREHI.A
UtoSGD LYVKEKY.NNPLIYITTENGIDEV.....ND.E.NLTMWEALYDTQRISYHKQHL.E
VmiSGD1 VYTKETY.HVPVLYVTTESGVVEE.N.....KT.NVLLSEARRDIHRMEYHQKHL.A
VmiSGD3 VYTKKTY.NVPVIYITTENGKDEE.N.....NT.KIVLQEALQDGDRITFIQQHL.A
VunSGD LYTKDKY.NNPLIYITTENGVNEY.N.....EP.SLSLEESLMDTFRIDYHYRHL.Y

Fig. 12 L

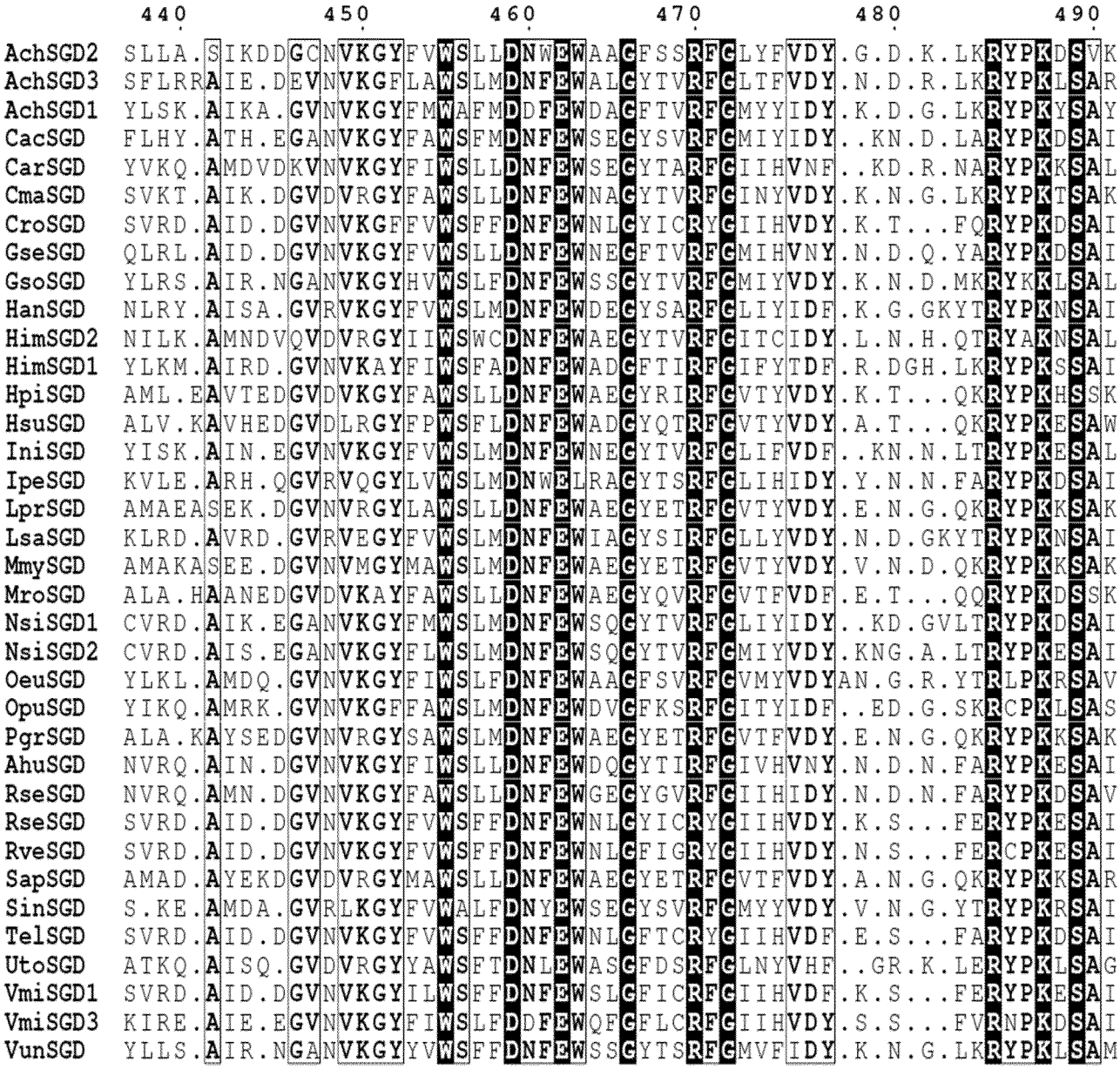
440
450
460
470
480
490
AchSGD2 SLLA.SIKDDGCNVKGYFVWSLLDNWEWAAGFSSRFGLYFVDY.G.D.K.LKRYPKDSVK
AchSGD3 SFLRRAIE.DEVNVKGFLAWSLMDNFEWALGYTVRFGLTFVDY.N.D.R.LKRYPKLSAR
AchSGD1 YLSK.AIKA.GVNVKGYFMWAFMDDFEWDAGFTVRFGMYYIDY.K.D.G.LKRYPKYSAY
CacSGD FLHY.ATH.EGANVKGYFAWSFMDNFEWSEGYSVRFGMIYIDY..KN.D.LARYPKDSAI
CarSGD YVKQ.AMDVDKVNVKGYFIWSLLDNFEWSEGYTARFGIIHVNF..KD.R.NARYPKKSAL
CmaSGD SVKT.AIK.DGVDVRGYFAWSLLDNFEWNAGYTVRFGINYVDY.K.N.G.LKRYPKTSAK
CroSGD SVRD.AID.DGVNVKGFFVWSFFDNFEWNLGYICRYGIIHVDY.K.T...FQRYPKDSAI
GseSGD QLRL.AID.DGVNVKGYFVWSLLDNFEWNEGFTVRFGMIHVNY.N.D.Q.YARYPKDSAI
GsoSGD YLRS.AIR.NGANVKGYHVWSLFDNFEWSSGYTVRFGMIYVDY.K.N.D.MKRYKKLSAL
HanSGD NLRY.AISA.GVRVKGYFVWSLMDNFEWDEGYSARFGLIYIDF.K.G.GKYTRYPKNSAI
HimSGD2 NILK.AMNDVQVDVRGYIIWSWCDNFEWAEGYTVRFGITCIDY.L.N.H.QTRYAKNSAL
HimSGD1 YLKM.AIRD.GVNVKAYFIWSFADNFEWADGFTIRFGIFYTDF.R.DGH.LKRYPKSSAI
HpiSGD AML.EAVTEDGVDVKGYFAWSLLDNFEWAEGYRIRFGVTYVDY.K.T...QKRYPKHSSK
HsuSGD ALV.KAVHEDGVDLRGYFPWSFLDNFEWADGYQTRFGVTYVDY.A.T...QKRYPKESAW
IniSGD YISK.AIN.EGVNVKGYFVWSLMDNFEWNEGYTVRFGLIFVDF..KN.N.LTRYPKESAL
IpeSGD KVLE.ARH.QGVRVQGYLVWSLMDNWELRAGYTSRFGLIHIDY.Y.N.N.FARYPKDSAI
LprSGD AMAEASEK.DGVNVRGYLAWSLLDNFEWAEGYETRFGVTYVDY.E.N.G.QKRYPKKSAK
LsaSGD KLRD.AVRD.GVRVEGYFVWSLMDNFEWIAGYSIRFGLLYVDY.N.D.GKYTRYPKNSAI
MmySGD AMAKASEE.DGVNVMGYMAWSLMDNFEWAEGYETRFGVTYVDY.V.N.D.QKRYPKKSAK
MroSGD ALA.HAANEDGVDVKAYFAWSLLDNFEWAEGYQVRFGVTFVDF.E.T...QQRYPKDSSK
NsiSGD1 CVRD.AIK.EGANVKGYFMWSLMDNFEWSQGYTVRFGLIYIDY..KD.GVLTRYPKDSAI
NsiSGD2 CVRD.AIS.EGANVKGYFLWSLMDNFEWSQGYTVRFGMIYVDY.KNG.A.LTRYPKESAI
OeuSGD YLKL.AMDQ.GVNVKGYFIWSLFDNFEWAAGFSVRFGVMYVDYAN.G.R.YTRLPKRSAV
OpuSGD YIKQ.AMRK.GVNVKGFFAWSLMDNFEWDVGFKSRFGITYIDF..ED.G.SKRCPKLSAS
PgrSGD ALA.KAYSEDGVNVRGYSAWSLMDNFEWAEGYETRFGVTFVDY.E.N.G.QKRYPKKSAK
AhuSGD NVRQ.AIN.DGVNVKGYFIWSLLDNFEWDQGYTIRFGIVHVNY.N.D.N.FARYPKESAI
RseSGD NVRQ.AMN.DGVNVKGYFAWSLLDNFEWGEGYGVRFGIIHIDY.N.D.N.FARYPKDSAV
RseSGD SVRD.AID.DGVNVKGYFVWSFFDNFEWNLGYICRYGIIHVDY.K.S...FERYPKESAI
RveSGD SVRD.AID.DGVNVKGYFVWSFFDNFEWNLGFIGRYGIIHVDY.N.S...FERCPKESAI
SapSGD AMAD.AYEKDGVDVRGYMAWSLLDNFEWAEGYETRFGVTFVDY.A.N.G.QKRYPKKSAR
SinSGD S.KE.AMDA.GVRLKGYFVWALFDNYEWSEGYSVRFGMYYVDY.V.N.G.YTRYPKRSAI
TelSGD SVRD.AID.DGVNVKGYFVWSFFDNFEWNLGFTCRYGIIHVDF.E.S...FARYPKDSAI
UtoSGD ATKQ.AISQ.GVDVRGYYAWSFTDNLEWASGFDSRFGLNYVHF..GR.K.LERYPKLSAG
VmiSGD1 SVRD.AID.DGVNVKGYILWSFFDNFEWSLGFICRFGIIHVDF.K.S...FERYPKESAI
VmiSGD3 KIRE.AIE.EGVNVKGYFIWSLFDDFEWQFGFLCRFGIIHVDY.S.S...FVRNPKDSAI
VunSGD YLLS.AIR.NGANVKGYYVWSFFDNFEWSSGYTSRFGMVFIDY.K.N.G.LKRYPKLSAM

Fig. 12 M

```
AchSGD2  W........FKNF.L.T.....................................
AchSGD3  W........YKNF.L.Q.........K...........................
AchSGD1  W........YKKF..LQ.........T...........................
CacSGD   W........YKN..F.....................LTKTEKTKKRQLDHKELDNIPQKK
CarSGD   W........FMNF.L.A.........K......SNLSPTKTTKRALDNGGLADLENPKKK
CmaSGD   W........FKSF.L.H.........K...........................
CroSGD   W........YKNF.I.S.....................................EGFV
GseSGD   W........LMNN.F.H.........K..........KFSGPPVKRSVEENQETDSRKRS
GsoSGD   W........FKNF.L.K.........K..................ESRLYGTSK.........
HanSGD   W........YKHF..LGYSNKQKTEKKKNLARERTCKSSEKTTKFELELENNCYCLDLLS
HimSGD2  W........FC..........................................
HimSGD1  WWTRFLNNKLMKSGSFK.........R.....................LTQNQCEDDTDS
HpiSGD   F........LKEW.F.A.....................................
HsuSGD   F........LVNW.F.K.....................................
IniSGD   W........FANF.L.A.........K.............................EDE...
IpeSGD   W........FRNA.F.H.........K................RLRIHVNKARPQEDDGA
LprSGD   S........LKPL.F.D.....................................
LsaSGD   W........YMNF..LK.SPKKLGEQKKIPKCVPNKPIAKTQSTETSTKTSRVLAEVVLI
MmySGD   S........LKAL.F.D.....................................
MroSGD   F........LAEW.Y.R.....................................
NsiSGD1  W........FMN..F.L.........K...NVIPTSRKRPLPSASPAKPAKKR.......
NsiSGD2  W........FMN..F.M.........K.......NAIPNSRKRPLPSASSAKPSKKR...
OeuSGD   W.....................................................WRNFLTKPTAVP
OpuSGD   W........FKYF.........................L.......EN.............
PgrSGD   A........MKPL.F.D.....................................
AhuSGD   W........LMNS.F.N.........K.....................KHSKIPVKRSI
RseSGD   W........LMNS.F.H.........K..............NISKLPAVKRSIREDDEEQ
RseSGD   W........YKNF.I.A.........G.....K....STTSPAKRRREEAQVELVKRQKT
RveSGD   W........YKNF.I.A.........G............VSTTSPAKRRREEAEGVELVK
SapSGD   S........LKPL.F.D.....................................
SinSGD   W........FMNF.L.N.........K............NILPRPKRQIEEIEDDNASAKR
TelSGD   W........YKNF.I.Y.........G.....KSLTLPVKRPRDEDREVELVKRQKKREL
UtoSGD   W........FKFF.........................LENGKSASFCWSIIGNNICLNKR
VmiSGD1  W........YKNF.I.A.........G.....K.STTLPLKRRRLEAQEVESVKMQKV..
VmiSGD3  W........YKNF.I.G.........G.....K.SPTSPPKRPREEEAAGIVEVVKKRRT
VunSGD   W........YKNF.L.K.........K..............ETRLYASSK.........
```

Fig. 12 N

```
                      500
AchSGD2  ..............SA..................................
AchSGD3  ..............SA..................................
AchSGD1  ..............SA..................................
CacSGD   ..............SA..................................
CarSGD   ILKT..........SA..................................
CmaSGD   ..............SA..................................
CroSGD   TNTAKKRFR.....SA..EEDKLVELVKKQKY..................
GseSGD   RK............SA..................................
GsoSGD   ..............SA..................................
HanSGD   FLLPRINMK.....SA..VNYKFGGVKLKDEQR.................
HimSGD2  KFLKSKKSQIQSSNK.....SA..RQIENNSENVLAKRYKV.........
HimSGD1  QKK...........SA..................................
HpiSGD   ..............A.....SA..HI........................
HsuSGD   ..............E.....SA..NVNSPKSSGEPRTSRIPNGAVPNGHI
IniSGD   ..............SA..................................
IpeSGD   FDTPRKRLR.....SA..KY..............................
LprSGD   ..............S.....SA..LI....KTD.................
LsaSGD   MILSILCIV.....SA..MFIFDYKMKIGCIY..................
MmySGD   ..............S.....SA..LI....KKD.................
MroSGD   ..............S.....SA..SLAK......................
NsiSGD1  ..............SA..................................
NsiSGD2  ..............SA..................................
OeuSGD   LK.....NEPEKSEDRRKRLRGSTSA..................... 
OpuSGD   ..............SA..................................
PgrSGD   ..............S.....SA..LI....EKD.................
AhuSGD   QDEDQEQVS.....SA..NKKSRK........................
RseSGD   VSSKRLRK........SA..............................
RseSGD   ..............SA..................................
RveSGD   RQKT..........SA..................................
SapSGD   ..............S.....SA..LI....KKD.................
SinSGD   KK.....GR.....SA..................................
TelSGD   RRKIMKK.......SA..................................
UtoSGD   SRCTLVDCR.....SA..IYILLVIRIYVC....................
VmiSGD1  ..............SA..................................
VmiSGD3  ..............SA..................................
VunSGD   ..............SA..................................
```

Fig. 12 O

		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
AchSGD2	1																		
HimSGD2	2	33,4																	
AchSGD3	3	42,1	37,5																
HanSGD	4	38,1	35,4	38,3															
LsaSGD	5	37,7	36,4	36,8	57,4														
IpeSGD	6	36,9	36,8	38,1	40,1	40,1													
HimSGD1	7	39,9	38,4	44,8	43,9	40,9	40,6												
CacSGD	8	40,3	37,8	41,1	41,7	39,6	41,9	46,3											
AchSGD1	9	46,3	36,6	48,1	40,5	41,8	43,6	42,9	44,2										
NsiSGD1	10	40,3	37,7	43,9	43,5	43,3	43,9	45,2	53,1	52,0									
NsiSGD2	11	41,2	38,7	43,7	44,6	44,0	43,3	46,6	55,7	51,2	78,4								
CmaSGD	12	47,5	38,7	50,2	40,7	40,0	43,2	47,1	46,9	61,2	50,2	51,3							
GsoSGD	13	45,5	36,7	46,1	42,0	42,7	43,8	44,9	46,8	58,6	49,3	51,4	60,4						
VunSGD	14	45,7	38,4	49,9	41,8	43,0	44,7	45,4	47,8	60,3	50,1	51,7	60,9	78,2					
IniSGD	15	40,9	40,4	41,8	41,4	41,7	41,3	47,4	49,8	48,3	52,9	53,7	49,2	48,0	48,3				
CarSGD	16	39,5	38,2	42,3	42,7	43,5	43,6	47,7	50,5	47,3	52,7	54,3	47,3	48,1	48,0	53,0			
GseSGD	17	40,9	38,2	43,4	44,9	44,5	48,8	49,7	52,9	48,8	52,8	54,2	49,5	49,8	50,7	52,0	56,1		
AhuSGD	18	39,6	39,6	41,5	43,4	42,6	49,6	49,1	47,2	45,3	48,9	49,6	47,8	46,4	46,9	52,0	52,8	63,9	
RseSGD2	19	39,2	41,2	42,8	43,7	41,9	49,7	48,5	51,4	48,2	51,9	51,3	49,6	48,4	49,4	52,3	54,2	66,1	65,1
CroSGD	20	37,2	35,7	39,0	41,2	39,3	41,9	39,9	42,9	42,2	42,4	42,1	45,4	44,1	45,5	42,3	42,3	45,3	46,8
RseSGD	21	38,7	38,3	41,4	43,8	42,1	42,9	43,3	48,7	45,1	48,0	49,2	48,7	47,5	49,9	47,5	47,3	51,8	50,4
RveSGD	22	39,0	38,1	40,7	42,7	41,7	43,1	42,7	46,6	44,1	46,3	47,3	47,1	45,5	47,8	47,0	48,0	51,7	50,5
VmiSGD1	23	39,4	36,4	41,1	43,4	41,4	43,0	43,3	47,4	45,1	48,6	49,7	48,0	47,5	48,9	46,8	46,5	51,0	50,4
TelSGD	24	39,3	37,8	41,5	43,3	41,6	45,0	42,9	47,6	44,9	47,7	48,5	48,7	46,5	48,3	47,8	48,5	51,9	51,1
VmiSGD3	25	38,5	37,3	43,7	41,8	42,7	43,4	41,2	45,1	46,6	46,9	47,7	46,1	46,8	48,0	46,6	47,8	52,6	48,7
OeuSGD	26	37,5	38,0	39,0	40,3	38,5	37,8	45,2	43,1	41,1	39,2	41,3	40,5	40,3	39,6	42,2	41,8	44,8	43,8
SinSGD	27	40,1	38,5	41,4	41,1	40,2	39,7	48,3	45,6	43,2	43,8	46,8	45,6	45,5	44,5	45,2	48,0	49,6	44,4
UtoSGD	28	34,8	31,5	36,8	34,7	35,3	34,9	37,6	38,4	38,8	39,3	40,4	40,3	38,5	39,6	40,2	39,4	39,4	38,5
OpuSGD	29	33,9	31,6	34,5	32,6	33,0	31,5	36,4	34,9	38,0	36,1	35,6	38,1	37,2	37,0	37,0	38,3	37,6	36,2
HsuSGD	30	37,1	29,1	37,0	32,9	32,6	32,9	34,4	32,7	34,7	34,4	34,5	38,6	35,8	36,4	33,9	34,6	36,0	34,6
HpiSGD	31	40,0	30,0	38,9	36,1	35,3	33,1	36,7	35,4	38,2	37,3	37,3	40,6	38,0	38,3	36,5	38,2	39,2	37,3
MroSGD	32	40,7	30,3	37,0	35,9	35,2	33,3	35,8	36,5	38,5	37,7	37,3	40,9	38,3	37,0	39,0	37,2	38,8	37,5
MmySGD	33	39,6	31,5	34,2	34,1	32,0	32,8	34,5	34,8	38,2	36,6	36,7	41,3	38,2	37,5	35,4	35,2	37,3	35,4
LprSGD	34	38,6	31,3	34,8	33,5	32,9	31,5	34,2	33,6	36,3	35,5	35,1	40,1	36,0	35,2	33,8	34,2	35,6	33,9
SapSGD	35	37,7	31,4	35,2	33,2	32,8	31,0	34,1	33,4	36,4	35,5	35,4	39,7	35,5	35,1	33,8	34,4	35,4	33,7
PgrSGD	36	37,3	30,6	34,4	32,9	31,8	32,2	33,1	33,2	35,9	34,9	35,0	40,8	36,7	36,1	34,7	34,0	35,7	33,8

Fig. 13 A

19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36
45,3																	
49,9	65,2																
50,2	65,7	84,8															
50,4	62,1	74,6	73,5														
50,9	60,3	70,9	69,9	70,5													
50,2	56,9	64,2	63,0	64,3	62,4												
42,5	38,0	40,3	40,8	40,6	40,5	40,6											
47,1	39,6	43,1	41,7	42,5	42,8	41,9	46,4										
38,9	35,7	36,2	36,0	36,9	35,9	35,7	35,1	36,7									
36,5	30,3	32,6	32,3	32,5	31,8	34,4	32,0	34,7	37,0								
36,0	33,1	33,1	33,2	33,7	33,0	33,3	31,8	35,3	33,0	28,9							
37,4	35,0	35,6	35,0	36,7	35,9	35,7	33,8	34,4	33,2	32,8	55,1						
39,2	34,4	34,5	34,5	36,0	36,2	35,6	33,2	33,7	32,4	32,8	58,5	61,4					
36,5	32,7	34,3	34,5	35,0	34,5	34,8	31,4	34,9	31,3	32,5	51,1	51,2	49,2				
34,7	32,5	32,7	32,0	32,9	32,4	33,5	29,6	34,5	30,1	31,6	50,5	48,8	49,2	85,7			
34,9	33,0	33,0	32,6	33,6	32,2	33,9	29,7	34,1	30,5	31,8	50,9	49,4	50,4	83,0	92,5		
34,1	32,6	32,8	32,3	33,2	32,5	33,6	30,7	33,6	31,3	31,9	51,4	49,3	49,3	81,6	80,9	80,1	

Fig. 13 B

INTERNATIONAL SEARCH REPORT

International application No
PCT/EP2020/063283

A. CLASSIFICATION OF SUBJECT MATTER
INV. C12P19/44 C12N9/42 C12P17/18
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 C12N

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, 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 A	WO 2017/152273 A1 (VINDOLON INC [CA]) 14 September 2017 (2017-09-14) paragraph [00034]; claim 95; sequence 1 --- WO 00/42200 A1 (UNIV LEIDEN [NL]; VERPOORTE ROBERT [NL] ET AL.) 20 July 2000 (2000-07-20) cited in the application page 18, line 19 - page 20, line 12; claims 1-19 --- -/--	1,2,18, 20-22 3-17,19, 23-26 1-26

[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
1 July 2020	10/07/2020

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	Stoyanov, Borislav

1

INTERNATIONAL SEARCH REPORT

International application No
PCT/EP2020/063283

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	STEPHANIE BROWN ET AL: "De novo production of the plant-derived alkaloid strictosidine in yeast", PNAS, vol. 112, no. 11, 9 February 2015 (2015-02-09), pages 3205-3210, XP055556409, US ISSN: 0027-8424, DOI: 10.1073/pnas.1423555112 figure 2 -----	1-26

1

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No
PCT/EP2020/063283

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 2017152273 A1	14-09-2017	AU 2017228939 A1	25-10-2018
		CA 3016339 A1	14-09-2017
		EP 3423586 A1	09-01-2019
		US 2020002339 A1	02-01-2020
		WO 2017152273 A1	14-09-2017
WO 0042200 A1	20-07-2000	AU 1696000 A	01-08-2000
		WO 0042200 A1	20-07-2000