Exploring natural products: from herbal resources, microbial synthesis to animal models

Guan, Zheng

DOI:
10.33612/diss.846916968

Chapter 4

Production of squalene in *Bacillus subtilis* by squalene synthases screening and metabolic engineering

Yafeng Song, **Zheng Guan**, Ronald van Merkerk, Hegar Pramastya, Ingy I. Abdallah, Rita Setroikromo, Wim J. Quax*

Published in Journal of Agricultural and Food Chemistry.
2020, 68(15): 4447-4455

Abstract

Squalene synthase (SQS) catalyzes the conversion of two farnesyl pyrophosphates to squalene, an important intermediate in between isoprene and valuable triterpenoids. In this study we have constructed a novel biosynthesis pathway for squalene in *Bacillus subtilis* and performed metabolic engineering aiming at facilitating further exploitation and production of squalene-derived triterpenoids. Therefore, systematic studies and analysis were performed including selection of multiple SQS candidates from various organisms, comparison of expression vectors, optimization of cultivation temperatures and examination of rate-limiting factors within the synthetic pathway. We were for the first time able to obtain squalene synthesis in *B. subtilis*. Furthermore, we achieved a 29-fold increase of squalene yield (0.26 mg/L-7.5 mg/L) by expressing SQS from *Bacillus megaterium* and eliminating bottlenecks within upstream methylerythritol-phosphate pathway. Moreover, our findings showed that also *ispA* could positively affect the production of squalene.

Keywords: *Bacillus subtilis*, MEP pathway, squalene, squalene synthase

Introduction

Bacillus subtilis, Generally Recognized as Safe (GRAS) by the Food and Drug Administration (FDA), has long been investigated and widely used in various fields of industry ranging from food, feed additive, pharmaceuticals and fine chemicals[1-3]. In addition, *B. subtilis* has been reported to be a high isoprene producer,[4] which indicates its potential to become a cell factory for high-value terpenoids.[5, 6] Terpenoids, also referred to as isoprenoids, constitute a large class of natural products with a great diversity in both structural and biochemical properties. Moreover, many of them have health-enhancing properties and therapeutic potential, such as ginsenosides and artemisinin.[7] Recently, biosynthesis of these terpenoids in *B. subtilis*, has attracted ample attention due to the numerous advantages of this microbial cell factory.[6, 8] In *B. subtilis*, isoprenoid precursors are synthesized through 2C-methyl-D-erythritol-4-phosphate (MEP) pathway[9, 10] where the common building blocks of terpenoids, the two five-carbon precursors isopentenyl pyrophosphate (IPP) and its isomer dimethylallyl pyrophosphate (DMAPP) are produced.[11] The two isoprene units condense to form geranyl pyrophosphate (GPP, C_{10}), and farnesyl pyrophosphate (FPP, C_{15}) by addition of another IPP, and geranylgeranyl pyrophosphate (GGPP, C_{20}) by condensation of another two IPPs, respectively. Then they are cyclized, glycosylated and modified to produce various terpenoids where GPP produce monoterpenoids, FPP yield sesquiterpenoids and triterpenoids and GGPP produce diterpenoids and tetraterpenoids.

Squalene, an acyclic isoprenoid, is a crucial intermediate for the synthesis of many bioactive triterpenoids, such as hopanoids and sterols, which play vital cellular functions in organisms.[12] In addition, squalene itself shows very promising physiological activities such as anti-oxidant effects, decreasing cancer risks, and enhancing the immune system, which promotes its wide applications as additive, supplement or nutraceutical in food and personal care industry.[13] Squalene synthase (EC 2.5.1.21) catalyzes the head-to-head condensation of two molecules of FPP to form linear C_{30} squalene in a two-step reaction, which is the first committed step towards many triterpenoids (**Figure 1**).[14, 15] In the first step, two molecules of FPP sequentially enter into the catalytic center of SQS to form presqualene pyrophosphate (PSPP), which is a stable cyclopropylcarbinyl pyrophosphate intermediate.[15, 16] In the following step, further carbon-skeleton rearrangement including heterolysis and isomerization occur to form squalene, with the NADPH-dependent reduction happening at the same time.[17, 18] In the past several decades, SQSs from different eukaryotic species have been extensively characterized, and the

catalytic mechanism has been clarified supported by 3D-structure elucidation, truncation and site-directed mutagenesis.[15, 19-22]

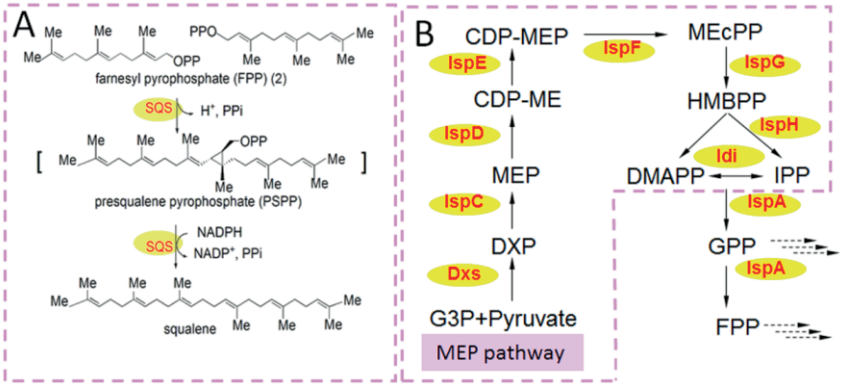


Figure 1. Squalene synthase reaction and 2C-methyl-D-erythritol-4-phosphate pathway. A. Scheme of squalene synthase (SQS) reaction steps. **B**. MEP pathway in *B. subtilis*. Dxs, 1-Deoxy-D-xylulose-5-phosphate synthase; IspC, 1-Deoxy-D-xylulose-5-phosphate reductoisomerase; IspD, 4-Pyrophospho-cytidyl-2-C-methyl-D-erythritol synthase; IspE, 4-Pyrophosphocytidyl-2-C-methyl- D-erythritol kinase; IspF, 2C-Methyl-D-erythritol 2,4-cyclopyrophosphate synthase; IspG, 1-Hydroxy-2- methyl-2-(E)-butenyl 4-pyrophosphate synthase; IspH, 1-Hydroxy-2-methyl-butenyl 4-pyrophosphate reductase; Idi, Isopentenyl pyrophosphate isomerase; IspA, Farnesyl pyrophosphate synthase; **Metabolite abbreviations**: G3P, Glyceraldehyde-3-phosphate; DXP, 1-Deoxy-D-xylulose 5-phosphate; MEP, 2-C-Methyl-D-erythritol 4-phosphate; CDP-ME, 4-(Cytidine 5'-pyrophospho)-2-C-methyl-D-erythritol; CDP-MEP, 2-Phospho-4-(cytidine 5'-pyrophospho)-2-C-methyl-D-erythritol; MEcPP, 2-C-Methyl-D-erythritol 2,4-cyclo-pyrophosphate; HMBPP; 1-Hydroxy-2-methyl-2-butenyl 4-pyrophosphate; IPP, Isopentenyl pyrophosphate; DMAPP, Dimethylallyl pyrophosphate; GPP, Geranyl pyrophosphate; FPP, Farnesyl pyrophosphate; PSPP: Presqualene pyrophosphate.

Since most studies focus their attention on eukaryotic SQSs, prokaryotic SQSs have rarely been explored.[19, 20, 23] In *B. subtilis*, no squalene producing capacity has been reported and the *yisP* gene, which was initially annotated as squalene synthase has now been characterized as a phosphatase with no squalene catalytic activity.[24, 25] Since the aim of this work is researching *B. subtilis* for the production of the most important committed triterpenoid intermediate (squalene), which has not yet been synthesized in *B. subtilis*, this study explored the synthesis of squalene directed by SQSs from different species and engineered the host strain to improve squalene production. Four representative SQS candidates from bacteria species, fungi and plants, were selected and expressed in *B. subtilis* to detect their squalene production. Different plasmids were employed to carry the SQSs and the upstream MEP pathway genes were also combined and overexpressed to explore their effect on squalene production. In addition, the

fermentation temperature was optimized to boost the production level of squalene. This paves the way for future metabolic engineering work in *Bacillus* to improve the production levels of other triterpenoids.

Materials and Methods

Strains and culture conditions

The plasmids and strains used in this study are listed in **Table S1** and **Table S2**. Single colonies of *B. subtilis* were picked up and inoculated into LB media with appropriate antibiotics and incubated at 37ºC overnight. Then overnight culture mixture was inoculated at a ratio of 1:100 (v/v) into 10 mL 2SR media (5% Yeast extract, 3% Tryptone and 0.3% K_2HPO_4) in 50 mL cellstar® cellreactor tube with filter screw cap (Greiner bio-one, Germany) for fermentation, in triplicates per strain. When the OD_{600} of the bacteria reached around 0.5-0.7, expression was induced by adding D-xylose to a final concentration of 1% (m/v), or IPTG at a final concentration of 1mM when necessary. Then bacterial cells were harvested after 48 hours of fermentation at 37ºC (unless indicated), 230 rpm. Antibiotics were added where appropriate (ampicillin at 100 μg/mL for *E. coli*, spectinomycin at 100 μg/mL, chloramphenicol at 5 μg/mL and erythromycin at 10 μg/mL for *B. subtilis*).

Plasmid Construction and Transformation

Candidate squalene synthase genes were obtained from *Bacillus acidocaldarius* (BaSQS) (Genbank: WP_012811689.1), *Bacillus megaterium* (BmSQS) (GenBank: ADF40697.1), *Panax ginseng* (PgSQS) (GenBank: AJV26445.1) and *Saccharomyces cerevisiae* (ScSQS) (GenBank: AAB68360.1), respectively. They were synthesized and codon optimized to *B. subtilis 168* (Eurofins, Netherlands). Eukaryotic SQSs normally contain their transmembrane regions (TMR) at the C-terminus of the proteins. Therefore, 25 and 24 amino acid residues at the C-terminus of PgSQS and ScSQS, which according to the literature reports and prediction results of TMHMM Server (v. 2.0) were supposed to be the TMRs, were truncated respectively (**Figure S1**).[19, 26, 27]

The plasmids were constructed by Prolonged Overlap Extension PCR (POE-PCR) method as described before.[28] The ribosome binding site (RBS) and the spacer between RBS and start codon (AAAGGGGG) were added at the N-terminus of the SQS candidates. The 6×His-tag (CATCATCATCATCATCAT) were placed at the C-terminus of the SQS candidates upstream

of the stop codon. The POE-PCR product was transformed directly to *E. coli*. The MEP pathway genes (*dxs, ispD, ispF, ispH*) were cut from previously constructed plasmid pHB201-SDFH by Xue *et al.*[5]with restriction enzymes XbaI and SpeI and ligated with linearized pBS0E (digested with the same enzymes XbaI and SpeI) by T4 Ligase (Thermo Fisher Scientific, USA), leading to the construction of pBS0E-SDFH. *ispC, ispE, ispG* and *ispA* amplified from pHCMC04G-CEGA were individually cloned into pBS0E-SDFH and placed downstream of *ispH* by the POE-PCR method, generating plasmids, pBS0E-SDFHC, pBS0E-SDFHE, pBS0E-SDFHG and pBS0E-SDFHA. Samples with positive colony-PCR results were further confirmed by sequencing the fragments (Macrogen, Netherlands). Plasmids were transformed to *B. subtilis* under the standard methods described by Kunst and Rapoport.[29] Primers used in this study are listed in Supplementary (**Table S3**).

Sample preparation for GC detection

Bacterial cells were harvested by centrifuging (11000×*g*) at 4 °C. Pre-chilled (-20°C) 50% methanol (methanol: Milli-Q water, 1:1, v/v) was added to the pellets to quench the cells. After centrifugation, the quenched cell pellets were quickly washed by 4 °C Milli-Q water. To lyse the cells, 1 mL of 50% cold methanol was added to the washed pellets and repeated freeze-thaw process five times by using liquid nitrogen.[30] The supernatants were collected in a new tube. In the following extraction procedure, acetone was used two times and ethyl acetate five times to extract the desired components. All the extracts of one sample were collected in the same tube. Then, samples were dried under nitrogen and dissolved in 250 μL of isopropanol (IPA)-acetonitrile (ACN) (7:3, v/v). Prior to use, all the samples were filtered through a 0.22-μm membrane.

Squalene detection and quantification

Sample analysis was performed on a Shimadzu GCMS-QP2010SE system equipped with a GC-2010 Plus gas chromatograph (GC) and AOC-20i autoinjector. Samples (4 μL) were injected in split mode onto the HP-5MS (5% Phenyl)-methylpolysiloxane GC column (Agilent J&W 0.25 mm inner diameter, 0.25 μm thickness, 30 m length), with helium as the carrier gas. The injector temperature was 280°C, and the column oven initial temperature was 210°C with an increase of 15°C per minute up to 260°C and then 5°C per minute till 280°C. Subsequently, the temperature was raised to 310°C with an increase of 25°C per minute and held for 8 min. The solvent cutoff was set at 8 min. Selected ion mode (SIM) was applied for acquisition,

monitoring m/z ion 136 and 384 for squalene and internal standard cholesterol, respectively. The integration tools in GCMSsolution 1.20 software (Shimadzu, Den Bosch, the Netherlands) was used to determine the chromatographic peak areas for squalene and cholesterol. To quantify the amount of squalene in the different samples, a calibration curve of standard squalene (Sigma-Aldrich S3626, purity≥98%) was generated with concentration range from 10 to 500 μg/mL and cholesterol (Sigma-Aldrich C8667, purity≥99%) as the internal standard at the concentration of 80 μg/mL.

Results

Candidate selection and analysis

To investigate the evolutionary relations of SQSs across multiple kingdoms of life, an unrooted phylogenetic tree was constructed using Neighbor-Joining method in MEGA 7.0. Apart from SQSs from *B. acidocaldarius* and *B. megaterium*, the other SQSs among the 28 candidates have been investigated and validated to maintain the capacity to convert FPP to squalene *in vitro* or *in vivo*. Results (**Figure 2A**) showed that these SQSs can be divided into four categories, including plants, bacteria, mammals and fungi.

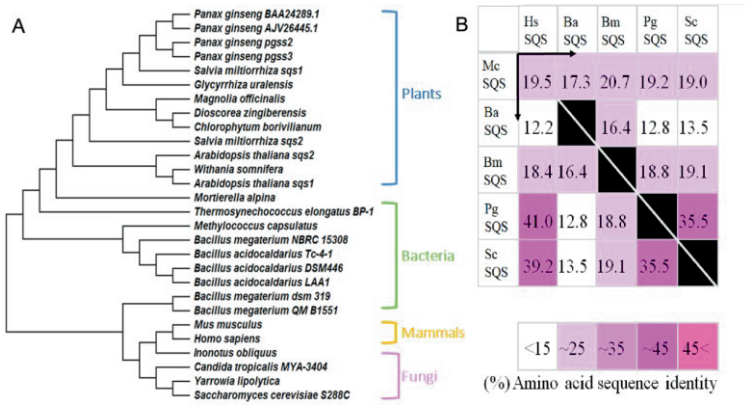


Figure 2. Phylogenetic tree analysis of squalene synthases and sequence alignment. A. Phylogenetic tree analysis of squalene synthases from different species constructed by MEGA 7.0. **B**. Squalene synthase sequence alignment. McSQS, HsSQS, BaSQS, BmSQS, PgSQS, and ScSQS represents squalene synthases originating from *Methylococcus capsulatus, Homo sapiens, Bacillus acidocaldarius*, *Bacillus megaterium, Panax ginseng* and *Saccharomyces cerevisiae*, respectively.

Squalene synthases exist both in prokaryotic and eukaryotic organisms. The candidates selected were either reported to be functional squalene synthases in their native hosts or have been annotated as squalene synthases (**Table 1**). After sequence alignment and literature comparison, candidates with a high sequence identity with reported functional SQSs or one with an available

crystal structure were chosen. Four candidates were selected for experimental analysis, i.e. BaSQS, BmSQS, PgSQS, and ScSQS, which originate from *Bacillus acidocaldarius*, *Bacillus megaterium*, *Panax ginseng*, and *Saccharomyces cerevisiae,* respectively. BaSQS was annotated as squalene synthase with its crystal structure available (4HD1)[31], which would facilitate further exploration if a high squalene synthesis capacity is being detected. The discovery of squalene cyclase in *B. megaterium* implied the existence of squalene synthase in this bacteria[32]. Hence the annotated squalene/phytoene synthase in *B. megaterium* was selected as a candidate. *P. ginseng* is famous for producing ginsenosides, the bioactive triterpenoids derived from squalene[27]. Squalene synthases from this plant were thought to possess high catalytic efficiency. The ScSQS in the candidate list was selected due to its high squalene synthesis capacity in both yeast and *E. coli*.[19, 33]

Table 1 Information of squalene synthase candidates.

SQS candidates	Original species	Characterization	Crystal structure	Amino acid length	Reference
BaSQS	*B. acidocaldarius*	Crystal structure available	Yes	291	PDB: 4HD1[31]
BmSQS	*B. megaterium*	Annotated as squalene /phytoene synthase	No	272	[34]
PgSQS	*P. ginseng*	Validated in *E. coli* and restores SQS function in plant	No	415	[26,27]
ScSQS	*S. cerevisiae*	Validated in *E. coli*	No	444	[19]

Then the four candidates BaSQS, BmSQS, PgSQS, and ScSQS were selected to align with the well-studied SQSs HsSQS and McSQS, originating from Eukaryotes (*Homo sapiens*) and Prokaryotes (*Methylococcus capsulatus*) to compare their sequences (**Figure 2B**).[15, 23] Results demonstrate that prokaryotic SQSs from different bacteria can show quite low amino acid identities among each other even within the same genus (up to 16.4 %). SQSs identities between prokaryotes and eukaryotes are even lower, with the percentages ranging from only 12.2% to 19.1%. Only PgSQS and ScSQS shared higher identities (41% and 39%) with HsSQS, which is consistent with the previous observation that SQSs from eukaryotic species are more conserved. The figures implied that SQSs from different species can be significantly distinctive, at least at the primary sequence level. This inspired us to investigate the squalene-synthesis capacities of different SQSs in *B. subtilis*.

Level of production of squalene by different SQSs in *B. subtilis* 168

E. coli-Bacillus shuttle vector pHCMC04G was introduced to express SQS candidates (BaSQS, BmSQS, PgSQS and ScSQS) by placing SQS downstream of an engineered *B. subtilis mntA* ribosome binding site.[5] The xylose-inducible promoter facilitated their precise expression. The constructs are shown in **Table S1**. All these constructs were transformed to *B. subtilis*, generating BA, BM, PG and SC (**Table S2**). The negative control strain BC was *B. subtilis* containing plasmid pHCMC04G without SQS. Since the eukaryotic SQS candidates (PgSQS and ScSQS) possess the transmembrane regions, which will anchor the protein to membranes[22, 27], these fragments were truncated (**Figure S1**).

B. subtilis strains were cultured in 2SR medium at 37°C. After 48 h incubation, metabolites were extracted and squalene production was quantified by GC-MS. The squalene production levels produced by different SQSs were compared (**Figure 3**). In the negative control, non-SQS containing strain BC, no squalene could be detected. Surprisingly, strain BA that is carrying the SQS candidate from *B. acidocaldarius* also showed no detectable squalene. For the other two eukaryotic SQSs, PgSQS and ScSQS originating from a plant and yeast, the conversion of precursor FPP into squalene could be measured in *B. subtilis* after the predicted TMR regions were truncated. SC produced a higher level of squalene than PG. BM that is containing the SQS from *B. megaterium* produced the highest level of squalene among the tested candidates, reaching 0.26 mg/L.

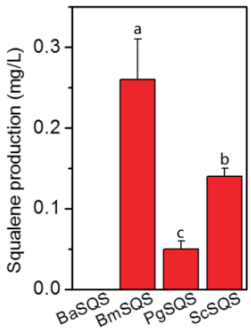


Figure 3. Production levels of squalene in *B. subtilis* strains expressing different squalene synthases. Error bars represent standard deviations of biological triplicates. Different letters indicate significant statistical differences (Scott Knott 5%). BaSQS, BmSQS, PgSQS, and ScSQS are squalene synthases originating from *Bacillus acidocaldarius, Bacillus megaterium, Panax ginseng* and *Saccharomyces cerevisiae*, respectively.

Effect of vector system on squalene production

BmSQS produced the highest level of squalene among the tested candidates when expressed in pHCMC04G. Subsequently, we explored multiple plasmids of maintaining the BmSQS genes in *Bacillus*. The first construct uses a rolling-circle replicating plasmid pHY300PLK (Strain HBM).[35] A second construct is pDR-BmSQS, in which the original integrative plasmid pDR111 containing the strong IPTG inducible promoter $P_{hyperspank}$ which is responsible for BmSQS expression upon insertion into the *amyE* locus of the *B. subtilis* genome (Strain DBM) (**Figure 4A**).[8] In the same way, pHY-PgSQS, pHY-ScSQS, pDR111-PgSQS and pDR111-ScSQS were constructed and transformed to *B. subtilis* 168 generating HPG, HSC, DPG and DSC (**Table S1, Table S2**). As shown in **Figure 4**, pDR-BmSQS produced the highest level of squalene at 0.4 mg/L. Whereas, pHY-BmSQS showed much lower squalene production with only 0.1 mg/L. Consistently, similar results displayed that DPG and DSC result in higher squalene production than using plasmids pHY300PLK and pHCMC04G (**Figure 4B**). Subsequently, SQSs integrated into the *B. subtilis* genome have been used for further experiments.

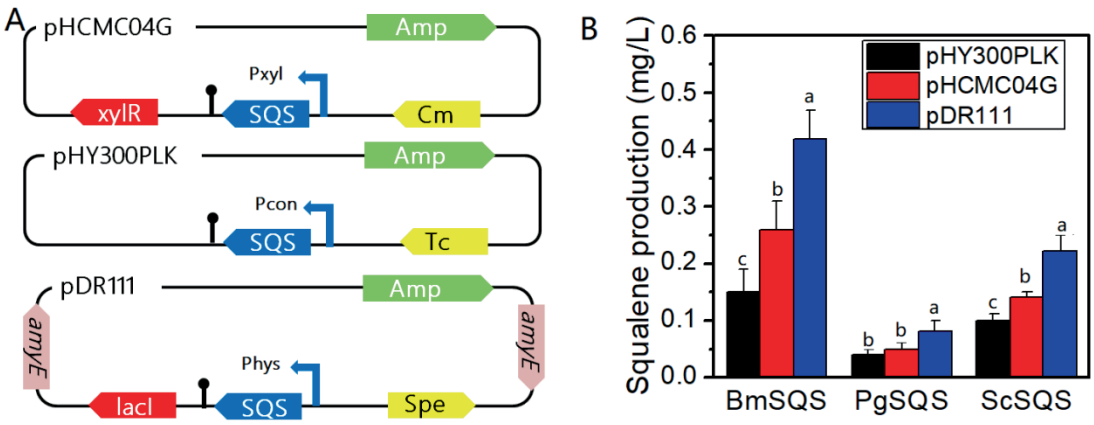


Figure 4. Effect of vector system on squalene production in *Bacillus subtilis*. **A**. Plasmids used for squalene synthase expression. Pxyl: xylose inducible promoter P_{xylA}; Pcon: constitutive promoter; Phys: IPTG inducible promoter $P_{hyperspank}$. **B**. Squalene production levels when expressed in different plasmids in *B. subtilis*. Error bars represent standard deviations of biological triplicates. Different letters indicate significant statistical differences (Scott Knott 5%). BmSQS, PgSQS, and ScSQS are squalene synthases originating from *Bacillus megaterium, Panax ginseng* and *Saccharomyces cerevisiae*, respectively.

Upregulation of MEP pathway genes increased the metabolic flux toward squalene

Previous literature indicates that *B. subtilis* is a high isoprene producer, and downstream terpenoids production could be enormously improved when the upstream pathway was boosted.[5] Hence, we tried to combine the upregulation of the MEP pathway with the expression

of squalene synthases and evaluated the downstream squalene production. In *B. subtilis*, the MEP pathway consists of *dxs, ispD, ispF, ispH, ispC, ispE, and ispG.* In addition *ispA,* encoding prenyltransferase, is responsible for subsequent elongation of isoprene units **(Figure 1B)**.[6] Part or all of these genes were combined as an entire operon in pHCMC04G. Four genes (*dxs, ispD, ispF* and *ispH*) were used to form pHCMC04G-SDFH. Additionally, another four genes *ispC, ispE, ispG* and *ispA* were assembled to form pHCMC04G-SDFHCEGA as described before.[8] Plasmids pHCMC04G-SDFH and pHCMC04G-SDFHCEGA were transformed to DBM, DPG and DSC, respectively. The resulting strains were tested using the same fermentation protocol. The results are displayed in **Figure 5**. Overall, all the three SQSs showed improved squalene production when co-expressed with pHCMC04G-SDFH. PgSQS showed the strongest increase possibly due to its low basal production and reached 0.44 mg/L. The BmSQS containing strain reached 0.6 mg/L and the ScSQS containing strain reached 0.85 mg/L squalene. When all 8 MEP pathway related genes were overexpressed, all the three different SQS strains had an around 4- to 10-fold increase of squalene production. DBM-MEP8 (containing pHCMC04G-SDFHCEGA) produced highest level of squalene, reaching around 1.6 mg/L.

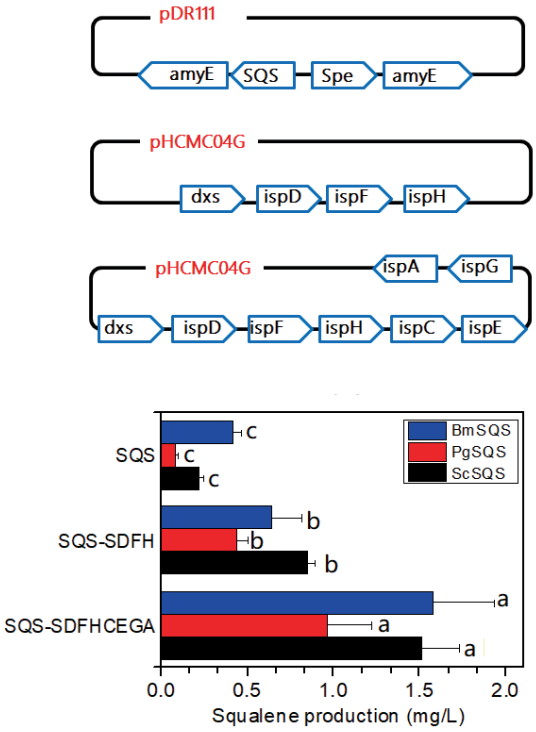


Figure 5. Effect of overexpressing MEP pathway genes on squalene production in *Bacillus subtilis.* *B. subtilis* strains were cultured at 37°C for 48 hours. SQSs were expressed in pDR111 and subsequently

integrated into genome of *B. subtilis*; MEP pathway related genes were expressed in pHCMC04G. S, C, D, E, F, G, H and A represents *dxs, ispC, ispD, ispE, ispF, ispG, ispH* and *ispA*, respectively. The second and third plasmids are pHCMC04G-SDFH and pHCMC04G-SDFHCEGA, respectively. Error bars represent standard deviations of biological triplicates. Different letters indicate significant statistical differences (Scott Knott 5%). BmSQS, PgSQS, and ScSQS are squalene synthases originating from *Bacillus megaterium, Panax ginseng* and *Saccharomyces cerevisiae*, respectively.

Improved production level of squalene by temperature optimization

Given that the SQS candidates originate from organisms living at different temperatures, it is useful to determine the optimal temperature for their expression and activity. To determine this, the influence of temperature on squalene production in *B. subtilis* was explored. Strains expressing SQSs, without and with MEP pathways genes, were fermented at 30ºC, 25ºC and 20ºC for 48 hours. Results (**Figure 6**) showed that, the squalene production increased when the culture temperature was decreased to 30 ºC or 25 ºC and performed best at 25ºC. The maximum squalene production was around 4 mg/L produced by DBM-MEP8, which is around 2.5-fold more than the yield found at 37ºC. When the temperature decreased to 20ºC, all of the strains showed similar squalene production as when being cultured at 37ºC. Considering that comparable squalene production was observed when strains were cultured at 30ºC and 25ºC, both of these temperatures were chosen for further experiments.

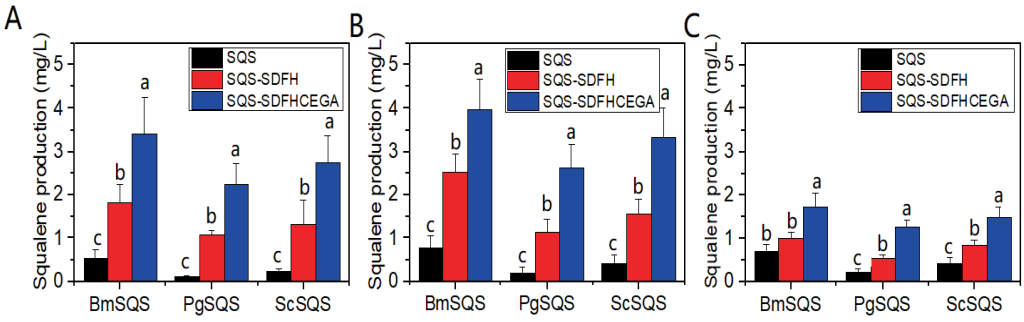


Figure 6. Effect of fermentation temperature on squalene production in *Bacillus subtilis*. A: Production of squalene in *B. subtilis* cultured 48 hours at 30ºC; **B**: Production of squalene in *B. subtilis* cultured 48 hours at 25ºC. C: Production of squalene in *B. subtilis* cultured 48 hours at 20ºC. SQSs were expressed in pDR111 and subsequently integrated into genome of *B. subtilis*; MEP pathway related genes were expressed in low-copy-number plasmid pHCMC04G. S, C, D, E, F, G, H and A represents *dxs, ispC, ispD, ispE, ispF, ispG, ispH* and *ispA*, respectively. Error bars represent standard deviations of biological triplicates. Different letters indicate significant statistical differences (Scott Knott 5%). BmSQS, PgSQS, and ScSQS are squalene synthases originating from *Bacillus megaterium, Panax ginseng* and *Saccharomyces cerevisiae*, respectively.

Improved production level of squalene by releasing rate-limiting factor in MEP pathway

In this step, investigation of whether further improvement of precursor level could improve squalene production was performed. Compared to pHCMC04G (5-6 units per chromosome), pBS0E has relatively high copy number (15-25 units per chromosome).[36-38] Hence, the vector pBS0E was employed to express MEP pathway genes and compare it to pHCMC04G. Four MEP pathway genes were first expressed in pBS0E (*dxs*, *ispD*, *ispF* and *ispH,* as pBS0E-SDFH) with SQSs. Results showed that overexpressing four MEP genes in high-copy-number plasmid pBS0E (pBS0E-SDFH) can lead to similar squalene production to the strain with eight genes overexpressed in low-copy-number plasmid pHCMC04G (pHCMC04-SDFHCEGA) (**Figure 6, 7**). To further evaluate contributions of *ispC, ispE, ispG* and *ispA* to terpenoids production, which have not been extensively investigated and discussed, each of them was combined individually with pBS0E-SDFH, respectively generating pBS0E-SDFHC, pBS0E-SDFHE, pBS0E-SDFHG and pBS0E-SDFHA. In this case, effects of each individual enzyme could be investigated, potential bottlenecks identified and released. Subsequently, these constructs were co-expressed with SQSs in *B. subtilis* and fermented for 48 hours at both 30ºC and 25ºC to test squalene production. Higher squalene productions were observed when strains were cultured at 30ºC (**Figure 7**). Results display that compared with pBS0E-SDFH containing strains, no enormous changes on squalene production were observed when pBS0E-SDFHC or pBS0E-SDFHE were overexpressed. Overexpression of pBS0E-SDFHA leads to 1.5- to 1.8-fold increase in squalene production, indicating that FPP concentration limited the synthesis of squalene. The maximum squalene production reached 7.5 mg/L acquired by BmSQS co-expressed with pBS0E-SDFHA. However, pBS0E-SDFHG overexpression decreases squalene production to 0.38- and 0.62- fold compared to pBS0E-SDFH strains.

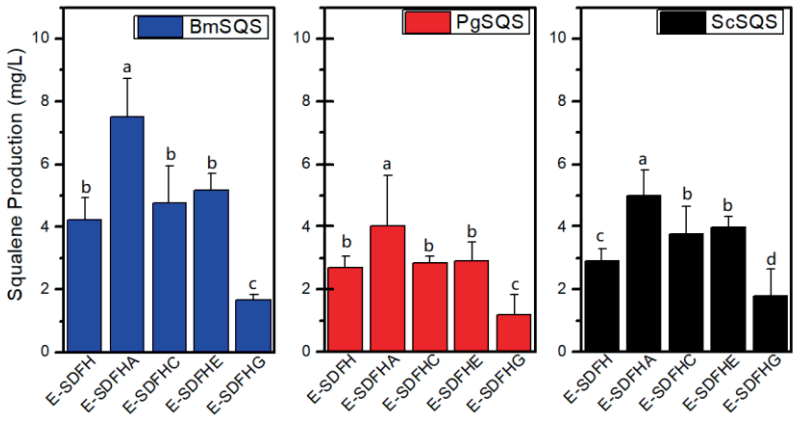


Figure 7. Effect of different combinations of MEP pathway genes in pBS0E plasmid on squalene production in *Bacillus subtilis* cultured 48 hours at 30°C. Error bars represent standard deviations of biological triplicates. Different letters indicate significant statistical differences (Scott Knott 5%). SQSs were expressed in pDR111 and subsequently integrated into genome of *B. subtilis*. MEP pathway related genes were overexpressed in high-copy-number plasmid pBS0E. S, C, D, E, F, G, Hand A represents *dxs, ispC, ispD, ispE, ispF, ispG, ispH* and *ispA*, respectively. BmSQS, PgSQS, and ScSQS are squalene synthases originating from *Bacillus megaterium, Panax ginseng* and *Saccharomyces cerevisiae*, respectively.

Discussion

Squalene is a pivot precursor for the biosynthesis of many triterpenoids and its synthesis is catalyzed by SQS. For decades, most researchers mainly focused on SQSs from eukaryotes such as human, yeast, rats and plants, where they were studied by recombinant expression, crystallization, and site-specific mutations to explore their catalytic sites. Limited efforts have been given to study prokaryotic SQSs, and only those from *Methylococcus capsulatus*, *Thermosynechococcus elongates*, and *Bradyrhizobium japonicum* have been cloned and expressed.[23,39,40] SQSs from multiple other microbial species have yet to be comprehensively and systematically investigated. Therefore, four SQS candidates from both prokaryotic and eukaryotic organisms were selected to compare and explore their capacity to synthesize squalene in *B. subtilis*.

In general, a lack of sequence homology among tested SQSs was observed after sequence alignment analysis (**Figure 2**). Canonical squalene synthases contain two conserved aspartate-rich motifs (DxxxD) associated with catalytic active sites,[23] and these two motifs could be observed in Ba (**Figure S2**). However, no squalene was detected in the extract from *B. subtilis* containing pHCMC04G-BaSQS or pDR111-BaSQS, even with the MEP pathway genes overexpressed. This result attracted our attention to “squalene synthase like” protein. Recently, it was discovered that FPP can be converted to squalene by three steps/three enzymes from the hopanoid biosynthesis pathway in the bacteria *Zymomonas mobilis* and *Rhodopseudomonas palustris*.[41] In this pathway, HpnD combines two molecules of FPP to form PSPP; then HpnC converts PSPP to hydroxysqualene (HSQ); and HpnE subsequently reduces HSQ to squalene (**Figure S3**). BaSQS is also annotated as squalene synthase HpnC according to the KEGG database and Uniprot, and sequence alignment of BaSQS with HpnC from *Z. mobilis* and *R. palustris* showed higher identities than when aligned with typical SQSs from *H. sapiens* and *M. capsulatus* (**Table S3**). Therefore, further exploring hydroxysqualene

synthase provides us new insights into understanding functions and characterizations of BaSQS. BmSQS is annotated as squalene/phytoene synthase, and it is the first SQS originating from a *Bacillus* species that has been validated to synthesize squalene. Interestingly, among the tested candidates expressed in *B. subtilis*, the highest squalene production was achieved by BmSQS. Considering that the first amino acid of the second aspartate-rich motif was not the conserved aspartate (**Figure S2**), mutation of this residue to aspartic acid provides a promising strategy to further improve the catalytic efficiency of BmSQS.

SQSs from eukaryotes (PgSQS and ScSQS) contain a TMR domain at C-terminus, which will target the protein to organelle membranes.[22,27] Thus, TMR regions of PgSQS and ScSQS were removed to permit their functional folding in the cytoplasm of bacteria. As expected, squalene could be readily detected in *B. subtilis* metabolites upon expression of truncated PgSQS or ScSQS. However, the squalene accumulations were not as high as in BmSQS strains. Identifying more non-essential domains of eukaryotic SQSs and truncating them proved to be vital for accumulating more squalene by improving enzyme expression, solubility, and activity.[20,42] This method has been demonstrated to be successful on human SQS where 30 N-terminal amino acids and 47 C-terminal amino acids have been deleted in *E. coli* leading to higher productions.[20, 43] In future, truncation both the N-terminal and C-terminal unnecessary residues could also be explored and tested on PgSQS and ScSQS in *B. subtilis*.

Compared to growth at 30ºC, the truncated human SQS obtained higher squalene production in *E. coli* when cultured at 37ºC.[41] In contrast, with the three tested SQSs expressed in *B. subtilis*, the highest squalene accumulations were observed at 25ºC or at 30ºC instead of culturing at 37ºC. And a similar observation was made on taxadiene synthase when expressed in *B. subtilis*.[8] *In vitro* experiments further demonstrated that both BmSQS and ScSQS showed highest catalytic activity at 30ºC instead of 37ºC (**Figure S4**). Taken together, properly decreasing cultivation temperature can serve as a candidate strategy to improve terpene synthases performance when expressed in *B. subtilis*.

Different levels of squalene production were observed, when three different types of plasmids were employed to express SQSs. It was reported that high level terpenoid production could be guaranteed with ample terpene synthesis pathway strength and minimized plasmid-born metabolic burden at the same time.[42, 43] In our study, the high copy number plasmid pHY300PLK might burden the growth of host cells, and its rolling circle replication made it unstable during long time cultivation, thus leading to the lowest level of squalene. The genome

integrative plasmid pDR111 performed best among the three tested plasmids. The stability of genome integrated expression cassette and the strong IPTG inducible promoter $P_{hyperspank}$ in pDR111 gave SQSs the advantages to reach higher squalene titers in *B. subtilis*. In addition, in-vitro assay showed that higher SQS activities were also measured in crude extracts from strains with higher squalene titers (**Figure S5**). Previous reports showed that overexpression of MEP pathway genes could dramatically increase terpenoid production in *B. subtilis*, including isoprene, carotenoids, amorphadiene, and taxadiene.[5,8,44,45] Similar consistent trends were observed with SQSs, regardless of fermentation temperature (37ºC, 30ºC, and 25ºC). The squalene production increased 3.4-, 5.7- and 3.8-fold, when four MEP pathway genes (*dxs, ispD, ispF* and *ispH*) were co-expressed with BmSQS, PgSQS and ScSQS and fermented at 25ºC, respectively. And these fold changes further increased to 5.3-, 13.2- and 8.3-fold when 8 MEP pathway related genes were overexpressed. These results indicate that enough supply of precursor is indispensable for high production of squalene. However, in *B. subtilis*, linearized C_{30} terpenoids (carotenoids) production could reach around 10 mg/g after four MEP pathway genes were overexpressed, and 20 mg/L amorphadiene could be achieved when additional *dxs* and *idi* were overexpressed. Therefore, it is deduced that the rate-limiting factor did not come from the precursor supply, instead it might exist within other metabolic factors or the step catalyzed by squalene synthase, for instance the insufficient supply or regeneration of NADPH in *B. subtilis*[46]. This hypothesis was confirmed by the *in vitro* assay result (Data not shown) demonstrating that squalene cannot be detected without additional NADPH added to the reaction samples, indicating that the NADPH concentration in the cell extract is insufficient to run the conversion.

Further experiments provided insights into effects of MEP pathway related genes on squalene production. First, IspA made significant contribution to precursor supply for squalene synthesis (**Figure 7, Figure S6**). IspA catalyzes isoprenoid chain elongation reactions, i.e. the formation of GPP and FPP. Previous results indicated that additional overexpression of *ispA* could increase C_{30} -terpenoids production, and it was demonstrated that this strategy also applies to squalene production improvement[6]. Second, more IspG overexpression led to negative effects on squalene production (**Figure 7**). IspG converts MEcPP to HMBPP, and subsequently HMBPP will form the basic isoprene precursor IPP and DMAPP catalyzed by IspH. The tremendous HMBPP accumulation could decrease terpenoids production in bacteria.[47] Li *et al.* reported that increased *ispG* gene expression led to decreased *β*-carotene production in *E. coli* due to toxicity of HMBPP, and this negative effect could be further eliminated by optimal

expression level of downstream gene *ispH* to consume HMBPP[47]. Next, *ispC* overexpression level should be screened and optimized to guarantee improved terpenoid production. IspC uses DXP as substrate to form MEP.[6] This study showed no sharp increase on squalene production when additional *ispC* was overexpressed. In contrast, previous results have been presented on effects of this enzyme. Xue *et al.* demonstrated IspC to be a rate-limiting factor in MEP pathway as a 5.5-fold increase of carotenoids was obtained when *ispC* was overexpressed in *B. subtilis 168*.[5] However, production level of isoprene remained unchanged with *ispC* overexpression in *B. subtilis DSM 10*.[44] It is theorized that different conclusions might be caused by varied expression levels of *ispC* in host strains, as effects of *ispC* overexpression could either increase or decrease lycopene production according to its overexpression levels in *E. coli*.[42] Overall, to further increase squalene production in *B. subtilis*, modulation of MEP pathway enzymes and fine-tuning of their expression levels, improving NADPH supply and regeneration, or protein engineering of SQS should be promising strategies in the future[48].

A significant increase of squalene production (approximately 29-fold) has been achieved in this study. However, this is still far behind the level produced by selected *Saccharomyces cerevisiae* strains, which yield around 2 g/L of squalene using 144 h fed-batch fermentation and optimized extraction.[33] Considering this is the first time that squalene being synthesized in *B. subtilis*, there are still many strategies that could be explored to further increase squalene production, such as metabolic engineering to release biosynthesis bottlenecks, squalene extraction methods, and optimization of fermentation conditions.In conclusion, this is the first time that squalene was synthesized in *B. subtilis* and different squalene synthases derived from plant and microbial sources were expressed and analyzed. Among the tested SQSs, the one from *B. megaterium* produced highest amount of squalene in *B. subtilis*. And when MEP pathway genes were overexpressed, the highest squalene production reached 7.5 mg/L after 48 hours of fermentation. IspA and IspG were shown to be critical factors that positively and negatively affect squalene production, respectively. This information provides important suggestions for further fine-tuning of the MEP pathway to increase production of squalene and its triterpenoid derivatives.

Abbreviations Used:

SQS, squalene synthase; IPP, isopentenyl pyrophosphate; DMAPP, dimethylallyl pyrophosphate; GPP, geranyl pyrophosphate; FPP, farnesyl pyrophosphate; GGPP,

geranylgeranyl pyrophosphate; MEP, 2-C-Methyl-D-erythritol 4-phosphate; HMBPP, 1-Hydroxy-2-methyl-2-butenyl 4-pyrophosphate; HSQ, hydroxysqualene

Competing Interests

The authors declare that they have no competing interests.

Acknowledgments

We thank P. Tepper for helpful suggestion on GC-MS analysis. Yafeng Song and Zheng Guan acknowledge funding from the China Scholarship Council. Hegar Pramastya is a recipient of Bernoulli sandwich scholarship from the University of Groningen.

References

(1) Schallmey, M.; Singh, A.; Ward, O. P., Developments in the use of *Bacillus* species for industrial production. *Can J Microbiol* **2004,** *50*, 1-17.

(2) Liu, L.; Liu, Y.; Shin, H. D.; Chen, R. R.; Wang, N. S.; Li, J.; Du, G.; Chen, J., Developing *Bacillus spp*. as a cell factory for production of microbial enzymes and industrially important biochemicals in the context of systems and synthetic biology. *Appl Microbiol Biotechnol* **2013,** *97*, 6113-27.

(3) Song, Y.; Fu, G.; Dong, H.; Li, J.; Du, Y.; Zhang, D., High-Efficiency Secretion of beta-Mannanase in *Bacillus subtilis* through Protein Synthesis and Secretion Optimization. *J Agric Food Chem* **2017,** *65*, 2540-2548.

(4) Kuzma, J.; Nemecek-Marshall, M.; Pollock, W. H.; Fall, R., Bacteria produce the volatile hydrocarbon isoprene. *Curr Microbiol* **1995,** *30*, 97-103.

(5) Xue, D.; Abdallah, II; de Haan, I. E.; Sibbald, M. J.; Quax, W. J., Enhanced C30 carotenoid production in *Bacillus subtilis* by systematic overexpression of MEP pathway genes. *Appl Microbiol Biotechnol* **2015,** *99*, 5907-15.

(6) Guan, Z.; Xue, D.; Abdallah, II; Dijkshoorn, L.; Setroikromo, R.; Lv, G.; Quax, W. J., Metabolic engineering of *Bacillus subtilis* for terpenoid production. *Appl Microbiol Biotechnol* **2015,** *99*, 9395-406.

(7) Ajikumar, P. K.; Tyo, K.; Carlsen, S.; Mucha, O.; Phon, T. H.; Stephanopoulos, G., Terpenoids: opportunities for biosynthesis of natural product drugs using engineered microorganisms. *Mol Pharm* **2008,** *5*, 167-90.

(8) Abdallah, II; Pramastya, H.; van Merkerk, R.; Sukrasno; Quax, W. J., Metabolic Engineering of *Bacillus subtilis* Toward Taxadiene Biosynthesis as the First Committed Step for Taxol Production. *Front Microbiol* **2019,** *10*, 218.

(9) Wagner, W. P.; Helmig, D.; Fall, R., Isoprene biosynthesis in *Bacillus subtilis* via the methylerythritol phosphate pathway. *J Nat Prod* **2000,** *63*, 37-40.

(10) Julsing, M. K.; Rijpkema, M.; Woerdenbag, H. J.; Quax, W. J.; Kayser, O., Functional analysis of genes involved in the biosynthesis of isoprene in *Bacillus subtilis*. *Appl Microbiol Biotechnol* **2007,** *75*, 1377-84.

(11) Withers, S. T.; Keasling, J. D., Biosynthesis and engineering of isoprenoid small molecules. *Appl Microbiol Biotechnol* **2007,** *73*, 980-90.

(12) Ghimire, G. P.; Thuan, N. H.; Koirala, N.; Sohng, J. K., Advances in Biochemistry and Microbial Production of Squalene and Its Derivatives. *J Microbiol Biotechnol* **2016,** *26*, 441-51.

(13) Gohil, N.; Bhattacharjee, G.; Khambhati, K.; Braddick, D.; Singh, V., Engineering Strategies in Microorganisms for the Enhanced Production of Squalene: Advances, Challenges and Opportunities. *Front Bioeng Biotechnol* **2019,** *7*, 50.

(14) Radisky, E. S.; Poulter, C. D., Squalene synthase: steady-state, pre-steady-state, and isotope-trapping studies. *Biochemistry* **2000,** *39*, 1748-60.

(15) Liu, C. I.; Jeng, W. Y.; Chang, W. J.; Shih, M. F.; Ko, T. P.; Wang, A. H., Structural insights into the catalytic mechanism of human squalene synthase. *Acta Crystallogr D Biol Crystallogr* **2014,** *70*, 231-41.

(16) Blagg, B. S.; Jarstfer, M. B.; Rogers, D. H.; Poulter, C. D., Recombinant squalene synthase. A mechanism for the rearrangement of presqualene diphosphate to squalene. *J Am Chem Soc* **2002,** *124*, 8846-53.

(17) Furubayashi, M.; Li, L.; Katabami, A.; Saito, K.; Umeno, D., Directed evolution of squalene synthase for dehydrosqualene biosynthesis. *FEBS Lett* **2014,** *588*, 3375-81.

(18) Furubayashi, M.; Li, L.; Katabami, A.; Saito, K.; Umeno, D., Construction of carotenoid biosynthetic pathways using squalene synthase. *FEBS Lett* **2014,** *588*, 436-42.

(19) LoGrasso, P. V.; Soltis, D. A.; Boettcher, B. R., Overexpression, purification, and kinetic characterization of a carboxyl-terminal-truncated yeast squalene synthetase. *Arch Biochem Biophys* **1993,** *307*, 193-9.

(20) Thompson, J. F.; Danley, D. E.; Mazzalupo, S.; Milos, P. M.; Lira, M. E.; Harwood, H. J., Jr., Truncation of human squalene synthase yields active, crystallizable protein. *Arch Biochem Biophys* **1998,** *350*, 283-90.

(21) Pandit, J.; Danley, D. E.; Schulte, G. K.; Mazzalupo, S.; Pauly, T. A.; Hayward, C. M.; Hamanaka, E. S.; Thompson, J. F.; Harwood, H. J., Jr., Crystal structure of human squalene synthase. A key enzyme in cholesterol biosynthesis. *J Biol Chem* **2000,** *275*, 30610-7.

(22) Linscott, K. B.; Niehaus, T. D.; Zhuang, X.; Bell, S. A.; Chappell, J., Mapping a kingdom-specific functional domain of squalene synthase. *Biochim Biophys Acta* **2016,** *1861*, 1049-57.

(23) Ohtake, K.; Saito, N.; Shibuya, S.; Kobayashi, W.; Amano, R.; Hirai, T.; Sasaki, S.; Nakano, C.; Hoshino, T., Biochemical characterization of the water-soluble squalene synthase from *Methylococcus capsulatus* and the functional analyses of its two DXXD(E)D motifs and the highly conserved aromatic amino acid residues. *FEBS J* **2014,** *281*, 5479-97.

(24) Hu, Y.; Jia, S.; Ren, F.; Huang, C. H.; Ko, T. P.; Mitchell, D. A.; Guo, R. T.; Zheng, Y., Crystallization and preliminary X-ray diffraction analysis of YisP protein from *Bacillus subtilis subsp. subtilis* strain 168. *Acta Crystallogr Sect F Struct Biol Cryst Commun* **2013,** *69*, 77-9.

(25) Feng, X.; Hu, Y.; Zheng, Y.; Zhu, W.; Li, K.; Huang, C. H., *et al*., Structural and functional analysis of *Bacillus subtilis* YisP reveals a role of its product in biofilm production. *Chem Biol* **2014,** *21*, 1557-63.

(26) Lee, M. H.; Jeong, J. H.; Seo, J. W.; Shin, C. G.; Kim, Y. S.; In, J. G.; Yang, D. C.; Yi, J. S.; Choi, Y. E., Enhanced triterpene and phytosterol biosynthesis in *Panax ginseng* overexpressing squalene synthase gene. *Plant Cell Physiol* **2004,** *45*, 976-84.

(27) Kim, T. D.; Han, J. Y.; Huh, G. H.; Choi, Y. E., Expression and functional characterization of three squalene synthase genes associated with saponin biosynthesis in *Panax ginseng*. *Plant Cell Physiol* **2011,** *52*, 125-37.

(28) You, C.; Zhang, X. Z.; Zhang, Y. H., Simple cloning via direct transformation of PCR product (DNA Multimer) to *Escherichia coli* and *Bacillus subtilis*. *Appl Environ Microbiol* **2012,** *78*, 1593-5.

(29) Kunst, F.; Rapoport, G., Salt stress is an environmental signal affecting degradative enzyme synthesis in *Bacillus subtilis*. *J Bacteriol* **1995,** *177*, 2403-7.

(30) Smart, K. F.; Aggio, R. B.; Van Houtte, J. R.; Villas-Boas, S. G., Analytical platform for metabolome analysis of microbial cells using methyl chloroformate derivatization followed by gas chromatography-mass spectrometry. *Nat Protoc* **2010,** *5*, 1709-29.

(31) Chang C., B. J., Li H., J. A., Crystal Structure of Squalene Synthase HpnC from *Alicyclobacillus Acidocaldarius*. **2012**.

(32) Sato, T.; Hoshino, H.; Yoshida, S.; Nakajima, M.; Hoshino, T., Bifunctional triterpene/sesquarterpene cyclase: tetraprenyl-beta-curcumene cyclase is also squalene cyclase in *Bacillus megaterium*. *J Am Chem Soc* **2011,** *133*, 17540-3.

(33) Han, J. Y.; Seo, S. H.; Song, J. M.; Lee, H.; Choi, E. S., High-level recombinant production of squalene using selected *Saccharomyces cerevisiae* strains. *J Ind Microbiol Biotechnol* **2018,** *45*, 239-251.

(34) Eppinger, M.; Bunk, B.; Johns, M. A.; Edirisinghe, J. N.; Kutumbaka, K. K.; Koenig, S. S., *et al*., Genome sequences of the biotechnologically important *Bacillus megaterium* strains QM B1551 and DSM319. *J Bacteriol* **2011,** *193*, 4199-213.

(35) Yoshida, K.; Ueda, S.; Maeda, I., Carotenoid production in *Bacillus subtilis* achieved by metabolic engineering. *Biotechnol Lett* **2009,** *31*, 1789-93.

(36) Nguyen, H. D.; Nguyen, Q. A.; Ferreira, R. C.; Ferreira, L. C.; Tran, L. T.; Schumann, W., Construction of plasmid-based expression vectors for *Bacillus subtilis* exhibiting full structural stability. *Plasmid* **2005,** *54*, 241-8.

(37) Titok, M. A.; Chapuis, J.; Selezneva, Y. V.; Lagodich, A. V.; Prokulevich, V. A.; Ehrlich, S. D.; Janniere, L., *Bacillus subtilis* soil isolates: plasmid replicon analysis and construction of a new theta-replicating vector. *Plasmid* **2003,** *49*, 53-62.

(38) Popp, P. F.; Dotzler, M.; Radeck, J.; Bartels, J.; Mascher, T., The *Bacillus* BioBrick Box 2.0: expanding the genetic toolbox for the standardized work with *Bacillus subtilis*. *Sci Rep* **2017,** *7*, 15058.

(39) Pan, J. J.; Solbiati, J. O.; Ramamoorthy, G.; Hillerich, B. S.; Seidel, R. D.; Cronan, J. E.; Almo, S. C.; Poulter, C. D., Biosynthesis of Squalene from Farnesyl Diphosphate in Bacteria: Three Steps Catalyzed by Three Enzymes. *ACS Cent Sci* **2015,** *1*, 77-82.

(40) Kang, J.; Zhang, Q.; Jiang, X.; Zhang, T.; Long, R.; Yang, Q.; Wang, Z., Molecular Cloning and Functional Identification of a Squalene Synthase Encoding Gene from Alfalfa (*Medicago sativa L.*). *Int J Mol Sci* **2019,** *20*.

(41) Katabami, A.; Li, L.; Iwasaki, M.; Furubayashi, M.; Saito, K.; Umeno, D., Production of squalene by squalene synthases and their truncated mutants in *Escherichia coli*. *J Biosci Bioeng* **2015,** *119*, 165-71.

(42) Kim, S. W.; Keasling, J. D., Metabolic engineering of the nonmevalonate isopentenyl diphosphate synthesis pathway in *Escherichia coli* enhances lycopene production. *Biotechnol Bioeng* **2001,** *72*, 408-15.

(43) Ajikumar, P. K.; Xiao, W. H.; Tyo, K. E.; Wang, Y.; Simeon, F.; Leonard, E.; Mucha, O.; Phon, T. H.; Pfeifer, B.; Stephanopoulos, G., Isoprenoid pathway optimization for Taxol precursor overproduction in *Escherichia coli*. *Science* **2010,** *330*, 70-4.

(44) Xue, J.; Ahring, B. K., Enhancing isoprene production by genetic modification of the 1-deoxy-d-xylulose-5-phosphate pathway in *Bacillus subtilis*. *Appl Environ Microbiol* **2011,** *77*, 2399-405.

(45) Zhou, K.; Zou, R.; Zhang, C.; Stephanopoulos, G.; Too, H. P., Optimization of amorphadiene synthesis in *Bacillus subtilis* via transcriptional, translational, and media modulation. *Biotechnol Bioeng* **2013,** *110*, 2556-61.

(46) Qi, H.; Li, S.; Zhao, S.; Huang, D.; Xia, M.; Wen, J., Model-driven redox pathway manipulation for improved isobutanol production in *Bacillus subtilis* complemented with experimental validation and metabolic profiling analysis. *PLoS One* **2014,** *9*, e93815.

(47) Li, Q.; Fan, F.; Gao, X.; Yang, C.; Bi, C.; Tang, J.; Liu, T.; Zhang, X., Balanced activation of IspG and IspH to eliminate MEP intermediate accumulation and improve isoprenoids production in *Escherichia coli*. *Metab Eng* **2017,** *44*, 13-21.

(48) Zhao, J.; Li, Q.; Sun, T.; Zhu, X.; Xu, H.; Tang, J.; Zhang, X.; Ma, Y., Engineering central metabolic modules of *Escherichia coli* for improving beta-carotene production. *Metab Eng* **2013,** *17*, 42-50.

Supporting Information

Materials and methods

GC-MS assay for catalytic activities of different crude SQS extracts

An *in-vitro* GC-MS assay was conducted to determine the catalytic activities of crude SQS enzymes. *B. subtilis* strains DBA, DBM, DPG and DSC, which contain the genes encoding respectively BaSQS, BmSQS, PgSQS, and ScSQS in the genome, were used to determine the catalytic activities of crude SQS extracts. Culture samples (1mL) were harvested after the *B. subtilis* strains had been cultured at 25ºC for 48 hours to obtain cell pellets. Then the cell pellets were lysed by lysis buffer (50µl lysis buffer per OD_{600}) containing: 50mM glucose, 25mM Tris-HCl (pH 8.0), 0.25mg/mL lysozyme, DNAse 0.01%, 2mM DTT, 1 cOmplete protease inhibitor (1 tablet per 50mL); and incubated for 1 hour at 37ºC. Then the supernatants were separated from the lyses by centrifugation (13000rpm, 10 min) and served as the crude enzyme extracts. For 0.5mL reaction of each sample containing crude extract enzymes 50µL in 10mM Tris-HCl buffer (pH 7.4), containing 10mM Mg^{2+}, 2mM DTT, 1 mM NADPH, and 46µM FPP substrate. The reaction samples were incubated at 30ºC (if not indicated otherwise) for 2 hours and stopped by addition of equal volume of cold methanol and 200µL of ethyl acetate containing cholesterol as internal standard. Then the reaction samples were centrifuged at 13000rpm for 2 minutes to obtain the supernatants. The supernatants were subsequently dried under nitrogen and dissolved in 100 µL of isopropanol (IPA)-acetonitrile (ACN) (7:3, v/v). Then samples were sent for GC-MS analysis.

Table S1. Plasmids used in this study.

Plasmids	Genotype and/or relevant characteristics	Sources/Reference
pHCMC04G	*B. subtilis* and *E. coli* shuttle vector; ori-pBR322; ori-pBS72; P_{xylA} xylose-inducible promoter; Cm^R; Amp^R	1
pHCMC04G-BaSQS	pHCMC04G derivative, squalene synthase originated from *Bacillus acidocaldarius*	This work
pHCMC04G-BmSQS	pHCMC04G derivative, squalene synthase originated from *Bacillus megaterium*	This work
pHCMC04G-PgSQS	pHCMC04G derivative, squalene synthase originated from *Panax ginseng*	This work
pHCMC04G-ScSQS	pHCMC04G derivative, squalene synthase originated from *Saccharomyces cerevisiae*	This work
pHY300PLK	*B. subtilis* and *E. coli* shuttle vector; ori-pACYC17; ori-pAMα1; Tc^R; Amp^R	2
pHY-BmSQS	pHY300PLK derivative, squalene synthase originated from *Bacillus megaterium*	This work
pHY-PgSQS	pHY300PLK derivative, squalene synthase originated from	This work

	Panax ginseng	
pHY-ScSQS	pHY300PLK derivative, squalene synthase originated from *Saccharomyces cerevisiae*	This work
pDR111	*B. subtilis* integration vector; ori-pBR322; $P_{hyperspank}$ IPTG-inducible promoter; Spe^R; Amp^R	3
pDR111-BaSQS	pDR111 derivative, squalene synthase originated from *Bacillus acidocaldarius*	This work
pDR111-BmSQS	pDR111 derivative, squalene synthase originated from *Bacillus megaterium*	This work
pDR111-PgSQS	pDR111 derivative, squalene synthase originated from *Panax ginseng*	This work
pDR111-ScSQS	pDR111 derivative, squalene synthase originated from *Saccharomyces cerevisiae*	This work
pHB201	*B. subtilis* and *E. coli* shuttle vector; ori-pUC19;ori-pTA1060 (rolling circle replication); P59 constitutive promoter; *cat86::lacZα*; Cm^R; Em^R	1
pHCMC04G-SDFH	pHCMC04G derivative, *dxs, ispD, ispF, ispH*	1
pHCMC04G-CEGA	pHCMC04G derivative, *ispC, ispE, ispG, ispA*	1
pHCMC04G-SDFHCEGA	pHCMC04G derivative, *dxs, ispD, ispF, ispH, ispC, ispE, ispG, ispA*	4
pBS0E	*B. subtilis* and *E. coli* shuttle vector; ori-1030 (theta replication); P_{xylA} xylose-inducible promoter; Erm^R; Amp^R	5
pBS0E-SDFH	pBS0E derivative, *dxs, ispD, ispF, ispH*	This work
pBS0E-SDFHC	pBS0E derivative, *dxs, ispD, ispF, ispH, ispC*	This work
pBS0E-SDFHE	pBS0E derivative, *dxs, ispD, ispF, ispH, ispE*	This work
pBS0E-SDFHG	pBS0E derivative, *dxs, ispD, ispF, ispH, ispG*	This work
pBS0E-SDFHA	pBS0E derivative, *dxs, ispD, ispF, ispH, ispA*	This work

Table S2. Strains used in this study.

Strains	Genotype and/or relevant characteristics	Sources
B. subtilis 168	*trpC2*	Lab stock
BC	*B. subtilis* 168 derivative, pHCMC04G, Cm^R	This work
BA	*B. subtilis* 168 derivative, pHCMC04G-BaSQS, Cm^R	This work
BM	*B. subtilis* 168 derivative, pHCMC04G-BmSQS, Cm^R	This work
PG	*B. subtilis* 168 derivative, pHCMC04G-PgSQS, Cm^R	This work
SC	*B. subtilis* 168 derivative, pHCMC04G-ScSQS, Cm^R	This work
HBM	*B. subtilis* 168 derivative, pHY-BmSQS, Tet^R	This work
HPG	*B. subtilis* 168 derivative, pHY-PgSQS,Tet^R	This work
HSC	*B. subtilis* 168 derivative, pHY-ScSQS, Tet^R	This work
DBA	*B. subtilis* 168 derivative, pDR111-BaSQS, Spe^R	This work
DBM	*B. subtilis* 168 derivative, pDR111-BmSQS, Spe^R	This work
DPG	*B. subtilis* 168 derivative, pDR111-PgSQS, Spe^R	This work
DSC	*B. subtilis* 168 derivative, pDR111-ScSQS, Spe^R	This work
DBA-MEP4	*B. subtilis* 168 derivative, pDR111-BaSQS, pHCMC04G-SDFH, Spe^R, Cm^R	This work
DBM-MEP4	*B. subtilis* 168 derivative, pDR111-BmSQS, pHCMC04G-SDFH, Spe^R, Cm^R	This work
DPG-MEP4	*B. subtilis* 168 derivative, pDR111-PgSQS, pHCMC04G-SDFH, Spe^R, Cm^R	This work
DSC-MEP4	*B. subtilis* 168 derivative, pDR111-ScSQS, pHCMC04G-SDFH, Spe^R, Cm^R	This work
DBA-MEP8	*B. subtilis* 168 derivative, pDR111-BaSQS, pHCMC04G-SDFHCEGA, Spe^R, Cm^R	This work
DBM-MEP8	*B. subtilis* 168 derivative, pDR111-BmSQS,	This work

	pHCMC04G-SDFHCEGA, Spe^R, Cm^R	
DPG-MEP8	*B. subtilis* 168 derivative, pDR111-PgSQS, pHCMC04G-SDFHCEGA, Spe^R, Cm^R	This work
DSC-MEP8	*B. subtilis* 168 derivative, pDR111-ScSQS, pHCMC04G-SDFHCEGA, Spe^R, Cm^R	This work
DBM-ESDFH	*B. subtilis* 168 derivative, pDR111-BmSQS, pBS0E-SDFH, Spe^R, Erm^R	This work
DPG-ESDFH	*B. subtilis* 168 derivative, pDR111-PgSQS, pBS0E-SDFH, Spe^R, Erm^R	This work
DSC-ESDFH	*B. subtilis* 168 derivative, pDR111-ScSQS, pBS0E-SDFH, Spe^R, Erm^R	This work
DBM-ESDFHC	*B. subtilis* 168 derivative, pDR111-BmSQS, pBS0E-SDFHC, Spe^R, Erm^R	This work
DPG-ESDFHC	*B. subtilis* 168 derivative, pDR111-PgSQS, pBS0E-SDFHC, Spe^R, Erm^R	This work
DSC-ESDFHC	*B. subtilis* 168 derivative, pDR111-ScSQS, pBS0E-SDFHC, Spe^R, Erm^R	This work
DBM-ESDFHE	*B. subtilis* 168 derivative, pDR111-BmSQS, pBS0E-SDFHE, Spe^R, Erm^R	This work
DPG-ESDFHE	*B. subtilis* 168 derivative, pDR111-PgSQS, pBS0E-SDFHE, Spe^R, Erm^R	This work
DSC-ESDFHE	*B. subtilis* 168 derivative, pDR111-ScSQS, pBS0E-SDFHE, Spe^R, Erm^R	This work
DBM-ESDFHG	*B. subtilis* 168 derivative, pDR111-BmSQS, pBS0E-SDFHG, Spe^R, Erm^R	This work
DPG-ESDFHG	*B. subtilis* 168 derivative, pDR111-PgSQS, pBS0E-SDFHG, Spe^R, Erm^R	This work
DSC-ESDFHG	*B. subtilis* 168 derivative, pDR111-ScSQS, pBS0E-SDFHG, Spe^R, Erm^R	This work
DBM-ESDFHA	*B. subtilis* 168 derivative, pDR111-BmSQS, pBS0E-SDFHA, Spe^R, Erm^R	This work
DPG-ESDFHA	*B. subtilis* 168 derivative, pDR111-PgSQS, pBS0E-SDFHA, Spe^R, Erm^R	This work
DSC-ESDFHA	*B. subtilis* 168 derivative, pDR111-ScSQS, pBS0E-SDFHA, Spe^R, Erm^R	This work
E. coli turbo	F' *proA B lacI*^{q} Δ*lacZM15 / fhuA2* Δ*(lac proAB) glnV galK16 galE15 R(zgb 210::Tn10)*Tet^S *endA1 thi-1* Δ*(hsdS-mcrB)5*	Lab stock

Table S3. Oligonucleotides used in this study.

Name	Sequences
Ba-F	gacaaatggtccaaactagtgataagaggaggagaaatatgggctcagttccggttgaactgag
Ba-R	catttccccctttgatttttagattcagtgatgatgatgatgatgtgctgatccgccttcgccttttgc
Bm-F	gacaaatggtccaaactagtgataagaggaggagaaatatgagcgttccgaataaactgcgcg
Bm-R	catttccccctttgatttttagattcagtgatgatgatgatgatgcatatcgacgacttcattgactg
Pg-F	gacaaatggtccaaactagtgataagaggaggagaaatatgggctcacttggcgcaattctgaaac
Pg-R	catttccccctttgatttttagattcagtgatgatgatgatgatgtgcgctattatggcctgattcgc
Sc-F	gacaaatggtccaaactagtgataagaggaggagaaatatgggcaaactgctgcaactggcactg
Sc-R	catttccccctttgatttttagattcagtgatgatgatgatgatgtttgtactcttcttcttgttgtgtc
SQS-F2	gggaaatgacaaatggtccaaactagtgataagaggaggagaaatatg
04-SQS-S	cattgaaataaacatttattttgtatatgatgagataaagttag
04-SQS-A	cctaataagccgatattagcctcgtatg
CO-SDFH-S	ccatttgtttaatctttaaattaagtatcaacatagtac
CO-SDFH-A	gattcattaatgcagctggcacgac
HYV-F	actagtcctctcttacggatcccc

HYV-R	gggagtagtctaagagaaagatgtgag
HYBm-F	ggggatccgtaagagaggactagtatgagcgttccgaataaactgcgcg
HYBm-R	cacatctttctcttagactactccctcagtgatgatgatgatgatgcatatcg
HYPg-F	ctcacatctttctcttagactactccctcagtgatgatgatgatgatgtgcgc
HYPg-R	ggggatccgtaagagaggactagtatgggctcacttggcgcaattctg
HYSc-F	ctcacatctttctcttagactactccctcagtgatgatgatgatgatgtttgtac
HYSc-R	cggggatccgtaagagaggactagtatgggcaaactgctgcaactggcac
DRV-F	taataatgagcactagtcaaggtcggc
DRV-R	gtttgtcctccttattagttaatcagctagc
DRBa-F	gccgaccttgactagtgctcattattagtgatgatgatgatgatgtgctgatccg
DRBa-R	gctgattaactaataaggaggacaaacatgggctcagttccggttgaactgagag
DRBm-F	gccgaccttgactagtgctcattattagtgatgatgatgatgatgcatatcgacg
DRBm-R	gctgattaactaataaggaggacaaacatgagcgttccgaataaactgcgcgataatg
DRPg-F	gctgattaactaataaggaggacaaacatgggctcacttggcgcaattctgaaac
DRPg-R	gccgaccttgactagtgctcattattagtgatgatgatgatgatgtgcgctattatg
DRSc-F	gccgaccttgactagtgctcattattagtgatgatgatgatgatgtttgtactcttc
DRSc-R	gctgattaactaataaggaggacaaacatgggcaaactgctgcaactggcactg
DRGV-F	catcatcatcatcatcactaataatgagcactagtc
DRGV-R	gccagaaccgcctttatacaattcatc
HY-SQS-S	cctatggaagttgatcagtcaacttatctg
HY-SQS-A	gcatgcgcaaccagttagatatgc
DR-SQS-S	gcacgaaaaaagcacccataagg
DR-SQS-A	gccgcgtttcggtgatgaagatc
DR-GSQS-S	gcacgaaaaaagcacccataagg
DR-GSQS-A	gatggtccagttttgttgccag
ESDFHV-F	gttttttgcttttacttttggaagtatttttttg
ESDFHV-R	cactagtagcggccgctgcaggca
ESDFH-F2	caaaaaaatacttccaaaagtaaaagcaaaaaactaacgcaagaggaggagaaat
ESDFHC-F	gtaaaagcaaaaaactaacgcaagaggaggagaaatatgaaaaatatttgtcttttag
ESDFHC-R	gcatgcctgcagcggccgctactagtgtgtgagtattgaattgacgtatccccg
ESDFHE-F	gtaaaagcaaaaaactaacgcaagaggaggagaaatatgcgtattttagaaaaagc
ESDFHE-R	tgcctgcagcggccgctactagtgatcaagagcgttctgttcgccgatc
ESDFHG-F	gtaaaagcaaaaaactaacgcaagaggaggagaaatatgcaagtgagtgaaatc
ESDFHG-R	atgcctgcagcggccgctactagtgagctttttgtgtttcttctttttaattttgc
ESDFHA-F	aaaagcaaaaaactaacgcaagaggaggagaaatatgacaaataaattaacgagc
ESDFHA-R	catgcctgcagcggccgctactagtggtgatctcttgccgcaattaaatcac
ESDFH-S	caggctttacactttatgcttccgg
ESDFH-A	gcagtttgatcacgaagatccatc

Table S4. Sequence alignment result of squalene synthase (SQS)/HpnC from different species.

Subject	Query	Identity (%)
Bacillus acidocaldarius	SQS from *Homo sapiens*	12.2
Bacillus acidocaldarius	SQS from *Methylococcus capsulatus*	17.3
Bacillus acidocaldarius	SQS from *Bacillus megaterium*	16.4
Bacillus acidocaldarius	SQS from *Panax ginseng*	12.8
Bacillus acidocaldarius	SQS from *Saccharomyces cerevisiae*	13.5
Bacillus acidocaldarius	HpnC from *Rhodopseudomonas Palustris*	24.9
Bacillus acidocaldarius	HpnC from *Zymomonas mobilis*	27.5

Figure S1

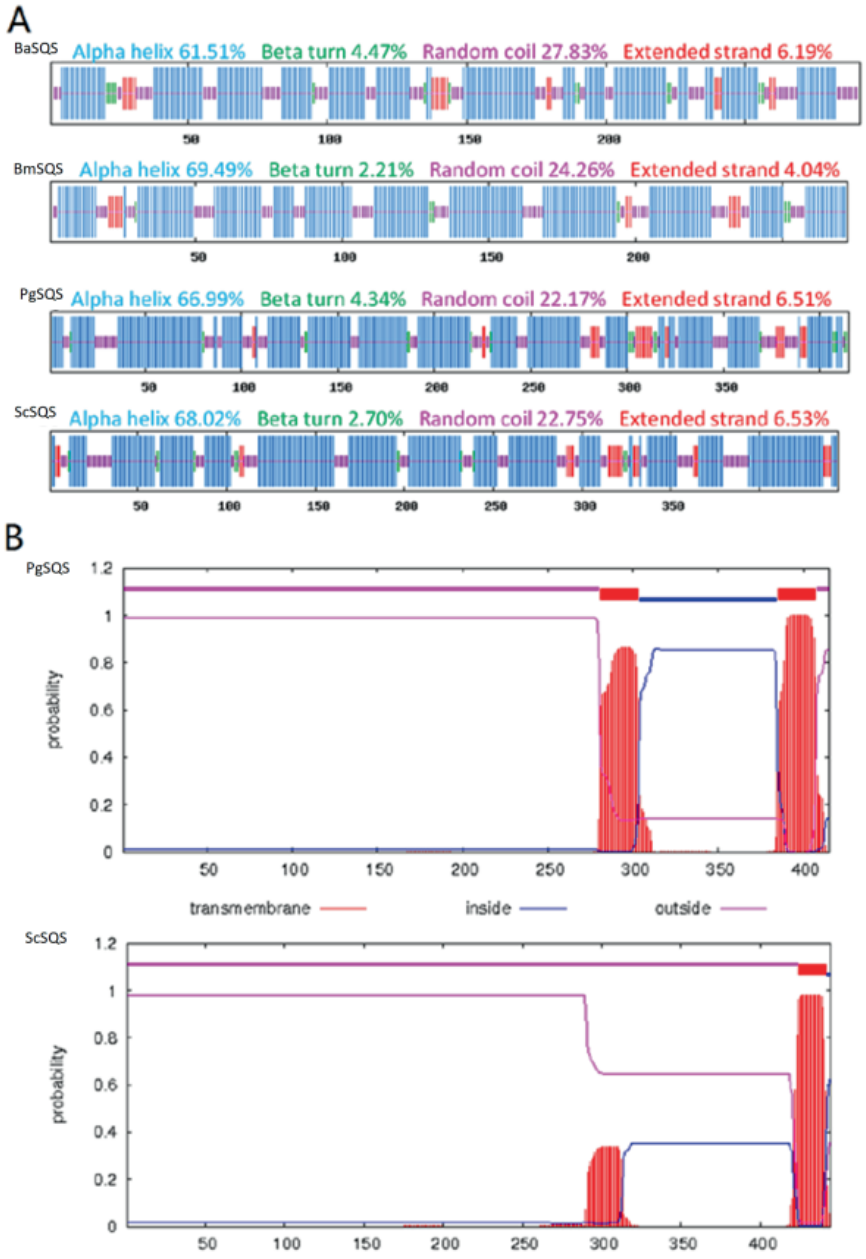


Figure S1. Analysis of the secondary structures of SQSs. A. Annotation of the secondary structures of SQS using SOPMA online server (https://npsa-prabi.ibcp.fr/cgi-bin/npsa_automat.pl?page=npsa_sopma.html). **B**. The predicted transmembrane regions of SQS using TMHMM Server (v. 2.0) (http://www.cbs.dtu.dk/services/TMHMM/). PgSQS, ScSQS, and HsSQS represents squalene synthase originating from *Panax ginseng, Saccharomyces cerevisiae*, *and Homo sapiens,* respectively.

Figure S2

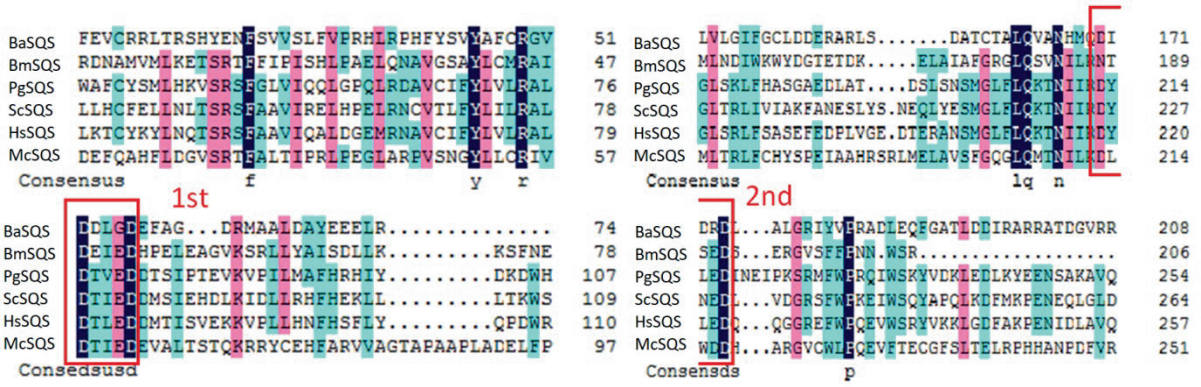


Figure S2. Sequence alignment of the squalene synthase candidates from different species. 1st and 2nd represent two conserved (predicted) aspartate-rich motifs "DxxxD". BaSQS, BmSQS, PgSQS, ScSQS, HsSQS, and McSQS represents squalene synthase originating from *Bacillus acidocaldarius*, *Bacillus megaterium, Panax ginseng, Saccharomyces cerevisiae*, *Homo sapiens and Methylococcus capsulatus,* respectively.

Figure S3

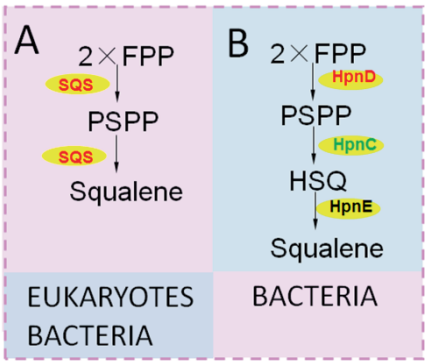


Figure S3. Biosynthesis pathway of squalene. A: Conversion of FPP to squalene in a two steps/one enzyme reaction by eukaryotic squalene synthase or bacterial squalene synthase, such as *Homo sapiens* and *Methylococcus capsulatus,*. **B**: Conversion of FPP to squalene in three steps/two enzymes reaction by bacterial squalene synthase from *Zymomonas mobilis* and *Rhodopseudomonas Palustris*. SQS: squalene synthase; FPP, Farnesyl pyrophosphate; PSPP: Presqualene pyrophosphate; HSQ: hydroxysqualene.

Figure S4

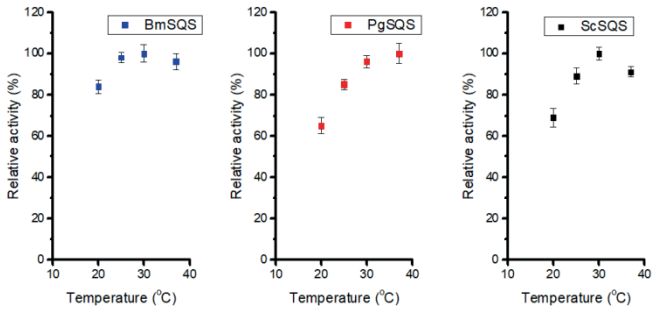


Figure S4. Effect of incubation temperature on the activity of crude SQS extracts. The crude enzyme extracts were prepared after the *B. subtilis* strains had been cultured 48 h at 25ºC. The *in vitro* reaction samples were incubated for 2 hours at 20ºC, 25 ºC, 30 ºC and 37 ºC, respectively. Error bars represent standard deviations of biological triplicates. Strains DBM, DPG and DSC, which contain the genes encoding respectively BmSQS, PgSQS, and ScSQS in the genome, were tested. BmSQS, PgSQS, and ScSQS are squalene synthases originating from *Bacillus megaterium, Panax ginseng* and *Saccharomyces cerevisiae*, respectively.

Figure S5

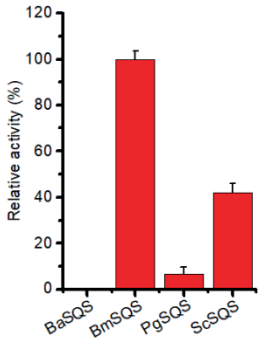


Figure S5. *In vitro* relative activity of crude SQS extracts. The crude enzyme extracts were prepared after the *B. subtilis* strains had been cultured 48 h at 25ºC. The *in vitro* reaction samples were incubated at 30ºC for 2 hours. Error bars represent standard deviations of biological triplicates. Strains DBA, DBM, DPG and DSC, which contain the genes encoding respectively BaSQS, BmSQS, PgSQS, and ScSQS in the genome, were tested. BaSQS, BmSQS, PgSQS, and ScSQS are squalene synthases originating from *Bacillus acidocaldarius, Bacillus megaterium, Panax ginseng* and *Saccharomyces cerevisiae*, respectively.

Figure S6

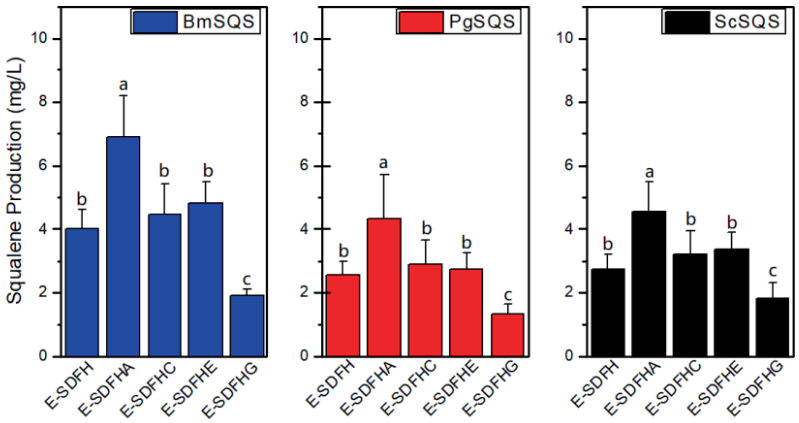


Figure S6. Effect of different combinations of MEP pathway genes in pBS0E plasmid on squalene production in *Bacillus subtilis* cultured 48 hours at 25ºC. Error bars represent standard deviations of biological triplicates. Different letters indicate significant statistical differences (Scott Knott 5%). MEP pathway related genes were overexpressed in pBS0E. S, C, D, E, F, G and A represents *dxs, ispC, ispD, ispE, ispF, ispG, ispH* and *ispA*, respectively. BmSQS, PgSQS, and ScSQS are squalene synthases originating from *Bacillus megaterium, Panax ginseng* and *Saccharomyces cerevisiae*, respectively.

References

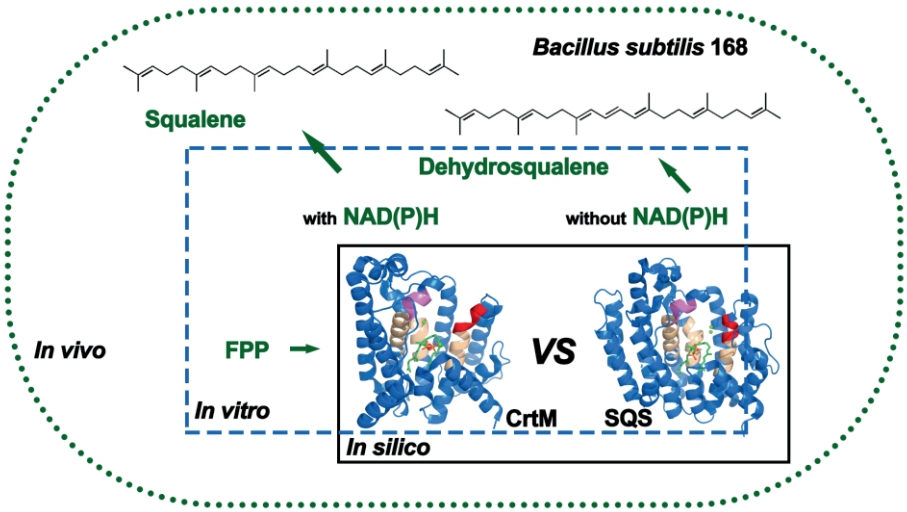
(1) Xue, D.; Abdallah, II; de Haan, I. E.; Sibbald, M. J.; Quax, W. J., Enhanced C30 carotenoid production in *Bacillus subtilis* by systematic overexpression of MEP pathway genes. *Applied microbiology and biotechnology* **2015**, 99, 5907-15.

(2) Yoshida, K.; Ueda, S.; Maeda, I., Carotenoid production in *Bacillus subtilis* achieved by metabolic engineering. *Biotechnology letters* **2009**, 31, 1789-93.

(3) Overkamp, W.; Beilharz, K.; Detert Oude Weme, R.; Solopova, A.; Karsens, H.; Kovacs, A.; Kok, J.; Kuipers, O. P.; Veening, J. W., Benchmarking various green fluorescent protein variants in *Bacillus subtilis, Streptococcus pneumoniae,* and *Lactococcus lactis* for live cell imaging. *Applied and environmental microbiology* **2013**, 79, 6481-90.

(4) Abdallah, II; Pramastya, H.; van Merkerk, R.; Sukrasno; Quax, W. J., Metabolic Engineering of *Bacillus subtilis* Toward Taxadiene Biosynthesis as the First Committed Step for Taxol Production. *Frontiers in microbiology* **2019**, 10, 218.

(5) Popp, P. F.; Dotzler, M.; Radeck, J.; Bartels, J.; Mascher, T., The *Bacillus* BioBrick Box 2.0: expanding the genetic toolbox for the standardized work with *Bacillus subtilis*. *Scientific reports* **2017**, 7, 15058.


Bacillus subtilis 168
Squalene
Dehydrosqualene
with NAD(P)H
without NAD(P)H
In vivo
FPP
In vitro
VS
CrtM
SQS
In silico