Supporting Information for:

A modular approach for facile biosynthesis of labdane-related diterpenes

Anthony Cyr, P. Ross Wilderman, Mara Determan, Reuben J. Peters*
Iowa State University, Department of Biochemistry, Biophysics, and Molecular Biology, Ames, IA 50011

General.

Unless otherwise noted all chemicals were purchased from Fisher Scientific (Loughborough, Leicestershire, UK) and molecular biology reagents from Invitrogen (Carlsbad, CA, USA).

Vector construction.

pGG

The pseudo-mature geranylgeranyl diphosphate synthase construct rAgGGPS[1] was PCR amplified using primers to introduce a 5' NdeI restriction site that overlapping with the initiating ATG codon and a 3' XhoI site immediately following the stop codon. The rAgGGPS PCR product and pACYCDuet (Novagen, San Diego, CA) were digested first with NdeI and then XhoI. The resulting large fragments were gel purified and then ligated together to create pGG. The inserted rAgGGPS, which falls within the second pACYCDuet multiple cloning site (MCS2), was verified by complete sequencing.

pGGAS

The pseudo-mature abietadiene synthase rAgAS[2] was inserted into MCS1 from pACYCDuet in pGG. A 5' CAT overhang at the 5' end of rAgAS, where the AT is part of the initiating ATG codon, was created by sticky-end PCR[3] and a 3' NotI site immediately following the stop codon also introduced. After digestion of the resulting sticky-end PCR product with NotI, this rAgAS construct was ligated into NcoI/NotI digested pGG to create pGGAS, and the presence of the rAgAS insert verified by sequencing the 5' and 3' ends.

pGGeC

The pseduo-mature *ent*-copalyl diphosphate synthase rAtCPS[4] was similarly inserted into MCS1 from pACYCDuet in pGG (i.e. using sticky-end PCR to create a 5' CAT overhang and 3' NotI site for digestion and ligation into NcoI/NotI digested pGG) to create pGG*e*C. The inserted rAtCPS was also verified by complete sequencing.

pGGsC

Three different pseduo-mature constructs, missing 70, 82, or 86 amino acid (aa) residues, of the *syn*-copalyl diphosphate synthase OsCPS4[5] were created (Δ70, Δ82, and Δ86, respectively). Preliminary analysis suggested that OsCPS4Δ70, which parallels rAtCPS in being truncated 32 aa residues upstream of a conserved SAYDT motif, exhibited the optimal combination of expression level and biochemical activity. Both NcoI sites (at nucleotides 903 and 1205) of OsCPS4Δ70 (rOsCPS4) were removed by synonymous site-directed mutagenesis using PCR amplification with overlapping mutagenic primers. The resulting rOsCPS4(-NcoI) construct was verified by complete sequencing, then PCR amplified with primers that introduced a 5' NcoI site

overlapping with the initiating ATG codon and a 3' NotI site immediately after the stop codon. This PCR product, along with pGG, was digested with NcoI and then NotI, and the resulting large fragments gel purified and then ligated together to create pGG*s*C. The inserted rOsCPS4 was verified by complete sequencing.

pGGnC

Both NcoI sites of rAgAS (at nucleotides 972 and 1630) were removed by synonymous site-directed mutagenesis using PCR amplification with overlapping mutagenic primers. The resulting rAgAS(-NcoI) construct was verified by complete sequencing and the D621A mutant[6] (re)created by overlapping PCR site-directed mutagenesis. This rAgAS:D621A construct was PCR amplified with primers that introduced a 5' NcoI site overlapping with the initiating ATG codon and a 3' BamHI site immediately after the stop codon. This PCR product, along with pGG, was digested with NcoI and then BamHI, and the resulting large fragments gel purified and then ligated together to create pGGnC. The inserted rAgAS:D621A was verified by complete sequencing.

Recombinant strains.

Analysis

Gas chromatography-mass spectrometry (GC-MS) was performed on organic extracts using an HP1-MS column on an Agilent (Palo Alto, CA) 6890N GC instrument with 5973N mass selective detector in electron-ionization mode (70eV) located in the W.M. Keck Metabolomics Research laboratory at Iowa State University, much as previously described.[5] Briefly, 5 µL of sample was injected at 40°C in splitless mode, the oven temperature held at 40°C for 3 min., then raised at 20°C/min. to 300°C, and held there for 3 min. MS data was collected from 50 to 500 *m/z* during the temperature ramp and final hold. The biosynthetically produced diterpenes were identified by comparison of retention time and mass spectra to authentic samples (see Supporting Figure). The amount of diterpene produced was determined by comparison of the biosynthetic organic extracts to a standard curve constructed with the use of known quantities of cembrene (Acros Organics, Geel, Belgium).

Transformation and expression

Previous work has established that the OverExpress C41 strain of *E. coli* (Lucigen, Middleton, WI) is well suited to T7 promoter based expression of plant derived labdane-related diterpene synthases.[4] Thus, all work in this study has been carried out with the use of this C41 strain. Chemically competent C41 cells (10 µL) were typically co-transformed with 1 µL of each plasmid and grown with dual antibiotic (34 µg/mL chloramphenicol and 50 µg/mL carbenicillin) selection on NZY media. Expression of the encoded genes was induced by the addition of IPTG.

Optimization

A number of factors were examined for their effect on diterpene production in the utilized shake flask culture growths. In particular, with the *ent*-kaur-16-ene (**6**) producing C41 strain transformed with pGG*e*C + pDEST15/AtKS. First, while all cultures were grown to mid-log phase (0.4-0.6 A_{600}) at 37°C, the effect of varying temperature from 15-30°C for expression of the introduced metabolic pathway was examined, and 20°C found to be optimum. Second, the amount of IPTG inducer, as a proxy for enzymatic expression level, was varied between 0.05 - 1

mM, and 0.5 mM found to be optimal. Thus, all cultures were grown to mid-log phase at 37°C, then shifted to 20°C for 1 hour prior to induction of the introduced metabolic pathway with 0.5 mM IPTG. Finally, the addition of glycerol to a final concentration of 5 g/L was found to also increase diterpene yield.

Secretion of diterpenes enables ready isolation from mixed phase culture growths

Expecting the produced hydrocarbon compounds to partition into the membrane, our initial analyses of diterpene production solely focused on organic solvent extraction of cell pellets from shake flask growths. However, more detailed analysis of the *ent*-kaur-16-ene (**6**) producing C41 strain transformed with pGG*e*C + pDEST15/AtKS revealed that ~90% of the diterpene resides in the spent culture media (i.e. the liquid remaining after the cells were removed by centrifugal pelleting). The spent media also provided a cleaner source of diterpene than cell pellets. In addition, the use of mixed phase cultures, specifically liquid media growth cultures containing adsorbent resin (i.e. liquid and solid phases), for harvesting a similarly secreted sesquiterpenoid has been previously reported.[7] We also found that inclusion of hydrophobic Diaion HP-20 beads (Supleco, Bellefonte, PA) enabled ready isolation of the biosynthetically produced diterpenes. Upon analysis of such mixed phase cultures with 1-5% (wt/vol) Diaion HP-20 beads, growths with 2% Diaion HP-20 beads were found to be optimal, increasing the observed yield ~2-fold from that obtained with liquid only cultures (presumably by preventing the loss of diterpene to volatilization, e.g. kaurene is emitted by several plant species,[8] and similar loss to volatilization has been reported for *E. coli* engineered to produce the sesquiterpene amorpha-4,11-diene[9]). The Diaion HP-20 beads from a typical 50 mL mixed phase culture growth were removed by filtration through a #100 sieve (Hogentogler & Co, Columbia, MD), washed in 25 mL dH_2O, again isolated by filtration, and then washed in 25 mL EtOH, followed by transfer to a glass wool plugged pasteur pipet. The isolated beads were then further washed with EtOH until all pigmentation is removed, i.e. the eluant is clear (10-15 column volumes). The beads were then dried by the application of a gentle flow of nitrogen gas prior to elution with 6 mL hexanes. A ~1 mL silica gel column with ~0.1 mL anhydrous magnesium sulfate overlay pre-washed with ~2 mL hexanes was prepared, and the organic extract passed over this column, which was then further washed with an additional ~2 mL hexanes. The pooled extract and wash was then dried under nitrogen gas and resuspended in 1 mL of hexanes for GC-MS analysis. Typical yields from this protocol were 5-100 μg/L of mixed phase culture growth (see Supporting Table), with the diterpene generally representing the most abundant extracted compound (e.g. Figure 1). Analysis of strains carrying pGG*e*C + either a mutant or wild-type rAtKS indicate that product yield corresponds to KS activity measured in vitro,[10] suggesting that the variation in yield observed here may largely reflect underlying changes in total activity, due to differences in catalytic rate and/or expression level, of the various class I diterpene synthases utilized in these studies. Notably, the yields at the higher end of the range observed here are similar to that previously reported for *E. coli* engineered to produce the sesquiterpene amorpha-4,11-diene prior to any further optimization of isoprenoid precursor supply (ca ~100 μg/L in similar shake-flask production runs), although a >100-fold in yield was observed upon such optimization.[11] Thus, future work will be focused on similarly increasing isoprenoid precursor supply within the context of the modular approach to labdane-related diterpene biosynthesis demonstrated here.

Literature References

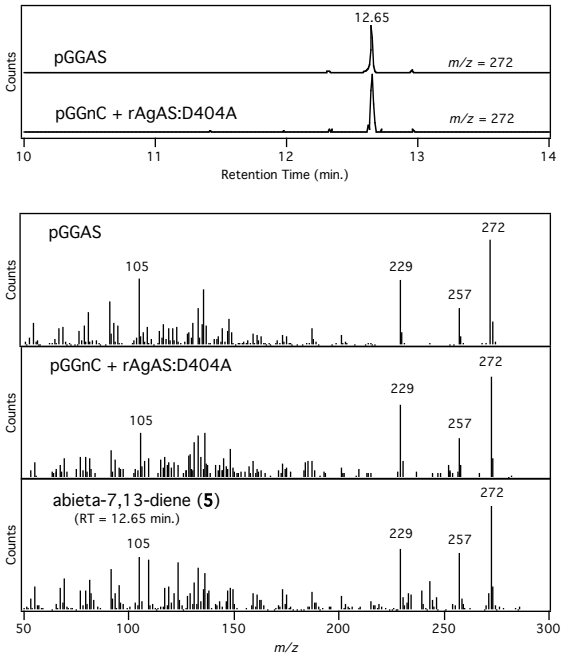
(1) Burke, C.; Croteau, R. *J. Biol. Chem.* **2002**, *277*, 3141-3149.
(2) Peters, R. J.; Flory, J. E.; Jetter, R.; Ravn, M. M.; Lee, H.-J.; Coates, R. M.; Croteau, R. B. *Biochemistry* **2000**, *39*, 15592-15602.
(3) Zeng, G. *Biotechniques* **1998**, *25*, 206-208.
(4) Prisic, S.; Peters, R. J. *Plant Physiol.* **2007**, *in press.*
(5) Xu, M.; Hillwig, M. L.; Prisic, S.; Coates, R. M.; Peters, R. J. *Plant J.* **2004**, *39*, 309-318.
(6) Peters, R. J.; Ravn, M. M.; Coates, R. M.; Croteau, R. B. *J. Am. Chem. Soc.* **2001**, *123*, 8974-8978.
(7) Jackson, B. E.; Hart-Wells, E. A.; Matsuda, S. P. T. *Org. Lett.* **2003**, *5*, 1629-1632.
(8) Otsuka, M.; Kenmoku, H.; Ogawa, M.; Okada, K.; Mitsuhashi, W.; Sassa, T.; Kamiya, Y.; Toyomasu, T.; Yamaguchi, S. *Plant Cell Physiol.* **2004**, *45*, 1129-1138.
(9) Newman, J. D.; Marshall, J.; Chang, M.; Nowroozi, R.; Paradise, E.; Pitera, D.; Newman, K. L.; Keasling, J. D. *Biotechnol. Bioeng.* **2006**, *95*, 684-691.
(10) Xu, M.; Wilderman, P. R.; Peters, R. J. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *in press.*
(11) Martin, V. J. J.; Pitera, D. J.; Withers, S. T.; Newman, J. D.; Keasling, J. D. *Nat. Biotechnol.* **2003**, *21*, 796-802.

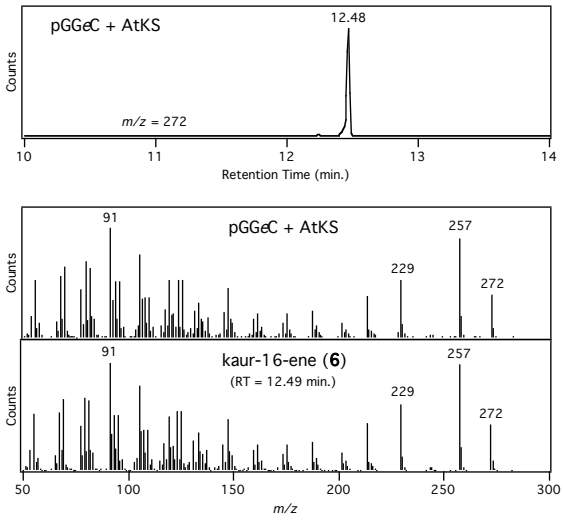
Supporting Table: Characterization of engineered *E. coli.*

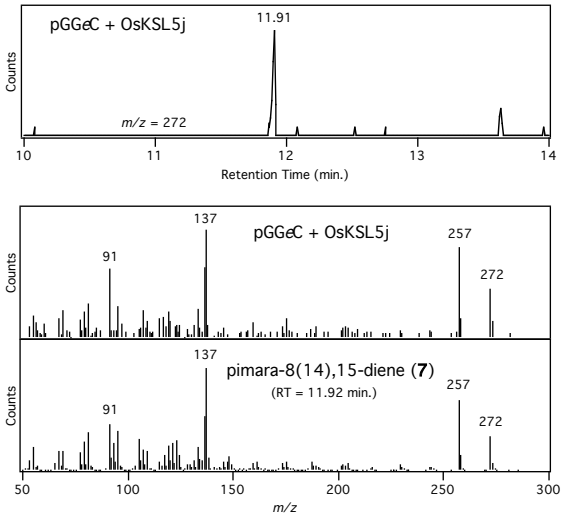
Plasmid(s)	Product	Yield (μg/L culture)*
pGGAS	abieta-7,13-diene (**5**)	50
pGG*e*C + pDEST15/rAtKS	*ent*-kaur-16-ene (**6**)	95
pGG*e*C + pDEST15/OsKSL5j	*ent*-pimara-8(14),15-diene (**7**)	11
pGG*e*C + pTH1/OsKSL6	*ent*-isokaur-15-ene (**8**)	81
pGG*e*C + pDEST15/OsKSL7	*ent*-cassa-12,15-diene (**9**)	67
pGG*e*C + pDEST15/OsKSL10	*ent*-sandaracopimaradiene (**10**)	90
pGG*s*C + pDEST15/OsKSL4	*syn*-pimara-7,15-diene (**11**)	32
pGG*s*C + pTH8/OsKSL8	*syn*-stemar-13-ene (**12**)	20
pGG*s*C + pDEST15/OsKSL11	*syn*-stemod-13(17)-ene (**13**)	70
pGGnC + pDEST14/rAgAS:D404A	abieta-7,13-diene (**5**)	5

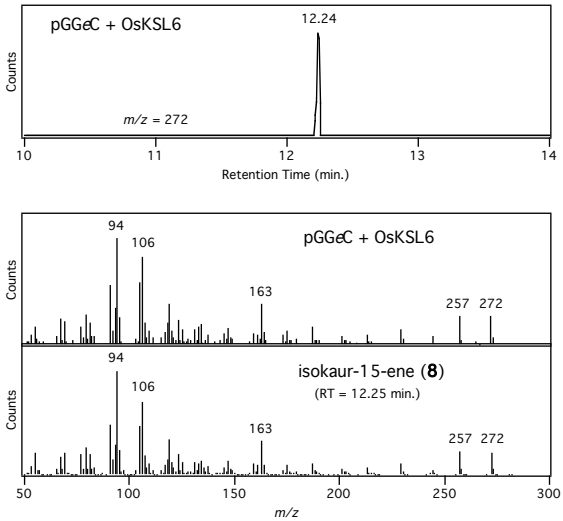
*Yield from repeated runs is typically within 20% of that reported here.

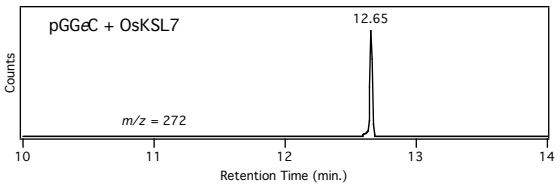
Supporting Figure: Selective ion chromatograms from GC-MS analysis of metabolically engineered *E. coli*, with the retention time (RT) of the biosynthetically produced diterpene indicated. Also shown are the mass spectra for the biosynthetically produced diterpenes and authentic reference compounds (with their RT indicated in parentheses).

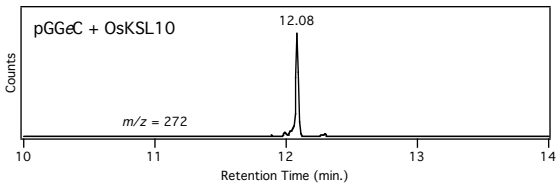


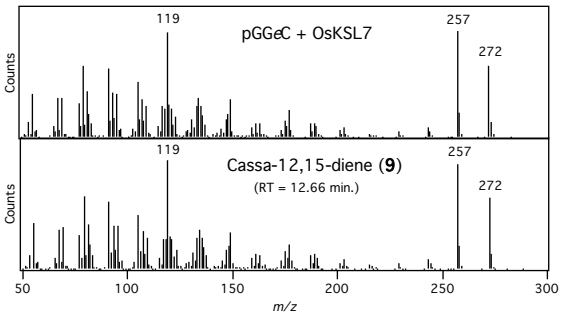


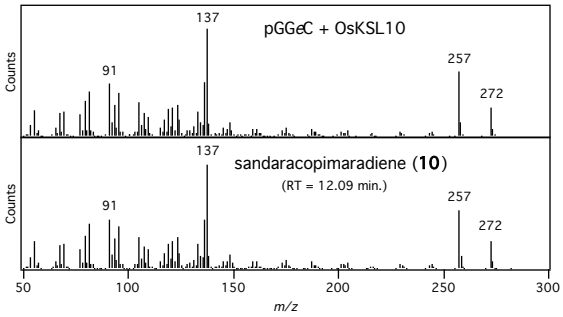


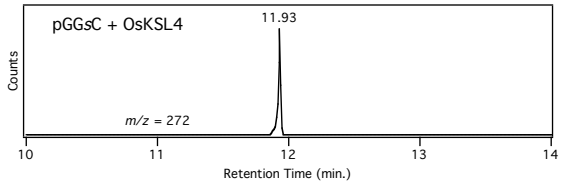


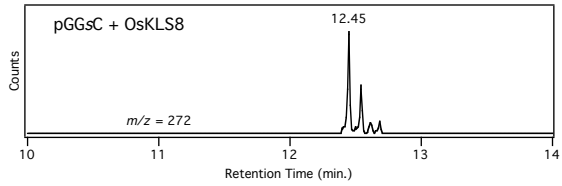

pGGeC + OsKSL7
12.65
Counts
m/z = 272
10
11
12
13
14
Retention Time (min.)

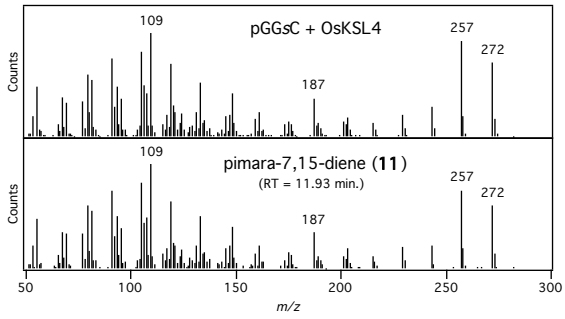

pGGeC + OsKSL10
12.08
Counts
m/z = 272
10
11
12
13
14
Retention Time (min.)

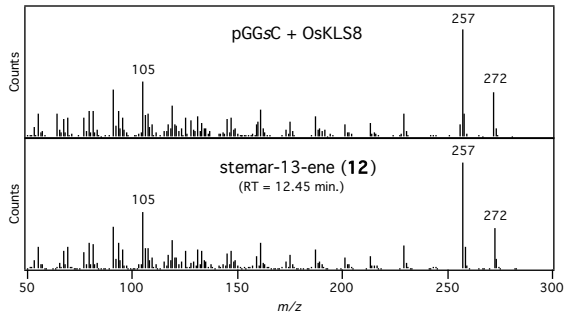

pGGeC + OsKSL7
119
257
272
Counts
Cassa-12,15-diene (9)
(RT = 12.66 min.)
119
257
272
Counts
50
100
150
200
250
300
m/z

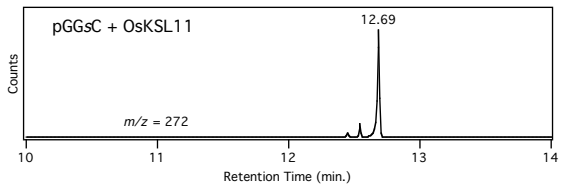

pGGeC + OsKSL10
137
91
257
272
Counts
sandaracopimaradiene (10)
(RT = 12.09 min.)
137
91
257
272
Counts
50
100
150
200
250
300
m/z

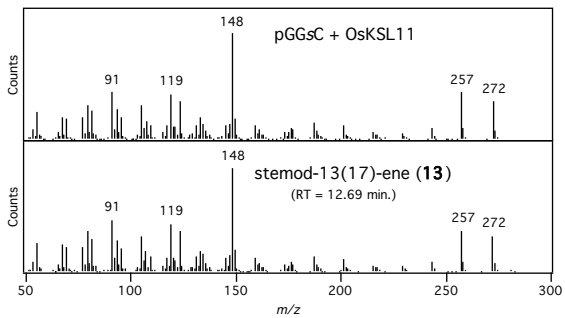

pGGsC + OsKSL4
11.93
Counts
m/z = 272
10
11
12
13
14
Retention Time (min.)


pGGsC + OsKLS8
12.45
Counts
m/z = 272
10
11
12
13
14
Retention Time (min.)


pGGsC + OsKSL4
109
187
257
272
Counts
pimara-7,15-diene (11)
(RT = 11.93 min.)
109
187
257
272
Counts
50
100
150
200
250
300
m/z


pGGsC + OsKLS8
105
257
272
Counts
stemar-13-ene (12)
(RT = 12.45 min.)
105
257
272
Counts
50
100
150
200
250
300
m/z


pGGsC + OsKSL11
12.69
Counts
m/z = 272
10
11
12
13
14
Retention Time (min.)


pGGsC + OsKSL11
148
91
119
257
272
Counts
stemod-13(17)-ene (13)
(RT = 12.69 min.)
148
91
119
257
272
Counts
50
100
150
200
250
300
m/z