Article

Unveiling the Bacterial Sesquiterpenome of *Streptomyces* sp. CBMAI 2042 Discloses Cyclases with Versatile Performances

Douglas C. Sachito [a] and Luciana G. de Oliveira *,[a]

aInstituto de Química, Universidade Estatual de Campinas, CP 6154, 13083-970 Campinas-SP, Brazil

Terpenes are the most abundant class of natural product that exist in nature. They possess a myriad of industrial applications including pharmaceutical, perfumery and flavors, bulk chemicals, and fuel. Intriguingly, until today, the vast majority of characterized terpenoids have been isolated from plants and fungi, and only in recent years bacteria were found to generate a representative reservoir of terpenoids molecules. Mining *Streptomyces* sp. CBMAI 2042 genome data has revealed the presence of five terpene cyclase genes. Chemical analysis of mycelium extract of this bacteria strain has unveiled at least 28 volatile terpenes molecules, where three encoding sesquiterpene cyclase (STC) genes are apparently responsible for their biosynthesis. The cyclic products obtained by incubation of these three purified recombinant STCs with farnesyl diphosphate (FPP) were analyzed by gas chromatography-mass spectrometry (GC-MS) and identified using the Van den Dool and Kratz equation.

Keywords: bacterial terpenome, genome mining, terpene cyclase, sesquiterpenes

Introduction

Due to the great chemical and structural diversity, studies involving specialized metabolites (SMs) are of great relevance and help humanity understand the dynamics of nature with its surroundings. In addition, development of new drugs inspired by SMs to control, mainly, human infectious diseases resulted in a better quality of life for society. Currently, research in natural products continues to be an abundant source for the discovery of new chemical entities with a great chemical diversity and industrial applications.[1,2]

Among the SMs, terpene metabolites are the most abundant. Terpene SMs have a range of applications and are part of society's daily life. About 80,000 terpenoids have been identified and cataloged in the natural products dictionary to date.[3] These metabolites have a myriad of applications including pharmaceutical (taxol and artemisinin), perfumery and flavors (valencene, menthol and linalool), bulk chemicals (isoprene) and fuels (farnesene, bisabolane).[4,5]

Despite the variety of applications and high demand for terpene products, there is a huge difficulty in accessing this class of molecules through chemical synthesis. The production of these molecules requires numerous synthetic steps and, consequently, increased production costs.[6] Furthermore, the extraction of terpene compounds from natural sources is not only laborious, but also results in low yields and excessive consumption of valuable resources. In this way, to overcome this problem, approaches involving biotechnological processes have been making progress in the last two decades and are proving to be a more sustainable and profitable alternative in the production of terpenes with industrial applications.[4]

Terpene molecules are derived from C_5 precursor molecules isopentenyl diphosphate (IDP) and dimethylallyl diphosphate (DMADP) which are assembled by prenyltransferase (PTs) enzymes to produce elongated polyisoprene. Enzymes known as terpene cyclases (TCs) convert these linear oligoprenyl diphosphate, via cationic intermediates, into a multitude of (poly)cyclic hydrocarbon molecules.[7,8]

Interestingly, only in recent years bacteria have been learnt to be a primary terpene producers, and therefore the vast majority of terpenoids characterized to date are derived from plants and fungi.[7,9,10] This breakthrough is due to the extensive mining of the bacterial genome in the last decade, showing that genes encoding terpene cyclase are abundant among these organisms, including the family Streptomycetaceae.[11]

*e-mail: lucianag@unicamp.br

Yet, studies involving bacterial terpene cyclases (BTCs) have proven to be an even greater challenge. Amino acid sequences of these enzymes do not show general similarity with TC enzymes of plant or fungi origin, in addition to low amino acid sequence identity and similarity with other BTCs. To investigate cryptic TCs in the bacterial genome, Yamada *et al.*[9] applied an alternative genome mining procedure known as HMMs (hidden Markov models),[12] which, in addition to aligning conserved regions in proteins, also perform searches in the Pfam database (protein family database)[13] making possible to correlate a tertiary structure with characterized BTCs.[14] This approach allowed to expand from 140 previously known BTCs to 262, permitting to enlarge the knowledge of this class of enzymes.[9,15]

Certainly, unraveling the bacterial terpenome can be a huge challenge. Therefore, genome mining approaches can guide the search for terpene compounds, nonetheless, so far, it is not possible to predict only from genome information the chemical nature of the (poly)cyclic terpenoid.

In this genome-guided study, three recombinant sesquiterpene cyclase (STC) enzymes were produced in *E. coli* Rosetta(DE3), purified, and incubated with farnesyl pyrophosphate (FPP). Enzyme-substrate assays revealed that all three recombinant enzymes are active, and their incubation products are cyclic sesquiterpenes. In addition, the analysis of mycelium oil extract from *Streptomyces* sp. CBMAI 2042, unveiled the presence of at least 28 cyclic sesquiterpenes. In parallel, to expand the knowledge of the biosynthetic machinery of this strain, the study of the sesquiterpenome made it possible to characterize a distinct group of metabolites produced by non-linear pathways whose main enzymes (cyclases) are of broad biotechnological interest and may allow comprehensive addition for this important class of compounds in actinobacteria isolated in the Brazilian territory.[16-18]

Experimental

Reagents and solvents

Unless otherwise stated, chemicals and high-performance liquid chromatography (HPLC) solvents were purchased from Sigma-Aldrich (São Paulo, Brazil) and primers from Exxtend (Paulínia, Brazil). Type 1 ultrapure water was obtained via the Milli-Q purification system (Millipore, São Paulo, Brazil). Tryptone soya broth, yeast extract, peptone, meat extract, malt extract and agar were purchased from Oxoid (Hampshire, UK); sucrose, CaCO$_3$, Na$_2$SO$_4$, NaOH, maltose and NaCl were purchased from Synth (São Paulo, Brazil).

Bacterial strain and fermentation

The endophytic strain *Streptomyces* sp. CBMAI 2042 was isolated from *Citrus sinensis* branches.[19] For transcription analysis assay TSBY medium (30 g L^{-1} tryptone soya broth; 5 g L^{-1} yeast extract; 103 g L^{-1} sucrose; pH 7.1) was used. For the study of volatile terpene metabolites, dense spores (c. 10^9 spores mL^{-1}) of CBMAI 2042 were inoculated in vegetative medium (5 g L^{-1} peptone; 20 g L^{-1} corn starch; 2 g L^{-1} meat extract; 3 g L^{-1} yeast extract; 2 g L^{-1} soybean extract; 1 g L^{-1} CaCO$_3$; pH 7) with nalidixic acid (25 µg mL^{-1}) and cultivated for 2 days. Then, 1% of the seed culture was used to inoculate production medium TSBY or A medium (4 g L^{-1} peptone; 4 g L^{-1} meat extract; 2 g L^{-1} yeast extract; 2 g L^{-1} soybean extract; 20 g L^{-1} maltose; 10 g L^{-1} dextrin; pH 6.2) or ISP2 medium (4 g L^{-1} yeast extract; 10 g L^{-1} malt extract; 4 g L^{-1} dextrose, pH 6.1). Fermentation was carried out in 500 mL baffled flasks equipped with cotton plugs containing 100 mL of production medium and nalidixic acid (25 µg mL^{-1}), at 28 °C and 200 rpm for 5 days.

E. coli DH10B strain was used for maintenance and replication of the constructed plasmids. *E. coli* Rosetta(DE3) strain was used for production of recombinant terpene cyclase proteins. For the cultivation of any strain of *E. coli*, lysogeny broth (LB) medium was used (LB: 5 g L^{-1} yeast extract; 10 g L^{-1} tryptone; 10 g L^{-1} NaCl).

Transcription analysis assay

For transcription analysis, a pre-inoculum was made in 5 mL of TSBY medium containing spores of *Streptomyces* sp. CBMAI 2042 and the bacteriostatic nalidixic acid (25 µg mL^{-1}) and cultured for 2 days at 200 rpm at 28 °C. To the inoculum of 50 mL TSBY and nalidixic acid (25 µg mL^{-1}) 1% of the seed culture was added and fermentation conducted for 5 days at 200 rpm at 28 °C. Aliquots were collected at 48, 72, 96 and 120 h of fermentation times. tRNA (transfer ribonucleic acid) was isolated using the Rneasy Mini KIT QIAGEN (São Paulo, Brazil) according to the manufacturer's protocols. The tRNA was treated with deoxyribonuclease (DNAse) and then reacted with the reverse transcriptase enzyme from the commercial SuperScript III First-Strand Synthesis kit Invitrogen (São Paulo, Brazil). The resulting cDNA (complementary deoxyribonucleic acid) was used in polymerase chain reaction (PCR) and amplified with specific primers.

DNA extraction and manipulation

Genomic DNA of the strain *Streptomyces* sp. CBMAI

2042 was isolated according to the protocol described and adapted from Sharma and Singh.[20] Plasmid pET28b(+) was extracted from the *E. coli* strain DH10B/pET28b(+) and purified using the commercial QIAprep Spin Miniprep kit QIAGEN (São Paulo, Brazil) following the manufacturer's protocol. Phusion GC buffer Master Mix from Thermo Scientific (São Paulo, Brazil) was used in PCR for plasmid construction. For colony PCR and screening the enzyme My RedTaq polymerase Bioline (São Paulo, Brazil) was employed. The restrictions endonucleases FastDigest® (*NdeI*, *NheI*, *EcoRI*, *HindIII*, *XbaI*) and T4 DNA ligase were acquired from Thermo Scientific (São Paulo, Brazil).

Sesquiterpene cyclase gene cloning

The genes *ts-1*, *ts-2* and *ts-3* were amplified from the *Streptomyces* sp. CBMAI 2042 genome. For the construction of pET28b(+)::*ts-1* plasmid the oligonucleotides *ts-1*_**F** (5'-GGCAGCCATATGGCACAGCCCTTCTCACTGC-3') and *ts-1*_**R** (5'-GAGTTCGAATTCTACGCCCTGCTG-CCCGCCA-3') were used, for pET28b(+)::*ts-2* plasmid construction the oligonucleotides *ts-2*_**F** (5'-GGCAGCCATATGGTGGACAGCGAACTGCCG-3') and *ts-2*_**R** (5'-GAATTAGAATTCAGCCCTGCC-GCGGCGCCT-3') and for pET28b(+)::*ts-3* plasmid construction *ts-3*_**F** (5'-TATGTAGCTAGCCAGATA-ACGTTACCGGCG-3') and *ts-3*_**R** (5'-TATTCA-AGCTTCAGGCGGCGAAATACCC-3'). For ligation reaction, previously digested insert and vector were added in a molar ratio of 5:1 to 7:1 and mixed with 2 μL of ligation buffer containing adenosine triphosphate (ATP) and 1 μL of T4 DNA ligase in a 20 μL reaction volume. The ligation reaction was carried during 16 h at 16 °C and the enzyme was inactivated at 65 °C for 10 min, 1 μL of the reaction mixture was used to transform 40 μL of electrocompetent DH10B cells. The transformation was spread in LB agar plate containing kanamycin (50 μg mL^{-1}) and incubated for 16 h at 37 °C. Colony PCRs were carried to find positive constructions, colonies were selected and prepped to confirm plasmid construction by PCR and double restriction assays using appropriate restriction endonucleases.

Recombinant STCs productions and purification

Constructed plasmids under study were transformed into *E. coli* Rosetta(DE3). A single colony was used to inoculate a pre-inoculum in LB liquid medium with selective antibiotics kanamycin (50 μg mL^{-1}) and chloramphenicol (34 μg mL^{-1}) and shaken at 250 rpm, 16 h and 37 °C. Then, 1 L of LB liquid medium with selective antibiotics (5 × 1 L Erlenmeyer flask containing 200 mL LB) was inoculated

with 1% of the overnight pre-inoculum and shaken at 250 rpm at 37 °C until optical density at 600 nm (OD$_{600}$) 0.6, the Erlenmeyer's were placed on ice and added 0.4 μmol L^{-1} of isopropyl β-D-1-thiogalactopyranoside (IPTG). The flasks were shaken at 250 rpm for 16 h at 18 °C. The harvested cells were resuspended in a lysis buffer (20 mM Tris-HCl; 20 mM imidazole; 1 mM MgCl$_2$; pH 7) and lysed by sonication (Eco-Sonics, São Paulo, Brazil). The filtered supernatant containing the recombinant *His$_6$Tag*-STCs enzyme was purified by affinity chromatography on a ÄKTA pure (GE Healthcare, Umeå, Sweden) coupled with a 1 mL Histrap FF crude or 1 mL Histrap HP column, using wash buffer (20 mM Tris-HCl; 20 mM imidazole; 1 mM MgCl$_2$; pH 7) and elution buffer (20 mM Tris-HCl; 500 mM imidazole; 1 mM MgCl$_2$; pH 7). Buffer exchange was accomplished by applying the purified enzyme to a 5 mL HiTrap Desalting column and eluted by reaction buffer (50 mM Tris-HCl; 1 mM MgCl$_2$; pH 7) suitable for carrying out enzymatic assays. Bradford method was used to quantify purified proteins. The steps from production to purification were visualized by SDS-PAGE (sodium dodecyl sulfate-polyacrylamide gel electrophoresis) stained with commercial Coomassie Bio-Safe BioRad (São Paulo, Brazil). All centrifugations' steps were carried at centrifuge 5810R (Eppendorf, Hamburg, Germany).

Enzyme assays

The enzymatic assay was based on the protocol described by Rabe *et al.*,[21] and Felicetti and Cane.[22] Enzyme-substrate incubation experiments were performed using purified recombinant enzymes. Typically, 0.1 mg of enzyme was incubated with buffer (50 mM Tris-HCl; 1 mM MgCl$_2$; pH 7) supplemented with 50 μM farnesyl pyrophosphate ammonium salt from Cayman Chemical (Paulínia, Brazil) for 16 h at 30 °C. The reaction was overlaid with 1 mL of HPCL grade hexane extracted and analyzed using gas chromatograph-mass spectrometry (GC-MS).

GC-MS analysis

The extracts obtained were analyzed in a mass spectrometer coupled to a gas chromatograph Agilent 7890 (Agilent, São Paulo, Brazil) equipped with an HP-5MS column, injector temperature 220 °C, auxiliary temperature 240 °C, He flow 1 mL min^{-1}, injection volume 1 μL, split 300:1 to 20:1 and a mass range of *m/z* 50-300. The ramp profile was programmed as 60-165 °C in a gradient temperature of 3 °C min^{-1}, 165-280 °C in a gradient temperature of 20 °C min^{-1}, finally the ramp remained at 280 °C for 4 min.

Extraction of terpenes produced by the *Streptomyces* sp. CBMAI 2042

After the fermentation, the cultivation was centrifuged. The mycelium was washed 2× with autoclaved Milli-Q H$_2$O, centrifuged and the harvested cell mass was taken to the Biofreezer (Sanyo, Osaka, Japan) at –80 °C for at least 1 h. Then, it was extracted with 200 mL of methanol in an ice bath for 90 min under magnetic stirring. Subsequently, a filtration in a Büchner funnel was performed to remove the mycelium. The methanol extract (200 mL) was transferred to a separatory funnel and extracted with HPLC grade hexane (2 × 100 mL) to obtain the fraction teeming terpenes and other volatile hydrocarbons. To remove free fatty acids from extraction procedure, the hexane extract was washed with 2 × 50 mL 0.5 M NaOH in 50% methanol solution. Finally, the organic phase was washed with brine, dried with Na$_2$SO$_4$ and concentrated at a pressure of 130 mbar at 20 °C until obtaining 1 mL of a yellowish oil.

Results and Discussion

As previously described,[16] the actinobacteria *Streptomyces* sp. CBMAI 2042, an endophytic isolated from *Citrus sinensis*, was fully sequenced using Illumina technology (MiSeq). The linear sequence was grouped into three contigs and the genome estimated at approximately 8 Mbp.

The input data from *S.* sp. CBMAI 2042 genome sequencing was uploaded on the bioinformatics platform antiSMASH (version 6.0.1)[23] and the output results revealed the presence of 35 biosynthetic gene clusters (BGCs) linked to specialized metabolite production. The annotated clusters are involved in the biosynthesis of butyrolactone, t1-polyketide synthase (PKS), t3PKS, non-ribosomal peptide synthetase (NRPS),[24,25] isorenieratene, terpene, lantipeptide, lassopeptide, melanin, siderophore, ectoin, and hybrid BGCs NRPS-transATPKS-t1PKS,[17] melanin-NRPS, t3PKS, thiopeptide-t1PKS-NRPS-, t2PKS-butyrolactone (Table S1, Supplementary Information (SI) section).[18]

The antiSMASH platform crosses genome information with several *in silico* analysis tools, including analysis using hidden Markov models (HMMer 3).[26] The analysis highlighted the presence of five bacterial terpene clusters (BTCs) related to the production of bacterial terpene metabolites. Analysis throughout the NCBI BLAST tool[27] of the genes contained in terpene clusters identified a core gene encoding terpene cyclase enzyme in each BTC. From the five BTCs predicted, three of them (*ts-1*, *ts-2* and *ts-3*) displayed high similarity to sesquiterpene cyclase encoding genes. Careful bioinformatics analysis

indicated that the *ts-1* gene was probably associated with the production of the degraded terpene geosmin and its precursor germacradienol, *ts-3* could be related to caryophyllene alcohol-type metabolites and there were no correlations of metabolites with the *ts-2* gene, although it was associated with 1-10 cyclization pattern of terpenes derivatives metabolites.

Following the *in silico* study, transcriptional analysis was performed to determine whether the genes encoding the biosynthesis of sesquiterpenes from *Streptomyces* sp. CBMAI 2042 were active under normal growth conditions in TSBY medium. To assess the functionality of the *ts-1*, *ts-2*, and *ts-3* genes, tRNA extraction was performed at different times of bacterial culture in TSBY medium, profiling the transcriptome for the behavior of the selected biosynthetic genes (Figure S1, SI section). The assay showed that the *ts-1* gene was inactive under the conditions described above, whereas transcription of the *ts-2* and *ts-3* genes was observed between 72 and 120 h (Figure 1).

Consequently, the endophytic *S.* sp. CBMAI 2042 was cultured in different culture media, including TSBY, ISP2, and A media, and the volatile organic profile was evaluated (Figure 2). The analysis revealed that the oils extracted from the mycelium of CBMAI 2042 had a greater diversity of peaks associated with terpene metabolites when cultured in ISP2 and A media (Figure S2, SI section). In contrast, when cultured in TSBY, the analyzed oil showed a smaller range of terpenes, possibly due to the absence of peaks corresponding to metabolites associated with *ts-1* gene expression. This result is consistent with the data observed during transcriptional analysis in TSBY medium (Figure 2).

Using GC-MS analysis we identified at least 28 sesquiterpenes in the mycelial extracts when the bacterium was cultured on A medium, and 21 sesquiterpenes when cultured on ISP2 medium. Although the terpene peaks identified in the chromatogram had different retention times, they had a similar fragmentation pattern, making it difficult to classify their chemical signature by searching only the NIST11 and FFNSC2 libraries.

To characterize the sesquiterpene molecules it was required to obtain the arithmetic index (AI). As the Kováts index[28] is calculated under isothermal conditions, the method described by Van den Dool and Kratz,[29] equation 1, was adopted.[30] In this method, a temperature ramp was used to separate the terpene compounds, and the AI was calculated through an arithmetic relationship of the retention indices of the *n*-alkanes standard as well as for the unidentified terpenes compounds (Table 1).

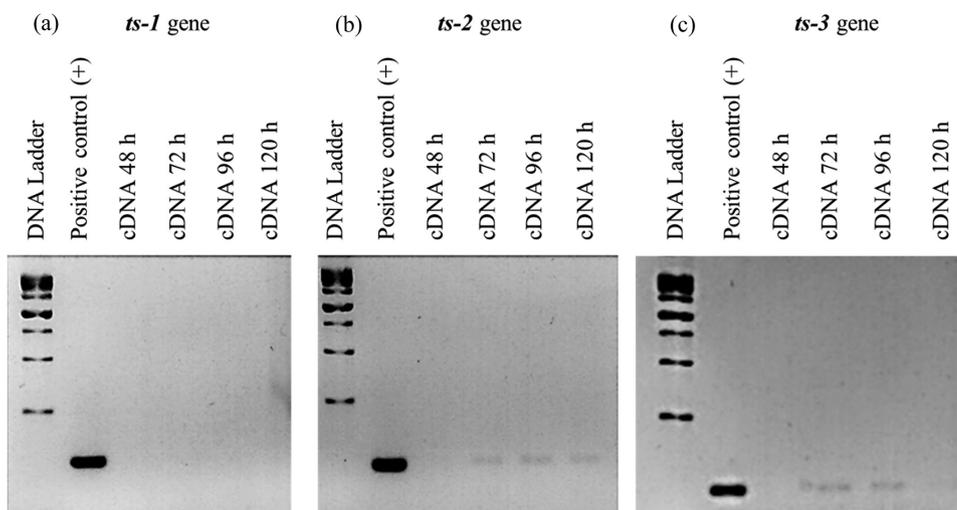$$AI = 100 \times \left(\frac{t_R a - t_R pz}{t_R p_{z+1} - t_R pz} + Pz \right) \qquad (1)$$

Figure 1. Transcription analysis of genes (a) *ts-1*; (b) *ts-2* and (c) *ts-3* during times of 48, 72, 96 and 120 h; for positive control, genomic DNA of *Streptomyces* sp. CBMAI 2042 was used as template for the PCR reaction.
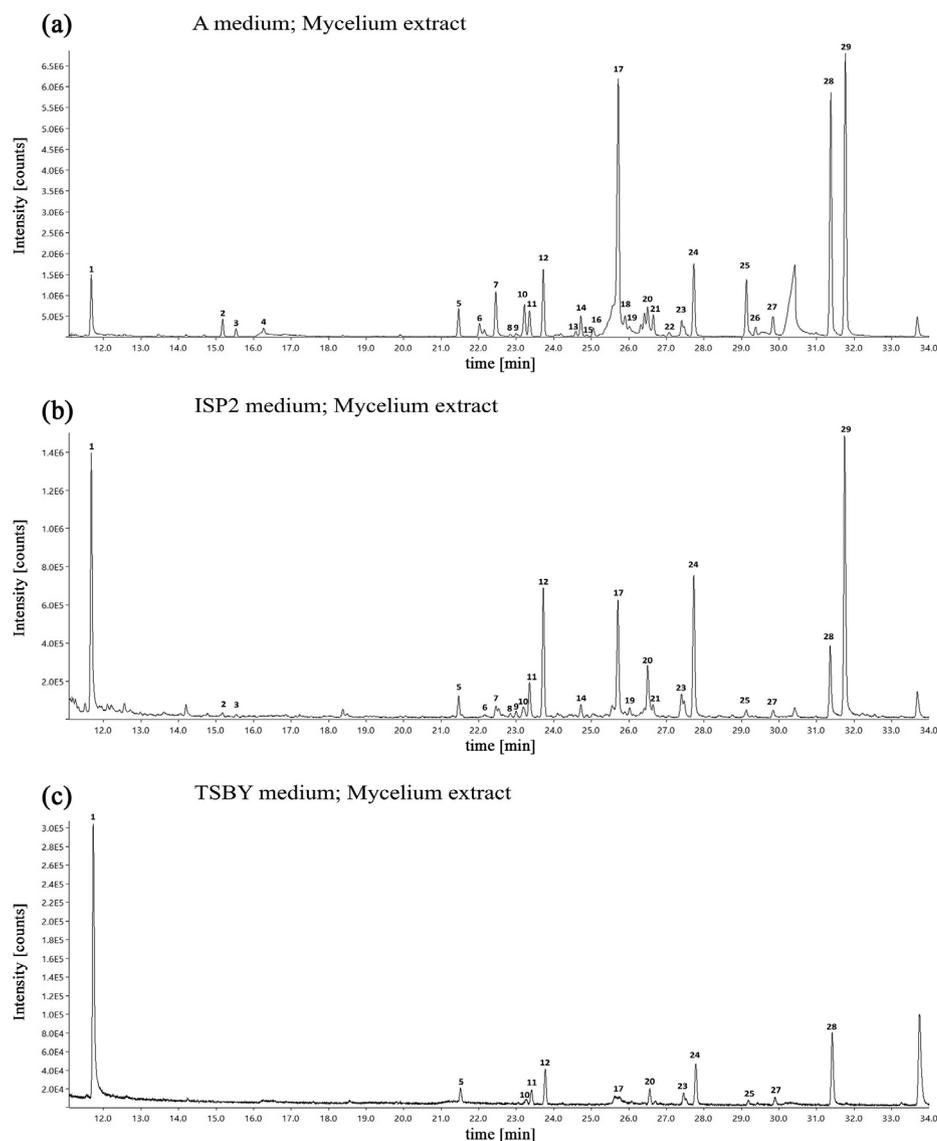

Figure 2. Chromatographic profile of the hexane extract from *Streptomyces* sp. CBMAI 2042 mycelium analyzed by GC-MS in (a) A medium; (b) ISP2 medium; and (c) TSBY medium. Compounds numbers were assigned in ascending order of retention time as shown in Table 1.

where, $t_R a$ represents the sample retention time, $t_R pz$ the retention time of the *n*-alkane standard peak that precedes the sample peak, $t_R p_{z+1}$ the retention time of the *n*-alkane standard peak that follows the sample peak and Pz represents the carbon number of the *n*-alkane with retention time $t_R pz$.

To relate the mining information to the terpene chemotypes, all STC identified gene cores *ts-(1-3)* were then amplified by PCR and cloned into the pET28b(+) system (Figures S3 and S4, SI section). For production of the recombinant terpene cyclases enzTS-(1-3), pET28b(+)::*ts-(1-3)* plasmids constructions were transformed into *E. coli* strains BL21(DE3) and Rosetta(DE3) (Table S2, SI section). Enzyme production experiments were performed using the *E. coli* Rosetta(DE3)

system, which has a pRARE plasmid encoding rare tRNAs with chloramphenicol resistance, as *in silico* analysis via the ATGme platform[33] confirmed the presence of a high percentage of rare codons in *E. coli* contained in *ts-(1-3)* genes. This system allowed us to obtain appreciable amounts of the recombinant enzymes in the soluble phase.

The purified enzTS-(1-3) enzymes (Figure S5, SI section) were conditioned in reaction buffer and incubated in an assay in the presence of the elongated precursor FPP; the blank assay was performed without the addition of recombinant enzymes in the reaction mixture. Examination of the chemotype profile using GC-MS analysis confirmed that the enzymes enzTS-(1-3) actively converted FPP to cyclic terpene derivatives (Figures S6-S8, SI section). After

Table 1. Bacterial terpenes uncovered in the GC-MS analysis of the hexane extracts of *Streptomyces* sp. CBMAI 2042 mycelium cultivated in A, ISP2 and TSBY medium

No.[a]	Compound[31]	A medium		ISP2 medium		TSBY medium		AI lit.[d]
		t_R[b] / min	AI exp[c]	t_R / min	AI exp.	t_R / min	AI exp.	
1	(+)-*trans*-limonene-1,2-epoxide[e]	11.66	1135	11.67	1135	11.73	1136	1137
2	8,10-dimethyl-1-octalin[32]	15.17	1222	15.18	1222	–	–	–
3	8,10-dimethyl-1(9)-octalin[32]	15.53	1230	15.54	1231	–	–	–
4	unidentified	16.26	1248	–	–	–	–	–
5	α-copaene	21.46	1373	21.46	1373	21.52	1374	1373
6	β-elemene[f]	22.16	1389	22.15	1389	–	–	1389
7	geosmin	22.46	1397	22.46	1397	–	–	1399
8	α-gurjunene	22.85	1406	22.84	1406	–	–	1409
9	unidentified	22.99	1410	22.99	1410	–	–	–
10	β-caryophyllene	23.22	1415	23.22	1415	23.26	1416	1417
11	unidentified	23.36	1419	23.35	1419	23.41	1420	–
12	β-gurjunene	23.77	1429	23.72	1428	23.77	1429	1431
13	humulene	24.60	1450	–	–	–	–	1452
14	unidentified	24.72	1453	24.72	1453	–	–	–
15	alloaromadendrene	24.89	1457	–	–	–	–	1458
16	unidentified	25.05	1461	–	–	–	–	–
17	germacrene D	25.72	1477	25.71	1477	–	–	1480
18	unidentified	25.90	1482	–	–	–	–	–
19	unidentified	26.01	1485	26.02	1485	–	–	–
20	α-muurolene	26.51	1497	26.51	1497	26.56	1498	1500
21	germacrene A	26.65	1501	26.65	1501	–	–	1508
22	unidentified	27.08	1512	–	–	–	–	–
23	δ-cadinene	27.41	1520	27.4	1520	27.46	1522	1522
24	*trans*-cadina-1,4-diene	27.74	1529	27.73	1529	27.79	1530	1533
25	caryophyllenyl alcohol	29.13	1565	29.13	1565	29.19	1567	1570
26	germacrene D-4-ol	29.38	1572	–	–	–	–	1574
27	4-epi-cubedol	29.85	1584	29.84	1583	29.90	1585	1576
28	1-epi-cubenol	31.38	1625	31.36	1624	31.42	1626	1627
29	1(10)*E*,5*E*-germacradien-11-ol	31.77	1635	31.74	1634	–	–	1633

[a]Compounds numbers were assigned in ascending order of retention time; [b]retention time; [c]experimental arithmetic index; [d]arithmetic index in literature; [e]internal standard; [f]generated from germacrene A by Cope rearrangement in the injection port of the GC apparatus.

searching the NIST11 and FFNSC2 mass spectral libraries, all biotransformation products of the predicted terpene cyclases in this study showed characteristic fragmentation patterns of sesquiterpenes (Figure S9, SI section). Similarly, AIs were calculated for the diagnostic cyclic products (Table 2).

As predicted by blastp, the incubation assay between the enzyme enzTS-1 and FPP resulted in the production of six cyclic terpenes, including the degraded terpenes 8,10-dimethyl-1-octalin (**2**), 8,10-dimethyl-1(9)-octalin (**3**), and geosmin (**7**), and the sesquiterpenes germacrene D (**17**) and 1(10)*E*,5*E*-germacradien-11-ol (**29**), plus an unidentified product (**18**). According to these results, the terpene cyclase enzTS-1 can be classified as "germacradienol/geosmin synthase".[32,34,35]

By analyzing the GC-MS results involving the reaction between enzTS-2 and FPP, it was possible to observe the presence of at least 12 volatile sesquiterpenes products, including nonpolar terpenes as β-gurjunene (**12**) and *trans*-cadina-1,4-diene (**24**) and the alcohols germacrene D-4-ol (**26**) and 1-epi-cubenol (**28**). The sesquiterpenes β-(**6**) and δ-(**30**) elemenes were possibly produced via Cope rearrangement due to temperature injection conditions used (above 200 °C).[36]

As the main products of the biotransformation reaction are alcohols, enzTS-2 can be described as "germacrene D-4-ol/1-epi-cubenol synthase", with germacrene D-4-ol (**26**) being the main product followed by 1-epi-cubenol (**28**) from *in vitro* enzyme-substrate assays. As predicted by *in silico* analysis, all products formed are derived from type 1-10 cyclization. Nevertheless, **26** was observed in low concentration in *Streptomyces* sp. CBMAI 2042 mycelium extract only when cultivated in A medium, which indicates that the formation of this sesquiterpene *in vitro* might be related to the reaction conditions such as culture supplements, pH and temperature.[37] Furthermore, enzTS-2 proved to be a terpene cyclase with promiscuity towards the formation of cyclic products, a suitable and promising target enzyme for biotechnological applications.

On the other hand, the incubation between enzTS-3 and FPP showed only one peak, with β-caryophyllene (**10**) being the major product of the reaction. However, formation of caryophyllene alcohol (**25**) with a calculated arithmetic index 1565 as observed *in vivo* production was expected. In spite of that, enzTS-3 can be classified as a "β-caryophyllene synthase".[38] Interestingly, among the studied sesquiterpene cyclases, enzTS-3 produces a unique product, β-caryophyllene (**10**), with 1-11 type cyclization.

Performing *in vitro* assays with recombinant terpene cyclase enzymes endorsed the identification of 18 sesquiterpenes, and 10 terpenes less than observed in the analysis of mycelial extracts of the wild-type strain. Nevertheless, the *in vitro* experiments could not always

Table 2. GC-MS analysis of *in vitro* incubation of recombinant enzymes enzTS-(1-3) with farnesyl pyrophosphate

Compound[32]	AI exp[a]			AI lit.[c]
	enzTS-1[b]	enzTS-2[b]	enzTS-3[b]	
8,10-Dimethyl-1-octalin[31] (**2**)	1221	–	–	–
8,10-Dimethyl-1(9)-octalin[31] (**3**)	1230	–	–	–
α-Copaene (**5**)	–	1371	–	1372
β-Elemene[d] (**6**)	–	1388	–	1389
Geosmin (**7**)	1395	–	–	1399
β-Caryophyllene (**10**)	–	–	1433	1417
β-Gurjunene (**12**)	–	1425	–	1431
Germacrene D (**17**)	1475	1475	–	1480
Unidentified (**18**)	1481	–	–	–
Unidentified (**19**)	–	1490	–	–
α-Muurolene (**20**)	–	1495	–	1500
δ-Cadinene (**23**)	–	1519	–	1522
trans-Cadina-1,4-diene (**24**)	–	1527	–	1533
Germacrene D-4-ol (**26**)	–	1569	–	1574
1-Epi-cubenol (**28**)	–	1621	–	1627
1(10)*E*,5*E*-Germacradien-11-ol (**29**)	1632	–	–	1633
δ-Elemene[d] (**30**)	–	1333	–	1335
Unidentified (**31**)	–	1580	–	–

[a]Experimental arithmetic index; [b]terpene cyclase recombinant enzyme; [c]arithmetic index in literature; [d]generated from germacrene A by Cope rearrangement in the injection port of the GC apparatus.

reproduce exactly the behavior of the native organism, due to the absence of assistant enzymes or even to the different physiological environment characteristic of heterologous expression. For instance, *in vitro* experiments demonstrated the absence of caryophyllenyl alcohol (**25**) and the formation of germacrene D-4-ol (**26**) at higher concentrations compared with *in vivo* experiments, supporting this hypothesis. This bias, although not fully elucidated, may have been influenced by a subtle difference in the folding of recombinant proteins when performing *in vitro* assays as well as by a different chemical environment.[11,37,39,40]

The current results reinforce the fundamental importance of STCs studies due to their promiscuity across the natural substrate qualifying such enzymes for industrial applications. In a study by Oberhauser *et al.*,[41] eight recombinant STCs were tested against six unnatural FPP containing heteroatoms, the assays resulted in the production of six unnatural heteroatom-modified cyclic terpenoids, in which one of the compounds obtained is a tricyclic terpenoid with interesting olfactory properties.

Additionally, studies performed with the recombinant enzyme germacrene D synthase (GDS) in the presence of FPP analogues generated unnatural SMs, including (*S*)-14,15-dimethylgermacrene D. Interestingly, unlike the repellent property of germacrene D against aphids, (*S*)-14,15-dimethylgermacrene D acted as an attractant against these insects.[42]

Unfortunately, the generation of unnatural terpenoids ended up being limited by the complexity of chemical synthesis and further purification of the products. However, recently, Johnson *et al.*[43] reported the production of FPP analogues using a cell free system. In this approach, promiscuous kinase enzymes were used to obtain pyrophosphate derivatives of prenol and isoprenol and the isoprene chain was elongated using prenyl transferases. The use of prenol analogues in different proportions made it possible to obtain FPP analogues in excellent yields, avoiding complex synthetic steps and difficulties with purification.

Advances in this area have allowed the expansion of terpenoid repertoire, allowing the use of these compounds as starting material in semi-synthesis applications. Improvements in cell free systems usage for obtaining unnatural precursors enables the application of this strategy on larger scales, which can be applied in the flavors and food, agrochemical, and pharmaceutical industries. In this sense the expansion of the terpenome of recombinant STCs from *Streptomyces* sp. CBMAI 2042 introduced in this study can be explored by performing incubation assays with unnatural precursors, as well as site-directed mutagenesis studies approaches for enzyme optimization, making it possible to obtain new derivatives of cyclic terpenes SMs.

Conclusions

The endophytic bacteria *Streptomyces* sp. CBMAI 2042 was fully sequenced revealing 35 biosynthetic gene clusters involved in the biosynthesis of secondary metabolites, with five of these clusters involved in the biosynthesis of terpene specialized metabolites. *In silico* analysis revealed the presence of three genes encoding sesquiterpene cyclase enzymes, qualifying the strain as a potential producer of volatile sesquiterpenes SMs.

To deepen the knowledge into the biosynthetic machinery of this strain and to expand the knowledge of bacterial sesquiterpene cyclases, we performed fermentation experiments in different culture media and extraction of the volatile fraction. As a result, we observed the production of at least 28 sesquiterpenes by the native strain. Moreover, the recombinant mapped STCs enzymes enzTS-(1-3) were produced in their active form and evaluated *in vitro* in the presence of the natural substrate farnesyl pyrophosphate. Using calculated arithmetic indices, mass spectrometry and comparison with literature data, we identified the sesquiterpenes chemotype produced *in vivo* and *in vitro*.

The technical-scientific knowledge generated by unveiling the bacterial terpenome of *Streptomyces* sp. CBMAI 2042 contributes in numerous ways to the development of this large area of biotechnology. In this work, it was possible to acquire knowledge about the functioning of three bacterial terpene cyclases, in addition through sequencing and careful mining of STCs in the bacterial genome we contributed to feed databases regarding the sequence of enzymes that can be further explored.

Moreover, with these recombinant enzymes in hand, incubation assays with unnatural precursors can be performed, allowing unprecedented generation and expansion of terpenoid backbones. Finally, deepening the knowledge about STCs enzymes will enable the development of sustainable systems with application in several areas including the agrochemical and medicinal sector.

Supplementary Information

Supplementary information (tables, agarose and SDS-PAGE gels, GC-MS chromatograms, and mass spectra) is available free of charge at http://jbcs.sbq.org.br as PDF file.

Acknowledgments

References

1. Ikeda, H.; *Biosci. Biotechnol. Biochem.* **2017**, *81*, 13.

2. Bauer, A.; Brönstrup, M.; *Nat. Prod. Rep.* **2014**, *31*, 35.

3. Harms, V.; Kirschning, A.; Dickschat, J. S.; *Nat. Prod. Rep.* **2020**, *37*, 1080.

4. Moser, S.; Pichler, H.; *Appl. Microbiol. Biotechnol.* **2019**, *103*, 5501.

5. Dickschat, J. S.; *Nat. Prod. Rep.* **2016**, *33*, 87.

6. Huynh, F.; Grundy, D. J.; Jenkins, R. L.; Miller, D. J.; Allemann, R. K.; *ChemBioChem* **2018**, *19*, 1834.

7. Reddy, G. K.; Leferink, N. G. H.; Umemura, M.; Ahmed, S. T.; Breitling, R.; Scrutton, N. S.; Takano, E.; *PLoS One* **2020**, *15*, e0232220.

8. Christianson, D. W.; *Curr. Opin. Chem. Biol.* **2008**, *12*, 141.

9. Yamada, Y.; Kuzuyama, T.; Komatsu, M.; Shin-Ya, K.; Omura, S.; Cane, D. E.; Ikeda, H.; *Proc. Natl. Acad. Sci. U. S. A.* **2015**, *112*, 857.

10. Bian, G.; Han, Y.; Hou, A.; Yuan, Y.; Liu, X.; Deng, Z.; Liu, T.; *Metab. Eng.* **2017**, *42*, 1.

11. Helfrich, E. J. N.; Lin, G. M.; Voigt, C. A.; Clardy, J.; *Beilstein J. Org. Chem.* **2019**, *15*, 2889.

12. Eddy, S. R.; *Bioinformatics* **1998**, *14*, 755.

13. *Pfam 35.0*; http://pfam.xfam.org, accessed in December 2021.

14. Moody, S. C.; Zhao, B.; Lei, L.; Nelson, D. R.; Mullins, J. G. L.; Waterman, M. R.; Kelly, S. L.; Lamb, D. C.; *FEBS J.* **2012**, 1640.

15. Cane, D. E.; Ikeda, H.; *Acc. Chem. Res.* **2012**, *45*, 463.

16. de Oliveira, L. G.; Sigrist, R.; Paulo, B. S.; Samborskyy, M.; *Microbiol. Resour. Announc.* **2019**, *8*, DOI 10.1128/MRA.01426-18.

17. Sigrist, R.; Luhavaya, H.; McKinnie, S. M. K.; da Silva, A. F.; Jurberg, I. D.; Moore, B. S.; de Oliveira, L. G.; *ACS Chem. Biol.* **2020**, *15*, 1067.

18. Paulo, B. S.; Sigrist, R.; Angolini, C. F. A.; Eberlin, M. N.; de Oliveira, L. G.; *J. Braz. Chem. Soc.* **2018**, *30*, 673.

19. Araújo, W. L.; Marcon, J.; Maccheroni, W.; van Elsas, J. D.; van Vuurde, J. W. L.; Azevedo, J. L.; *Appl. Environ. Microbiol.* **2002**, *68*, 4906.

20. Sharma, A. D.; Singh, J.; *Anal. Biochem.* **2005**, *337*, 354.

21. Rabe, P.; Samborskyy, M.; Leadlay, P. F.; Dickschat, J. S.; *Org. Biomol. Chem.* **2017**, *15*, 2353.

22. Felicetti, B.; Cane, D. E.; *J. Am. Chem. Soc.* **2004**, *126*, 7212.

23. Blin, K.; Shaw, S.; Kloosterman, A. M.; Charlop-Powers, Z.; van Wezel, G. P.; Medema, M. H.; Weber, T.; *Nucleic Acids Res.* **2021**, *49*, W29.

24. Paulo, B. S.; Sigrist, R.; Angolini, C. F. F.; de Oliveira, L. G.; *ChemistrySelect* **2019**, *4*, 7785.

25. Sigrist, R.; Paulo, B. S.; Angolini, C. F. F.; de Oliveira, L. G.; *J. Visualized Exp.* **2020**, DOI 10.3791/60825.

26. Eddy, S. R.; *PLoS Comput. Biol.* **2011**, *7*, e1002195.

27. McGinnis, S.; Madden, T. L.; *Nucleic Acids Res.* **2004**, *32*, W20.

28. von Kováts, E.; *Helv. Chim. Acta* **1958**, *XLI*, 1915.

29. van Den Dool, H.; Kratz, P. D.; *J. Chromatogr. A* **1963**, *11*, 463.

30. Viegas, M. C.; Bassoli, D. G.; *Quim. Nova* **2007**, *30*, 2031.

31. Adams, R. P.; *Identification of Essential Oil Components by Gas Chromatography/Mass Spectrometry*; 4th ed.; Allured Publishing: Carol Stream, 2007.

32. Nawrath, T.; Dickschat, J. S.; Müller, R.; Jiang, J.; Cane, D. E.; Schulz, S.; *J. Am. Chem. Soc.* **2008**, *130*, 430.

33. *ATGme*; http://atgme.org/, accessed in December 2021.

34. Cane, D. E.; He, X.; Kobayashi, S.; Ōmura, S.; Ikeda, H.; *J. Antibiot. (Tokyo)* **2006**, *59*, 471.

35. Jiang, J.; Cane, D. E.; *J. Am. Chem. Soc.* **2008**, *130*, 428.

36. Calvert, M. J.; Ashton, P. R.; Allemann, R. K.; *J. Am. Chem. Soc.* **2002**, *124*, 11636.

37. Dickschat, J. S.; Pahirulzaman, K. A. K.; Rabe, P.; Klapschinski, T. A.; *ChemBioChem* **2014**, *15*, 810.

38. Nakano, C.; Horinouchi, S.; Ohnishi, Y.; *J. Biol. Chem.* **2011**, *286*, 27980.

39. Rabe, P.; Schmitz, T.; Dickschat, J. S.; *Beilstein J. Org. Chem.* **2016**, 1839.

40. Baer, P.; Rabe, P.; Fischer, K.; Citron, C. A.; Klapschinski, T. A.; Groll, M.; Dickschat, J. S.; *Angew. Chem., Int. Ed.* **2014**, *53*, 7652.

41. Oberhauser, C.; Harms, V.; Seidel, K.; Schröder, B.; Ekramzadeh, K.; Beutel, S.; Winkler, S.; Lauterbach, L.; Dickschat, J. S.; Kirschning, A.; *Angew. Chem., Int. Ed.* **2018**, *57*, 11802.

42. Touchet, S.; Chamberlain, K.; Woodcock, C. M.; Miller, D. J.; Birkett, M. A.; Pickett, J. A.; Allemann, R. K.; *Chem. Commun.* **2015**, *51*, 7550.

43. Johnson, L. A.; Dunbabin, A.; Benton, J. C. R.; Mart, R. J.; Allemann, R. K.; *Angew. Chem., Int. Ed.* **2020**, *59*, 8486.