

DR. MICHEL MATTHIJS (Orcid ID : 0000-0002-6051-3671)

Article type : Research Article

## Engineering The Unicellular Alga *Phaeodactylum tricornutum* For High-Value Plant

### Triterpenoid Production

Sarah D'Adamo<sup>1</sup>, Gino Schiano di Visconte<sup>1</sup>, Gavin Lowe<sup>1</sup>, Joanna Szaub-Newton<sup>1</sup>, Tracey Beacham<sup>2</sup>, Andrew Landels<sup>2,3</sup>, Michael J. Allen<sup>2,4</sup>, Andrew Spicer<sup>1</sup>, Michiel Matthijs<sup>1\*</sup>

<sup>1</sup>Algenuity, Eden Laboratory, MK43 9ND Stewartby, United Kingdom

<sup>2</sup>PML: Plymouth Marine Laboratory, Prospect Place, The Hoe, Plymouth, PL1 3DH, United Kingdom

<sup>3</sup>Rothamsted Research, Harpenden, AL5 2JQ, United Kingdom

<sup>4</sup>Biosciences, College of Life and Environmental Sciences, University of Exeter, Exeter EX4 4QD, UK

\*Corresponding author: Eden Laboratory/Broadmead Rd, Bedford MK43 9ND, United Kingdom +44 1234 765773, [mm@algenuity.com](mailto:mm@algenuity.com)

**Keywords:** triterpenoid biosynthesis, natural product, algal synthetic biology, diatoms, microalgae, lupeol, betulin, blue biotechnology

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1111/pbi.12948

This article is protected by copyright. All rights reserved.

**Abbreviations:** BA, Betulinic acid; MVA, Mevalonate Pathway; MEP, non-mevalonate pathway; LjLUS, *Lotus japonicus* Lupeol synthase; AtLUS, *Arabidopsis thaliana* Lupeol Synthase; OSC, oxidosqualene cyclase; IPP, Isopentyl Pyrophosphate; CYP, cytochrome P450 monooxygenase; CPR, cytochrome b5 reductase; lcPUFA, long chain polyunsaturated fatty acids

## ABSTRACT

Plant triterpenoids constitute a diverse class of organic compounds that play a major role in development, plant defense and environmental interaction. Several triterpenes have demonstrated potential as pharmaceuticals. One example is betulin, which has shown promise as a pharmaceutical precursor for the treatment of certain cancers and HIV. Major challenges for triterpenoid commercialization include their low production levels and their cost-effective purification from the complex mixtures present in their natural hosts. Therefore, attempts to produce these compounds in industrially relevant microbial systems such as bacteria and yeasts have attracted great interest. Here we report the production of the triterpenes betulin and its precursor lupeol in the photosynthetic diatom *Phaeodactylum tricornutum*, a unicellular eukaryotic alga. This was achieved by introducing three plant enzymes in the microalga: a *Lotus japonicus* oxidosqualene cyclase and a *Medicago truncatula* cytochrome P450 along with its native reductase. The introduction of the *L. japonicus* oxidosqualene cyclase perturbed the mRNA expression levels of the native mevalonate and sterol biosynthesis pathway. The best performing strains were selected and grown in a 550L pilot scale photobioreactor facility. To our knowledge, this is the most extensive pathway engineering undertaken in a diatom and the first time that a sapogenin

has been artificially produced in a microalga, demonstrating the feasibility of the photo-bio-production of more complex high-value, metabolites in microalgae.

## INTRODUCTION

Plants produce a wide variety of secondary metabolites that are involved in development, defense and interaction with the environment (Moses & Pollier 2013; Sawai & Saito 2011).

Alkaloids, flavonoids and terpenoids are the major classes of plant secondary metabolites with members of all three classes already being exploited as high-value therapeutics, flavors, fragrances and colorants (Moses et al. 2013; Bourgaud et al. 2001; Bansal et al. 2016; Paddon et al. 2013; Rouck et al. 2017).

Terpenoids are a structurally diverse group, which includes carotenoids, sterols and tocopherols. The number of isoprene units that are used in their synthesis determines their classification, e.g. diterpenoids consist of four isoprene units and have twenty carbon atoms (C<sub>20</sub>), while triterpenoids consist of six isoprenoid units and have thirty carbon atoms (C<sub>30</sub>).

Triterpenoids constitute a wide and diverse class of plant natural products. Basic triterpenoids are termed sapogenins that often undergo chemical modifications such as oxidation, (de)methylation and the addition of nitrogen or sulfur atoms (Augustin et al. 2011; Kvasnica et al. 2015; Moses et al. 2013). One of the most common modifications is glycosylation, which converts sapogenin backbones into amphipathic saponins. Simple and conjugated triterpenes have a wide range of applications in the food, health, and industrial biotechnology sectors (Thimmappa et al. 2014; Augustin et al. 2011).

The triterpenoid saponins derive from intermediates of the phytosterol pathway and typical biosynthesis in plants involves a minimum of three enzyme classes: oxidosqualene cyclases (OSC), which construct the basic triterpenoid scaffolds, cytochrome P450 monooxygenases (CYP), which mediate oxidations, and uridine diphosphate dependent glycosyltransferases, which catalyze glycosylations (Sawai & Saito 2011). In nature, more than 100 distinct triterpenoid scaffolds have been identified; some of the best characterized include  $\alpha$ -amyrin,  $\beta$ -amyrin, dammarenediol and lupeol (Moses et al. 2015).

In this study, we focused on lupeol, betulin and betulinic acid (BA) triterpenoids. Derivatives of these compounds have shown potential for the treatment of HIV and certain cancers (Cichewicz & Kouzi 2004; Drag et al. 2009; Kommera et al. 2011).

Although there is substantial industrial interest in triterpenoids, large-scale extraction from their natural hosts is often costly, as these compounds are usually present in low amounts and in complex mixtures of related compounds, making their isolation difficult. Moreover, increasing the yield of target triterpenoids in the plants that naturally produce them, is often unfeasible, as most medicinal plants do not have an established transformation procedure or cannot be grown on a suitable commercial scale. Additionally the transfer of the relevant biosynthetic enzymes into other plant species is often complicated by the presence of native enzymes that can make undesirable modifications to intermediates or to the end product, as seen in *Nicotiana benthamiana* and *Oryza sativa* (Khakimov et al. 2015; Huang et al. 2015). Therefore, the investigation of biotechnological routes to produce these compounds in industrially relevant microbial hosts has become an active field of research with significant efforts being undertaken to isolate the key enzymes involved in these plant-

Accepted Article

specific biosynthetic pathways. In general, eukaryotic hosts like *Saccharomyces cerevisiae* seem preferable for the production of triterpenoids as they can tap into the existing sterol biosynthesis pathway that is present in all eukaryotes but rare in prokaryotes (Wei et al. 2016). Additionally, cytochrome P450 enzymes generally do not function well in bacteria such as *Escherichia coli* and extensive engineering is often required to obtain high titers (Pateraki et al. 2015; Rouck et al. 2017).

Although the most popular heterologous hosts for terpenoid production are bacteria and yeasts (Møller et al. 2008; Zhou et al. 2016; Leavell et al. 2016; Kirby et al. 2008; Moses et al. 2015), there has been increasing interest in photosynthetic microbial platforms, as they can utilize sunlight and CO<sub>2</sub> to obtain energy and carbon for growth, thereby minimizing the environmental impact of production (Davies et al. 2015; Kyle J. Lauersen et al. 2016; Work et al. 2013; Wijffels et al. 2013; Work et al. 2012; Ruiz et al. 2016; Gimpel et al. 2015). Moreover, algal biomass can be a useful source of a variety of value-added molecules such as pigments, oils and omega 3 fatty acids, which can find application in the food, cosmetics, aquaculture and animal feeds industries.

Advances in synthetic biology have seen important breakthroughs in terpenoid engineering with some photosynthetic hosts, including cyanobacteria and moss tissue cultures, being successfully exploited for heterologous terpenoid production (Davies et al. 2014; Anterola et al. 2009; Arendt et al. 2016).

Most of microalgal research has been performed on green algae such as *Chlamydomonas reinhardtii* as they have been historically adopted as model systems to understand photosynthesis and aspects of physiology such as circadian clocks (Noordally & Millar 2015; Calvin 1962). Recently a sesquiterpenoid (C15) was produced in this alga (Kyle J Lauersen et al. 2016). However, successful engineering of green algae involves overcoming the poor expression of heterologous genes that have been integrated into the nuclear genome. Often these genes are prone to gene silencing mechanisms (Cerutti et al. 1997; Rasala et al. 2014). Therefore, extensive screening needs to be undertaken to weed out lines that only express the resistance marker and identify those which express the transgene to sufficient levels (Shin et al. 2016; Erpel et al. 2016; Kyle J Lauersen et al. 2016). While *C. reinhardtii* has proven to be a good model organism for the elucidation of physiological processes, other algal systems are better suited for both the expression of heterologous pathways and for deployment in industrial processes. In recent years, the heterokont algae have become increasingly attractive for biotechnology and metabolic engineering applications because of their high lipid content and existing production for aquaculture (P. Kroth 2007; Pulz & Gross 2004; Nrel 1998; Wikfors & Ohno 2001; Hamilton et al. 2016). Two examples of brown unicellular algae that are genetically tractable and have publically available sequenced nuclear genomes include the Eustigmatophyte *Nannochloropsis oceanica* and the diatom *Phaeodactylum tricornutum* (Vieler et al. 2012; Bowler et al. 2008; Apt et al. 1996). The diatom, *P. tricornutum*, has already shown industrial potential as it can grow photoautotrophically and stably express functional enzymes, although molecular tools for genetic modification are limited. Commercially interesting products that have been produced in *P. tricornutum* include antibodies and biodegradable plastics (Franziska et al.

2011; Hempel et al. 2011). To the best of our knowledge, the only terpenoid engineering performed in this organism has been on the carotenoid pathway (Eilers et al. 2016).

Cyanobacteria and green algae form isopentenyl diphosphate (IPP) exclusively through the non-mevalonate pathway (MEP). IPP is the key precursor for sterol and other isoprenoid production. Conversely, both higher plants and diatoms such as *P. tricornutum* contain both the cytosolic mevalonate (MVA) and the chloroplastic MEP pathway. Here, we report the efficient production of two plant triterpenoids (lupeol and betulin) in *P. tricornutum* by the heterologous expression of plant enzymes: an oxidosqualene cyclase and a cytochrome P450 along with its reductase. In the plant betulinic acid (BA) biosynthetic pathway (Fig. 1) a lupeol synthase OSC produces lupeol, which is subsequently oxidized at the C28 position by a CYP716A family cytochrome P450 to BA. Most of the CYP716A family enzymes also produce partial C28 oxidation products, which, in the case of lupeol oxidation, are betulin and betulin aldehyde. This is the first report of genetic engineering of a eukaryotic microalga for triterpenoid (C30) production and opens the way for photo-bioproduction of high-value plant secondary metabolites.

## **RESULTS and DISCUSSION**

### **Expression of two lupeol synthases in *P. tricornutum* results in lupeol accumulation.**

Triterpenes, such as sterols, are synthesized via the 30-carbon intermediate, 2,3-oxidosqualene, which is cyclized by members of the oxidosqualene cyclase (OSC) family. These enzymes generally localize to the ER and contain a membrane binding domain. The genomes of the microalgae *Chlamydomonas reinhardtii*, *Phaeodactylum tricornutum* and the moss *Physcomitrella patens* contain a single oxidosqualene cyclase (OSC) gene for

tetracyclic sterol biosynthesis. Higher plants, in contrast, can contain a remarkable amount of chemical diversity derived from the single substrate 2,3-oxidosqualene. Typically plant genomes contain from nine to sixteen OSC genes e.g. the model plant *Arabidopsis thaliana* genome contains thirteen OSC gene family enzymes and can produce a diverse array of triterpene skeletons (Vincken et al. 2007; Xue et al. 2012). Among these, a lupeol synthase, At1g78970 / LUP1 has been characterized (Herrera et al. 1998), successfully expressed and used for lupeol and betulinic acid production in yeasts (Husselstein–Muller et al. 2001; Czarnotta et al. 2017; Lang & Lewandowski 2016; Zhou et al. 2016).

The *Arabidopsis thaliana* lupeol synthase (from now on designated as AtLUS) catalyzes the production of lupeol and the related triterpenoid, 3 $\beta$ , 20-dihydroxylupane, as a by-product (Segura et al. 2000; Salmon et al. 2016).

To select an OSC enzyme for introduction in *P. tricornutum*, we performed a phylogenetic analysis of selected plant OSCs with known functions (Fig. S2); interestingly, AtLUS clusters with beta-amyrin synthases such as the *Artemisia annua* OSC2 and the *Catharantus roseus* beta-amyrin synthase (BAS), while all lupeol synthases (LUS) cluster separately i.e. *Bupleurum lanceolate* LUS, *Lotus japonicus* OSC3, *Olea europaea* LUS. It has been suggested previously that either two branches of lupeol synthase genes have been generated during higher plant evolution or that, AtLUS represents an evolutionary transition state between lupeol synthase and other unknown triterpene synthases (Shibuya et al. 1999).



To test the feasibility of *P. tricornutum* as a lupeol production platform, we ectopically expressed two representative OSCs: AtLUS or *Lotus japonicus* OSC3 (from now on LjLUS).

Both enzymes have been shown to be functional in *S. cerevisiae*, with AtLUS being the most prevalent in literature (Sawai et al. 2006; Zhou et al. 2016; Czarnotta et al. 2017). The native coding sequences of *AtLUS* and *LjLUS* were integrated into the *P. tricornutum* nuclear genome under the control of the strong fucoxanthin chlorophyll a/c binding protein A (*FCPA/LHCF1*) promoter, along with zeocin resistance gene marker (*ble<sup>r</sup>*) (SF1, A and B).

Single transformant lines grown on selective solid medium (zeocin antibiotic), were transferred to liquid medium and grown until early stationary phase. Thirty lines were subjected to GC-MS analysis and lupeol was detected in both *LjLUS* and *AtLUS* expressing lines (Fig. 2) with the identification supported by the MS electron impact spectra (Fig. S3A-R)

Based on the initial GC-MS screening we selected those lines with the highest relative production of lupeol (Supplementary table 1). On average, we were able to detect lupeol in 30% of the screened lines for each construct. This established that the enzymes are functionally expressed and that *P. tricornutum* lines expressing heterologous genes can be identified using a relatively easy and low throughput screening method. This compares favorably with the *C. reinhardtii* platform used for sesquiterpene patchoulol (C15) production, where more intricate vector design and screening methods need to be employed to select positive lines, including the preparation of a construct harboring three copies of the transgene, one of which was fused with a reporter gene (YFP) to allow colonies to be screened for fluorescence on primary transformant plates (Kyle J Lauersen et al. 2016).

In the transformant lines expressing AtLUS, we were able to identify lupeol and the 3 $\beta$ , 20-dihydroxylupane by-product, often produced at similar titers, but also dependent on culturing conditions, while LjLUS expressing lines produced just lupeol (Fig. 2). We, therefore, focused our attention on line LjLUS-25. To the best of our knowledge, this is the first successful demonstration of pentacyclic triterpenoid production in an engineered microalga.

Interestingly, no canonical squalene epoxidase enzyme has been identified in *P. tricornutum* based upon sequence or structural similarity (Fabris et al. 2014), although its product, the 2-3 oxidosqualene, was predicted to be present as the substrate of the well annotated cycloartenol synthase (Fig. 1). This result affirms the presence of 2-3 oxidosqualene in *P. tricornutum*, as the introduced lupeol synthases require this precursor to synthesize lupeol.

#### **High lupeol productivity correlates with high *LjLUS* expression but is independent on the number of integrated expression cassettes**

Nuclear transformation events in *P. tricornutum* lead to random insertions of a given transgene into the nuclear genome. In order to test whether there is any correlation between higher lupeol production and relative transgene expression levels, we performed qRT-PCR on the three LjLUS lines that showed the lowest lupeol accumulation and the three lines that showed the highest lupeol accumulation. A one way ANOVA analysis of LjLUS expression in the selected lines showed that there was a significant difference in expression  $F(7,20) = 7.17$ ,  $p < 0.01$ . The line with the best lupeol accumulation, namely LjLUS-25, was significantly different as supported by a *post hoc* Tukey test ( $\alpha = 0.01$ ) while the other lines grouped together (Fig. 3A).

To investigate whether mRNA expression levels are mainly determined by locus of integration or due to transgene copy number resulting from single or multiple nuclear genome integration events, the approximate transgene copy number was quantified using qRT-PCR for the three low and three high lupeol producing strains. Results were normalized by amplifying regions within two different wild-type genomic loci (*RP3a Phatr3\_J13566* and *UBQ Phatr3\_J 28620*), for which the copy number is assumed to be two in a diploid cell. The copy number of the *LjLUS* coding sequence and the zeocin resistance marker (*ble<sup>r</sup>*) were determined. Both transgenes were present on a single transformation vector and were found to have similar abundance: two to three transgene copies for each transformant line (Fig. 3B). This is consistent with what has been previously reported for *P. tricornutum* integration events (Falciatore et al. 1999). An ANOVA was conducted to compare the main effects of the individual lines and the integrated transgene. This analysis confirmed that there was no difference between the number of integrated *ble<sup>r</sup>* and *LjLUS* copies ( $p=0.29$ ) but that there were significant differences between the lines ( $p<0.001$ ,  $F(6,29)=26.49$ ). Post hoc comparisons were performed for the comparison of the lines and groups significantly different from each other are indicated in figure 3B ( $\alpha=0.01$ ). There was no correlation between mRNA expression levels and the number of integrated *LjLUS* cassettes (Pearson correlation score of  $-0.03$ ,  $t = 0.07$ ,  $p > 0.9$ ). These results suggest that variations in relative mRNA expression and lupeol accumulation are likely due to positional (site of integration) effects rather than the absolute number of integrated transgene copies.

## Production of lupeol is growth phase dependent and reduces brassicasterol levels in the cell

In order to monitor the lupeol productivity during the different growth phases, we performed experiments in an Algem® lab-scale photobioreactor, collecting and analyzing the biomass at days 0, 1, 2, and 5 of growth. Cells reached maximum productivity of both brassicasterol (the main sterol) and lupeol during mid-exponential growth phase (as indicated by the arrows in Figure. 4), suggesting that cells are actively engaged in sterol metabolism at the early stage of their growth. We could also observe a reduction of brassicasterol per cells in the LjLUS-25 line when compared to WT, consistent with the amount of lupeol produced per cell (Fig. 4A). The maximum yield of lupeol in the cell was reached after two days of growth.

No brassicasterol was observed in either the LjLUS or WT media, suggesting cellular lysis at this stage was minimal. It appears that while native sterols such as brassicasterol are retained within the cell, lupeol is able to actively or passively translocate into the medium (Fig. 4B).

We also tracked gene expression of *LjLUS* and the zeocin resistance confirming protein *ble<sup>r</sup>* for five days during which the cells transitioned from exponential to stationary phase. Both transcripts showed maximum expression levels during mid-exponential growth phase, in line with the lupeol productivity obtained (Fig. 4C). An ANOVA analysis showed that the difference in time was significant ( $F(3)= 6.32$ ,  $p=0.005$ ) but there was no significant difference between the expression of *LjLUS* and *ble<sup>r</sup>* ( $F(1)=0.7$ ,  $p=0.41$ ). This result is also consistent with the *FCPA/LHCF1* promoter activity, which peaks during exponential growth.

While the *FCPA/LHCF1* promoter is widely used, a stronger and growth phase independent promoter could potentially enhance LjLUS activity.

Strain LjLUS-25 was selected as the best producing line. LjLUS-25 reached a maximum lupeol yield (C30) of  $0.1 \text{ mg}\cdot\text{L}^{-1}$  over two days of culturing. This reported yield is similar to the  $0.35 \text{ mg/L}$  yield of lupeol achieved in *C. reinhardtii* engineered for the production of the sesquiterpene patchoulol (C15), grown in comparable conditions, over a seven day period (Kyle J Lauersen et al. 2016).

For comparison: the highest yield for betulinic acid production ( $182 \text{ mg}\cdot\text{L}^{-1}$ ) has been achieved in a *S. cerevisiae* strain (Czarnotta et al. 2017). This strain is the result of multiple optimization rounds and fed-batch cultivations. The first-generation strain showed betulinic acid titers of  $0.1 \text{ mg/L/OD}$  (Huang et al. 2012), and with further genetic improvements, increased flux into the mevalonate pathway and elevated cofactor supply, betulinic acid titers reached  $12.1 \text{ mg/g}$  dry biomass (Li and Zhang 2014; Li and Zhang 2015). In our microalgal system we obtained titers  $> 0.1 \text{ mg/g}$  dry biomass with no prior optimization, this is higher than the initial concentrations reported for the primary yeast strains.

In addition, this first-generation *P. tricornutum* strain can be grown photoautotrophically and is rich in fucoxanthin and long-chain poly-unsaturated fatty acids (lcPUFAs), adding value to the platform. Further optimizations in genetics and culture conditions, and extraction of multiple value-added molecules are expected to substantially improve the overall process economics. In Figure 5 we report the extraction procedure of triterpenes and a proposed biorefinery approach. Sources for this analysis can be found in

supplementary table 4. The extraction procedure for triterpene is the same for yeast and microalgae and estimated costs are equal. Refining the microalgal biomass from this purification procedure allows the separation of other valuable molecules. Examples of these co-products include eicosapentaenoic acid (market estimated to reach \$3.79 billion by 2022), carotenoids such as fucoxanthin (currently \$95 million global market) and phytosterols such as brassicasterol (\$62.5 million global market by 2020). The defatted biomass (protein and sugars) can still be used as a low value animal food. Therefore, further optimizations in genetics and culture conditions, and extraction of multiple value-added molecules, are expected to increase the competitiveness of microalgal platforms.

#### **Triterpenoid production affects the expression levels of native sterol metabolism enzyme transcripts in the LjLUS-25 transformant line**

In eukaryotic photosynthetic organisms different terpenoid classes are produced in different cellular compartments. Both carotenoids and sterols are produced from the same precursor isopentenyl pyrophosphate (IPP), but the former is made in the chloroplast while the latter is made in the ER. Higher plants and *P. tricornutum* contain both the cytosolic MVA and the chloroplast localized MEP pathway (Fig. 1). While both pathways produce IPP, there appears to be no or minimal IPP interchange between the cytosol and chloroplast in higher plants (Rodríguez-Concepción 2006). Indeed, disruption of key genes in either the MVA or MEP pathway results in distinct and severely deleterious phenotypes in *Arabidopsis thaliana* (Suzuki et al. 2004; Hsieh & Goodman 2005). Green algae such as *C. reinhardtii* only contain the chloroplast localized MEP pathway which supplies IPP for both the ER localized sterol biosynthesis and photosynthetic pigments.

The introduction of a non-native OSC such as the LjLUS in *P. tricornutum* could potentially affect sterol biosynthesis or the IPP producing pathways. Therefore we measured the expression of a selected set of sterol biosynthetic enzymes, including enzymes from the chloroplast localized MEP pathway and cytosolic MVA pathway, by qPCR to determine whether ectopic *LjLUS* influences native sterol biosynthesis expression.

Samples for RNA extraction were taken from cultures of wild type and LjLUS-25 line during five days of culturing. The algae started entering stationary phase by the fifth day (Fig. S4). It was immediately apparent that expression of sterol biosynthesis enzymes is higher during exponential phase and nearly all investigated genes reached their expression minimum on day five. This holds true for both the wild type and the LjLUS overexpression line and correlates well with the lupeol and brassicasterol levels measured (Fig. 4A-B).

Expression of some of the mevalonate pathway encoding genes increased markedly in the LjLUS-25 line compared to the wild type as shown in figure 6. The statistical significance of the investigated genes was tested using a repeated measures ANOVA model (Supplementary Table 5). Genes coding for key enzymes such as HMGR (*Phatr3\_J16649*) and the isopentenyl diphosphate isomerase - squalene synthase fusion enzyme (*Phatr3\_EG02290*) are consistently upregulated, while HMG-CoA synthase (*Phatr3\_J16649*) increased only slightly.

Since the levels of brassicasterol are lower in the LjLUS-25 line, we expected to see marked upregulation of genes involved in brassicasterol synthesis in an attempt to compensate. Surprisingly, no upregulation was seen for the PtOSC transcript (*Phatr3\_EG02293* + *Phatr3\_J46726* fusion) even though it competes with the LjLUS enzyme for 2,3-oxidosqualene. However, enzyme transcripts downstream of PtOSC show upregulation. This includes a methylsterol oxygenase (*Phatr3\_J10852*), a sterol dehydrogenase (*Phatr3\_J48864*), a C24 sterol reductase (*Phatr3\_J48260*) and C22 reductase (*Phatr3\_J51757*) with all of these enzymes showing a similar gene expression pattern. The MEP pathway transcripts *IspE* (*Phatr3\_EG02383*) and *ISPD2* (*Phatr3\_J21829*) did not show a consistent upregulation pattern in transgenic lines and its role in sterol biosynthesis, if any in *P. tricornutum*, remains unclear. Overall these results suggest that, as in higher plants, the MVA pathway is the major source of IPP for sterol biosynthesis and triterpenoid production in *P. tricornutum*. This is in agreement with earlier studies, which showed that terpenoids in the *P. tricornutum* cytosol and chloroplast do not have the same source of IPP (Cvejić & Rohmer 2000).

### **Introducing a CYP716A12 cytochrome P450 and a P450-NADPH reductase leads to the oxidation of lupeol to betulinic acid**

In several higher plant species (e.g. *Betula platyphylla*), lupeol is oxidized at the C28 position by a cytochrome P450-dependent monooxygenase (CYP) to form betulinic acid (Fig. 1) (Hamberger & Bak 2013). Recently, a number of studies have reported enzymes of the CYP716A family to be responsible for this modification (Carelli et al. 2011; Fukushima et al. 2011; Andre et al. 2016; Moses et al. 2014). Based on the literature we selected the



*Medicago truncatula* CYP716A12 and CPR for expression in *P. tricornutum*. In-house activity assays in yeast confirmed that the *Medicago truncatula* CYP716A12 enzyme is a multifunctional enzyme with  $\beta$ -amyrin 28-oxidase,  $\alpha$ -amyrin 28-oxidase and lupeol 28-oxidase activities.

Cytochrome P450 enzymes work in tandem with specific cytochrome P450 reductases (CPR), which shuttle electrons from NAD(P)H to the cytochrome P450. The CPR–P450 interacting domains are well conserved, as demonstrated by the ability of CPRs from different species to at least partially complement in functional terms. Although eukaryotes such as *S. cerevisiae* and *P. tricornutum* contain a native CPR, CYP enzymes usually require a CPR from the same or a related species for maximum activity when introduced into a heterologous host, to ensure an adequate supply of electrons to the CYP and potentially minimizing the release of reactive oxygen species (Jennewein et al. 2005; Kim et al. 2009; Kim et al. 2005; Jensen & Møller 2010; Zangar et al. 2004). This reduction step can be rate limiting and increased catalytic efficiency has been obtained by fusing one or more CPRs directly to the CYP enzymes (Schückel et al. 2012; Leonard et al. 2006; Munro et al. 2007).

In this work, we tested the expression in *P. tricornutum* of the *Medicago truncatula* CYP716A12 (MtCYP716A12), co-expressed along with the *M. truncatula* CPR (MtCPR). In order to increase catalytic efficiency of the CYP–CPR system, the MtCYP716A12 sequence was fused to its native MtCPR co-enzyme. We therefore employed two approaches:

1. A co-transformation approach wherein *P. tricornutum* was simultaneously transformed with the LjLUS + zeocin resistance gene marker (*ble<sup>r</sup>*), MtCYP716A12, and MtCPR coding sequences.
2. An iterative approach, where *P. tricornutum* was first transformed with an OSC construct and the *ble<sup>r</sup>* gene marker. The resulting lines were screened for lupeol production and the best strain was further transformed with a fused version of MtCYP716A12-CPR and the nourseothricin resistance selection marker (*nat<sup>r</sup>*).

For the co-transformation approach, the native genes encoding LjLUS, MtCYP716A12 and MtCPR were integrated into the genome of *P. tricornutum* in a single transformation, using three plasmids (Supplemental figure 1).

In the iterative approach, we fused the *M. truncatula* CYP716A12 and CPR enzyme, following a strategy inspired by the work of Leonard and Koffas, where they functionally expressed the *Glycine max* P450 isoflavone synthase (IFS) fused to a CPR from *Catharantus roseus* in *E. coli* (Leonard & Koffas 2007; Leonard et al. 2006). The full coding sequence of MtCYP716A12 without the stop codon was fused in frame to the MtCPR coding sequence with the exclusion of the first 71 amino acids, which are predicted to correspond to the transmembrane domain ( $\Delta 71$ CPR). The membrane anchor was deleted to avoid unwanted membrane insertion or association, which could compromise the enzyme activity. The two coding sequences were spaced by an additional short stretch of DNA coding for a flexible linker  $\lambda$  (GSTSSGSG), to prevent the formation of secondary structures, which could disrupt the native 3D structure of the two enzymes. The resulting final coding sequence,

MtCYP716A12- $\lambda$ - $\Delta$ 71CPR, was cloned under the control of the *P. tricornutum* FCPA/LHCF1 promoter. This vector was used to transform the best lupeol producing line: LjLUS-25 (Fig. 2).

For both approaches, the transformant lines were grown on selective solid medium, then transferred into liquid medium, and metabolites analyzed by GC-MS in early stationary phase.

All combinations and strategies resulted in transformant lines that produced betulin, the first intermediate in betulinic acid production (Fig. 7). Peak identity was confirmed by comparing the electron impact fragmentation patterns with those of known standards (Fig. S3E-I). Of the thirty independent lines that were screened per transformation, one co-transformation line and two fusion lines showed clear peaks for betulin. This intermediate is also the main product identified in recombinant yeast expressing this enzyme (Fukushima et al. 2011). As shown in Fig. 7, the peak of betulin is very low compared to lupeol and brassicasterol, making the quantification by GC-analysis challenging.

In the co-transformation approach, we showed that *P. tricornutum* is able to integrate three different plasmids in a single co-transformation event and functionally express the transgenes. This approach was mainly aimed to reduce time and cost of transformation, compared to the step-wise transformation, and to side-step the limited resistance markers available for selection, with only two well characterized antibiotic selection markers currently available for *P. tricornutum* (*ble<sup>r</sup>* and *nat<sup>r</sup>*). This compares favorably to *C. reinhardtii* where the frequencies for the successful expression of a single transgene, such as

GFP or luciferase, from the nuclear genome vary around 10% and it has been shown that strong expression can only be obtained by translation linkage with an antibiotic resistance marker or the usage of mutagenized strains (Neupert et al. 2009; Fuhrmann et al. 2004; Rasala et al. 2013). This would preclude a similar approach being applied in *C. reinhardtii* without extensive screening as performed in the 2016 study by Lauersen et al.

In the iterative approach, we demonstrated the possibility of successfully applying protein engineering approach to re-design the *M. truncatula* CYP716A12-CPR system, for lupeol modification. This allowed us to economize on promoters and antibiotic markers in systems with limited molecular biology tools. Although more effort is needed to improve the efficiency of betulin and betulinic acid production, these results establish the feasibility of using existing genetic engineering strategies for plant pathway reconstitution in *P. tricornutum*, making this microorganism a promising system for industrial biotechnology.

#### **Growth performance analysis and scaling up process**

The repurposing of sterol pathway intermediates and random nuclear genome insertion could have a detrimental effect on growth speed and biomass accumulation of the transformant lines. Therefore, we compared growth performances of lupeol and betulin producing lines with the wild type strain in an Algem® lab-scale photobioreactor. As shown in Fig. S4 and S5, no significant differences were observed between the wild type and any of the transformant strains. This suggests that the introduction of the OSC, cytochrome P450 and reductase genes did not affect cell growth and that the products of these enzymes are well tolerated at the levels produced.

Following these smaller-scale trials in Algem<sup>®</sup> photobioreactors, we compared the biomass productivity and lupeol content of three transgenic strains: AtLUS-6, LjLUS-25 and LjLUS+MtCYP716A12+MtCPR in a fence-style tubular photobioreactor (PBR). Transgenic cultures were grown in a 550 liter working volume photobioreactor, which offers control over pH, lighting, improved agitation and CO<sub>2</sub> delivery. Cultures were monitored over a period of 25 days for the first harvest, once they reached early stationary phase, with a cell density in line with the flask shake cultures.

In this large scale trial, despite displaying expected growth kinetics, line LjLUS-25 line did not recover following the initial harvest (fig. S6). This was unexpected as growth curves did not show any difference in behavior compared to the two other lines tested at the large scale.

This was likely due to an unidentified technical performance issue in the bioreactor system immediately prior, during or post-harvesting, since one of the well-performing lines (LjLUS+MtCYP716A12+MtCPR) is a derivative of the LjLUS25. The other two lines showed the expected growth performance, allowing multiple harvests (30 days for AtLUS6 and 20 days for LjLUS+MtCYP716A12+MtCPR). We performed a small scale extraction from the dry biomasses harvested from this trial, using an amount of dry biomass equivalent to approximately 1L culture. Supplementary table 2 shows the lupeol quantification for the AtLUS6 and LjLUS-25 lines, as expected from the growth fitness, we extracted lower amount of lupeol in LjLUS-25 line, compared to AtLUS6; this result might be due to the status of LjLUS-25 culture at the time of harvest. Further analysis on the scale up system and the extraction optimization from dry biomass will enable us to investigate the feasibility of this algal system for commercialization.

## CONCLUSIONS

Herein, we have demonstrated the feasibility of introducing plant transgenes in *P. tricornutum*, for the production of high-value sapogenins, without evident impact on growth and fitness. We show for the first time the possibility of introducing in *P. tricornutum* a plant pathway that requires the expression of genes corresponding to three different membrane-bound enzymes: a lupeol synthase, a P450 enzyme and a NADPH reductase. This is the most complex terpenoid pathway expressed to date in photosynthetic microorganisms. Previously, mono- and sesquiterpenoid production has been achieved in these organisms through the introduction of single enzymes. We have, therefore, demonstrated the feasibility and relative ease in terms of screening and genetic approach, of introducing multiple heterologous enzymes, to reconstitute a plant secondary metabolite pathway in a microalga. The results from this study provide novel insights on the regulation of the sterol biosynthesis pathway, which can be employed in future terpenoid pathway engineering. Gene expression patterns suggest that there is limited precursor availability, which could be overcome by the overexpression of key MVA pathway enzymes or the introduction of heterologous enzymes that bypass native regulation. A prime candidate would be the HMGR enzyme, as it has been shown to result in higher sterol and triterpenoid levels in several organisms (Chappell et al. 1995; Moses et al. 2014).

Further investigation to test the feasibility of having multiple enzyme fusions or using a synthetic scaffold approach to create metabolon-like complexes (Dueber et al. 2009), will allow improved triterpenoid production in *P. tricornutum*. The betulinic acid (BA) pathway is well suited for the latter approach, as the plant CYPs localize to the membrane and can assemble in dynamic complexes (Bassard et al. 2017). Moreover, we showed the feasibility

of growing this microalga at large scale. Although there is still substantial room for improving the product yield through bioprocess optimization and metabolic pathway engineering in *P. triornutum*, this first-generation strain and our results open a new perspective and are instructive with regard to the establishment of a new realm of investigations towards industrial biotechnology exploitation for this photosynthetic microorganism. Substantial advances will need to be made for algae with regard to both gene expression control and cultivation before this platform could become commercially feasible for the production of small molecules. To date the regulation of the lipid and sterol metabolism of diatoms is largely unknown; a recent study has shown that sterol biosynthesis can be chemically induced in a lipid independent manner (Prioretti et al. 2017). This study, therefore, represents pioneering work validating the potential of this eukaryotic microalgae as a chassis for the photo production of value-added molecules. Future studies, that include enzyme screening and engineering, modification on the carbon sink, semi-continuous batch system analyses, as well as use of a biorefinery approach for extraction of value-added molecules such as fucoxanthin and long chain PUFAs, could likely improve the overall productivity and economics towards the establishment of a renewable triterpenoid producing platform.

## **MATERIALS AND METHODS**

### **Strain and growth conditions**

*P. triornutum* CCAP1055/1 was grown in F/2 medium. Cells were transferred from plate to 5mL of F/2 medium for 5 days and then inoculated in 25ml as preculture until reaching mid-log phase (~3 days) and then transferred to fresh medium with a 1:5 dilution. Cells were all

grown at 20°C in New Brunswick™ incubator agitated at 120 rpm under white fluorescent light ( $60 \mu\text{mol m}^{-2}\text{s}^{-1}$ ).

### **Algem growth analysis**

For strain performance analysis, cells were grown in Algem® lab-scale photobioreactor systems (Algenuity, UK). Light was set to  $150 \mu\text{mol photons m}^{-2}\text{s}^{-1}$  using the 'sunlight' LED profile with temperature and shaking kept constant at 21 °C and 120 rpm respectively. The pH was kept at 8 by controlled bubbling with 3% CO<sub>2</sub> in air. OD<sub>740nm</sub> measurements were taken every 10 minutes using the in-built sensor.

To monitor lupeol and brassicasterol productivity and for gene expression analysis, cells were sub-cultured twice until they reached mid-exponential phase, and then inoculated at a cell culture density of  $2 \times 10^6 \text{ cells}\cdot\text{mL}^{-1}$  in fresh F/2 medium and grown in Algem® lab-scale photobioreactors. Light was set to  $400 \mu\text{mol photons m}^{-2}\cdot\text{s}^{-1}$  using the 'sunlight' LED profile with temperature and shaking pH control and OD measurements as described above.

### **Large scale cultivation**

For large scale growth, a fence-style photobioreactor (Bouygues Energies and Services, Manchester, UK) with a working volume of approximately 550 L was used. A photostage consisting of an array of horizontal polycarbonate 50 mm diameter tubes (36 tubes in total, in a 6 tube manifold formation) was illuminated at a light intensity of  $450 \mu\text{mol photons m}^{-2}\cdot\text{s}^{-1}$ , by eight 600W high pressure sodium lamps. Continuous circulation of the culture via centrifugal pump (up to  $30 \text{ m}^3 \cdot\text{h}^{-1}$ ) was allowed by stainless steel (316S) pipework connected to a holding tank (200 L). The holding tank was continually sparged with a flux of



air ( $\sim 10 \text{ L} \cdot \text{min}^{-1}$ ) for oxygen removal, while a pH-stat system controlled carbon dioxide delivery. Culture monitoring, which included oxygen saturation, pH, conductivity and temperature measurement was performed via a Profilux 3 interface. For media composition, F/2 nutrients (Varicon Aqua, United Kingdom) were used in combination with artificial sea water (Instant Ocean, USA).

### Plasmid design and transformation

pPhaT-1 (GenBank accession number: AF219942) vector including *AtLUS* (GenBank accession number: NP178018.1) and  $\text{ble}^r$  marker gene sequence was a gift from the group of Alain Goossens (PSB-VIB, Belgium). Sequences for *LjLUS: OSC3* (GenBank accession number: AB181245), *MtCYP716A12* (GenBank accession number DQ335781), *MtCPR*: MTR\_3g100160 (NCBI reference sequence XM\_003602850.2), *nat<sup>r</sup>*: (GenBank accession number: X73149) were synthesized as gBLOCKs (IDT, Leuven, Belgium) and introduced by Golden Gate cloning in final recipient plasmids, according to the design rules of the Open Plant standard (Patron & et al 2015) (Supplemental Fig. 1). Where necessary, PCRs were performed with Q5 polymerase (New England Biolabs, Hitchin, UK). All primers used are listed in supplementary table 3.

For the fusion of *MtCYP716A12* with *MtCPR*, firstly the sequence encoding for the predicted membrane-spanning domain of *MtCPR* (1-71 AA, 1-213bp) was removed from the original sequence and a sequence coding for a flexible linker GSTSSGSG ( $\lambda$ ) was added in-frame to the remaining sequence (72-692AA, 214-20179bp) creating a cDNA encoding for a  $\lambda\Delta 71\text{CPR}$ . The cDNA coding for *MtCYP716A12* without the stop codon was added in-frame by Gibson

assembly to the sequence  $\lambda\Delta 71\text{CPR}$ , in order to obtain a chimeric cDNA encoding a fusion version of MtCYP716A12 $\lambda\Delta 71\text{CPR}$ .

Biolistic transformation of *P. tricornutum* was performed according to methods previously described (P. G. Kroth 2007). Bombarded cells were transferred onto F/2 medium (with 1.7% of NaCl) agar plates containing either 75 $\mu\text{g}/\text{ml}$  zeocin (Invivogen) or 50 $\mu\text{g}/\text{ml}$  nourseothricin (Roth, Germany) depending on selection method. The selective medium plates were placed in 24hr illumination under fluorescent lights (60 $\mu\text{mol}\cdot\text{m}^{-2}\cdot\text{s}^{-1}$ ) and incubated at 20 °C for 3 weeks. Individual antibiotic-resistant colonies were transferred to fresh selective plates and subsequently transferred into liquid medium.

#### **GC-MS analyses**

All cell lines were grown until early stationary phase (~5 days) and harvested by centrifugation 4,200x *g* for 10 mins at 4°C. For screening of transformant lines, beta methyl cyclodextrin (Sigma Aldrich, DE) was added to 25 mL cell culture to a final concentration of 5mM, 24 hr prior to the harvest, cells were collected by centrifugation and supernatant containing beta methyl cyclodextrin was extracted with pentane and analyzed as previously described (Moses et al. 2015).

For quantification, analysis of lupeol cell pellets derived from 100 mL of cells were lysed by saponification and analyzed as previously described (Zhou et al. 2016). Derivatization was performed using N-methyl-N-(trimethylsilyl)trifluoroacetamide (Sigma Aldrich, UK). Standards were obtained from ExtrasyntheseSAS (GenayCedex, FR) and Molekula, (UK).

For betulin detection, after quick screen method with beta methyl cyclodextrin, as described above, 400 mL of the positive lines were grown until stationary phase (~5 days) and harvested and processed as described above for lupeol quantification analysis.

All GC-MS samples were run on a Thermo TRACE-GC DSQII quadrupole with a SLB-5MS column (Sigma Aldrich, UK). The injector temperature was set to 280°C and the MS-transfer line was set to 320°C. Oven temperature was set to 120°C and subsequently increased to 290°C (50°C/minute ramp), after which it was ramped to 300°C at (1°C/ minute).

#### **Copy number analysis and mRNA expression analysis**

RNA and DNA were extracted using TriReagent (Thermo) following the manufacturers protocol. First strand cDNA synthesis was performed using MMuLV (New England Biolabs, UK) with 500ng of total RNA as input. For gene expression analysis, cDNA was diluted 1/20 and the RP3a and UBQ genes were used as reference genes as described earlier (Matthijs et al. 2017). The same primers were used as normalizers for the gDNA CNV analysis.

Quantitative PCR was performed with JumpStart SYBR green ReadyMix (Sigma Aldrich, UK). Analysis was performed using the  $\Delta\Delta C_t$  method as implemented in CFX Manager (Livak & Schmittgen 2001) (Biorad, UK). All primers used are listed in supplementary table 3. ANOVA and Tukey tests were computed in R 3.43 using the agricolae 1.2-8 package. The significance

of correlation was determined by calculating a t-statistic using the formula:  $t = |r| \sqrt{\frac{n-2}{1-r^2}}$

## ACKNOWLEDGMENTS

We thank Sam Sizer and Denise Pallister for technical support. We are also grateful to Tessa Moses and Jacob Pollier for many helpful discussions. The research leading to these results has been partly funded from the Theme [KBBE.2013.3.1.01] Plant High Value Products – from discovery to final product of the European Union's Seventh Framework Program FP7/2007-2013/ under REA grant agreement n° [613692]

## Conflicts of interest

At the time of the study SD, JS, GS, GPL, APS and MM were employed by Algeniuty, a division of Spicer Consulting Limited, UK, which manufactures the Algem photobioreactor used for parts of this study.

## REFERENCES

- Andre, C.M. et al., 2016. Multifunctional oxidosqualene cyclases and cytochrome P450 involved in the biosynthesis of apple fruit triterpenic acids. *New Phytologist*, 211(4), pp.1279–1294.
- Anterola, A. et al., 2009. Production of taxa-4 (5), 11 (12)-diene by transgenic *Physcomitrella patens*. *Transgenic research*, 18(4), p.655.
- Apt, K.E., Kroth-Pancic, P.G. & Grossman, A.R., 1996. Stable nuclear transformation of the diatom *Phaeodactylum tricornutum*. *Molecular & general genetics : MGG*, 252(5), pp.572–579.
- Arendt, P. et al., 2016. Synthetic biology for production of natural and new-to-nature terpenoids in photosynthetic organisms. *The Plant Journal*, p.n/a-n/a. Available at:

<http://doi.wiley.com/10.1111/tpj.13138>.

Augustin, J.M. et al., 2011. Molecular activities, biosynthesis and evolution of triterpenoid saponins. *Phytochemistry*, 72(6), pp.435–457. Available at:

<http://www.ncbi.nlm.nih.gov/pubmed/21333312> [Accessed July 19, 2014].

Bansal, V. et al., 2016. Chapter 8 Diverse Utilization of Plant-Originated Secondary Metabolites. In *Plant Secondary Metabolites: Volume 3: Their Roles in Stress Ecophysiology*. Apple Academic Press, pp. 215–238.

Bassard, J.-E., Møller, B.L. & Laursen, T., 2017. Assembly of Dynamic P450-Mediated Metabolons—Order Versus Chaos. *Current Molecular Biology Reports*, pp.37–51.

Bourgaud, F. et al., 2001. Production of plant secondary metabolites: a historical perspective. *Plant science*, 161(5), pp.839–851.

Bowler, C. et al., 2008. The Phaeodactylum genome reveals the evolutionary history of diatom genomes. *Nature*, 456(7219), pp.239–44. Available at:

<http://www.ncbi.nlm.nih.gov/pubmed/18923393> [Accessed July 17, 2014].

Calvin, M., 1962. The path of carbon in photosynthesis. *Angewandte Chemie International Edition in English*, 1(2), pp.65–75.

Carelli, M. et al., 2011. Medicago truncatula CYP716A12 is a multifunctional oxidase involved in the biosynthesis of hemolytic saponins. *The Plant cell*, 23(8), pp.3070–81.

Cerutti, H. et al., 1997. Epigenetic silencing of a foreign gene in nuclear transformants of Chlamydomonas. *The Plant cell*, 9(6), pp.925–945.

Chappell, J. et al., 1995. Is the Reaction Catalyzed by 3-Hydroxy-3-Methylglutaryl Coenzyme A Reductase a Rate-Limiting Step for Isoprenoid Biosynthesis in Plants? *Plant*

*Physiology*, 109(4), p.1337 LP-1343. Available at:

<http://www.plantphysiol.org/content/109/4/1337.abstract>.

Cichewicz, R.H. & Kouzi, S.A., 2004. Chemistry, biological activity, and chemotherapeutic potential of betulinic acid for the prevention and treatment of cancer and HIV infection. *Medicinal Research Reviews*, 24(1), pp.90–114.

Cvejić, J.H. & Rohmer, M., 2000. CO<sub>2</sub> as main carbon source for isoprenoid biosynthesis via the mevalonate-independent methylerythritol 4-phosphate route in the marine diatoms *Phaeodactylum tricornutum* and *Nitzschia ovalis*. *Phytochemistry*, 53(1), pp.21–28. Available at:  
<http://www.sciencedirect.com/science/article/pii/S0031942299004653>.

Czarnotta, E. et al., 2017. Fermentation and purification strategies for the production of betulinic acid and its lupane-type precursors in *Saccharomyces cerevisiae*. *Biotechnology and Bioengineering*, 114(11), pp.2528–2538. Available at:  
<http://dx.doi.org/10.1002/bit.26377>.

Davies, F.K. et al., 2014. Engineering Limonene and Bisabolene Production in Wild Type and a Glycogen-Deficient Mutant of *Synechococcus* sp. PCC 7002. *Frontiers in bioengineering and biotechnology*, 2, p.21.

Davies, F.K., Jinkerson, R.E. & Posewitz, M.C., 2015. Toward a photosynthetic microbial platform for terpenoid engineering. *Photosynthesis research*, 123(3), pp.265–284.

Drag, M. et al., 2009. Comparison of the cytotoxic effects of birch bark extract, betulin and betulinic acid towards human gastric carcinoma and pancreatic carcinoma drug-sensitive and drug-resistant cell lines. *Molecules*, 14(4), pp.1639–1651.

Dueber, J.E. et al., 2009. Synthetic protein scaffolds provide modular control over metabolic flux. *Nature biotechnology*, 27(8), pp.5–8.

Eilers, U. et al., 2016. Limitations in the biosynthesis of fucoxanthin as targets for genetic engineering in *Phaeodactylum tricornutum*. *Journal of applied phycology*, 28(1),

pp.123–129.

Erpel, F., Restovic, F. & Arce-Johnson, P., 2016. Development of phytase-expressing *Chlamydomonas reinhardtii* for monogastric animal nutrition. *BMC biotechnology*, 16(1), p.29.

Fabris, M. et al., 2014. Tracking the sterol biosynthesis pathway of the diatom *Phaeodactylum tricornutum*. *The New phytologist*. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/24996048>.

Falciatore, a et al., 1999. Transformation of Nonselectable Reporter Genes in Marine Diatoms. *Marine biotechnology (New York, N.Y.)*, 1(3), pp.239–251. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/10383998>.

Franziska, H. et al., 2011. Microalgae as bioreactors for bioplastic production.

Fuhrmann, M. et al., 2004. Monitoring dynamic expression of nuclear genes in *Chlamydomonas reinhardtii* by using a synthetic luciferase reporter gene. , pp.869–881.

Fukushima, E.O. et al., 2011. CYP716A subfamily members are multifunctional oxidases in triterpenoid biosynthesis. *Plant and Cell Physiology*, 52(12), pp.2050–2061.

Gimpel, J.A. et al., 2015. In *Metabolic Engineering of Eukaryotic Microalgae : Potential and Challenges Come with Great Diversity.* , 6(December), pp.1–14.

Hamberger, B. & Bak, S., 2013. Plant P450s as versatile drivers for evolution of species-specific chemical diversity. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, 368(1612), p.20120426. Available at: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3538417&tool=pmcentrez&rendertype=abstract>.

Hamilton, M.L. et al., 2016. Heterotrophic production of omega-3 long-chain polyunsaturated fatty acids by trophically converted marine diatom *Phaeodactylum*

tricornutum. *Marine Drugs*, 14(3).

Hempel, F. et al., 2011. Algae as protein factories: expression of a human antibody and the respective antigen in the diatom *Phaeodactylum tricornutum*. *PLoS One*, 6(12), p.e28424.

Herrera, J.B. et al., 1998. Cloning and characterization of the *Arabidopsis thaliana* lupeol synthase gene. *Phytochemistry*, 49(7), pp.1905–1911.

Hsieh, M.-H. & Goodman, H.M., 2005. The *Arabidopsis* IspH Homolog Is Involved in the Plastid Nonmevalonate Pathway of Isoprenoid Biosynthesis. *Plant Physiology*, 138(2), p.641 LP-653. Available at: <http://www.plantphysiol.org/content/138/2/641.abstract>.

Huang, Z. et al., 2015. Production of dammarane-type saponins in rice by expressing the dammarenediol-II synthase gene from *Panax ginseng* CA Mey. *Plant Science*, 239, pp.106–114.

Husselstein–Muller, T., Schaller, H. & Benveniste, P., 2001. Molecular cloning and expression in yeast of 2, 3–oxidosqualene–triterpenoid cyclases from *Arabidopsis thaliana*. *Plant molecular biology*, 45(1), pp.75–92.

Jennewein, S. et al., 2005. Coexpression in yeast of *Taxus* cytochrome P450 reductase with cytochrome P450 oxygenases involved in Taxol biosynthesis. *Biotechnology and bioengineering*, 89(5), pp.588–598.

Jensen, K. & Møller, B.L., 2010. Plant NADPH-cytochrome P450 oxidoreductases. *Phytochemistry*, 71(2–3), pp.132–141.

Khakimov, B. et al., 2015. Identification and genome organization of saponin pathway genes from a wild crucifer, and their use for transient production of saponins in *Nicotiana benthamiana*. *The Plant Journal*, 84(3), pp.478–490.

Kim, D.H. et al., 2005. Enhancement of isoflavone synthase activity by co-expression of P450



reductase from rice. *Biotechnology letters*, 27(17), pp.1291–1294.

Kim, D.H. et al., 2009. Production of genistein from naringenin using *Escherichia coli* containing isoflavone synthase-cytochrome P450 reductase fusion protein. *Journal of Microbiology and Biotechnology*, 19(12), pp.1612–1616.

Kirby, J. et al., 2008. Engineering triterpene production in *Saccharomyces cerevisiae*-beta-amyrin synthase from *Artemisia annua*. *The FEBS journal*, 275(8), pp.1852–1859.

Kommerer, H. et al., 2011. Lupane triterpenoids—Betulin and betulinic acid derivatives induce apoptosis in tumor cells. *Investigational new drugs*, 29(2), pp.266–272.

Kroth, P., 2007. Molecular biology and the biotechnological potential of diatoms. *Advances in experimental medicine and biology*, 616, pp.23–33.

Kroth, P.G., 2007. Genetic transformation: a tool to study protein targeting in diatoms. *Methods in molecular biology (Clifton, N.J.)*, 390, pp.257–267.

Kvasnica, M. et al., 2015. Pentacyclic triterpenoids with nitrogen- and sulfur-containing heterocycles: synthesis and medicinal significance. *Natural product reports*, 32(9), pp.1303–1330.

Lang, C. & Lewandowski, A., 2016. Yeast Strain And Microbial Method For Production Of Pentacyclic Triterpenes And/or Triterpenoids. Available at: <https://lens.org/094-683-954-150-436>.

Lauersen, K.J. et al., 2016. Efficient phototrophic production of a high-value sesquiterpenoid from the eukaryotic microalga *Chlamydomonas reinhardtii*. *Metabolic Engineering*, 38, pp.331–343.

Lauersen, K.J. et al., 2016. Efficient phototrophic production of a high-value sesquiterpenoid from the eukaryotic microalga *Chlamydomonas reinhardtii*. *Metabolic engineering*, 38, pp.331–343.

- Leavell, M.D., McPhee, D.J. & Paddon, C.J., 2016. Developing fermentative terpenoid production for commercial usage. *Current Opinion in Biotechnology*, 37, pp.114–119.
- Leonard, E. & Koffas, M.A.G., 2007. Engineering of artificial plant cytochrome P450 enzymes for synthesis of isoflavones by *Escherichia coli*. *Applied and Environmental Microbiology*, 73(22), pp.7246–7251.
- Leonard, E., Yan, Y. & Koffas, M.A.G., 2006. Functional expression of a P450 flavonoid hydroxylase for the biosynthesis of plant-specific hydroxylated flavonols in *Escherichia coli*. *Metabolic Engineering*, 8(2), pp.172–181.
- Livak, K.J. & Schmittgen, T.D., 2001. Analysis of relative gene expression data using real-time quantitative PCR and the 2<sup>-Delta Delta C(T)</sup> Method. *Methods (San Diego, Calif.)*, 25(4), pp.402–408.
- Matthijs, M. et al., 2017. The transcription factor bZIP 14 regulates the TCA cycle in the diatom *Phaeodactylum tricornutum*. , pp.1–18.
- Møller, K. et al., 2008. Production of Plant Sesquiterpenes in *Saccharomyces cerevisiae* : Effect of ERG9 Repression on Sesquiterpene Biosynthesis. , 99(3), pp.666–677.
- Moses, T. et al., 2013. Bioengineering of plant (tri)terpenoids: from metabolic engineering of plants to synthetic biology in vivo and in vitro. *The New phytologist*, 200(1), pp.27–43.
- Moses, T. et al., 2014. Combinatorial biosynthesis of sapogenins and saponins in *Saccharomyces cerevisiae* using a C-16 $\alpha$  hydroxylase from *Bupleurum falcatum*. *Proceedings of the National Academy of Sciences of the United States of America*, 111(4), pp.1634–9. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/24434554> [Accessed August 7, 2014].
- Moses, T. et al., 2015. Unraveling the triterpenoid saponin biosynthesis of the african shrub *maesa lanceolata*. *Molecular Plant*, 8(1), pp.122–135.

- Moses, T. & Pollier, J., 2013. Bioengineering of plant (tri) terpenoids: from metabolic engineering of plants to synthetic biology in vivo and in vitro. *New Phytologist*, 200, pp.27–43. Available at: <http://onlinelibrary.wiley.com/doi/10.1111/nph.12325/full>.
- Munro, A.W., Girvan, H.M. & McLean, K.J., 2007. Cytochrome P450-redox partner fusion enzymes. *Biochimica et Biophysica Acta - General Subjects*, 1770(3), pp.345–359.
- Neupert, J., Karcher, D. & Bock, R., 2009. Generation of Chlamydomonas strains that efficiently express nuclear transgenes. , pp.1140–1150.
- Noordally, Z.B. & Millar, A.J., 2015. Clocks in algae. *Biochemistry*, 54(2), pp.171–183.
- Nrel, 1998. A look back at the U. S. Department of Energy's aquatic species program: biodiesel from algae. *Report*, 328, p.291 p. Available at: [http://www.biodiesel.pl/uploads/media/A\\_Look\\_Back\\_at\\_the\\_U.S.\\_Department\\_of\\_Energy\\_s\\_Aquatic\\_Species\\_Program\\_Biodiesel\\_from\\_Algae\\_July\\_1998.pdf](http://www.biodiesel.pl/uploads/media/A_Look_Back_at_the_U.S._Department_of_Energy_s_Aquatic_Species_Program_Biodiesel_from_Algae_July_1998.pdf).
- Paddon, C.J. et al., 2013. High-level semi-synthetic production of the potent antimalarial artemisinin. *Nature*, 496(7446), pp.528–32. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/23575629>.
- Pateraki, I., Heskes, A.M. & Hamberger, B., 2015. Cytochromes P450 for terpene functionalisation and metabolic engineering. *Advances in biochemical engineering/biotechnology*, 148, pp.107–139.
- Patron, N. & et al, 2015. Standards for Plant Synthetic Biology: A Common Syntax for Exchange of DNA Parts. *New Phytologist*, pp.13–19.
- Prioretti, L. et al., 2017. The inhibition of TOR in the model diatom Phaeodactylum tricornutum promotes a get-fat growth regime. *ALGAL RESEARCH-BIOMASS BIOFUELS AND BIOPRODUCTS*, 26, pp.265–274. Available at: <https://hal.archives-ouvertes.fr/hal-01620276>.

Pulz, O. & Gross, W., 2004. Valuable products from biotechnology of microalgae. *Applied microbiology and biotechnology*, 65(6), pp.635–648.

Rasala, B.A. et al., 2014. Enhanced Genetic Tools for Engineering Multigene Traits into Green Algae. , 9(4).

Rasala, B., Barrera, D. & Ng, J., 2013. Expanding the spectral palette of fluorescent proteins for the green microalga *Chlamydomonas reinhardtii*. *The Plant ...*, 74(4), pp.545–556.

Rodríguez-Concepción, M., 2006. Early Steps in Isoprenoid Biosynthesis: Multilevel Regulation of the Supply of Common Precursors in Plant Cells. *Phytochemistry Reviews*, 5(1), pp.1–15. Available at: <https://doi.org/10.1007/s11101-005-3130-4>.

Rouck, J.E. et al., 2017. Heterologous expression and characterization of plant Taxadiene-5 $\alpha$ -Hydroxylase (CYP725A4) in *Escherichia coli*. *Protein expression and purification*, 132, pp.60–67.

Ruiz, J. et al., 2016. Towards industrial products from microalgae. *Energy Environ. Sci.*, 24, pp.405–413.

Salmon, M. et al., 2016. A conserved amino acid residue critical for product and substrate specificity in plant triterpene synthases. , pp.1–8.

Sawai, S. et al., 2006. Functional and structural analysis of genes encoding oxidosqualene cyclases of *Lotus japonicus*. *Plant Science*, 170(2), pp.247–257. Available at: <http://www.sciencedirect.com/science/article/pii/S0168945205003018>.

Sawai, S. & Saito, K., 2011. Triterpenoid Biosynthesis and Engineering in Plants. *Frontiers in Plant Science*, 2(June), pp.1–8.

Schückel, J. et al., 2012. A Gene-Fusion Approach to Enabling Plant Cytochromes P450 for Biocatalysis. *ChemBioChem*, 13(18), pp.2758–2763.

Segura, M.J., Meyer, M.M. & Matsuda, S.P., 2000. *Arabidopsis thaliana* LUP1 converts

oxidosqualene to multiple triterpene alcohols and a triterpene diol. *Organic letters*, 2(15), pp.2257–2259.

Shibuya, M. et al., 1999. Two branches of the lupeol synthase gene in the molecular evolution of plant oxidosqualene cyclases. , 307, pp.302–307.

Shin, S. et al., 2016. OPEN CRISPR / Cas9-induced knockout and knock-in mutations in *Chlamydomonas reinhardtii*. *Nature Publishing Group*, (April), pp.1–15.

Suzuki, M. et al., 2004. Loss of function of 3-hydroxy-3-methylglutaryl coenzyme A reductase 1 (HMG1) in *Arabidopsis* leads to dwarfing, early senescence and male sterility, and reduced sterol levels. *The Plant Journal*, 37(5), pp.750–761. Available at: <http://dx.doi.org/10.1111/j.1365-313X.2004.02003.x>.

Thimmappa, R. et al., 2014. Triterpene biosynthesis in plants. *Annual review of plant biology*, 65, pp.225–257.

Vieler, A. et al., 2012. Genome, functional gene annotation, and nuclear transformation of the heterokont oleaginous alga *Nannochloropsis oceanica* CCMP1779. *PLoS genetics*, 8(11), p.e1003064.

Vincken, J.-P. et al., 2007. Saponins, classification and occurrence in the plant kingdom. *Phytochemistry*, 68(3), pp.275–297.

Wei, J.H., Yin, X. & Welander, P. V, 2016. Sterol Synthesis in Diverse Bacteria. *Frontiers in Microbiology*, 7, p.990.

Wijffels, R.H., Kruse, O. & Hellingwerf, K.J., 2013. Potential of industrial biotechnology with cyanobacteria and eukaryotic microalgae. *Current opinion in biotechnology*, 24(3), pp.405–413.

Wikfors, G.H. & Ohno, M., 2001. IMPACT OF ALGAL RESEARCH IN AQUACULTURE. *Journal of Phycology*, 37(6), pp.968–974.

Work, V.H. et al., 2013. Biocommodities from photosynthetic microorganisms.

*Environmental Progress and Sustainable Energy*, 32(4), pp.989–1001.

Work, V.H. et al., 2012. Improving photosynthesis and metabolic networks for the competitive production of phototroph-derived biofuels. *Current opinion in biotechnology*, 23(3), pp.290–7. Available at:

<http://www.ncbi.nlm.nih.gov/pubmed/22172528> [Accessed July 11, 2014].

<http://www.ncbi.nlm.nih.gov/pubmed/22172528> [Accessed July 11, 2014].

Xue, Z. et al., 2012. Divergent evolution of oxidosqualene cyclases in plants. *New Phytologist*, 193(4), pp.1022–1038.

*Phytologist*, 193(4), pp.1022–1038.

Zangar, R.C., Davydov, D.R. & Verma, S., 2004. Mechanisms that regulate production of reactive oxygen species by cytochrome P450. *Toxicology and applied pharmacology*, 199(3), pp.316–331.

*Toxicology and applied pharmacology*, 199(3), pp.316–331.

Zhou, C. et al., 2016. Improvement of betulinic acid biosynthesis in yeast employing multiple strategies. *BMC Biotechnology*, pp.1–9.

*BMC Biotechnology*, pp.1–9.

## FIGURE LEGENDS

### **Figure 1. Engineering *P. tricornutum* by introducing a plant sapogenin pathway.**

Schematic representation of sterol metabolism in diatoms starting from the precursor isopentenyl pyrophosphate (IPP) and leading to brassicasterol. In the plant sapogenin pathway the 2,3-oxidosqualene precursor is cyclized to lupeol, by a lupeol synthase; the lupeol can be further modified to betulinic acid through a cytochrome P450 family enzyme, along with its coenzyme NADPH reductase. The plant sapogenin pathway enzymes introduced in *P. tricornutum* will, therefore, compete with the native cycloartenol synthase for the common precursor 2,3-oxidosqualene.

**Figure 2. Lupeol production in *P. tricornutum* transformant lines.** GC-MS analysis of transformant lines AtLUS-6 (yellow) and LjLUS-25 (red) with the wild type (blue). Lupeol standard in black. 1. Brassicasterol (major sterol in *P. tricornutum*). 2. Lupeol; 3. 3 $\beta$ , 20-dihydroxylupane by-product of AtLUS enzyme only. Additional peaks: a. campesterol; b. unidentified diterpene molecule; \* unidentified terpenoid molecules, mass spectra for these peaks with their best database hits are shown in Supplementary Figure 3.

**Figure 3. Expression and copy number of lupeol producing strains. (A)** qPCR analysis of six lines showing expression of the *L. japonicus* lupeol synthase (*LjLUS*) mRNA relative to the geometric average of the *RP3a* and *UBQ* reference genes **(B)** Copy number of integrated *LjLUS* expression cassettes (black) and the *ble<sup>r</sup>* resistance marker (grey). Error bars represent the standard deviation from three biological replicates. Identical letters denote groups where means are not statistically different according to a *post hoc* Tukey test with  $\alpha = 0.01$ .

**Figure 4. Lupeol productivity in 400mL batch cultures.** Triplicate cultures were grown in F/2 medium in lab-scale Algem<sup>®</sup> photobioreactors for 5 days. **Panel A.** Cell density and productivity per cell of brassicasterol for the wild type (diamonds) and LjLUS-25 (squares). Lupeol was only detected in the LjLUS-25 line (triangles). **Panel B.** Brassicasterol extracted from cell pellet of WT (diamonds) and the LjLUS-25 line (squares). Lupeol extracted from LjLUS-25 is shown by the triangles in the cell (full line) and the medium (dashed line). No brassicasterol was detected in the medium of either the WT or LjLUS-25. **Panel C.** Gene expression of *LjLUS* and *ble<sup>r</sup>* resistance marker for LjLUS-25 transformant line (WT as

negative control) monitored during culturing time and relative to the reference genes *RP3a* and *UBQ*. Red arrows indicate the day when maximum lupeol productivity is observed corresponding to mid-exponential growth phase as described in the text. All error bars represent the standard deviation from 3 biological replicates. Identical letters denote groups where means are not statistically different according to a *post hoc* Tukey test with  $\alpha = 0.01$ .

**Figure 5.** Schematic representation of proposed model for extraction procedure with a biorefinery approach of triterpenes from *P. tricornutum* and *S. cerevisiae* biomass. The extraction procedure for triterpenoids from *P. tricornutum* simultaneously isolates more commercially relevant co-products compared to the yeast *S. cerevisiae*. Cost of extraction are estimated to be assuming equal levels of extraction efficiency. Cost of biomass for *P. tricornutum* is higher (produced in Europe, Necton), but market sizes of lipids omega 3 and carotenoids (unique products in microalgae) create cost offsets.

References for panel A:

**A** \*W. Yongmanitchai and O. P. Ward Growth of and Omega-3 Fatty Acid Production by *Phaeodactylum tricornutum* under Different Culture Conditions Appl. Env. Micro., Feb. 1991, p. 419-425 0099-2240/91/020419-07\$02.00/0.

**B** Yu-Hong Yang, Lei Du, Masashi Hosokawa, Kazuo Miyashita, Yume Kokubun, Hisayoshi Arai and Hiroyuki Taroda. Fatty Acid and Lipid Class Composition of the Microalga *Phaeodactylum tricornutum* J. Oleo Sci. 66, (4) 363-368 (2017) DOI: 10.5650/jos.ess16205

**C this paper**

**D** S. M. Tibbetts, J. E. Milley, P. Santosh and J. Lall Chemical composition and nutritional



properties of freshwater and marine microalgal biomass cultured in photobioreactors J Appl Phycol (2015) 27:1109–1119

References for panel B:

E International Journal of Environment, Agriculture and Biotechnology (IJEAB) Vol-2, Issue-2, Mar-Apr- 2017 <http://dx.doi.org/10.22161/ijeab/2.2.2> ISSN: 2456-1878 Page | 558

F M. Lamac Ka and J. S Ajbidor Ergosterol determination in *Saccharomyces cerevisiae*. Comparison of different methods. Biotech. Tech. 11, 723–725 (1997)

G E. Czarnotta, M. Dianat, M. Korf, F. Granica, J. Merz, J. Maury, S. A. Baallal Jacobsen, J. Förster, B. E. Ebert, and L. M. Blank. Fermentation and purification strategies for the production of betulinic acid and its lupane- type precursors in *Saccharomyces cerevisiae*. Biotech. Bioeng. 114, 2528-2538 (2017) DOI 10.1002/bit.26377

**Figure 6. mRNA Expression of sterol biosynthesis enzymes during a five day time course in the wild type (WT) and the LjLUS-25 line.** Schematic representation of sterol metabolism in

*P. tricornutum*, indicating the gene selected for mRNA expression analysis and relative heat map for WT and LjLUS-25 line. Values are the expression relative to the geometric average of

the *RP3a* and *UBQ* reference genes. Transcripts not significantly different between wt and

LjLUS are indicated in grey. Enzymes involved in the cytosolic MVA pathway: HMGS:

Hydroxymethylglutaryl-CoA synthase; HMGR: 3-hydroxy-3-methyl-glutaryl-coenzyme A

reductase; IDI-SQS: isopentenyl diphosphate isomerase-squalene synthase. Enzymes

involved in brassicasterol biosynthesis: PtOSC: oxidosqualene cyclase, methylsterol monooxygenase, sterol dehydrogenase, C24 sterol reductase, C22 sterol desaturase.

Enzymes involved in the plastidial MEP pathway: ISPD2: 2-C-methyl-D-erythritol 4-phosphate cytidyltransferase, ISP-E: 4-diphosphocytidyl-2-C-methyl-D-erythritol kinase. In

the yellow box, the non-native lupeol synthase from *Lotus japonicus*, introduced in our engineered strain LjLUS25. Dashed lines indicate multiple reactions. Gene accession numbers are mentioned in the results section. Precursor abbreviations: IPP: Isopentyl pyrophosphate, MEP: 2-C-methyl-D-erythritol 4-phosphate, , CDP-ME: 4-diphosphocytidyl-2-C-methylerythritol, CDP-MEP: 4-diphosphocytidyl-2-C-methyl-D-erythritol 2-phosphate, DMAP: Dimethylallyl pyrophosphate, G3P: Glycerol 3-Phosphate, Pyr: Pyruvate.

**Figure 7. Betulin production in *P. tricornutum* transformant lines. Panel A.** GC-MS analysis of WT (blue), LjLUS-25 containing only the *L. japonicus* lupeol synthase (LjLUS) (red); transformant line with constructs harboring *L. japonicus* lupeol synthase (LjLUS), *M. truncatula* cytochrome P450 (MtCYP716A12) and cytochrome P450 reductase (MtCPR) in WT background (purple); transformant line with construct harboring fused P450-reductase protein MtCYP716A12 $\Delta$ 72CPR in LjLUS-25 background (light blue); standards (black). Peak as indicated: 1. Brassicasterol; 2. Lupeol; 3. Betulin. **Panel B.** Detail of CYP716A enzyme reaction. The oxidation at C28 leads to betulin intermediate, which is converted to betulin aldehyde, and betulinic acid.

#### SUPPLEMENTAL LEGENDS

**Supplemental Figure 1. Illustration of plasmids used for *P. tricornutum* transformation.** All expressed gene were under control of native fucoxanthin chlorophyll a/c binding protein A (FCPA/LHCF1) promoter. **A.** Expression vector containing lupeol synthase from *A. thaliana* (*AtLUS*) and zeocin resistance (*ble<sup>r</sup>*) encoding genes **B.** Expression vector containing lupeol synthase from *L. japonicus* (*LjLUS*) and zeocin resistance (*ble<sup>r</sup>*) encoding genes. **C.** Expression

vector containing *M. truncatula* cytochrome P450 (*MtCYP716A12*) encoding gene. **D.** Expression vector containing *M. truncatula* cytochrome P450 reductase (*MtCPR*) encoding gene. **E.** Expression vector containing nourseothricin (*nat'*) resistance encoding gene. **F.** Expression vector containing engineered encoding gene of *M. truncatula* cytochrome P450 fused to *M. truncatula* cytochrome P450 reductase (*MtCYP716A12* $\Delta$ *71MtCPR*) and nourseothricin (*nat'*) resistance encoding gene.

**Supplemental Figure2. Molecular Phylogenetic tree of selected plant OSCs with the highest log likelihood by Maximum Likelihood method of Whelan and Goldman model [1].**

The percentage of trees in which the associated taxa clustered together is shown next to the branches. Initial tree(s) for the heuristic search were obtained automatically by applying Neighbour-Join and BioNJ algorithms to a matrix of pairwise distances estimated using a JTT model, and then selecting the topology with superior log likelihood value. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. All positions with less than 95% site coverage were eliminated. That is, fewer than 5% alignment gaps, missing data, and ambiguous bases were allowed at any position. There were a total of 708 positions in the final dataset. Evolutionary analyses were conducted in MEGA7 [2]. AaOSC2: *Artemisia annua*, AtLUS: *Arabidopsis Thaliana*, BpLUS: *Bupleurum lanceolate*, CrBAS: *Catharantus roseus*, ERG7: *S. cerevisiae* Lanosterol cyclase, GgBAS: *Glycyrrhiza Glabra*, LjLUS: *Lotus japonicus* OeLUS: *Olea europea*, ToLUS: *Taraxacum officinale*

1. Whelan, S. and Goldman, N. (2001). A general empirical model of protein evolution derived from multiple protein families using a maximum-likelihood approach. *Molecular Biology and Evolution* 18:691-699.

2. Kumar S., Stecher G., and Tamura K. (2016). MEGA7: Molecular Evolutionary Genetics

Analysis version 7.0 for bigger datasets. *Molecular Biology and Evolution* 33:1870-1874.

3. Felsenstein J. (1985). Confidence limits on phylogenies: An approach using the bootstrap. *Evolution* 39:783-791

**Supplemental Figure 3. GC-MS ionization pattern for lupeol and betulin peaks. A.** Lupeol peak from LjLUS-25. **B.** Lupeol peak of standard. **C.** Lupeol peak of AtLUS-6. **D.** 2-3 $\beta$  hydroxylupane from AtLUS-6. **E.** Lupeol peak of LjLUS+MtCYP716A12+MtCPR. **F.** Betulin peak from LjLUS+MtCYP716A12+MtCPR. **G.** Lupeol peak of LjLUS\_MtCYP716A12 $\Delta$ 71MtCPR. **H.** Betulin peak of LjLUS\_MtCYP716A12 $\Delta$ 71MtCPR. **I.** Betulin peak of standard. **L.** Peak **a** from WT, LjLUS-25 and AtLUS-6, identified as campesterol. **M.** Peak **b** from WT, LjLUS-25 and AtLUS-6 (unidentified diterpene-like molecule). **N.** First peak \* after 1 in AtLUS-6 (unidentified sterol-like molecule). **O.** Second peak \* after 1 in AtLUS-6 (unidentified sterol-like molecule). **P.** Third peak \* after 1 in AtLUS-6 (unidentified sterol-like molecule). **Q.** Fourth peak \* after 1 in AtLUS-6 (unidentified sterol-like molecule). **R.** Fifth peak \* after 1 in AtLUS-6 (unidentified sterol-like molecule).

**Supplemental Figure 4. Growth performance of the LUS transgenic lines in Algem<sup>®</sup> photobioreactors.**

**A.** Growth performance in Algem<sup>®</sup> photobioreactors of the AtLUS-6 line compared to the UTEX646 wild type strain (WT). **B.** Growth performance in Algem<sup>®</sup> photobioreactors of the LjLUS-25 strain compared to the CCAP1055/1 wild type strain (WT). Error bars represent standard deviation from three biological replicates.

**Supplemental Figure 5. Growth performance of the Betulin transgenic lines in Algem® photobioreactors.**

**A.** Growth performance of the selected MtCYP-MtCPR line transformed into the LjLUS-25 background. **B.** Growth performance of the MtCYP716A12- $\lambda$ - $\Delta$ 71CPR LjLUS co-transformation. Error bars represent standard deviation from three biological replicates.

**Supplemental Figure 6. Growth performance in 550L tubular PBR.** Growth curves for the three lines grown in 550L tubular photobioreactor. Red dots indicate harvest time point, green dots the restart culture. For AtLUS-6 approximately 454L was harvested at day 17, and the same volume replenished with fresh F/2 medium into the PBR for a second harvest after 21 days. For LjLUS\_25 285L have been harvested at day 29, then the culture started dying after restarting the PBR with fresh medium. For LjLUS\_MtCYP716A12\_MtCPR\_6, 373L were first harvested after 12 days, and the same volume replenished with fresh F/2 medium into the PBR for a second harvest (455L) after 3 days.

**Supplementary Table 1. Quantification of lupeol accumulation in the best performing transformant lines.** Lupeol extracted from cell pellet of best performing strains, selected from previous semi-quantitative analyses. Quantities were normalized per cell number, per liter of culture, and per gram of dry weight. Error bars represent standard deviation from three biological replicates.

**Supplementary Table 2. Scale-up conditions and specification.** Growth parameters of the 550L tubular photobioreactor (PBR) used for pilot scale production. Strain, liter of biomass harvested, cell densities at harvest and total cell dry weight harvested are shown. Lupeol quantification per gram dry weight was also determined.

**Supplementary Table 3. Oligonucleotides used in this study.** List of oligonucleotides used as PCR primers for domestication of golden gate parts and QPCR.

**Supplementary Table 4.** Sources for the economic analysis.

**Supplementary Table 5.** Repeated measures significance of the genes shown in Figure 6.















