Engineering triterpene metabolism in the oilseed of Arabidopsis thaliana

Chase Kempinski¹,² and Joe Chappell¹,²,*

¹Plant Biology Program, University of Kentucky, Lexington, KY, USA
²Department of Pharmaceutical Sciences, University of Kentucky, Lexington, KY, USA

Summary
Squalene and botryococcene are linear, hydrocarbon triterpenes that have industrial and medicinal values. While natural sources for these compounds exist, there is a pressing need for robust, renewable production platforms. Oils are an excellent target for heterologous production because of their roles as natural storage repositories and their capacity to produce precursors from photosynthetically-derived carbon. We generated transgenic Arabidopsis thaliana plants using a variety of engineering strategies (subcellular targeting and gene stacking) to assess the potential for oils to produce these two compounds. Constructs used seed-specific promoters and evaluated expression of a Triterpene synthase alone and in conjunction with a farnesyl diphosphate synthase (FPS) plus 1-deoxyxylulose 5-phosphate synthase (DXS). Constructs directing biosynthesis to the cytosol to harness isoprenoid precursors from the mevalonic acid (MVA) pathway were compared to those directing biosynthesis to the plastid compartment diverting precursors from the methylerythritol phosphate (MEP) pathway. On average, the highest accumulation for both compounds was achieved by targeting the triterpene synthase, FPS and DXS to the plastid (526.84 µg/g seed for botryococcene and 227.30 µg/g seed for squalene). Interestingly, a higher level accumulation of botryococcene (a non-native compound) was observed when the biosynthetic enzymes were targeted to the cytosol (>1000 µg/g seed in one line), but not squalene (natively produced in the cytosol). Not only do these results indicate the potential for engineering triterpene accumulation in oils seeds, but they also uncover some the unique regulatory mechanisms controlling triterpene metabolism in different cellular compartments of seeds.

Keywords: isoprenoid, triterpene, metabolic engineering, oilseed, seed-specific.

Introduction
Isoprenoids or terpenoids comprise one of the largest classes of natural products with tens of thousands of unique molecules currently identified and estimates that these represent 1%–10% of the total number of possible species. This large catalog of compounds arises, in part, from regio- and stereo-chemistry diversification imposed during their biosynthesis, as well secondary modifications like hydroxylation, acylation, aroylation and glycosylation. While the biosynthesis of this chemical family has recently been reviewed (Hemmerlin et al., 2012; Kempinski et al., 2015; Vranová et al., 2013), isoprenoid biosynthesis proceeds via the generation of the universal 5-carbon building blocks, isopentenyl diphosphate (IPP) and dimethylallyl diphosphate (DMAPP), which are then polymerized together to yield families of molecules. The condensation of two isoprenyl units, IPP and DMAPP, yields the ten carbon prenyl chain geranyl diphosphate (GPP) by GPP synthase (GPS). GPP is the precursor to monoterpenes. Condensation of two IPPs with one DMAPP generates the fifteen carbon prenyl chain farnesyl diphosphate (FPP) by FPP synthase (FPS), and FPP is the precursor to the sesquiterpene family of compounds. The head-to-head condensation of two FPPs by squalene synthase gives rise to squalene, the precursor to a complex array of triterpenes including sterols. Three IPP and one DMAPP can be condensed to produce the twenty carbon intermediate geranylgeranyl diphosphate (GGPP) by GGPP synthase (GGPPS), the precursor to the diterpenes. The forty carbon tetraterpenes are derived from the condensation of two GGPPs to yield phytoene by the squalene synthase-like reaction catalysed by phytoene synthase. In plants, an interesting subcellular division of isoprenoid biosynthesis exists: the MVA pathway resides in the cytosol, whereas the MEP pathway is localized to the plastids, both of which give rise to the initial prenyl units, IPP and DMAPP. However, sesqui- and tri-terpenes are derived from FPP biosynthesized in the cytosol nearly exclusively from the MVA pathway, whereas mono-, di- and tetra-terpenes are biosynthesized in the plastids from precursors provided by the MEP pathway. These two pathways are understood to normally operate independently of one another, but there have been documented cases of cross-talk and possible sharing of intermediates or basic prenyl units (Arigoni et al., 1997; Flügge and Gao, 2005; Hemmerlin et al., 2003; Opitz et al., 2014). Triterpenes arise from the head-to-head (1′–1 bond) condensation of two molecules of FPP, and in all eukaryotes this is catalysed by the enzyme squalene synthase (SQS) to produce the symmetrical, linear compound squalene. Squalene can be then be epoxidized by squalene epoxidase (SQE) to produce 2,3-oxidosqualene, the next committed precursor to all sterols and triterpenes derived from oxidosqualene synthases (OSCs) (Nes, 2011; Xue et al., 2012). In the green algae, Botryococcus braunii (race B), FPP can be dimerized in an alternative fashion (1′–3 bond) by squalene synthase-like (SSL) enzymes, to produce the C30 compound, botryococcene. More specifically SSL-1 generates the presqualene diphosphate intermediate that is then reductively rearranged by SSL-3 using NADPH to yield botryococcene (Bell et al., 2014; Niehaus et al., 2011). In B. braunii,
botryococcene and squalene can be further methylated in a SAM-dependent manner to generate up to tetramethylated, linear triterpenes (Metzger and Largeau, 2005; Niehaus et al., 2012). Both squalene and botryococcene (and especially their carbon-dense methylated derivatives) are of particular interest because of their capacity to be catalytically or hydro-cracked and distilled into hydrocarbons that can be used as fuels in current internal combustion engines or as feedstocks for the petrochemical industry (Hillen et al., 1982; Tracy et al., 2011). In addition, squalene is also of considerable interest due to its use in cosmetics, vaccine adjuvants and purported nutritional benefits (Spanova and Daum, 2011).

Oilseeds comprise an important part of human diets throughout the world and are an important part of the world economy. World oilseed production has increased more than 80 million metric tons from 2011 to 2016 and world demand for oilseed crops for food and industrial uses are expected to increase following increases in population and gross domestic product (USDA, 2016). The increased demand for oilseeds will increase the need to obtain higher yielding or added value seeds in the future, irrespective of whether these seeds will be used for food or industrial uses. Arabidopsis thaliana is a member of the Brassicaceae family and is closely related to other important oilseed crops, such as the well-established and important crop, Brassica napus and the emerging oilseed crop, Camelina sativa (Bansal and Durret, 2016; Dalal et al., 2015; Li et al., 2006). Arabidopsis is an excellent model to evaluate seed-specific engineering schemes due to its physically desirable qualities as a model organism (i.e. small stature, rapid lifespan and prolific seed set) and ease of genetic transformation (Bechtold and Pelletier, 1998; Clough and Bent, 1998; Harrison et al., 2006; Zhang et al., 2006).

Because of their dietary and nutritional roles, attempts at seed-specific engineering have primarily focused on two main areas: enhancing oil yieldaltering oil composition and enhancing/ introducing carotenoid biosynthesis in seed tissue (Burkhhardt et al., 1997; Fujisawa et al., 2009; Paine et al., 2005; Shewmaker et al., 1999, Ye et al., 2000). Phytosterol levels have also been enhanced through seed engineering, but this has been demonstrated in Nicotiana tabacum rather than an oilseed (Holmberg et al., 2002, 2003). The efforts to alter phytosterols in the N. tabacum seeds was undertaken by introducing the sterol methyltransferase 1 (SMT1) and/or the 3-hydroxy-3-methylglutaryl-CoA reductase (HMGR). HMGR is an important rate-limiting step of the MVA pathway operating in the cytosol. SMT1 has been demonstrated to control phytosterol composition in N. tabacum leaves, but not overall phytosterol levels (Sitbon and Jonsson, 2001). In contrast, when over-expressed in a seed-specific manner total phytosterols were increased up to 44% in N. tabacum (Holmberg et al., 2002). Tocopherols and tocotrienols are important nutritional antioxidants, both derived from the MEP pathway, whose levels have also been enhanced through seed-specific engineering of a homogentisate phytol transferase (the rate limiting step in tocopherol production; Savidge et al., 2002) and over-expression of a heterologous homogentic acid geranylgeranyl transferase (the committed step in tocotrienol biosynthesis) in Zea mays seeds resulting in up to sixfold elevated levels of tocotrienols and tocopherols (Cahoon et al., 2003).

While phytosterols, tocopherols/tocotrienols and carotenoids are all biosynthesized utilizing isoprenoid precursors, up to this point there have been few published attempts to introduce the biosynthesis of ‘non-native’ isoprenoids in seeds. However, Augustin et al. (2015) introduced production of the novel monoterpene, (4S)-limonene and the novel sesquiterpene, (+)-α-cadinene, into Camelina seeds by expressing limonene synthase (LS) and (+)-α-cadinene synthase (CDNS). Each of these genes was under a seed-specific promoter, NAPIN (Josefsson et al., 1987) and GLYCININ (Nielsen et al., 1989) respectively. They did this by targeting the terpene synthase enzymes to the cytosol or the plastid compartments in conjunction with the corresponding prenyltransferase, LS + GPS or CDNS + FPS resulting in 1.5–3 mg of limonene/g seed and 2.4–3.0 mg of cadinene/g seed. They also generated plastid-targeted constructs that included a DXS in conjunction with the prenyl transferase and terpene synthase which boosted terpene production by approximately twofold to ~5 mg/g seed for both terpenes. Borghi and Xie (2016) also used seed-specific promoters (BANYULS and FRUITFULL) to drive LS expression in Camelina seeds. They expressed only LS and obtained average titres much lower than that reported by Augustin et al. (2015): ~20 ng/g seed for LS under the BANYULS promoter. This could partly be explained by the difference in expression patterns of the seed-specific promoters: BANYULS is expressed in the seed coat (Nesi et al., 2009), whereas NAPIN and GLYCININ are expressed in developing embryos (Fernandez et al., 1991; Nielsen et al., 1989) where each cell type may have critical differences in the availability of metabolic precursors.

We wished to evaluate the feasibility of utilizing an oilseed platform to produce the triterpenes botryococcene and squalene. Because these compounds are pure hydrocarbons it seems that an oil-rich cell would be a natural repository and it may also have benefits in their extraction methodology for the future—due to the nature of in which oil can be expressed from oilseeds. We built on engineering strategies developed previously (Augustin et al., 2015; Jiang et al., 2016; Wu et al., 2006, 2012) in which we targeted the triterpene synthase, SQS, for squalene biosynthesis or a fusion enzyme of SSL-1 + SSL-3 (Niehaus et al., 2011) (which will be referred to here as botryococcene synthase [BS]), for botryococcene biosynthesis, to either the cytosol or the plastid to take advantage of IPP/DMAPP derived from the MVA or MEP pathways, respectively. Hence, the work reported here evaluates design principles for engineering efficient, high-level production of triterpenes in seeds, paving the way for seeds as possible production platforms for high-value triterpenes.

Results
Plastid-targeted engineering of botryococcene and squalene results in accumulation in seeds, which can be increased by including a heterologous DXS

We designed constructs that introduced a heterologous triterpene synthase only (BS or SQS for the cytosol and tpBS or tpSQS for the plastid) or in combination with a heterologous FPS (BS + FPS, tpBS + tpFPS, SQS + FPS, tpSQS + tpFPS), as well as adding a plastid-targeted heterologous DXS into the plastid-targeted constructs (tpBS + tpFPS + tpDXS or tpSQS + tpFPS + tpDXS) (see Figure 1 for construct schematics). These enzymes were all under their own seed-specific promoter and transformed into Arabidopsis (Figure 2). We allowed plants to self and examined segregation ratios of T2 plants germinated on hygromycin B (Tables S1 and S2), and moved resistant plants to soil. The T2 seed lots that these individual T2 plants produced were screened for triterpene accumulation (see Figure 2 for a diagram of the experimental design). We observed botryococcene accumulation in plants with biosynthesis targeted to the cytosol and plastid, with the highest amounts on average being in the plastid in conjunction with FPS

Engineering triterpene metabolism in Arabidopsis seeds 387
and DXS (Figure 3). To help verify that we assayed enough independently generated lines to account for genomic position effect variation in transgene expression, we created normal score plots for botryococcene accumulating lines (Figure 4). We hypothesized that if botryococcene accumulation truly represented the capability to redirect isoprenoid metabolism in these conditions, then we would expect normal distributions in the botryococcene titers of the populations examined. This would account for lower and higher than expected expression amounts due to position effect variation (or even multiple gene copies). In contrast to botryococcene, we only observed squalene accumulation when SQS was targeted to the plastid in conjunction with FPS, but similar to botryococcene, the highest average titres were observed in the tpSQS + tpFPS + tpDXS engineered lines (Figure 5).

Engineering the triterpene synthases with a heterologous FPS into the plastid resulted in substantial accumulation for both botryococcene and squalene (Figures 3 and 5); although squalene yields were lower than that for botryococcene. Accumulation of both triterpenes in the plastid was dependent on the inclusion of a co-expressed FPS which is consistent with previous reports (Augustin et al., 2015; Jiang et al., 2016; Wu et al., 2006, 2012). This is due to very low levels of endogenous FPP being produced within the plastid. While there was some botryococcene accumulation (average 6.22 ± 0.851 μg/g seed) in seeds engineered with only tpBS (where none was expected) this could be due to low levels of FPP in the plastid or mis-targeted enzyme (which could access endogenous FPP produced elsewhere within the cell) which is similar to the results observed by Jiang et al. (2016) where they expressed tpBS in N. tabacum leaves. In order to verify the seed-specific nature of our constructs, we extracted several T1 plant lines’ leaf tissue and assayed for botryococcene accumulation (we chose botryococcene as it is a foreign molecule whereas squalene would be native to the leaf tissue and a small fraction should be present). In some of the T1 lines assayed we did see botryococcene accumulating in the leaves, but only in very minor amounts (Table 1). We also observed minor ectopic expression of the various transgenes in leaf tissue using RT-PCR but none of this expression followed a specific pattern and appeared stochastic, which we interpret as probably due to position-dependent effects (data not shown).

It has been previously reported that DXS may be a rate-limiting step in carbon flux through the MEP pathway (Hemmerlin et al., 2012; Paetzold et al., 2010; Wright et al., 2014) and isoprenoid engineering in the plastids of the related oilseed plant, Camelina sativa, showed increased levels of target isoprenoids with the addition of a co-expressed DXS (Augustin et al., 2015). Because of these observations, we also introduced a plastid-targeted DXS (tpDXS) into the plastid-targeted gene constructs of tpBS + tpFPS and tpSQS + tpFPS. Similar to what was previously seen with seed engineering in Camelina, including tpDXS caused an increase in target triterpenoid accumulation for both squalene and botryococcene. The inclusion of tpDXS seemed to boost triterpene content slightly over twofold for both types of triterpene, again with botryococcene accumulating to a higher amount than squalene (~500 μg botryococcene/g seed vs. ~200 μg squalene/g seed; Figures S1 and S2).

Cytosolic-targeted engineering in seeds results in botryococcene accumulation, but only marginal squalene accumulation

In order to further evaluate the capabilities of producing linear triterpene in oilseeds we also directed our enzymes into the cytosol to utilize endogenous, MVA pathway produced, FPP. We also introduced genetic cassettes that co-expressed an FPS enzyme which could utilize IPP/DMAPP produced by the MVA pathway to increase the total FPP available to the triterpene synthase. We observed substantial accumulation of botryococcene in the cytosol with BS alone (~250 μg/g seed) and this
increased on average with inclusion of an FPS (325 μg botryococcene/g seed; Figure 3). Although the inclusion of FPS did not appear to universally increase the amount of botryococcene in the cytosol in all the transgenic lines evaluated, the maximum capacity for botryococcene biosynthesis using MVA-derived carbon was improved (Figure S1, compare A and B). These results were unexpected given previous work on engineering botryococcene in the cytosol of N. tabacum leaves (Jiang et al., 2016). However, engineering of the FPP-derived sesquiterpene, (−)-δ-cadinene, in the Camelina seeds did result in substantial product accumulation (400–500 μg/g seed), although this was 10-fold lower than their plastidic-targeting strategy which included co-expressing the sesquiterpene synthase, FPS and DXS enzymes (Augustin et al., 2015).

Interestingly, attempting to target squalene accumulation to the cytosol did not result in any substantial squalene accumulation beyond what is normally found in wild-type Col-0 seeds (~5 μg/g seed; Figure 6). Squalene is normally formed within the cytosol and in association with the Endoplasmic Reticulum (ER) as it is the precursor to phytosterols (and other triterpenes) and thus any squalene formed may be acted upon by squalene monooxygenase and other integral-membrane, enzymes resulting in.

**Figure 2** Schematic showing study set up and plant generation designations. All seed generations after transformation were germinated and selected on hygromycin B. The initial plant number is the independent transformation event for that particular construct, with each daughter plant in the following generation indicated by iterative addition of a hyphen and number. Thus, for example plant 3-2-5 would be plant 5 (T3 generation), a daughter from plant 2 (T2 generation), which was a daughter of a T1 plant from independent transformation event 3.

**Figure 3** Average botryococcene accumulation (μg/g seed; ±SE) obtained with each construct design. Values are the averages of the highest yielding T3 seed lots out of the four assayed (like that shown in Figure S1), representing the average of independently transformed homozygous seed lots obtained with each gene cassette. The number of independently generated transgenic lines (n) per construct is noted as well. N.D., none detected.
accumulation of downstream phytosterols and triterpenes. In order to test this, we examined the levels of major phytosterols (campesterol, stigmasterol and β-sitosterol) in seeds of T₄ transgenic lines engineered with BS, BS + FPS, tpBS + tpFPS, SQS + FPS, tpSQS + tpFPS and wild type (Figure 6). Consistent with Nguyen et al. (2013), high level expression of a heterologous SQS in Arabidopsis seeds resulted in modestly elevated levels of total major phytosterols in only one line examined (Figure 6b). The cytosolic BS and BS + FPS lines accumulating high levels of botryococcene did not show any significant difference in total phytosterol levels compared to wild type. However, there was a slight overall reduction in phytosterol levels in the BS only line examined, suggesting that FPP normally utilized for phytosterol biosynthesis might be rerouted to botryococcene biosynthesis, reducing phytosterol levels in this line. However, the phytosterol levels in the BS + FPS lines accumulating high amounts of

Table 1  Botryococcene accumulation in Arabidopsis leaf tissue of T₁ lines that were transformed with the indicated genetic construct. The number of independent lines extracted is indicated and the number of those independent lines which accumulated botryococcene is shown. The average (µg/g FW) of those lines which accumulated botryococcene is also presented.

| Engineered enzymes | T₁ lines extracted | T₁ lines with botryococcene | Average (µg/g FW) | SD | SE |
|--------------------|--------------------|-----------------------------|------------------|----|----|
| BS                 | 18                 | 8                           | 1.37             | 0.94 | 0.33 |
| BS + FPS           | 18                 | 1                           | 3.03             | NA | NA |
| tpBS + tpFPS + tpDXS | 9                  | 0                           | ND               | NA | NA |
botryococcene were approximately the same as wild type, indicating that flux through the MVA pathway was possibly increased in proportion to the IPP/DMAPP diverted to botryococcene biosynthesis. This hints at the metabolic plasticity of the MVA pathway in Arabidopsis seeds, which may hinge more on IPP/DMAPP availability than FPP, and can compensate for a drain on IPP/DMAPP by the introduced FPS.

Both of these results suggest a flexibility within the MVA pathway carbon flux to FPP which can be utilized by downstream enzymes. However, the capability of the plant to accumulate large amounts of squalene in the cytosol does not seem possible as it is most likely metabolized further, whereas botryococcene can be accumulated as it is not a native substrate for any enzymes present within Arabidopsis (or any organism besides Botryococcus braunii).

Accumulation of triterpene has no effect on total seed weight per se, but redirection of plastid metabolism might

Considering the potential applied uses of this engineering, and since seed weight is a critical agronomic yield parameter in such applications, we evaluated if engineering triterpene metabolism and accumulation impacted seed weights in various transgenic T3 seed lines. Seed weights from multiple independent transgenic events were determined. It is important to note that seed weight can vary depending on the environmental conditions in which the plants are grown, with light intensity playing an important role in seed oil content (and thus, weight) in Arabidopsis (Li et al., 2006). The plants in this study were hence grown to minimize differences in light irradiation, which varied less than 10% across a shelf and between the various shelves. Plant positions on shelves were also occasionally rotated. While we attempted to survey homozygous lines to present the most accurate values for our analysis, ambiguous resistance to the hygromycin selection marker occasionally complicated zygosity determinations. This was evident in seed lots from an individual plant that produced multiple phenotypes (Figure S3, bottom). T3 seed lines randomly selected for high level triterpene accumulation (Figures S1 and S2) were weighed (except for SQS, SQS + FPS and tpSQS lines which did not show accumulation). As seen in Figure 7, it does not appear that accumulation of either triterpene directly correlates with any impact on seed weight. However, in both engineering strategies a significant reduction in average seed weight was more often observed when engineering was directed to the plastid, and this reduction depended on presence of a plastid-targeted FPS. However, our analysis of crude seed oil as a percent of seed weight did not show any significant differences when compared to wild type (Figure S4). Some cytosolic-targeted lines
did demonstrate a reduction in seed weight [BS 10-11 and 13-9 (Figure 7a) and SQS + FPS 13-3 and SQS 1-11 (Figure 7b)]. To check if seed oil content might be altered leading to this altered seed weight we evaluated crude seed oil from select lines (Figure S4). We did not see any significant difference in crude seed oil when expressed as a percent of seed weight in any of the lines, regardless if they exhibited significantly reduced total seed weight (compare selected lines from Figure 7 in Figure S4).

Discussion

In these experiments we have demonstrated the feasibility of producing the triterpene hydrocarbons, botryococcene and squalene, in the oilseeds of Arabidopsis. We obtained the highest average yields when targeting our metabolism to the plastid in conjunction with expression of a putative rate-limiting step in the MEP pathway, catalysed by DXS. We have also shown the ability to produce botryococcene using a cytosolic-targeted strategy, but not squalene. We hypothesize that the inability to accumulate squalene in the cytosol is due to its metabolism, either by the conventional sterol biosynthetic machinery or alternative machinery. The latter result could come about by the form of the heterologous expressed squalene synthase used. The SQS used in the current work originates from Saccharomyces cerevisiae and its native, carboxy-terminal membrane-spanning domain was deleted and hence, encodes for a soluble SQS and not one in association with the ER membrane. The molecular fate of the squalene produced by this soluble SQS enzyme could be entirely different from that generated by ER membrane tethered enzyme. Previous reports indicate the inability of a heterologously expressed SQS to complement and participate in sterol biosynthesis in a cross-kingdom manner, which may be occurring here (Linscott et al., 2016). Nguyen et al. (2013) previously reported significantly elevated phytosterols in Arabidopsis seeds due to seed-specific expression of a full-length, ER-targeted soybean SQS. In contrast, we observed only a single transgenic event where total sterol content was significantly enhanced with a truncated, soluble SQS. More typically, we did not observe a significant change in total sterol levels in the soluble SQS engineered lines. As stated above, this could be, in part, due to the inability of the introduced S. cerevisiae SQS to integrate with the native Arabidopsis sterol machinery, which may require protein-protein interactions that are dictated by plant-specific motifs present on those proteins (Linscott et al., 2016).

A second limitation of our work is that the absolute levels of squalene produced in the cytoplasm might not be sufficient to significantly alter the observable sterol levels. When targeted to the plastid compartment, upwards of 200 μg of squalene per g seed accumulated. If a similar amount of production occurred in the cytosolic engineered lines, this would represent an extra 200 μg in addition to the 3000 μg of total phytosterols. Reliably observing less than a 10% increase in total phytosterols represents a challenge beyond our current technical capabilities. Hence, we conclude that extra squalene produced in transgenic Arabidopsis seeds harboring a cytosolic, soluble SQS could be funneled into the phytosterol pools and escape detection because shear mass is not significant, or it could be subject to alternative metabolism (catabolism?) due to its unusual intracellular production.

Equally intriguing, we saw some botryococcene accumulation in lines engineered with only tpBS (no co-expressed tpFPS), but no squalene accumulation in lines engineered with only tpSQS. Because FPP is not natively produced within the plastid (and believed to be nearly absent in that compartment), these results suggest that botryococcene present in these plants might be the result of a mis-targeted enzyme (i.e. tpBS residing within the plastid) and subsequent metabolism of a cytosolic precursor by the plastid machinery.
cytosol or mitochondria—both locations where FPP is normally produced), similar to results reported by Jiang et al. (2016). Other studies have shown the availability of MEP-derived prenyl units to cytosolic enzymes but not vice-versa (Dudareva et al., 2005; Gutensohn et al., 2013), hence we do think a plastid targeted SQS could be receiving its substrate (FPP) from an MVA origin. We also cannot eliminate the possibility that a minor amount of FPP could be released from the plastidic GPPs during GPP formation. However, this possibility also seems unlikely because no squalene was observed in the tpsQS only engineered lines. Hence, the most likely explanation is that a small portion of the tpsBS synthesized in the cytoplasm does not find its way to the plastid compartment and thus has cytosolic FPP available for catalysis.

We have demonstrated the ability to produce the linear triterpenes, botryococcene and squalene in the oilseeds of Arabidopsis thaliana using various subcellular targeting strategies and gene stacking. Targeting biosynthesis to the plastid in combination with a heterologous DXS gene resulted in the highest average accumulation for both compounds. However, cytosolic botryococcene production was substantial, and illustrates the capacity of the MVA pathway to produce nonnative terpenes. It may be worthwhile in the future to express a soluble, heterologous HMGR in combination with FPS and BS to evaluate this production potential. Similarly, higher levels of production in the plastid might be obtained by expressing the last two MEP pathway enzymes 4-hydroxy-3-methylbut-2-enyl-diphosphate synthase and/or 4-hydroxy-3-methylbut-2-enyl-diphosphate reductase, as these have been speculated to also control MEP flux (Banerjee and Sharkey, 2014). A striking difference between our work and that of engineering sesquiterpene metabolism in Camelina (Augustin et al., 2015) is the higher titres of sesquiterpene (also FPP-derived) they were able to obtain in both the cytosol and plastid compared to what we have observed with our triterpene engineering in Arabidopsis. Because their experiments also involved seed-specific promoters, inherent physiological differences might account for the production differences between these two species. Considering this, further seed engineering efforts may wish to evaluate a broader array of oilseed species (and accessions) to identify those best suited for each isoprenoid engineering project. Regardless, because of the ease to store and extract seeds, seeds appear to be an ideal tissue for just-in-time manufacturing of highly desired, commercially relevant chemical targets.

**Experimental procedures**

**Plant growth and transformation**

All transgenic lines were generated in the Col-0 ecotype background. Plants were grown on Pro-Mix BX soil (PremierTech) in 8 cm wide × 8 cm long × 10 cm deep pots in flats and were sub-irrigated as necessary with fertilizing approximately every other watering using 20 : 20 : 20 (N : P : K) general purpose fertilizer (The Scotts Company) at a nitrogen amount of 300 ppm. Plants were grown under 16 h light 8 h dark cycles using Sylvania OCTRON 6500K bulbs (Sylvania), with an approximate photosynthetic light intensity of 50 μmolm⁻²s⁻¹. Plants were grown sub-irrigated as necessary with fertilizing approximately every 4 weeks. Plants were grown on Pro-Mix BX soil (PremierTech) (Augustin et al., 2015) is the higher titres of sesquiterpene (also FPP-derived) they were able to obtain in both the cytosol and plastid compared to what we have observed with our triterpene engineering in Arabidopsis. Because their experiments also involved seed-specific promoters, inherent physiological differences might account for the production differences between these two species. Considering this, further seed engineering efforts may wish to evaluate a broader array of oilseed species (and accessions) to identify those best suited for each isoprenoid engineering project. Regardless, because of the ease to store and extract seeds, seeds appear to be an ideal tissue for just-in-time manufacturing of highly desired, commercially relevant chemical targets.

**Normal score calculation**

The botryococcene accumulation amounts obtained from each construct were evaluated to see if they followed a normal distribution. To do this, the highest botryococcene accumulating T₃ lines (n) from independent transformation events (the same that were used to generate the averages shown in Figure 3) were used to create plotting points from 0 to 1 by listing the n accumulation amounts in ascending order and taking the rank in this list of n and using it in the formula: (rank + (1/6))/n + 1/3. The resulting value was used to generate a normal score in Excel (Microsoft) with the NORMSINV function. This normal score was plotted on the y-axis with the corresponding accumulation value used to obtain the normal score plotted on the x-axis (Figure 4). A linear trendline was fitted to the data to show the expected normal distribution of the data.

**Metabolite extraction and quantification**

The triterpenes squalene and botryococcene were extracted from leaves by harvesting tissue in a preweighed 4 mL glass vial and grinding in liquid nitrogen. The tissue was allowed to warm to room temperature then weighed to obtain the fresh weight. A 1 : 1 : 1 mixture of hexane:acetone:water (1 mL each) was added to the sample (plus 100 μL of a 100 ng/μL hexadecane internal standard). For seed tissue, the seeds were weighed (~10–25 mg) moved to a clean mortar, 100 μL of a 100 ng/μL hexadecane internal standard was added, followed by 1 mL acetone and 1 mL water. The seeds were then ground using a clean pestle and 1 mL of hexane was then added and the slurry
was ground further before being moved to a 4 mL glass vial. Another 1 mL of hexane was added to the mortar to rinse and collect any remaining seed tissue and combined with the above slurry. The remainder of the extraction process was the same for both types of tissue. The solutions was allowed to shake (~220 rpm) at room temperature, in a horizontal position, for at least 20 min. The phases were then allowed to separate (or centrifuged briefly to separate the organic and aqueous phases), and the hexane phase was aspirated off and saved. Two more rounds of extraction with 1 mL hexane were completed and pooled with the initial extract. This n-hexane extraction was dried under nitrogen gas, resuspended in iso-octane and transferred to a GC vial. Samples were analysed on either an Agilent 7890 GC (HP-SMS column, 30 m x 0.25 mm, 0.25 µm film, 250 °C inlet temperature; oven temperature was 150 °C for 1 : 00 min, then 10 °C/min to 280 °C at, then 5 °C/min to 310 °C for 1 : 00 min; 0.9 mL/min He flow rate) connected to a 5975C Agilent mass spectrometer (run in positive ionization mode, 70 eV, scanning 50–500 amu) or an Agilent 7890 GC (HP-5 column, 30 m x 0.32 mm, 0.25 µm film; 60 °C for 1 : 00 min, then 30 °C/min to 230 °C, then 2 °C/min to 280 °C; 5.7494 mL/min He flow rate) equipped with a flame ionization detector. Botryococcene and squalene in experimental samples was verified by comparing to external standards. Sterterpene accumulation was calculated based on an external standard curve run in tandem with experimental samples.

Sterol levels in seeds were analysed by grinding tissue (~100 mg) in a mortar and pestle then moving to a pre-weighed 24 mL glass vial and weighed. The samples were saponified and the total sterols (and sterterpenes) were extracted as described by Du and Ahn (2002). The hexane extraction was dried to completion under nitrogen gas and samples were derivatized using a 1 : 1 mixture of pyridine and MSTFA + 1% TMCS at 50 °C for 1 h. Samples were analysed by GC-MS (same parameters as above) except the oven temperature was 200 °C initially, 10 °C/min to 270 °C, then 3 °C/min to 320 °C and held at 320 °C for 10 min. Major sterols were compared to known standards that had been derivatized as above.

Seed weight and crude oil calculations

Determination of seed weights was conducted by averaging the weight of three replicates of 100 seeds from a seed lot. In order to determine the wild-type (Col-0) seed weight, three replicates were used from wild-type seed lots bulked up at different times from various areas of the growth carts. This should reflect the inherent variation in the seed weights due to environmental factors. Seed crude oil was determined by conducting an extraction for sterterpene as described above, but the hexane phases were collected in a pre-weighed vial and after all phases had been collected the hexane was evaporated and the remaining oil in the vial was weighed. This is what we denoted as the seed crude oil.

Acknowledgements

We thank Dr. Edgar Cahoon’s laboratory at the University of Nebraska for kindly supplying the seed-specific promoters used in this study. We also thank all members of the Chappell lab for their constructive comments and help. This work was supported by grant 201006141622 from the National Institute of Food and Agriculture – USDA.

Conflicts of interest

Intellectual property has been secured by the University of Kentucky for some of this work and CK and JC are actively pursuing commercialization of this technology.

Author contributions

J.C. and C.K. conceived the research plan and wrote the article; C.K. conducted the experiments.

References

Arioglu, D., Sagner, S., Latzel, C., Eisenreich, W., Bacher, A. and Zenk, M.H. (1997) Terpenoid biosynthesis from 1-deoxy-D-xylulose in higher plants by intramolecular skeletal rearrangement. Proc. Natl. Acad. Sci. USA, 94, 10600–10605.

Augustin, J.M., Higashi, Y., Feng, X. and Kutchan, T.M. (2015) Production of mono- and sesquiterpenes in Camelina sativa oilseed. Planta, 242, 693–708.

Banerjee, A. and Sharkey, T.D. (2014) Methylerythritol 4-phosphate (MEP) pathway metabolic regulation. Nat. Prod. Rep. 31, 1043–1055.

Bansal, S. and Durrett, T.P. (2016) Camelina sativa: an ideal platform for the metabolic engineering and field production of industrial lipids. Biochimie, 120, 9–16.

Bechtold, N. and Pelletier, G. (1998) In Plantago Agrobacterium Mediated Transformation of Adult Arabidopsis thaliana Plants by Vacuum Infiltration, pp. 259–266. Totowa, NJ: Arab. Protoc. Humana Press.

Bell, S.A., Niehaus, T.D., Nybo, S.E. and Chappell, J. (2014) Structure-function mapping of key determinants for hydrocarbon biosynthesis by squalene and squalene synthase-like enzymes from the green alga Botryococcus braunii race B. Biochemistry, 53, 7570–7581.

Borgfi, M. and Xie, D.Y. (2016) Tissue-specific production of limonene in Camelina sativa with the Arabidopsis promoters of genes BANYULS and FRUITFULL. Planta, 243, 549–561.

Burrhardt, P.K., Beyer, P., Wann, J., Kloti, A., Armstrong, G.A., Schledz, M., Lintig, J. et al. (1997) Transgenic rice (Oryza sativa) endosperm expressing daffodil (Narcissus pseudonarcissus) phytoene synthase accumulates phytoene, a key intermediate of provitamin A biosynthesis. Plant J. 11, 1071–1078.

Cahoon, E.B., Hall, S.E., Ripp, K.G., Ganzke, T.S., Hitz, W.D. and Coughlan, S.J. (2003) Metabolic redesign of vitamin E biosynthesis in plants for tocotrienol production and increased antioxidant content. Nat. Biotechnol. 21, 1082–1087.

Chen, S., Songkumarn, P., Liu, J. and Wang, G.-L. (2009) A versatile zero background T-vector system for gene cloning and functional genomics. Plant Physiol. 150, 1111–1121.

Clough, S.J. and Bent, A.F. (1998) Floral dip: a simplified method for Agrobacterium-mediated transformation of Arabidopsis thaliana. Plant J. 16, 735–743.

Dalal, J., Lopez, H., Vasani, N.B., Hu, Z., Swift, J.E., Yalamanchili, R., Dvora, M. et al. (2015) A photorespiratory bypass increases plant growth and seed yield in biofuel crop Camelina sativa. Biotechnol. Biofuels, 8, 175.

Du, M. and Ahn, D.U. (2002) Simultaneous analysis of tocopherols, cholesterol, and phytosterols using JPS Food Chem. Toxicol. 40, 1696–1700.

Dudareva, N., Andersson, S., Orlova, I., Gatto, N., Reichelt, M., Rhodes, D., Boland, W. et al. (2005) The nonmevalonate pathway supports both monoterpene and sesquiterpene formation in snapdragon flowers. Proc. Natl. Acad. Sci. USA, 102, 933–938.

Fernandez, D.E., Turner, F.R. and Crouch, M.L. (1991) Insitu localization of embryos. Plant Biol. 16, 735–743.

Flügge, U.-I. and Gao, W. (2005) Transport of isoprenoid intermediates across chloroplast envelope membranes. Plant Biol. 7, 91–97.

Fujiwara, M., Takita, E., Harada, H., Sakurai, N., Suzuki, H., Ohyama, K., Shibata, D. et al. (2009) Pathway engineering of Brassica napus seeds using...
multiple key enzyme genes involved in ketocarotenoid formation. J. Exp. Bot. 60, 1319–1332.

Gutensohn, M., Orlova, I., Nguyen, T.T.H., Davidovich-Rikanati, R., Ferruzzi, M.G., Sitrit, Y., Levinsohn, E. et al. (2013) Cytosolic monoterpene biosynthesis is supported by plastid-generated geranyl diphosphate substrate in transgenic tomato fruits. Plant J. 75, 351–363.

Harrison, S.J., Mott, E.K., Parsley, K., Aspinall, S., Gray, J.C. and Cottage, A. (2006) A rapid and robust method of identifying transformed Arabidopsis thaliana seedlings following floral dip transformation. Plant Methods 2, 19.

Hemmerlin, A., Hoefl er, I.-F., Meyer, O., Trisch, D., Kagan, I.A., Grosdemange-Billard, C., Rohmer, M. et al. (2003) Cross-talk between the cytosolic mevalonate and the plastidial mevalyeryl triphosphate pathways in tobacco bright yellow-2 cells. J. Biol. Chem. 278, 26666–26676.

Hemmerlin, A., Harwood, J.L. and Bach, T.J. (2012) A raison d'être for two distinct pathways in the early steps of plant isoprenoid biosynthesis? Prog. Lipid Res. 51, 95–148.

Hillen, L.W., Pollard, G., Wake, L.V. and White, N. (1982) Hydrocracking of the oils of Botryococcus braunii to transport fuels. Biotechnol. Bioeng. 24, 193–205.

Holmberg, N., Harker, M., Gibbard, C.L., Wallace, A.D., Clayton, J.C., Rawlins, S., Helyer, A. et al. (2002) Sterol C-24 methyltransferase type 1 controls the flux of carbon into sterol biosynthesis in tobacco seed. Plant Physiol. 130, 303–311.

Holmberg, N., Harker, M., Wallace, A.D., Clayton, J.C., Gibbard, C.L. and Safford, R. (2003) Co-expression of N-terminal truncated 3-hydroxy-3-methylglutaryl CoA reductase and C24-sterol methyltransferase type 1 in transgenic tobacco enhances carbon flux towards end-product sterols. Plant J. 36, 12–20.

Jiang, Z., Kempinska, C., Bush, C.J., Nybo, S.E. and Chappell, J. (2016) Engineering triterpene and methylated triterpene production in plants provides biochemical and physiological insights into terpene metabolism. Plant Physiol. 170, 702–716.

Josefsson, L.G., Lenman, M., Ericson, M.L. and Rask, L. (1987) Structure of a gene encoding the 1.7 S storage protein, napin, from Brassica napus. J. Biol. Chem. 262, 12196–12201.

Kempinska, C., Jiang, Z., Bell, S. and Chappell, J. (2015) Metabolic engineering of higher plants and algae for isoprenoid production. Adv. Biochem. Eng. Biotechnol. 148, 161–199 Springer Berlin Heidelberg.

Lee, D.W., Lee, S., Lee, G.-J., Lee, K.H., Kim, S., Cheong, G.-W. and Hwang, I. (2006) Functional characterization of sequence motifs in the transit peptide of Arabidopsis small subunit of ribulose. Plant Physiol. 140, 466–483.

Li, Y., Beisson, F., Pollard, M. and Ohlrogge, J. (2006) Oil content of Arabidopsis plants: a rich source for Golden Rice through increased pro-vitamin A content. Nat. Biotechnol. 23, 482–487.

Sivadige, B., Weiss, J.D., Wong, Y.H., Lassner, M.W., Mitsky, T.A., Shewmaker, C.K., Post-Beittenmiller, D. et al. (2002) Isolation and characterization of homogentisate phytyltransferase genes from Synechocystis sp. FCC 6803 and Arabidopsis. Plant Physiol. 129, 321–332.

Shewmaker, C., Sheehy, J., Daley, M., Colburn, S. and Ke, D. (1999) Seed-specific overexpression of phytoene synthase: increase in carotenoids and other metabolic effects. Plant J. 20, 401–412X.

Sim, T.L. and Goldberg, R.B. (1989) Nucleic acids research the glycinin Gy1 gene from soybean. Nucleic Acids Res. 17, 4386.

Sitbon, F. and Jonsson, L. (2001) Sterol composition and growth of transgenic tobacco plants expressing type-1 and type-2 sterol methyltransferases. Planta 212, 568–572.

Spanova, M. and Daum, G. (2011) Squalene - biochemistry, molecular biology, process biotechnology, and applications. Eur. J. Lipid Sci. Technol. 113, 1299–1320.

Tardos, L.C., Yan, M., Pouler, C.D. and Sacchettini, J.C. (1994) Crystal structure of recombinant farnesy l diphosphate synthase at 2.6 A-resolution. Biochemistry, 33, 10871–10877.

Tracy, N.I., Crunkleton, D.W. and Price, G.L. (2011) Catalytic cracking of squalene to gasoline-range molecules. Biomass Bioenergy 35, 1060–1065.

USDA (2016) Oilseeds: World Markets and Trade. Vranova, E., Coman, D. and Grusswein, V. (2013) Network analysis of the MVA and MEP pathways for isoprenoid synthesis. Annu. Rev. Plant Biol. 64, 665–700.

Wright, L.P., Rohwer, J.M., Ghirardo, A., Hammerbacher, A., Ortiz-Alcaide, M., Raguschtze, B., Schnitzler, J.-P. et al. (2014) Deoxyxylulose 5-phosphate synthase controls flux through the mevalyeryl triphosphate 4-phosphate pathway in Arabidopsis. Plant Physiol. 165, 1488–1504.

Wu, S., Schalk, M., Clark, A., Miles, R.B., Coates, R. and Chappell, J. (2006) Redirection of cytosolic or plastidic isoprenoid precursors elevates terpene production in plants. Nat. Biotechnol. 24, 1441–1447.

Xue, Z., Duan, L., Liu, D., Guo, J., Ge, S., Dicks, J., ÔMaille, P. et al. (2012) Divergent evolution of oxidosqualene cyclases in plants. New Phytop. 193, 1022–1038.

Ye, X., Al-Babili, S., Klüt, A., Zhang, J., Lucca, P., Beyer, P. and Potrykus, I. (2000) Engineering the provitamin A (beta-carotene) biosynthetic pathway into (carotenoid-free) rice endosperm. Science, 287, 303–305.

Zhang, X., Henriques, R., Lin, S.-S., Niu, Q.-W. and Chua, N.-H. (2006) Engineering triterpene metabolism in Arabidopsis seeds. J. Biol. Chem. 281, 867–877.

Supporting information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Table S1 Segregation ratios for T2 Arabidopsis plants engineered with various squalene biosynthesis constructs.
Table S2 Segregation ratios for T2 Arabidopsis plants engineered with various botryococcene biosynthesis constructs.

Figure S1 Botryococcene content (µg/g seed) of four independent T3 seed lots.

Figure S2 Squalene content (µg/g seed) of four independent T3 seed lots.

Figure S3 Example of phenotypes complicating segregation analyses.

Figure S4 Crude seed oil (% seed weight) and triterpene values from indicated T3 seed lots.