Supporting Information

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

Yafeng Song[1], Zheng Guan[1], Ronald van Merkerk[1], Hegar Pramastya[1,2], Ingy I. Abdallah[1,3], Rita Setroikromo[1], Wim J. Quax[1,]*

[1]Department of Chemical and Pharmaceutical Biology, Groningen Research Institute of Pharmacy, University of Groningen, Antonius Deusinglaan 1, 9713 AV Groningen, The Netherlands

[2]Pharmaceutical Biology Research Group, School of Pharmacy, Institut Teknologi Bandung, 40132, Bandung, Indonesia

[3]Department of Pharmacognosy, Faculty of Pharmacy, Alexandria University, Egypt

Corresponding author: Prof. Dr. Wim J. Quax

Table S1 Plasmids used in this study

Table S2 Strains used in this study

Table S3 Oligonucleotides used in this study

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

Figure S1 Analysis of the secondary structures of SQSs

Figure S2 Sequence alignment of the squalene synthase candidates from different species

Figure S3 Biosynthesis pathway of squalene

Figure S4 Effect of incubation temperature on the activity of crude SQS extracts

Figure S5 *In vitro* relative activity of crude SQS extracts

Figure S6 Effect of different combinations of MEP pathway genes in pBS0E plasmid on squalene production in *Bacillus subtilis* cultured 48 hours at 25°C

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 25°C 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 37°C. 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 30°C (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 *Panax ginseng*	This work
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, <u>*dxs, ispD, ispF, ispH*</u>	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, pHCMC04G-SDFHCEGA, Spe^R, Cm^R	This work
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 ΔlacZM15 / 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	catttccccctttgattttttagattcagtgatgatgatgatgatgtgctgatccgccttcgcctttttgc
Bm-F	gacaaatggtccaaactagtgataagaggaggagaaatatgagcgttccgaataaactgcgcg
Bm-R	catttcccccctttgattttttagattcagtgatgatgatgatgatgcatatcgacgacttcattgactg
Pg-F	gacaaatggtccaaactagtgataagaggaggagaaatatgggctcacttggcgcaattctgaaac
Pg-R	catttcccccctttgattttttagattcagtgatgatgatgatgatgtgcgctattatggcctgattcgc
Sc-F	gacaaatggtccaaactagtgataagaggaggagaaatatgggcaaactgctgcaactggcactg
Sc-R	catttcccccctttgattttttagattcagtgatgatgatgatgatgtttgtactcttcttcttgttgtgtc
SQS-F2	gggaaatgacaaatggtccaaactagtgataagaggaggagaaatatg

04-SQS-S	cattgaaataaacatttatttttgtatatgatgagataaagttag
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	gttttttgcttttactttttggaagtattttttttg
ESDFHV-R	cactagtagcggccgctgcaggca
ESDFH-F2	caaaaaaatacttccaaaagtaaaagcaaaaaactaacgcaagaggaggagaaat
ESDFHC-F	gtaaaagcaaaaaactaacgcaagaggaggagaaatatgaaaaatatttgtcttttag
ESDFHC-R	gcatgcctgcagcggccgctactagtgtgtgagtattgaattgacgtatccccg
ESDFHE-F	gtaaaagcaaaaaactaacgcaagaggaggagaaatatgcgtattttagaaaaagc
ESDFHE-R	tgcctgcagcggccgctactagtgatcaagagcgttctgttcgccgatc
ESDFHG-F	gtaaaagcaaaaaactaacgcaagaggaggagaaatatgcaagtgagtgaaatc
ESDFHG-R	atgcctgcagcggccgctactagtgagctttttgtgtttcttcttttaattttgc
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
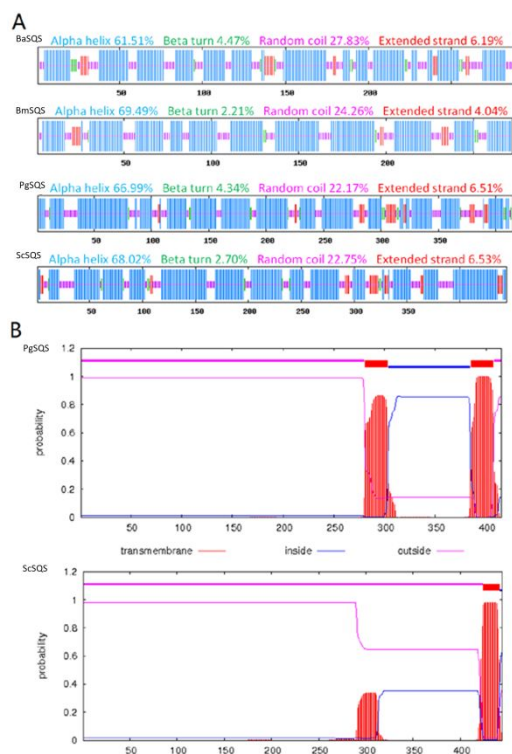
Figure S1 Analysis of the secondary structures of SQSs. **A**. Annotation of

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,*
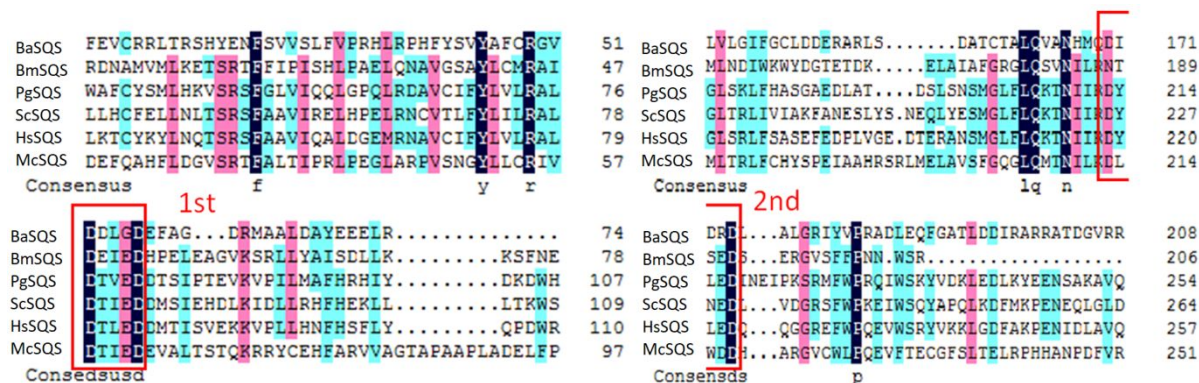
Saccharomyces cerevisiae, and Homo sapiens, respectively.

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.
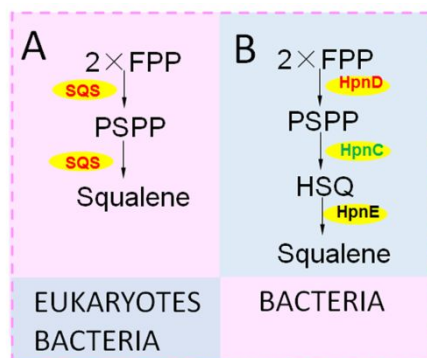
Figure S3

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.
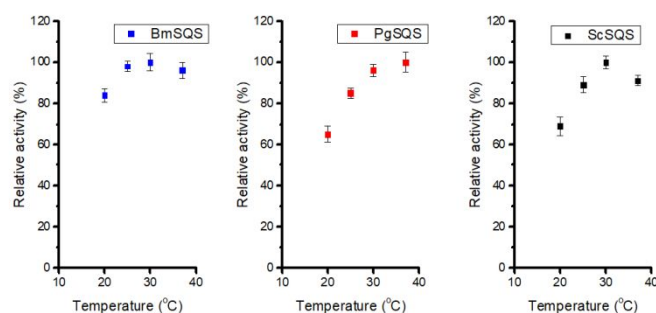
Figure S4

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 25°C. The *in vitro* reaction samples were incubated for 2 hours at 20°C, 25 $^\circ$C, 30 $^\circ$C and 37 $^\circ$C, 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.
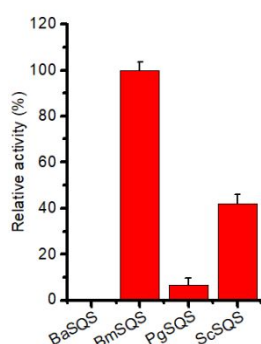
Figure S5

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 25°C. The *in vitro* reaction samples were incubated at 30°C 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.
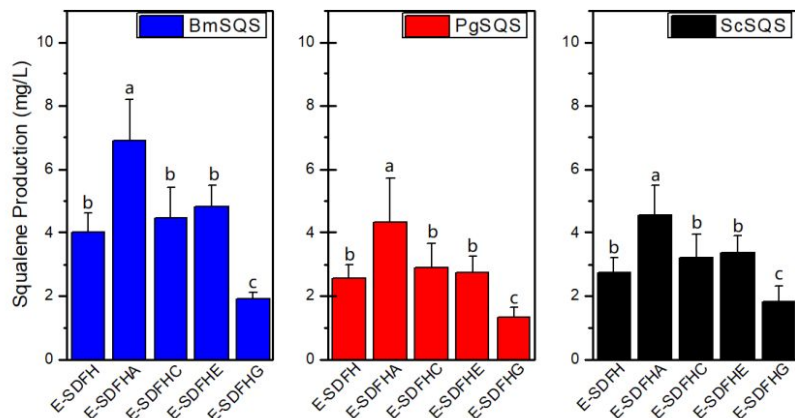
Figure S6

Figure S6

Effect of different combinations of MEP pathway genes in pBS0E plasmid on squalene production in *Bacillus subtilis* cultured 48 hours at 25°C. 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.