Contents lists available at ScienceDirect

### Algal Research



journal homepage: www.elsevier.com/locate/algal

# N-terminal sequences affect expression of triterpene biosynthesis enzymes in *Chlamydomonas* chloroplasts



Shih-Chi Hsu<sup>a,2</sup>, Daniel R. Browne<sup>b,1</sup>, Mehmet Tatli<sup>b,1,3</sup>, Timothy P. Devarenne<sup>b</sup>, David B. Stern<sup>a,\*</sup>

<sup>a</sup> Boyce Thompson Institute, Ithaca, NY 14853, USA

<sup>b</sup> Department of Biochemistry and Biophysics, Texas A&M University, College Station, TX 77843, USA

| ARTICLE INFO   | A B S T R A C T   |
|--|---|
| Keywords:<br>Chloroplast<br>Protein expression<br>N-terminal sequence<br>Algae<br>Chlamydomonas reinhardtii<br>Botryococcene | Metabolic engineering is an emerging technology to modify the biochemical properties of living cells. In mi-<br>croalgae, metabolic engineering has often been directed towards optimizing the production of desirable lipids or<br>related bioproducts. Here we describe efforts to engineer the green alga <i>Chlamydomonas reinhardtii</i> for the<br>production of botryococcene, a drop-in biofuel precursor. Genes encoding farnesyl diphosphate synthase (FPS)<br>and squalene synthase-like (SSL)-1 and -3, were introduced into the chloroplast genome using biolistic trans-<br>formation. Through a series of construct modifications, we identified intergenic sequences that promote ex-<br>pression of stable, discrete transcripts. We also found amino acids that dramatically increased the accumulation<br>of SSL-3 when they were inserted at the N-terminal penultimate position, and similar manipulation of the N-<br>terminal sequence of FPS appeared to improve its protein level as well. However, SSL-1 only accumulated to<br>detectable levels when expressed as a chimera with SSL-3. <i>In vitro</i> assays showed that chloroplast-expressed SSL-<br>3 was enzymatically active, but not SSL-1, although the SSL-1-SSL-3 chimeras were active when expressed in<br>yeast. Taken together, our results suggest that the N-terminal sequence and other cellular factors are important |

when heterologous proteins are expressed in this model algal species.

#### 1. Introduction

The adaptation of microalgae as factories for useful molecules has received much attention in the past decade, whether for biomass production, oil production, or as a means to produce antigens or other health-related products. While there have been recent advances in the mutagenesis and transformation of various algal species (reviewed in [1,2]), the main species for proof of concept for potential production has been the green alga *Chlamydomonas reinhardtii. C. reinhardtii* is often selected because of its extensive toolbox including a high-quality genome sequence, simple transformability of each of its genetic compartments, facile genetic crosses, and rapid growth under laboratory conditions [3,4]. It has challenges for scaled production, however, which has limited its adoption for commercial platforms [5].

We have been using *C. reinhardtii* as a model for optimizing lipid production, based on using a lab-on-chip platform to screen for desired growth and production properties [6]. This initial screen was to detect and isolate strains with mutagen-induced properties of increased

https://doi.org/10.1016/j.algal.2019.101662

growth rate and/or lipid production, and demonstrated the potential of a microfluidic approach. Other related approaches have also been used to screen *C. reinhardtii* populations for variation in growth rate or lipid content [7,8]. Together, published results suggest that an approach of beginning with genetically modified cells, followed by high-throughput screening for enhanced traits, is a timely strategy.

The current work describes attempts to modify *C. reinhardtii* chloroplast metabolism to divert molecules from the isoprenoid pathway into production of specialized hydrocarbons called botryococcenes. Botryococcenes are a class of long-chain triterpenes found in the green colonial microalga *Botryococcus braunii*. The colonies of *B. braunii* may contain up to 80% hydrocarbon content (typically 30–50%; [9,10]). However, they are relatively slow growing, in their infancy in terms of genetic manipulation, and face challenges for large-scale production of biomolecules [10,11]. Analysis of *B. braunii* genes and the encoded proteins has shown that the C<sub>30</sub> botryococcene found in this species is generated through the action of two enzymes; squalene synthase-like 1 (SSL-1) produces presqualenediphosphate (PSPP) from two molecules

<sup>\*</sup> Corresponding author.

E-mail address: ds28@cornell.edu (D.B. Stern).

<sup>&</sup>lt;sup>1</sup> These authors contributed equally.

<sup>&</sup>lt;sup>2</sup> Current address: Synthetic Genomics, 11149 N Torrey Pines Rd, La Jolla, CA 92037.

<sup>&</sup>lt;sup>3</sup> Current address: Department of Bacteriology, University of Wisconsin, Madison, WI 53706, USA.

Received 9 July 2019; Received in revised form 9 September 2019; Accepted 9 September 2019 2211-9264/ © 2019 Elsevier B.V. All rights reserved.



Fig. 1. Construct design for the transformation of C. reinhardtii chloroplasts with the botryococcene biosynthetic pathway.

(A) The botryococcene biosynthetic pathway in B. braunii race B. Isopentenyl diphosphate is synthesized in the chloroplast and exported to the cytosol, where FPS sequentially condenses three isopentenvl diphosphate molecules/isomers into FPP. SSL-1 condenses two FPP molecules into PSPP, SSL-3 converts PSPP to C<sub>30</sub> botryococcene, and TMT-3 transfers the methyl donor group from S-adenosylmethionine to C<sub>30</sub> botryococcene to form dimethylated botryococcene. (B) Chloroplast transformation vector p16SOlacAF. The vector contains an inducible gene expression system (promoters and terminators in white boxes; GOI, gene of interest). The expression system consists of the C. reinhardtii 16S promoter (Prm16, striped box) and rbcL 5' UTR (stippled box), E. coli lac operators (open circles), C. reinhardtii psbA terminator following the GOI, and the Lac repressor cassette (lacl driven by the C. reinhardtii atpA promoter and flanked by the E. coli rRNA terminator). The nonphotosynthetic DEVL mutant, which has an aadA cassette replacing most of the rbcL gene, was used as the transformation recipient; the sequences for homologous recombination are indicated by broken lines. (C) Construct design. B. braunii SSL-1 and SSL-3, and chicken FPS coding regions were codon-optimized for C. reinhardtii chloroplasts and incorporated into the p16SOlacAF vector (only the GOI region is illustrated). The biosynthesis genes were designed to encode epitope tags (light gray boxes, not to scale) at the enzyme N- or C- terminus: FPS-Myc, 6xHis-SSL-1, and FLAG-SSL-3. Promoters are shown as striped boxes, 5' UTR sequences as stripped boxes; lac operators are depicted as ovals. The pV2-U construct contains a single promoter (16S promoter flanked by two lac operators), and has rbcL, atpB, and psbD 5' UTRs preceding the FPS, SSL-1, and SSL-3 genes, respectively. The terminators for FPS, SSL-1, and SSL-3 are E. coli thrA 3' UTR, and C. reinhardtii petD and psbA 3' UTRs, respectively; the C. reinhardtii chloroplast trnE1 (between thrA 3' UTR and atpB 5' UTR) and trnR1 (between petD 3' UTR and psbD 5' UTR) genes are drawn as diamonds. pV2-P was generated from pV2-U by adding atpB and psbD promoters to atpB 5' UTR-SSL-1 and psbD 5' UTR-SSL-3, respectively. A third construct (pV3) has the same design as pV2-P, except a Glu or Val codon (marked by an asterisk) is inserted after the start codon. The N-terminal sequence of each enzyme is shown below pV3.

of farnesyl diphosphate (FPP), while SSL-3 converts PSPP to  $C_{30}$  botryococcene (Fig. 1A; [12]).  $C_{30}$  botryococcene is then methylated by the enzyme triterpene methyltransferase-3 (TMT-3) to produce monoand dimethylbotryococcenes (Fig. 1A) [13], and by an unidentified enzyme(s) to produce tri- and tetramethylbotryococcenes.

The ability to divert isoprenoid precursors into botryococcenes was demonstrated in tobacco and the monocot *Brachypodium distachyon* 

through both cytosolic and chloroplast-localized pathways [14–16]. Additionally, botryococcene production has been targeted to the seeds of *Arabidopsis thaliana* [17]. In the case of tobacco, the chloroplast pathway was far more efficient, and transgenic plants produced significant levels of the desired compounds. However, plant growth and development was severely affected, possibly through an impact on membrane structure. The chloroplast pathway in *B. distachyon*,

however, severely affected plant development and reproduction, likely due to sensitivities of chloroplast isoprenoid metabolism, and resulted in lower botryococcene levels than those obtained through the cytosolic pathway. Botryococcene production has also been realized in yeast and several other microbes [12,18,19]. The conceptual basis exists, therefore, to extend this approach to the genetically tractable *C. reinhardtii*, which in addition to the attributes mentioned above, is a facultative heterotroph and therefore insulated against major impacts on photosynthesis. Here we have engineered the *C. reinhardtii* chloroplast genome to express proteins required for botryococcene production. Our results illustrate both the possibilities and limitations of this approach, and highlight opportunities for further refinement and optimization.

#### 2. Results

#### 2.1. Gene cassette design and selection of intergenic regions

We elected to conduct our attempts to engineer *C. reinhardtii* for botryococcene production through the introduction of chloroplast transgenes. In *C. reinhardtii*, chloroplast transgenes are placed at precise locations through homologous recombination and unlike nuclear transgenes, are not subject to silencing [20]. Furthermore, the chloroplast is polyploid, with an estimated 83 copies of the chloroplast genome per haploid cell [21], often conferring high levels of transgene expression. Additionally, the chloroplast is a major site of lipid biosynthesis, and thus is an appropriate compartment for engineered oil production. Finally, the upstream isoprenoid pathway in green algae is localized to the chloroplast [22].

We initially attempted to express all necessary exogenous enzymes (Fig. 1A) under the control of an IPTG-inducible promoter (pV1, Fig. S1A). The efficacy of this lac operator system, consisting of the lacI repressor gene (expressed constitutively) and two LacI binding sites (operators) flanking a chloroplast rRNA promoter (Figs. 1B and S1A), has been previously reported [23]. This regulated promoter was designed to drive expression of farnesyl diphosphate synthase (FPS), squalene synthase-like-1 and -3 (SSL-1/SSL-3), and triterpene methyltransferase-3 (TMT-3). FPS was overexpressed to increase the FPP pool for utilization by SSL-1 and subsequent conversion to botryococcene by SSL-3. Additionally, using online cellular localization prediction tools, the C. reinhardtii FPS (Uniprot A8IX41) is predicted to be localized in the cytosol. TMT-3 was used to methylate botryococcene in order to increase the amount of carbon in the biofuel precursor. The operon was designed to be inserted downstream of the rbcL gene, which was used as a selectable marker for transformation. The recipient strain, DEVL, has a deletion spanning most of *rbcL*, and recovery of photoautotrophic growth can be used for transformant selection [24]. Furthermore, DEVL possesses an aadA transgene, which confers streptomycin and spectinomycin resistance (Fig. 1B). Insertion of the botryococcene operon should be accompanied by deletion of *aadA*, and thus the absence of the aadA gene can be used to verify that transformants are homoplasmic (see Experimental procedures).

FPS, SSL-1, SSL-3 and TMT-3 coding sequences genes were synthesized to be in agreement with *C. reinhardtii* chloroplast codon usage, and interspersed with intergenic spacers derived from *Arabidopsis* chloroplasts (see Experimental procedures and Fig. S1A). Although intercistronic expression elements (IEEs; [25,26]) have been used to ensure processing of heterologous polycistrons into monocistrons and translation of heterologous proteins in land plant chloroplasts, it is unclear whether these IEEs can be correctly recognized and processed in an algal chloroplast, given that land plants and *C. reinhardtii* have distinct groups of RNA binding proteins [27]. Instead of seeking algal IEEs, which were little studied when we began this project, we decided to take an alternative approach to maintain a translatable heterologous polycistron. The chosen transcribed intergenic regions, *psbL/psbJ*, *psaB/ rps14*, and *psaC/ndhD*, were well studied in our laboratory; these RNA sequences remain unprocessed in *Arabidopsis*, resulting in polycistron accumulation in the chloroplast [28]. This observation suggested that these regions are likely free of RNase recognition sites, or may bind proteins that protect the polycistrons from being processed in these regions.

In addition, these intergenic sequences solely connect proteincoding genes and should therefore include ribosome-binding sites (RBS) for the downstream genes. While interchangeability of chloroplast RBS between species may not be universal, prokaryotic characters of gene expression are well conserved among chloroplasts, and both C. reinhardtii and artificial 5' UTRs have been successfully used in transplastomic constructs in plants [29,30]. If these sequences promote RNA stability and translation, they would make excellent building blocks for algal synthetic biology as they are short (less than 150 bp) and may promote equimolar production of each protein in a polycistron. Each coding region was also flanked by a single epitope tag (Myc, 6xHis, FLAG, or HA; Fig. S1A). Epitope tags were attached to either the N- or C-terminus of these enzymes based on data indicating locations that did not disrupt enzyme activities (J. Chappell, personal communication). The constructs were initially expressed in E. coli and after IPTG induction, bacteria carrying the pV1 construct produced FPS (Fig. S1B), demonstrating the efficacy of the inducible system. SSL-1 and SSL-3 were also detected, suggesting the RBS derived from A. thaliana psbL/ psbJ and psaB/rps14 were recognized by E. coli ribosomes, and would likely function in the algal chloroplast. We did not detect TMT-3 in bacteria (data not shown).

Having successfully tested functionality in E. coli, algal chloroplast transformation was conducted. Once homoplasmic transformants were obtained, cultures were grown to early log phase and then divided in half, with IPTG added to one aliquot. We tested induction at the RNA level for up to 24 h in preliminary experiments, and determined that induction peaked within 24 h. A more extensive time course was then performed to detail the induction and turnover of transcripts, as shown in Fig. S1C. The results showed a near-total repression of transcript accumulation in the absence of IPTG, and substantial RNA accumulation in its presence. Thus, the lac operator was functioning as expected. Total RNA was analyzed using gel blots with probes for each gene, all of which identified a ~5 kb transcript of the size expected for an mRNA initiating at the rrn16 promoter and terminating between TMT-3 and lacI (Fig. S1C). Numerous shorter bands were also observed, however, suggesting that the full-length transcript was unstable, and/or that undesired transcription termination was occurring at various points in the operon.

To check for expression of operon-encoded proteins, antibodies that recognized each of the four epitope tags were used. Multiple experiments, however, failed to provide evidence for expression at the protein level (Fig. S1D). This suggested either that the transcripts were not efficiently translated, and/or that the synthesized proteins were subject to rapid proteolysis.

Under the working hypothesis that instability and/or untranslatability of the polycistronic transcript led to the results shown in Fig. S1C and D, we created a second generation of constructs that would produce individual transcripts. As shown in Fig. 1C, pV2-U was designed to express FPS, SSL-1 and SSL-3 from a single, inducible promoter, but added intergenic RNA processing sites in the form of C. reinhardtii chloroplast tRNA genes, as previously employed in another transgenic context [31]. These tRNAs were preceded by inverted repeats, which are intended to form stem-loop structures that should stabilize the transcripts following tRNA excision. To further protect transcripts and promote successful translation, C. reinhardtii atpB and psbD 5' UTRs were placed before SSL-1 and SSL-3, respectively. This design is similar to tobacco synthetic operon modules containing 3' UTR-IEE-5' UTR intergenic sequences [25]. Plasmid pV2-P was similar to pV2-U, except that each cistron was preceded by its own promoter (atpB and psbD promoters). In this case, FPS should be inducible but SSL-1 and SSL-3 should be expressed constitutively. TMT-3 was removed from the construct as the gene encodes a methyltransferase that



Fig. 2. Selecting intergenic sequences for robust production of discrete transcripts.

Transcript profiles of representative V2-U (A) or V2-P (B) strains. Cultures were untreated (-) or treated (+) with 5 mM IPTG to induce transgene expression; samples were collected before (0 day) and one day after the addition of IPTG. Transcript levels were analyzed by RNA blot, and the 25S rRNA band is shown to reflect loading. The major processed product of each transcript is indicated by an arrowhead.

modifies botryococcene, but is not essential for producing unmodified  $C_{30}$  botryococcene (Fig. 1A). A third construct, pV3, will be discussed below, but is identical to pV2-P except that an extra amino acid is inserted following the initial Met codon.

Fig. 2A and B show RNA expression from strains V2-U and V2-P, respectively. In V2-U, FPS and SSL-1 were IPTG-inducible and correctly processed from a polycistron. A ~3.5-kb transcript identified by the SSL-1 probe was also induced by IPTG, and likely represents a dicistron containing FPS and SSL-1. The SSL-3 transcript failed to accumulate, and the non-inducible signals for SSL-3 in Fig. 2A are likely non-specific rRNA hybridization. It is possible that the polycistron was not fully transcribed and/or that the processed SSL-3 mRNA was degraded (see Discussion). When SSL-3 was provided its own promoter in strain V2-P, its RNA abundance mirrored that of FPS (after induction) and SSL-1, establishing a three-cassette design that yields accumulating transcripts of three foreign genes in the algal chloroplast (in addition to the lacI cassette). Similar to V2-U, long transcripts were observed in V2-P, suggesting incomplete transcription termination of 3' UTRs: a 3.5-kb transcript, likely a FPS-SSL-1 dicistron in the FPS blot, and a 5 kb tricistron identified by the SSL-1 and SSL-3 probes. In spite of producing readily detectable transcripts, transgene-encoded proteins were extremely difficult to detect in the V2 strains (Fig. S2). Only SSL-3 could be detected by immunoblot, at a low level, as shown for strain V2-P (Fig. S2C). Therefore, the failure to accumulate the desired foreign proteins in strain V1 could not be ascribed to the operon design, leaving open the question of whether translatability or protein instability accounted for the observed results.

### 2.2. The penultimate N-terminal residue can promote protein accumulation and functional SSL-3

Because the *cis* elements in our pV2 constructs consisted of wellcharacterized sequences that should promote transgene expression, we hypothesized that protein instability was an important limitation in our work. The N-end rule is a well-studied mechanism in prokaryotes and eukaryotes in which the protein N-terminus, often the second encoded amino acid after N-terminal methionine excision, or a residue revealed through other post-translational modifications, may be stabilizing, destabilizing, or neutral [32]. There is little experimental evidence bearing on an N-end rule in chloroplasts, however a series of GFP-based transgenes introduced into tobacco chloroplasts suggested a protein stability hierarchy consistent with such a rule [33]. N-terminal sequencing of *Arabidopsis* chloroplast proteins also showed a significant bias in the N-terminal amino acid, again consistent with an N-end rule [34].

We compared the second encoded amino acids of chloroplast-encoded proteins from *C. reinhardtii* with those of tobacco (Fig. 3A). With a few exceptions the frequencies in the two species were similar and furthermore, Ser, Thr, and Ala are also the most frequently encoded second amino acids in *Arabidopsis* chloroplasts [34]. In terms of putative stabilizing function we relied on experimental data from tobacco, which identified Glu, Met and Val as the most stabilizing residues, after taking into account RNA accumulation and translation rate [33]. On the other hand, the pV1 and pV2 constructs encoded His and Asp as second amino acids, which in tobacco were inferred to be among the most destabilizing.

To test whether transgene expression could be improved by utilizing this knowledge, we inserted Val or Glu codons immediately downstream of the initiating Met in *FPS*, *SSL-1* and *SSL-3*, as shown in Fig. 1C (pV3). RNA analysis of V3 strains carrying various N-terminal combinations showed that as in the progenitor V2 strain, *FPS* was inducible and the *SSL* genes were constitutively expressed (Fig. 3B; the threeletter abbreviations denote the inserted amino acids at the second positions in FPS, SSL-1, and SSL-3, respectively). Protein analysis showed that SSL-3 was readily detected whether the second encoded amino acid was Val (EVY, VVY) or Glu (VEE, VVE; Fig. 3C). A direct comparison of V2 and V3 RNA and protein levels is shown in Fig. 3D, and suggests that the single amino acid insertion dramatically improves SSL-3 accumulation. Nevertheless, there are additional factors determining the levels of FPS and SSL-1 in the chloroplast, as these proteins remained



(A) Frequency of amino acid occurrence at the second position of chloroplast-encoded ORFs in tobacco (Nt) and *C. reinhardtii* (Cr). (B) and (C) Transcript (B) and protein (C) profiles of representative V3 strains. Each strain is shown as a three-letter abbreviation indicating the amino acids inserted at the second positions of FPS, SSL-1, and SSL-3, respectively. Algal cultures were split into untreated (-) and IPTG-induced (+) fractions; both fractions were collected at the 24-h time point after the addition of IPTG. Transcripts were analyzed by RNA blot, and the 25S rRNA is shown to reflect loading. Proteins were analyzed by immunoblot, with the Rubisco large subunit (LS) shown to reflect loading. (D) Comparison of SSL-3 transcript and protein levels in V2 and V3 strains. V2-P and V3-VVE were grown in tandem to  $\sim 1 \times 10^7$  cells/mL, and equal numbers of cells were analyzed by immunoblot. Equal amounts of total RNA were analyzed by gel blot. (E) *In vitro* activity of SSL-3 produced in *C. reinhardtii*. Cell lysates were prepared from wild-type (CC-125) and the V3-VVE strain, and incubated with (+ SSL-1) and without recombinant SSL-1 (- SSL-1) and radiolabeled FPP. A reaction with recombinant SSL-1 and SSL-3 (SSL-1 + SSL-3) was included as a positive control. The average and standard deviation of duplicate data are shown.

undetectable with either Glu or Val insertions (Fig. S3).

To further explore the effect of the second amino acid on SSL-3 accumulation, an *SSL-3*-only construct was used to test five different amino acids at the N-terminal penultimate position. Glu and Val insertions again resulted in SSL-3 accumulation as did Met, consistent with results in tobacco chloroplasts (Fig. S4). Though His was reported

to be most destabilizing for GFP, SSL-3 was little affected by this insertion (Fig. S4). Most surprisingly, when an Asp was inserted, resulting in a Met-Asp-Asp-Tyr N-terminal sequence compared to the original Met-Asp-Tyr (in V1 and V2 strains), SSL-3 also accumulated to an equivalent level (Fig. S4). This result suggests that instead of protein abundance being determined by a particular amino acid at the second position, a key factor is the distance of a downstream element(s) in the SSL-3 RNA or protein sequence from the initial Met/start codon (see Discussion).

To determine whether the SSL-3 protein produced in C. reinhardtii chloroplasts is enzymatically active, a cell lysate was prepared from a V3-VVE line for an in vitro activity assay (Fig. 3E). To do so, radiolabeled FPP was incubated with the lysate along with E. coli-produced SSL-1, and the production of labeled botryococcene was analyzed. Additionally, the production of labeled squalene was analyzed since SSL-3 can also convert SSL-1-produced PSPP into squalene [12]. E. coliexpressed and purified SSL-1 plus SSL-3 proteins were used as a positive control. Only the VVE lysate plus SSL-1 successfully converted radiolabeled FPP into botryococcene (Fig. 3E, VVE + SSL-1), as did bacterial recombinant SSL-1 and SSL-3 (Fig. 3E, SSL-1 + SSL-3), suggesting that the C. reinhardtii-produced SSL-3 protein is functional. Radiolabeled squalene was produced by wild-type (CC-125) and VVE lysates since C. reinhardtii has an endogenous squalene synthase [35]. In the absence of bacterial SSL-1, the VVE lysate synthesized only a low level of labeled squalene but not botryococcene (Fig. 3E), in accord with the lack of SSL-1 signal in the immunoblots.

## 2.3. Botryococcene was not detected in the presence of chloroplast FPS and SSL-1-SSL-3

Compared to co-expressing SSL-1 and SSL-3 as individual proteins, SSL-1-SSL-3 and SSL-3-SSL-1 fusions, connected by a triplet repeat linker of GGSG, synthesized more than twice as much botryococcene in transgenic yeast [12]. Since our V3 strains with modified SSL-3 N-termini produced functional SSL-3 and reasonable protein accumulation, we hypothesized that SSL-3-SSL-1 chimeras would benefit from this Nterminal sequence, allowing the accumulation of SSL-1 sequences and its attendant activity. The same peptide linker, with its coding sequence adjusted according to C. reinhardtii chloroplast codon bias, was used to attach SSL-1 to the C-terminus of Glu-inserted SSL-3 (Fig. 4A, pV4a). We also designed an inverted version of the chimera in which SSL-1 is at the N-terminus, followed by the GGSG repeat and Glu-inserted SSL-3 (pV4b). The SSL-1 polypeptide in both chimeras was no longer epitopetagged, while SSL-3 maintained its N-terminal FLAG tag. Both chimeric sequences were placed in the SSL-3 cassette (psbD promoter and 5' UTR and psbA 3' UTR found in pV2, pV3, and SSL-3-only constructs). Additionally, since Val and Glu did not improve FPS and SSL-1 levels in V3 lines, Met, another strong stabilizing amino acid at the second position in tobacco [33] that also yielded SSL-3 accumulation (Fig. S4) was inserted to replace the second Val/Glu of FPS and the SSL-1 moiety of SSL-1-SSL-3. We also swapped the C-terminal Myc-tag of FPS for a FLAG-tag, which had worked well for SSL-3.

We screened more than 50 putative transformants of each construct for chimeric protein levels and identified two SSL-1-SSL-3 strains, #24 and #37, with the highest amount of fusion protein (Fig. 4B). RNA gel blot analysis showed that the FPS transcript was induced by IPTG, as expected, and that the SSL-1-SSL-3 chimeric transcript was constitutively expressed from the psbD promoter (Fig. 4C). After 36 h of induction, a small amount of FPS protein was detected (Fig. 4B), indicating that the Met insertion and/or a C-terminal FLAG tag improved FPS accumulation relative to Val or Glu and a Myc tag at the same positions. SSL-1-SSL-3 was present as a band estimated at 92 kD, the size expected for a fusion protein. This represented our first success in detecting SSL-1 protein sequences, which we ascribe to some or all of the following: 1) engineering the second encoded amino acid as Met; 2) changing non-coding sequences surrounding the SSL-1-containing transgene from the atpB 5' UTR and petD 3' UTR to those of psbD and psbA, respectively; and/or 3) fusing SSL-3 to the C-terminus of SSL-1. SSL-3-SSL-1 transformants were also screened, however none showed equivalent protein expression to strains #24 and #37 derived from pV4b (data not shown).

vitro assays of cell lysates (Fig. S5). We found that while there was weak SSL-3 activity, SSL-1 activity was not detected as shown by the lack of botryococcene production in lysates without the addition of recombinant SSL-1 (panel A). IPTG-induced strains were further analyzed by GC-MS, and production of botryococcene was not detected, suggesting SSL-1 was inactive in vivo as well (Fig. 4D). It is also possible that the chloroplast FPS was inactive in vivo and did not produce FPP, the precursor of botryococcene. The activity of chloroplast FPS was not assayed due to the inability to distinguish the activity of native C. reinhardtii FPS from the introduced, heterologous enzyme. Since B. distachyon expressing a chloroplast-targeted SSL chimera but not FPS still produced botryococcene [16]. FPS is an unlikely culprit for undetectable botryococcene production in V4b. To test whether our constructs are able to produce functional fusion enzymes in a different context, we expressed the chimeric sequences in yeast following an established protocol [12]. The Chlamydomonas chloroplast SSL fusion genes were likely to function in yeast, since when they were analyzed for their codon adaptation index value [36] using yeast codon usage tables, a value of 0.891 was obtained, close to the maximum theoretical value of 1. GC-MS analysis of the yeast transformants showed that both chimeric designs resulted in the production of botryococcene, demonstrating that the encoded fusion proteins were successfully expressed and enzymatically active (Fig. S6). These results suggest that the SSL-1 activity was impaired in C. reinhardtii chloroplasts. Potential strategies to resolve this inactivity and advance metabolic engineering in algal chloroplasts are discussed below.

#### 3. Discussion

In this study, we advanced the state of C. reinhardtii chloroplast transformation for production of foreign proteins through an iterative series of modifications. First, we employed the lac operon system devised by Kato et al. [23], which to our knowledge had not been further tested. Both repression in the absence of IPTG, and strong RNA expression in its presence, were confirmed. Second, we found that creation of operons, a single promoter coupled with a tRNA punctuation model, and multiple promoters all showed promise, although RNA instability and transcription termination may be confounding factors. Finally, we found that codon-optimized transcript and/or peptide sequences of FPS, SSL-1, and SSL-3, did not necessarily lead to protein accumulation in the organelle. Through inserting an amino acid at the penultimate position of these proteins, and in the case of SSL-1, removing an N-terminal tag and linking it to the N-terminus of SSL-3 as well, we were able to detect all three proteins in the alga. All five amino acids tested, Met, His, Val, Glu, and Asp, were able to increase the level of SSL-3 when they were inserted as the second amino acid. With a Met insertion and C-terminal FLAG tag, FPS accumulated at a low level after its production was induced, while the original and Val/Glu-inserted FPS-Myc proteins were undetectable. Although the detailed mechanism is unknown, our results suggest that a single amino acid insertion at the N-terminus may be applicable as a rapid strategy to rescue poor transgenic protein accumulation in the chloroplast, a problem encountered in some other attempts to overexpress foreign proteins in C. reinhardtii chloroplasts [37,38].

An early report showed that transcripts with spinach and wheat chloroplast 5' UTRs were unstable in *C. reinhardtii* chloroplasts [39]. Though our V1 transcript contains a *C. reinhardtii* 5' UTR, we observed extensive processing/degradation of the polycistron, suggesting additional sequence element(s) may control transcript stability in the algal chloroplast. The instability may be caused by incompatibility between the transcript and *C. reinhardtii* RNA-binding proteins. Chloroplast transcripts often depend on sequence-specific nucleus-encoded proteins to define their termini and protect them from ribonucleases which may otherwise be indiscriminate [40]. In *C. reinhardtii*, the main protein family with this role appears to be octotricopeptide repeat (OPR) proteins, for which over 120 members are encoded [27,41]. The V1



Fig. 4. FPS and SSL-1-SSL-3 proteins are produced in the fourth-generation strains.

(A) Construct designs for SSL fusions. pV4a and pV4b contain the same controllable gene expression system as previous constructs, and the *lac* operators are depicted as ovals. Dotted lines are used to connect different DNA elements and do not represent actual sequences. Promoters are shown as striped boxes, and 5' UTR sequences as stippled boxes (16S promoter flanked by two *lac* operators and *rbcL* 5' UTR for *FPS*; *psbD* promoter and 5' UTR for SSL fusion genes). The terminators for *FPS* and the SSL fusion genes are *E. coli thrA* 3' UTR and *C. reinhardtii psbA* 3' UTRs, respectively; the *C. reinhardtii* chloroplast *tmE1* is drawn as a diamond and separates two gene cassettes. FLAG tag coding sequences are shown as light gray boxes (not to scale). An asterisk depicts the insertion of amino acid at the second codon position: a Met insertion in *FPS*, a Glu insertion in *SSL*-3 of pV4a, and Met and Glu insertions in *SSL*-3 of pV4b. The 3xGGSG linker coding sequences that connect SSL-1 and SSL-3 genes are shown as black boxes. (B) and (C) V4b protein and transcript profiles. Cultures were split into untreated (-) and IPTG-induced (+) fractions; both fractions were collected at the time points shown after IPTG addition. Equal numbers of cells were used for immunoblot analysis, and equal amounts of RNA were analyzed by gel blot. The major protein and transcript species are indicated by arrowheads. A FLAG antibody cross-reacting protein of 45 kD is marked by an asterisk. (D) GC–MS profiles of *n*-hexane extracts of the V4b strains. Panels 1-III, V4b strain #24. I, 12 h; II 36 h; III, 48 h after IPTG addition. VII, C<sub>30</sub> botryococcene standard.

transcript with *Arabidopsis* intergenic sequences might not be recognized in the algal chloroplast, leaving it vulnerable to endonucleolytic cleavage. Data from the *C. reinhardtii ncc1/ncc2* mutants also suggest that certain OPRs may promote transcript degradation through RNA interactions [42].

To promote processing in intergenic regions, we included tRNA sequences in the second-generation constructs (Fig. 1C), akin to a previous strategy we employed in the study of 3' RNA stability elements [31]. We also modified the UTR sequences of transgenes to improve transcript stability. So far, *C. reinhardtii* nuclear factors that are known to stabilize/mature chloroplast transcripts all target the 5' end, while stem-loop structures in 3' UTRs define and protect the 3' end of transcripts [27]. We therefore integrated *C. reinhardtii* chloroplast 5' UTRs and a 3' UTR that is predicted to form a stem-loop into the V2 constructs. One *E. coli* 3' stem-loop known to stabilize a transcript in the algal chloroplast [43] was adopted as well. The processing of the V2-U

polycistron into monocistrons resulted in two discrete and accumulating transcripts, *FPS* and *SSL-1* (Fig. 2A). Therefore, the intergenic sequence (*E. coli thrA* 3' UTR-*C. reinhardtii trnE1-atpB* 5' UTR) between these two genes can potentially be used for designing algal synthetic operons and its length (548 bp) is comparable to the recently tested algal IEE sequences, *psbN-psbH* (569 bp) and *tscA-chlN* (650 bp, [44]). Whether such transcripts can be translated, however, remain to be tested.

The failure to detect the *SSL-3* transcript in V2-U could in principle result from transcription termination at the *petD* 3' UTR flanking *SSL-1*, however no *C. reinhardtii* stem-loop has been found to terminate at greater than a 50% rate [45]. There are cases, however, where 3' processing at stem-loops is coupled to a vectorial degradation of the sequences downstream [46], a process that seems to involve interactions with nucleus-encoded factors [47]. We also cannot rule out that cryptic RNA instability elements exist within the *SSL-3* moiety. In

particular, SSL-3 in V2-U, where it fails to accumulate, vs. V2-P, where it does accumulate, contain different 5' sequences. The difference lies between the tRNA<sup>Arg</sup> moiety and the 74-nt psbD 5' UTR, where a 50-nt sequence downstream of trnR is adjacent to psbD 5' UTR in V2-U, but the *psbD* promoter is interspersed between *trnR* and the *psbD* 5' UTR in V2-P. Therefore, the SSL-3 transcript in V2-U likely contains a 5' UTR longer than 74 nt after tRNA processing compared to its counterpart in V2-P, where the 5' UTR would be restricted to psbD promoter-derived sequences. The *psbD* 5' UTR is known to be processed into a 47-nt-long sequence in the context of the predominant and actively translated form of psbD RNA [48]. This processed sequence contains a critical stabilizing element in its first 12 nucleotides, and the maturation of the short form from the long form seems to be essential for the stability of the short form [48]. Additionally, a 47-kD protein is known to bind the long but not short form of the psbD 5' UTR [49], and this protein and possibly other trans-acting factors may be required for the maturation and stability of the transcript. In V2-U, the additional sequence preceding the 74-nt UTR may interfere with the binding of such trans-factors and/ or mask the stabilizing element in the sequence, resulting in unstable transcripts.

While V2-P strains produced apparently abundant transcripts for each gene, we were unable to detect FPS and SSL-1 by immunoblotting, and SSL-3 only accumulated at a low level (Fig. S2). This implicates translation initiation, translation elongation, and/or protein instability as culprits. Based on a previous publication in tobacco chloroplasts using a GFP reporter [33], we inserted Glu or Val at the second position of transgenic proteins in V3 strains, a slightly different strategy than the second position substitution in the tobacco study. While both Glu and Val insertions increased the level of SSL-3 (Fig. 3C), FPS and SSL-1 remained undetectable (Fig. S3). Similarly, GFP proteins with different N-terminal fusions (8-9 residues of N-terminal peptides from certain chloroplast proteins), but with the same second amino acid, accumulated at distinct levels [33]. Together, our results and previous reports suggest that there are N-terminal elements other than the penultimate amino acid determining the protein level. Since the transcript levels of the second- and third-generation SSL-3 were comparable (Fig. 3D), it is possible that the translation efficiency of SSL-3 increased and/or the degradation rate of the protein decreased in the third-generation strain. A previous study has shown that the 5' coding region of a chloroplast transgene was critical for translation initiation, and a 30-bp insertion in the 5' coding region allowed translation of an otherwise untranslatable viral coding sequence [50]. An inhibitory sequence may also be in the 5' coding region of our transgenes. The effect of coding region determinants on foreign protein production was also encountered when C. reinhardtii chloroplasts were engineered to produce insecticidal toxins [51].

The N-end rule is an important protein degradation pathway found in the cytoplasm of bacteria and eukaryotes. Although mechanisms controlling this pathway are different in bacteria and eukaryotes, the protein degradation follows similar routes: (i) proteins with a primary destabilizing N-terminal residue are recognized and sent for degradation, and (ii) if the N-terminus is a secondary (and tertiary, in eukaryotes) destabilizing residue, it is modified and the new N-terminus with a primary residue is recognized for degradation [52]. Recent studies suggest that an N-end rule-like pathway may exist in chloroplasts [33,34], but its components and mechanism remain unclear. When a protein is synthesized in the chloroplast, the initiating methionine is deformylated and then often excised from the protein (Nterminal methionine excision, NME) [53]. As in bacteria, these processes are likely to take place during translation, before a protein is fully folded [53].

The rule of NME seems to be similar in all organisms, and the removal of Met is inhibited when the second amino acid possesses a bulky side chain. Apel et al. [33] were able to verify the N-terminus of one transgenic GFP protein: Met was retained when Glu was the second residue. It remained to be determined whether the N-end, after applying the rule of NME, played a major role in the stability of GFP. In our study, V2-encoded SSL-3 contains Asp at the penultimate position, which hinders the Met removal as shown in bacteria and chloroplasts [34,53,54]. We tested five different insertions following the initiating Met of SSL-3, and they all promoted protein accumulation (Fig. S4). The initiating Met should be present in mature proteins when Glu or Asp is inserted as the new penultimate residue. When Val is the second residue, NME is less predictable, as some proteins keep the N-terminal Met in bacteria and chloroplasts [34,54]. In bacteria, Met at the second position prevents the removal of the first Met [54], while the first and second Met of TIC214/Ycf1 were both absent in the Arabidopsis chloroplast N-terminome [34]. SSL-3 with a Met insertion likely maintains at least one N-terminal Met since the subsequent residue is Asp. In bacteria, His also prevents the cleavage of the N-terminal Met, but processing in chloroplasts is less clear since none of the Arabidopsis chloroplast-encoded proteins has His as the second residue. Although the C. reinhardtii chloroplast encodes one such protein, RpoC1a, its in vivo N-terminal sequence is unknown. It is possible that all five insertions lead to inhibition of NME in SSL-3. Therefore, the N-termini of V2and V3-derived SSL-3 are both likely to be Met, and the N-end rule is unlikely to account for the different SSL-3 protein levels.

We utilized epitope tagging for the detection of algal recombinant proteins. FPS-Myc, 6xHis-SSL-1/3, and N-terminal FLAG-tagged squalene synthase (a paralog of SSLs) remained enzymatically active in plants and bacteria (J. Chappell, personal communication), so we adopted the same tags in our initial construct design. An N-terminal FLAG tag indeed did not interfere SSL-3 function (Fig. 3E), but might have an adverse effect on SSL-3 protein levels unless an amino acid is inserted between the initiating Met and the tag (Fig. 3D). An N-terminal His-tag may be detrimental to protein production in chloroplasts, since we were unable to detect His-tagged SSL-1, with or without an insertion at the second position. When we replaced the His tag of SSL-1 in pV3 with Glu-FLAG, we identified strains producing FLAG-tagged proteins (we were unable to distinguish SSL-1 from SSL-3 as their sizes are similar), but only SSL-3 activity was detected by the in vitro enzyme assay (data not shown). Therefore, FLAG-SSL-1 likely was not produced and an unknown factor other than the N-terminal sequence was affecting its accumulation. Changing the V3 version by altering the C-terminal tag to FLAG and the inserted amino acid to Met resulted in detectable FPS accumulation in V4 strains. It is possible that Met is superior in promoting FPS production compared to Glu/Val. Altering the C-terminal tag may also increase protein levels since Apel et al. showed that Cterminal sequences altered GFP levels in tobacco chloroplasts, although the effect was much less pronounced than that of N-terminal sequences [33].

Although two V4b strains accumulated both FPS and the SSL-1-SSL-3 chimera, we were unable to detect botryococcene by GC-MS, nor in vitro activity of SSL-1 (Figs. 4D and S5A). It is possible that the SSL-1-SSL-3 level was too low to detect botryococcene biosynthesis in vivo or in vitro. Another possibility is that the SSL-1 moiety of the chimeric protein was non-functional as produced in the chloroplast, since the addition of recombinant SSL-1 conferred the ability to produce botryococcene in vitro (Fig. S5A). When the same SSL fusions were expressed in yeast, they were enzymatically active (Fig. S6). In genetically engineered tobacco plants that produced botryococcene, SSL fusions were expressed in the cytosol and then imported into the chloroplast [14]. Therefore, the chimeric enzyme, or at least the SSL-1 polypeptide, may need to be synthesized in the cytosol for proper folding and/or post-translational modifications to be enzymatically active. Since botryococcene production was detrimental for chloroplast and plant development in tobacco [14], the functional chimeric protein may also create a stress condition in C. reinhardtii chloroplasts. The inducible system used in this study was leaky, and we might have inadvertently selected V4 strains with inactive SSL-1 which was innocuous to the organelle and organism. Recently, Phaeodactylum tricornutum, a diatom, was engineered to produce a sapogenin in its cytosol, suggesting

production of triterpenoids in a different algal compartment is feasible [55]. Therefore, future work on algal metabolic engineering would require improvements in compartment-specific expression or a tight inducible system in the organelle, of which several have been created [56–59].

#### 4. Experimental procedures

#### 4.1. Plasmid construction

All constructs were created by Gibson assembly (New England Biolabs, Ipswich, MA), and have been deposited under accession numbers MK052916 (pV1), MK052917 (pV2-P), MK052918 (pV2-U), MK052919 (pV4a), and MK052920 (pV4b). The individual peptide or gene accession numbers are as follows: FPS (P08836), SSL-1 (HQ585058), SSL-3 (HQ585060), and TMT-3 (JN828964). The coding sequences of FPS-Myc, 6xHis-SSL-1, FLAG-SSL-3, and HA-TMT-3 were synthesized and optimized by Bio Basic (Amherst, NY) based on C. reinhardtii chloroplast codon bias [http://www.kazusa.or.jp/codon/cgibin/showcodon.cgi?species = 3055.chloroplast [60]]. The intergenic sequences of pV2-P, [E. coli thrA terminator [43] -C. reinhardtii trnE1 region (nucleotides 14,460-14,585, accession NC\_005353)-C. reinhardtii atpB promoter and 5' UTR (nucleotides 162,002-161,641; [61])] and [C. reinhardtii petD 3' UTR (including trnR1; nucleotides promoter 6,865-7,259)-psbD and 5' UTR (nucleotides 175,839-175,633; [48])] were synthesized by Integrated DNA Technologies (Skokie, Illinois). Synthetic DNA fragments were assembled and cloned into p16SOlacAF [23] replacing the uidA gene to create pV2-P. pV2-U was constructed by removing the *atpB* and *psbD* promoter sequences (nucleotides 162,002-161981 and 175,839-175,707, respectively; accession NC\_005353) of pV2-P. Degenerate primers were used to amplify transgenes and insert a valine (GTA) or glutamic acid (GAA) codon after the start codon of each transgene: these mutated sequences were assembled into pV3 constructs. For pV4 constructs, several changes were made in the coding region: by primer design and PCR amplification, the Myc tag of FPS was changed to a FLAG tag, the 6xHis tag of SSL-1 was removed, and instead of a valine or glutamic acid residue, a methionine was inserted at the second position of FPS and SSL-1. SSL-1-SSL-3 and SSL-3-SSL-1 fusions were created by PCR amplification with 3xGGSG linker-encoding primers and Gibson assembly. The FPS-thrA terminator-trnE1 fragment, psbD promoter and 5' UTR, and the SSL-1-SSL-3/SSL-3-SSL-3 fragment were then assembled into the same vector backbone as that of other constructs.

#### 4.2. C. reinhardtii strains, transformation, and growth conditions

*C. reinhardtii* wild-type (CC-125) and transgenic strains were grown in Tris-acetate-phosphate (TAP) or Tris-minimum (M) media [62] at 23 °C under constant illumination of 70–90  $\mu$ E/m<sup>2</sup>s unless otherwise noted. For chloroplast transformation, the *C. reinhardtii* DEVL strain [24] was grown to  $1-2 \times 10^6$  cells/ml in the presence of 0.5 mM 5fluorodeoxyuridine in TAP medium in the dark. Cells were harvested by centrifugation and resuspended to  $5 \times 10^7$  cells/ml in M medium. For plating, 400 µl of cell suspension was spread onto M agar as a ~6-cm circle and transformed by particle bombardment [63]. Gold particles (S550d, Seashell Technology, San Diego, CA or 1652262, Bio-Rad, Hercules, CA) were coated with plasmid DNA according to the protocol provided by the manufacturer. Bombardment was performed using a Bio-Rad PDS-1000/He system at 1350 psi with chamber vacuum of 27–28 inches Hg.

Transformants were screened for the presence of transgenes using gene-specific forward and reverse primers: *FPS*, 5'-ATTGTCGTCAACG TCCATACT-3' and 5'-TTGACGTTTGTAAATTTTCTGAGC-3'; *SSL-1*, 5'-CATGGGTAATGGTATGGCTGA-3' and 5'-TGCTAAACCGTCACGAC AAG-3'; *SSL-3*, 5'-TGCGTTTAGGTGCTGAAATG-3' and 5'-TGCTAAACC TGCTTTGCATAAA-3'; *TMT-3*, 5'-ACGAATCACGTATGGCATCA-3' and

5'-TTTCCATGATTCAGGAGGAAG-3'. The presence of untransformed chloroplast DNA (DEVL background) was detected by *aadA* primers 5'-TGATTTGCTGGTTACGGTGA-3' and 5'TACTGCGCTGTACCAAATGC-3'. Some transformants also carried wild-type DNA copies, possibly a result of gene rearrangement or a heteroplasmic parental strain, and the presence of wild-type *rbcL* was amplified by primers 5'-CTGCTGCATG TGAAGTTTGG-3' and 5'-CTTCGACTTCAGGTCGTTTCG-3'. DNA was extracted from transformant colonies using a Chelex 100 protocol [64]. For a 20-µl PCR reaction,  $1-2 \mu$ l of extracted DNA was amplified using the above primers and GoTaq DNA Polymerase (Promega, Madison, WI).

#### 4.3. RNA isolation and analysis

To isolate total RNA, at least  $1 \times 10^7$  cells were collected by centrifugation. RNA was extracted from cell pellets using TRI Reagent (Sigma-Aldrich, St. Louis, MO) per the manufacturer's instructions. Five µg of total RNA was electrophoresed in a 1.2% agarose-formaldehyde gel, transferred to a nylon membrane, and hybridized with <sup>32</sup>P- or biotin-labeled double-stranded DNA probes. Genotyping primers were used to generate templates for random-primed DNA labeling with the Klenow fragment (New England Biolabs) and <sup>32</sup>P-dCTP (PerkinElmer, Waltham, MA). Biotin probes were produced by PCR using genotyping primers and biotin-16-dUTP (Biotium, Fremont, CA). Radioactive signals were detected by a phosphor screen and Storm<sup>®</sup> imaging system (GE Healthcare, Pittsburgh, PA); biotin was detected by IRDye-labeled streptavidin and an Odyssey Infrared Imaging System (LI-COR, Lincoln, NE) following the RNA gel blot protocol provided by the manufacturer.

#### 4.4. Immunoblotting

For protein collection, the algal cell concentration was estimated by measuring optical density at 550 nm [cell concentration =  $OD_{550} \times 2.4 \times 10^7$  cells/ml [65]]. A BioMate<sup>TM</sup> 3S spectrophotometer (Thermo Fisher Scientific, Waltham, MA) was used for OD measurements, and the calculated cell concentration was similar to the value obtained by cell counting with a hemocytometer. After centrifugation, cell pellets were resuspended in protein sample buffer and heated to 100 °C for 7 min, and  $1 \times 10^6$  cells were analyzed by SDS-PAGE and immunoblotting. Various epitope antibodies were used to detect tagged transgenic proteins: anti-c-Myc 9E10 antibody (626801; BioLegend, San Diego, CA), THE<sup>™</sup> His-tag antibody (A00186; GenScript, Piscataway, NJ), anti-FLAG M2 antibody (F1804; Sigma), anti-FLAG polyclonal antibody (F7425; Sigma), and THE™ HA-tag antibody (A01244; Genscript). For X-ray film detection, horseradish peroxidase-conjugated anti-mouse or anti-rabbit IgG antibody (Promega) and Pierce ECL Plus Substrate (Thermo Fisher Scientific) were used.

#### 4.5. In vitro enzyme assay

Protein was extracted from each sample by resuspending the cell pellet in 600 µl of protein extraction buffer (50 mM MOPS pH 7.3, 20 mM MgCl<sub>2</sub>, 5 mM beta-mercaptoethanol, 5 mM EGTA, 20% glycerol). The resuspended cells were added to a 1.5 ml tube containing 200 µl of stainless-steel beads, followed by bead beating for 10 min. The cell homogenate was transferred to a fresh 1.5 ml tube and centrifuged at 9000 x g for 10 min at 4 °C. The supernatant was collected into a fresh 1.5 ml tube and the protein concentration was determined by Bradford assay. To determine enzyme activity, protein was added to 25 µl of 2x reaction buffer (100 mM MOPS pH 7.3, 5 mM beta-mercaptoethanol, 40 mM MgCl<sub>2</sub>), 5  $\mu$ l of <sup>3</sup>H-FPP mix (0.25  $\mu$ l <sup>3</sup>H-FPP, 0.22 µl FPP, 4.53 µl dH2O), and 5 µl of 20 mM NADPH, and the final volume was brought to 50 µl with dH<sub>2</sub>O. The reactions were incubated at 37 °C for 1 h and then terminated by extraction with 60 µl of nhexane containing authentic squalene and  $C_{30}$  botryococcene standards. Thin-layer chromatography (Silica gel 60 plates, n-hexane

mobile phase) was used to analyze 30  $\mu$ l of the reaction extracts. The standards in each sample were visualized by iodine vapor, samples scraped from the TLC plates, and analyzed with scintillation counting to measure <sup>3</sup>H incorporation.

#### 4.6. GC–MS analysis

For analysis of botryococcene accumulation in *C. reinhardtii* or yeast strains, a cell pellet was freeze-dried and extracted with *n*-hexane. The samples were centrifuged at 10,000 × *g* to remove debris and the *n*-hexane supernatant was evaporated. The samples were resuspended in 100 µl of *n*-hexane, and 5 µl was analyzed by GC–MS in a splitless mode. A Bruker 436-GC-SCION SQ Premium was used with a 5% Phenyl BR-5 MS capillary column (30 m × 0.25 mm, film thickness: 0.25 µm) in electron ionization (70 eV) mode. Initial oven temperature was 220 °C, held for 1 min, then increased to 280 °C at a rate of 5 °C min<sup>-1</sup>, then ramped to 300 °C at a rate of 2 °C min<sup>-1</sup> and held for 20 min. Helium was used as a carrier gas at a flow rate of 2.58 ml min<sup>-1</sup>. Temperatures of injection port, interface, and ion source were 280 °C, 250 °C and 200 °C, respectively.

#### Author contributions

S-C.H. Performed experiments except those presented in Figs. 3D, 4 E, S5 and S6. Those experiments were performed by D.B., M.T. and T.P.D. S-C.H. drafted the manuscript, which was edited by all authors and finalized by D.B.S. All authors performed data analysis and helped to prepare the respective figures.

#### Statement of informed consent, human/animal rights

No conflicts, informed consent, or human or animal rights are applicable to this study.

#### **Declaration of Competing Interest**

Authors declare no conflicts of interest for this work.

#### Acknowledgements

We thank Gwendolyn Beacham, Meiya Li, Katrina Arajs, Nathan Wojtyna, Kerrin Mendler, Colleen McLoughlin, and Gareth Holmes for assistance with molecular biology experiments, and Hsuan-Chung Ho for critical reading of the article. This work was supported by National Science Foundation award EFRI-PSBR 1240478 to T.P.D and D.B.S.

#### References

- [1] A. Ghosh, S. Khanra, M. Mondal, G. Halder, O.N. Tiwari, S. Saini, T.K. Bhowmick, K. Gayen, Progress toward isolation of strains and genetically engineered strains of microalgae for production of biofuel and other value added chemicals: a review, Energy Convers. Manage. 113 (2016) 104–118.
- [2] M. Hlavova, Z. Turoczy, K. Bisova, Improving microalgae for biotechnology—from genetics to synthetic biology, Biotechnol. Adv. 33 (2015) 1194–1203.
- [3] E.H. Harris, D.B. Stern, G. Witman (Eds.), The *Chlamydomonas* Sourcebook, 2nd ed., Academic Press-Elsevier, Oxford, UK, 2009.
- [4] R.E. Jinkerson, M.C. Jonikas, Molecular techniques to interrogate and edit the *Chlamydomonas* nuclear genome, Plant J. 82 (2015) 393–412.
- [5] F.J. Fields, J.T. Ostrand, S.P. Mayfield, Fed-batch mixotrophic cultivation of Chlamydomonas reinhardtii for high-density cultures, Algal Res. 33 (2018) 109–117.
- [6] H.S. Kim, S.-C. Hsu, S.-I. Han, H.R. Thapa, A.R. Guzman, D.R. Browne, M. Tatli, T.P. Devarenne, D.B. Stern, A. Han, High-throughput droplet microfluidics screening platform for selecting fast-growing and high lipid-producing microalgae from a mutant library, Plant Direct 1 (2017) e00011.
- [7] S.P. Damodaran, S. Eberhard, L. Boitard, J.G. Rodriguez, Y. Wang, N. Bremond, J. Baudry, J. Bibette, F.A. Wollman, A millifluidic study of cell-to-cell heterogeneity in growth-rate and cell-division capability in populations of isogenic cells of *Chlamydomonas reinhardtii*, PLoS One 10 (2015) e0118987.
- [8] M. Terashima, E.S. Freeman, R.E. Jinkerson, M.C. Jonikas, A fluorescence-activated cell sorting-based strategy for rapid isolation of high-lipid *Chlamydomonas* mutants,

Plant J. 81 (2015) 147-159.

- [9] A.C. Brown, B.A. Knights, E. Conway, Hydrocarbon content and its relationship to physiological state in the green alga *Botryococcus braunii*, Phytochemistry 8 (1969) 543–547.
- [10] P. Metzger, C. Largeau, *Botryococcus braunii*: a rich source for hydrocarbons and related ether lipids, Appl. Microbiol. Biotechnol. 66 (2005) 486–496.
- [11] H. Berrios, M. Zapata, M. Rivas, A method for genetic transformation of Botryococcus braunii using a cellulase pretreatment, J. Appl. Phycol. 28 (2016) 201–208.
- [12] T.D. Niehaus, S. Okada, T.P. Devarenne, D.S. Watt, V. Sviripa, J. Chappell, Identification of unique mechanisms for triterpene biosynthesis in *Botryococcus braunii*, Proc. Natl. Acad. Sci. U. S. A. 108 (2011) 12260–12265.
- [13] T.D. Niehaus, S. Kinison, S. Okada, Y.S. Yeo, S.A. Bell, P. Cui, T.P. Devarenne, J. Chappell, Functional identification of triterpene methyltransferases from *Botryococcus braunii* race B, J. Biol. Chem. 287 (2012) 8163–8173.
- [14] Z. Jiang, C. Kempinski, C.J. Bush, S.E. Nybo, J. Chappell, Engineering triterpene and methylated triterpene production in plants provides biochemical and physiological insights into terpene metabolism, Plant Physiol. 170 (2016) 702–716.
- [15] Z. Jiang, C. Kempinski, S. Kumar, S. Kinison, K. Linscott, E. Nybo, S. Janze, C. Wood, J. Chappell, Agronomic and chemical performance of field-grown tobacco engineered for triterpene and methylated triterpene metabolism, Plant Biotechnol. J. 16 (2018) 1110–1124.
- [16] C. Kempinski, Z. Jiang, G. Zinck, S.J. Sato, Z. Ge, T.E. Clemente, J. Chappell, Engineering linear, branched-chain triterpene metabolism in monocots, Plant Biotechnol. J. 17 (2019) 373–385.
- [17] C. Kempinski, J. Chappell, Engineering triterpene metabolism in the oilseed of Arabidopsis thaliana, Plant Biotechnol. J. 17 (2019) 386–396.
- [18] A. Khalid, H. Takagi, S. Panthee, M. Muroi, J. Chappell, H. Osada, S. Takahashi, Development of a terpenoid-production platform in *Streptomyces reveromyceticus* SN-593, ACS Synth. Biol. 6 (2017) 2339–2349.
- [19] N.E. Khan, S.E. Nybo, J. Chappell, W.R. Curtis, Triterpene hydrocarbon production engineered into a metabolically versatile host–*Rhodobacter capsulatus*, Biotechnol. Bioeng. 112 (2015) 1523–1532.
- [20] S. Purton, Tools and techniques for chloroplast transformation of *Chlamydomonas*, Adv. Exp. Med. Biol. 616 (2007) 34–45.
- [21] S.D. Gallaher, S.T. Fitz-Gibbon, D. Strenkert, S.O. Purvine, M. Pellegrini, S.S. Merchant, High-throughput sequencing of the chloroplast and mitochondrion of *Chlamydomonas reinhardtii* to generate improved de novo assemblies, analyze expression patterns and transcript speciation, and evaluate diversity among laboratory strains and wild isolates, Plant J. 93 (2018) 545–565.
- [22] M. Rohmer, The discovery of a mevalonate-independent pathway for isoprenoid biosynthesis in bacteria, algae and higher plants, Nat. Prod. Rep. 16 (1999) 565–574.
- [23] K. Kato, T. Marui, S. Kasai, A. Shinmyo, Artificial control of transgene expression in *Chlamydomonas reinhardtii* chloroplast using the *lac* regulation system from *Escherichia coli*, J. Biosci. Bioeng. 104 (2007) 207–213.
- [24] K. Ishikura, Y. Takaoka, K. Kato, M. Sekine, K. Yoshida, A. Shinmyo, Expression of a foreign gene in *Chlamydomonas reinhardtii* chloroplast, J. Biosci. Bioeng. 87 (1999) 307–314.
- [25] L.B. Scharff, R. Bock, Synthetic biology in plastids, Plant J. 78 (2014) 783–798.
- [26] F. Zhou, D. Karcher, R. Bock, Identification of a plastid intercistronic expression element (IEE) facilitating the expression of stable translatable monocistronic mRNAs from operons, Plant J. 52 (2007) 961–972.
- [27] M. Cavaiuolo, R. Kuras, F.A. Wollman, Y. Choquet, O. Vallon, Small RNA profiling in *Chlamydomonas*: insights into chloroplast RNA metabolism, Nucleic Acids Res. 45 (2017) 10783–10799.
- [28] A. Germain, S. Herlich, S. Larom, S.H. Kim, G. Schuster, D.B. Stern, Mutational analysis of *Arabidopsis* chloroplast polynucleotide phosphorylase reveals roles for both RNase PH core domains in polyadenylation, RNA 3'-end maturation and intron degradation, Plant J. 67 (2011) 381–394.
- [29] C.P. Boehm, R. Bock, Recent advances and current challenges in synthetic biology of the plastid genetic system and metabolism, Plant Physiol. 179 (2019) 794–802.
- [30] L.B. Scharff, M. Ehrnthaler, M. Janowski, L.H. Childs, C. Hasse, J. Gremmels, S. Ruf, R. Zoschke, R. Bock, Shine-dalgarno sequences play an essential role in the translation of plastid mRNAs in tobacco, Plant Cell 29 (2017) 3085–3101.
- [31] Y. Komine, E. Kikis, G. Schuster, D. Stern, Evidence for *in vivo* modulation of chloroplast RNA stability by 3'-UTR homopolymeric tails in *Chlamydomonas reinhardtii*, Proc. Natl. Acad. Sci. U. S. A. 99 (2002) 4085–4090.
- [32] N. Dissmeyer, S. Rivas, E. Graciet, Life and death of proteins after protease cleavage: protein degradation by the N-end rule pathway, New Phytol. 218 (2018) 929–935.
- [33] W. Apel, W.X. Schulze, R. Bock, Identification of protein stability determinants in chloroplasts, Plant J. 63 (2010) 636–650.
  [34] E. Rowland, J. Kim, N.H. Bhuiyan, K.J. van Wijk, The *Arabidopsis* chloroplast
- stromal N-terminome: complexities of amino-terminal protein maturation and stability, Plant Physiol. 169 (2015) 1881–1896.
- [35] M. Kajikawa, S. Kinohira, A. Ando, M. Shimoyama, M. Kato, H. Fukuzawa, Accumulation of squalene in a microalga *Chlamydomonas reinhardtii* by genetic modification of squalene synthase and squalene epoxidase genes, PLoS One 10 (2015) e0120446.
- [36] P. Puigbo, I.G. Bravo, S. Garcia-Vallve, CAIcal: a combined set of tools to assess codon usage adaptation, Biol. Direct 3 (2008) 38.
- [37] F. De Marchis, A. Pompa, M. Bellucci, Plastid proteostasis and heterologous protein accumulation in transplastomic plants, Plant Physiol. 160 (2012) 571–581.
- [38] R. Surzycki, K. Greenham, K. Kitayama, F. Dibal, R. Wagner, J.D. Rochaix, T. Ajam, S. Surzycki, Factors effecting expression of vaccines in microalgae, Biologicals 37 (2009) 133–138.

- [39] J. Nickelsen, Transcripts containing the 5' untranslated regions of the plastid genes psbA and psbB from higher plants are unstable in *Chlamydomonas reinhardtii* chloroplasts, Mol. Gen. Genet. 262 (1999) 768–771.
- [40] D.B. Stern, M. Goldschmidt-Clermont, M.R. Hanson, Chloroplast RNA metabolism, Annu. Rev. Plant Biol. 61 (2010) 125–155.
- [41] L. Kleinknecht, F. Wang, R. Stube, K. Philippar, J. Nickelsen, A.V. Bohne, RAP, the sole octotricopeptide repeat protein in *Arabidopsis*, is required for chloroplast 16S rRNA maturation, Plant Cell 26 (2014) 777–787.
- [42] A. Boulouis, D. Drapier, H. Razafimanantsoa, K. Wostrikoff, N.J. Tourasse, K. Pascal, J. Girard-Bascou, O. Vallon, F.A. Wollman, Y. Choquet, Spontaneous dominant mutations in *Chlamydomonas* highlight ongoing evolution by gene diversification, Plant Cell 27 (2015) 984–1001.
- [43] R. Rott, V. Liveanu, R.G. Drager, D.B. Stern, G. Schuster, The sequence and structure of the 3'-untranslated regions of chloroplast transcripts are important determinants of mRNA accumulation and stability, Plant Mol. Biol. 36 (1998) 307–314.
- [44] K.S. Macedo-Osorio, V.H. Perez-Espana, C. Garibay-Orijel, D. Guzman-Zapata, N.V. Duran-Figueroa, J.A. Badillo-Corona, Intercistronic expression elements (IEE) from the chloroplast of *Chlamydomonas reinhardtii* can be used for the expression of foreign genes in synthetic operons, Plant Mol. Biol. 98 (2018) 303–317.
- [45] D.B. Stern, K.L. Kindle, 3' End maturation of the *Chlamydomonas reinhardtii* chloroplast *atpB* mRNA is a two-step process, Mol. Cell. Biol. 13 (1993) 2277–2285.
- [46] A. Hicks, R.G. Drager, D.C. Higgs, D.B. Stern, An mRNA 3' processing site targets downstream sequences for rapid degradation in *Chlamydomonas* chloroplasts, J. Biol. Chem. 277 (2002) 3325–3333.
- [47] M. Goldschmidt-Clermont, M. Rahire, J.D. Rochaix, Redundant cis-acting determinants of 3' processing and RNA stability in the chloroplast rbcL mRNA of Chlamydomonas, Plant J. 53 (2007) 566–577.
- [48] J. Nickelsen, M. Fleischmann, E. Boudreau, M. Rahire, J.D. Rochaix, Identification of cis-acting RNA leader elements required for chloroplast *psbD* gene expression in *Chlamydomonas*, Plant Cell 11 (1999) 957–970.
- [49] J. Nickelsen, J. van Dillewijn, M. Rahire, J.D. Rochaix, Determinants for stability of the chloroplast *psbD* RNA are located within its short leader region in *Chlamydomonas reinhardtii*, EMBO J. 13 (1994) 3182–3191.
- [50] M. Nakamura, Y. Hibi, T. Okamoto, M. Sugiura, Cooperation between the chloroplast psbA 5'-untranslated region and coding region is important for translational initiation: the chloroplast translation machinery cannot read a human viral gene coding region, Plant J. 85 (2016) 772–780.
- [51] S. Kang, O.W. Odom, C.L. Malone, S. Thangamani, D.L. Herrin, Expression of a synthetic gene for the major cytotoxin (Cyt1Aa) of *Bacillus thuringiensis* subsp. *israelensis* in the chloroplast of wild-type *Chlamydomonas*, Biology (Basel) 7 (2018) 29.

- [52] D.A. Dougan, D. Micevski, K.N. Truscott, The N-end rule pathway: from recognition by N-recognins, to destruction by AAA + proteases, Biochim. Biophys. Acta 1823 (2012) 83–91.
- [53] C. Giglione, A. Boularot, T. Meinnel, Protein N-terminal methionine excision, Cell. Mol. Life Sci. 61 (2004) 1455–1474.
- [54] F. Frottin, A. Martinez, P. Peynot, S. Mitra, R.C. Holz, C. Giglione, T. Meinnel, The proteomics of N-terminal methionine cleavage, Mol. Cell Proteomics 5 (2006) 2336–2349.
- [55] S. D'Adamo, G. Schiano di Visconte, G. Lowe, J. Szaub-Newton, T. Beacham, A. Landels, M.J. Allen, A. Spicer, M. Matthijs, Engineering the unicellular alga *Phaeodactylum tricornutum* for high-value plant triterpenoid production, Plant Biotechnol. J. 17 (2018) 75–87.
- [56] S. Ramundo, J.D. Rochaix, Controlling expression of genes in the unicellular alga *Chlamydomonas reinhardtii* with a vitamin-repressible riboswitch, Methods Enzymol. 550 (2015) 267–281.
- [57] J.D. Rochaix, R. Surzycki, S. Ramundo, Tools for regulated gene expression in the chloroplast of *Chlamydomonas*, Methods Mol. Biol. 1132 (2014) 413–424.
- [58] S. Shamriz, H. Ofoghi, Outlook in the application of *Chlamydomonas reinhardtii* chloroplast as a platform for recombinant protein production, Biotechnol. Genet. Eng. Rev. 32 (2016) 92–106.
- [59] R. Surzycki, L. Cournac, G. Peltier, J.D. Rochaix, Potential for hydrogen production with inducible chloroplast gene expression in *Chlamydomonas*, Proc. Natl. Acad. Sci. U. S. A. 104 (2007) 17548–17553.
- [60] Y. Nakamura, T. Gojobori, T. Ikemura, Codon usage tabulated from international DNA sequence databases: status for the year 2000, Nucleic Acids Res. 28 (2000) 292.
- [61] A.D. Blowers, G.S. Ellmore, U. Klein, L. Bogorad, Transcriptional analysis of endogenous and foreign genes in chloroplast transformants of *Chlamydomonas*, Plant Cell 2 (1990) 1059–1070.
- [62] D.S. Gorman, R.P. Levine, Cytochrome f and plastocyanin: their sequence in the photosynthetic electron transport chain of *Chlamydomonas reinhardtii*, Proc. Natl. Acad. Sci. U. S. A. 54 (1965) 1665–1669.
- [63] J.E. Boynton, N.W. Gillham, E.H. Harris, J.P. Hosler, A.M. Johnson, A.R. Jones, B.L. Randolph-Anderson, D. Robertson, T.M. Klein, K.B. Shark, et al., Chloroplast transformation in *Chlamydomonas* with high velocity microprojectiles, Science 240 (1988) 1534–1538.
- [64] R. Werner, D. Mergenhagen, Mating type determination of Chlamydomonas reinhardtii by PCR, Plant Mol. Biol. Rep. 16 (1998) 295–299.
- [65] J.A. Myers, B.S. Curtis, W.R. Curtis, Improving accuracy of cell and chromophore concentration measurements using optical density, BMC Biophys. 6 (2013) 4.