

# Codon reassignment to facilitate genetic engineering and biocontainment in the chloroplast of *Chlamydomonas reinhardtii*

Rosanna E. B. Young\* and Saul Purton

Algal Research Group, Institute of Structural and Molecular Biology, University College London, London, UK

Received 15 May 2015;

revised 8 September 2015;

accepted 13 September 2015.

\*Correspondence (Tel +44 (0)

20 7679 2676; email r.young@ucl.ac.uk)

## Summary

There is a growing interest in the use of microalgae as low-cost hosts for the synthesis of recombinant products such as therapeutic proteins and bioactive metabolites. In particular, the chloroplast, with its small, genetically tractable genome (plastome) and elaborate metabolism, represents an attractive platform for genetic engineering. In *Chlamydomonas reinhardtii*, none of the 69 protein-coding genes in the plastome uses the stop codon UGA, therefore this spare codon can be exploited as a useful synthetic biology tool. Here, we report the assignment of the codon to one for tryptophan and show that this can be used as an effective strategy for addressing a key problem in chloroplast engineering: namely, the assembly of expression cassettes in *Escherichia coli* when the gene product is toxic to the bacterium. This problem arises because the prokaryotic nature of chloroplast promoters and ribosome-binding sites used in such cassettes often results in transgene expression in *E. coli*, and is a potential issue when cloning genes for metabolic enzymes, antibacterial proteins and integral membrane proteins. We show that replacement of tryptophan codons with the spare codon (UGG→UGA) within a transgene prevents functional expression in *E. coli* and in the chloroplast, and that co-introduction of a plastidial *trnW* gene carrying a modified anticodon restores function only in the latter by allowing UGA readthrough. We demonstrate the utility of this system by expressing two genes known to be highly toxic to *E. coli* and discuss its value in providing an enhanced level of biocontainment for transplastomic microalgae.

**Keywords:** *Chlamydomonas reinhardtii*, chloroplast, microalgae, non-sense suppression, transfer RNA.

## Introduction

The microalgal chloroplast has many advantages as a production platform for recombinant proteins and small molecules including low culturing costs, lack of toxins and ease of genetic manipulation. The presence of multiple copies of the chloroplast genome per cell and lack of gene silencing give the chloroplast an advantage over nuclear-encoded transgene expression (Bock, 2015). The green alga *Chlamydomonas reinhardtii* is the most widely used for recombinant protein expression, with products such as vaccines, immunotoxins, therapeutics and industrial enzymes (reviewed by Rasala and Mayfield (2015) and Scaife *et al.* (2015)). Chloroplasts evolved from a cyanobacterial endosymbiont (Timmis *et al.*, 2004) and many chloroplast genes in *C. reinhardtii* have retained bacterial features such as –35 and/or –10 promoter elements and 70S ribosome-binding sequences. This is the case for the promoter and 5' untranslated region (5' UTR) of exon 1 of *psaA*, encoding a core subunit of photosystem I. The *psaA* promoter/5' UTR is often used to drive robust expression of foreign genes in the *C. reinhardtii* chloroplast (Michelet *et al.*, 2011; Specht and Mayfield, 2013; Young and Purton, 2014), but it cannot be used for proteins that are detrimental to *Escherichia coli* as they will be expressed during cloning in this host and will prevent successful production of the plasmid vector for subsequent transfer to the microalga. For example the PanDaTox database lists over 40 000 microbial genes that are

predicted to be toxic to *E. coli* based on their failure to be propagated during genome sequencing projects (Amitai and Sorek, 2012). This lack of clonability can constrain the modification or introduction of biochemical pathways in *C. reinhardtii* for metabolic engineering due to alterations in carbon or nitrogen flux in *E. coli* or the generation of toxic intermediates, and may also prevent the cloning of genes for some antibacterial enzymes or integral membrane proteins.

Transfer RNAs (tRNAs) and their cognate aminoacyl-tRNA synthetases together determine the amino acid sequence that is encoded by messenger RNA, so manipulation of these components can alter the genetic code. In the standard genetic code, 61 of the 64 RNA triplet codons are translated as amino acids whereas the remaining three (UAA, UAG and UGA) are stop signals at which release factors aid the termination of translation. The *C. reinhardtii* chloroplast genome uses this standard genetic code; however, DNA sequencing revealed that there is a strong preference for UAA as the stop codon with 65 of the 69 protein-coding genes using this codon (Maul *et al.*, 2002). The remaining four genes use UAG, and UGA is not used at all, although early genetic evidence demonstrates that it can function as a stop codon in the *C. reinhardtii* chloroplast. For example non-photosynthetic mutants were isolated in which the chloroplast *rbcL* gene, encoding the large subunit of Rubisco, contained a TGG to TAG (amber) or TGA (opal) non-sense mutation (Spreitzer *et al.*, 1985). Chemical mutagenesis of the amber mutant, followed by selection for photosynthetic com-

petence, produced a cell line in which the wild-type *trnW<sub>CCA</sub>* gene and a mutated version with an amber-specific CUA anticodon coexisted as a heteroplasmic mix in the polyplid plastome, thus allowing both UGG and UAG codons to be translated as tryptophan (Yu and Spreitzer, 1992). A similar experiment using the opal mutant also produced heteroplasmic non-sense suppressors but the genetic basis of the suppression was not characterized (Spreitzer *et al.*, 1984). These results suggested that it would be possible to genetically engineer the *C. reinhardtii* chloroplast *trnW* to recognize amber, and possibly opal, codons instead of UGG.

Here, we address the challenge of cloning genes whose products are toxic to *E. coli* by exploiting the unused UGA codon to create a genetic system in which the gene of interest (GOI) is modified to carry opal mutations at one or more tryptophan codons (i.e. UGG to UGA), thereby preventing synthesis of the full-length protein in either *E. coli* or the chloroplast. Translational read-through is restored in the chloroplast, but not in *E. coli*, by combining the GOI with a plastidial *trnW* gene encoding a tRNA with a modified anticodon.

The existence of an unused codon in the *C. reinhardtii* chloroplast genetic code, together with our demonstration that it can be integrated into coding sequences and translated without the need to eliminate any plastidial release factors, provides opportunities for future genetic engineering of the microalgal chloroplast involving canonical or non-canonical amino acids. The use of a non-sense codon to interrupt the coding sequence also reduces the risk of transgenes being translated into full-length proteins were they to spread to other organisms by horizontal gene transfer, thereby providing informational containment of the transgenes.

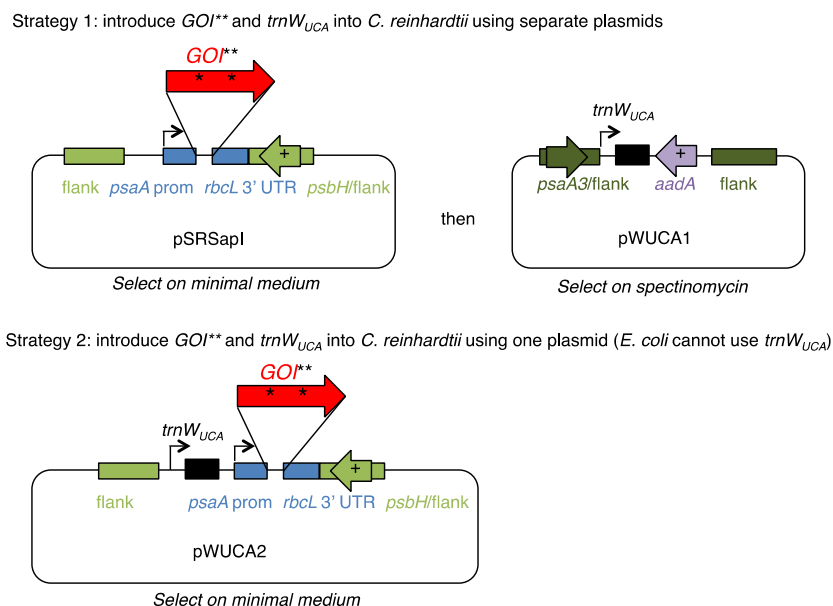
## Results

Our scheme for cloning genes that are toxic to *E. coli* into a *C. reinhardtii* chloroplast expression vector requires firstly a version of the gene with one or more TGG codons modified to TGA 'stop' codons, and secondly a synthetic tRNA gene to read the TGA codon/s as tryptophan. Strategies for introducing these two elements into *C. reinhardtii* are outlined in Figure 1. Plasmids used in this work are detailed in Table 1.

### Mutation of two TGG codons to TGA codons within a transgene (*crCD*) prevents accumulation of CrCD protein in *Chlamydomonas reinhardtii*

The first set of experiments was carried out using *crCD* as a test gene. This is an *E. coli* cytosine deaminase gene optimized for the *C. reinhardtii* chloroplast as a negative selectable marker (Young and Purton, 2014) and was chosen as a test gene due to the stability of CrCD protein in the chloroplast, ease of detection by Western blotting (via an added HA epitope) and clear phenotype of sensitivity to 5-fluorocytosine. CrCD is not toxic when expressed in *E. coli*, allowing appropriate control strains to be used.

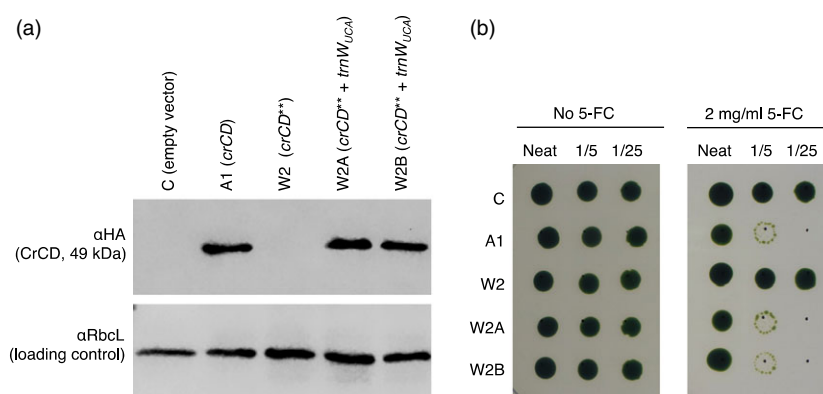
A chloroplast expression vector containing *crCD* under the control of the *C. reinhardtii* *psaA* exon 1 promoter (plasmid pCD, previously called pRY127d; (Young and Purton, 2014)) was modified so that two of the TGG codons, encoding tryptophan, were altered to TGA (Table S1). The resulting plasmid, pCD\*\*, was used to transform *C. reinhardtii* TN72 (a non-photosynthetic *psbH* mutant). The flanking region of pCD\*\* contains an intact copy of *psbH*, so homologous recombination into the chloroplast genome restores phototrophic growth and allows the selection of transformants on minimal medium in the light. Transgene



**Figure 1** Strategy to clone genes into a chloroplast expression vector whilst preventing their expression in *Escherichia coli*. The gene of interest (*GOI*) is redesigned such that one or more tryptophan (TGG) codons are altered to TGA, indicated by asterisks. These changes can be integrated into the codon-optimized gene design prior to ordering the synthetic gene and integrating it into the *Chlamydomonas reinhardtii* chloroplast genome. A tRNA gene based on the *C. reinhardtii* plastidial *trnW*, but with the anticodon sequence altered to recognize UGA, is also introduced (*trnW<sub>UCA</sub>*). This enables readthrough of the *GOI* in *C. reinhardtii*. Flanking regions amplified from chloroplast DNA allow targeted integration of the constructs into the chloroplast genome by homologous recombination; the target site is a neutral region downstream of either *psbH* or *psaA* exon 3, depending on the construct. The *psbH* gene can be used for selection in a *psbH* mutant recipient strain.

**Table 1** Plasmids used in this work. All selection plates were incubated in the light except where pPsaA\* was used. All synthetic protein-coding genes were codon-optimized for the *Chlamydomonas reinhardtii* chloroplast and encode a C-terminal HA tag

Plasmid	Synthetic gene	Protein encoded by gene	Expected protein size incl. HA tag (kDa)	Integration site in <i>C. reinhardtii</i> chloroplast	Selection
pCD	<i>crCD</i>	Cytosine deaminase (Young and Purton, 2014)	49	Downstream of <i>psbH</i>	Minimal medium ( <i>psbH</i> restored)
pCD**	<i>crCD</i> with 2 internal TGA stop codons	Cytosine deaminase	49 (if pWUCA1 also present)	Downstream of <i>psbH</i>	Minimal medium ( <i>psbH</i> restored)
pWUCA1	<i>trnW<sub>UCA</sub></i>	n/a	n/a	Downstream of <i>psaA</i> exon 3	Spectinomycin ( <i>aadA</i> cassette)
pWUCA2	<i>trnW<sub>UCA</sub></i>	n/a	n/a	Downstream of <i>psbH</i>	Minimal medium ( <i>psbH</i> restored or <i>psaA</i> * translated, depending on recipient cell line)
pWUCA2-CD**	<i>crCD</i> with 2 internal TGA stop codons + <i>trnW<sub>UCA</sub></i>	Cytosine deaminase	49	Plasmid not used in <i>C. reinhardtii</i>	Plasmid not used in <i>C. reinhardtii</i>
pPsaA*	None; introduces TGG to TGA mutation at W693 in native <i>psaA</i> gene	Core protein of photosystem I	n/a	<i>psaA</i> exon 3 region	Spectinomycin ( <i>aadA</i> cassette); incubate in dark
pSty	<i>Salmonella</i> Typhimurium phage gene SPN9CC_0043 with 1 internal TGA stop codon + <i>trnW<sub>UCA</sub></i>	Endolysin (lyses <i>Escherichia coli</i> ; Lim <i>et al.</i> , 2014)	19	downstream of <i>psbH</i>	Minimal medium ( <i>psbH</i> restored)
pSde	<i>Shewanella denitrificans</i> Sden_1266 with 2 internal TGA stop codons + <i>trnW<sub>UCA</sub></i>	Hypothetical protein (very toxic to <i>E. coli</i> ; Kimelman <i>et al.</i> , 2012)	42	Downstream of <i>psbH</i>	Minimal medium ( <i>psbH</i> restored)

**Figure 2** Introduction of a synthetic *trnW<sub>UCA</sub>* gene into the *Chlamydomonas reinhardtii* chloroplast genome allows the expression of full-length, active CrCD protein from the *crCD*\*\* gene. (a) Western analysis. Equalized cell lysates were subjected to SDS-PAGE and two identical blots were probed with antibodies as indicated on the left. (b) Growth tests demonstrating that cell lines A1, W2A and W2B contain active CrCD protein that leads to cell death on media containing 5-fluorocytosine (5-FC). *C. reinhardtii* cultures were adjusted to equal optical densities and spotted onto TAP agar containing no drug (left panel) or 5-FC (right panel), with serial fivefold dilutions left to right. Plates were incubated under 50  $\mu\text{E}/\text{m}^2/\text{s}$  light for 10 days.

integration and homoplasmy across the approximately 80 copies of the chloroplast genome per cell was confirmed by PCR (Figure S1a and Table S2). As expected, the introduction of TGA codons into the *crCD* gene prevented the accumulation of full length CrCD protein as detected by immunoblotting with an

antibody against the C-terminal HA tag (cell line W2 in Figure 2a). This suggests that no tRNA in the *C. reinhardtii* chloroplast is able to recognize the UGA codon and insert an amino acid at the corresponding position in the peptide chain, so the chain terminates prematurely.

### Introduction of a synthetic *trnW<sub>UCA</sub>* gene into the chloroplast genome allows the accumulation of full length, active CrCD protein and does not affect growth

There is a single codon for tryptophan in the standard genetic code (UGG) and a single copy of the tryptophan tRNA gene, *trnW<sub>CCA</sub>*, in the *C. reinhardtii* chloroplast genome (Cognat *et al.*, 2013; Maul *et al.*, 2002); see the PlantRNA database at <http://plantrna.ibmp.cnrs.fr>. A tRNA with a 5'-UCA-3' anticodon would be expected to recognize UGA codons within the mRNA, such as those transcribed from the TGA codons that had been inserted into *crCD*. To test this, a version of the *C. reinhardtii* chloroplast *trnW* gene with 100 bp of its flanking sequence each side, which included its promoter, was synthesized with a mutated anticodon (CCA to UCA) and cloned into a chloroplast targeting vector to make plasmid pWUCA1 (Figure 1). The vector inserts the *trnW<sub>UCA</sub>* gene downstream of *psaA* exon 3 in the chloroplast genome and contains an *aadA* spectinomycin resistance cassette. Homoplasmic transformants were recovered following selection on medium containing spectinomycin (Figure S1b and Table S2). The transformation of *C. reinhardtii* W2 with plasmid pWUCA1 was found to elicit the accumulation of full-length CrCD protein (Figure 2a, cell lines W2A and W2B), demonstrating that the synthetic *trnW<sub>UCA</sub>* gene is expressed and indeed recognizes the UGA codon. This also confirms that the 100 bp flanking sequences included around *trnW<sub>UCA</sub>* were sufficient for its transcription and any subsequent 5' and 3' processing by RNaseP and other RNases. There was no observable difference in CrCD protein yield between W2A and a control *C. reinhardtii* cell line that had been transformed with pCD (Figure S2).

Cytosine deaminase (CrCD) normally converts cytosine to uracil but can also convert the synthetic compound 5-fluorocytosine to a toxic product, 5-fluorouracil. The activity of the CrCD protein made using the synthetic tRNA was demonstrated by the reduced growth of cell lines W2A and W2B on medium containing 5-fluorocytosine (Figure 2b).

The expression of a synthetic tRNA that reassigns a stop codon sometimes slows the growth of the organism, presumably due to the lack of proper termination of endogenous proteins (Wang *et al.*, 2014). As UGA is not used as a stop or sense codon in the *C. reinhardtii* chloroplast, we did not expect to see a growth

defect in this case. Indeed, we found that the synthetic *trnW<sub>UCA</sub>* gene expressed in the chloroplast had no detrimental effect on the growth rate of *C. reinhardtii* when tested under mixotrophic or phototrophic conditions at 25 °C (Figure S3).

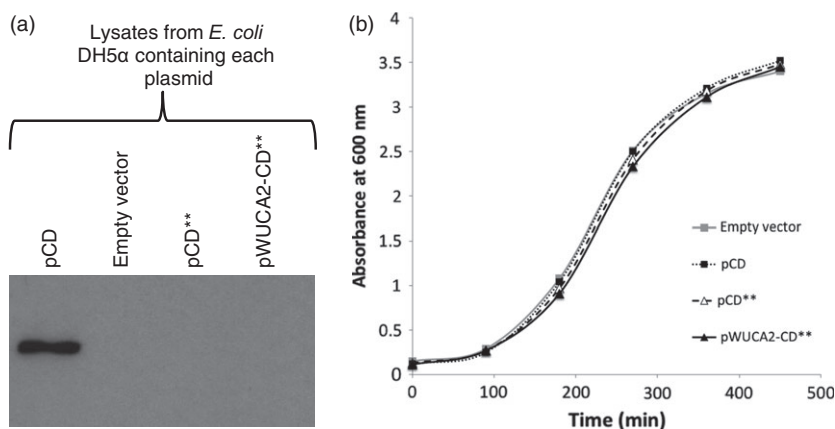
### The tRNA and gene of interest can be introduced into the chloroplast in a single homologous recombination step

The production of *C. reinhardtii* cell lines that express transgenes using the synthetic tRNA could be streamlined if both the transgene and tRNA gene were combined into one vector. This would require a single algal transformation step and would minimize the use of drug selection cassettes, especially if the restoration of *psbH* (i.e. growth on minimal medium) was used as the selection method. However, for this strategy to be useful for transgene containment and cloning toxic genes, *E. coli* must be unable to use the synthetic tRNA to translate the foreign protein. This was shown to be the case by inserting the *trnW<sub>UCA</sub>* gene and its 100 bp flanks upstream of the *crCD* expression cassette in plasmid pCD\*\*, generating pWUCA2-CD\*\*. No CrCD protein was detected in *E. coli* pWUCA2-CD\*\* lysates by immunoblotting with an anti-HA antibody (Figure 3a), indicating that *E. coli* cannot use the synthetic tRNA to read through the UGA codons within the *crCD\*\** mRNA. This is more likely to be due to a lack of tRNA function or to competition with other factors than to a lack of tRNA transcription, as the *trnW<sub>UCA</sub>* flank contains an exact bacterial consensus promoter (see Discussion and Appendix S1). The presence of a *trnW<sub>UCA</sub>* plasmid did not affect the growth rate of *E. coli* (Figure 3b).

A new *trnW<sub>UCA</sub>* chloroplast expression vector, pWUCA2, was then constructed into which a transgene can be cloned between the *psaA* exon 1 promoter/5' UTR and the *rbcl* 3' UTR using SapI and SphI restriction enzymes. The pWUCA2 plasmid carries the *trnW<sub>UCA</sub>* gene upstream of this expression cassette (see Strategy 2 in Figure 1).

### The *trnW<sub>UCA</sub>* gene rescues the mutation of an essential tryptophan codon to TGA in *psaA*

Tryptophan is the largest canonical amino acid and is the only one to carry an indole side-chain. The double ring structure of its indole moiety is often involved in stacking interactions that are important for substrate binding and catalysis in some enzymes



**Figure 3** In *Escherichia coli*, the synthetic *Chlamydomonas reinhardtii* chloroplast *trnW<sub>UCA</sub>* gene does not permit full length CrCD protein expression from the *crCD\*\** gene. (a) Western analysis using equalized lysates of *E. coli* DH5α carrying four different plasmids. The blot was probed with an anti-HA antibody to detect HA-tagged CrCD protein. (b) Growth curve of *E. coli* DH5α carrying the four different plasmids.

(Nakamura *et al.*, 2013; Zhang *et al.*, 2004a). To demonstrate that the synthetic  $\text{trnW}_{\text{UCA}}$  adds tryptophan rather than any other amino acid to the growing peptide chain, we carried out experiments on the *C. reinhardtii* chloroplast *psaA* gene, which encodes a core component of photosystem I (PSI).

The tryptophan at position 693 of PsaA is  $\pi$ -stacked through the indole moiety with the bound phylloquinone cofactor (Boudreaux *et al.*, 2001; Jordan *et al.*, 2001). *Chlamydomonas reinhardtii* strains in which W693 has been substituted for another amino acid have a functional PSI complex but are highly sensitive to oxygen during phototrophic growth, possibly from the formation of free radical species (Purton *et al.*, 2001).

We substituted the W693 TGG codon in *psaA* with TGA by homologous recombination (plasmid pPsaA\* in Table 1), using an *aadA* spectinomycin resistance cassette for the selection of *C. reinhardtii* transformants. Homoplasmic integration of *aadA* downstream of *psaA* exon 3 was confirmed by PCR (Figure 4a), and the introduction of the non-sense codon into *psaA* was confirmed by DNA sequencing (Figure S4). The resulting strain, *cw15* + pPsaA\*, shows the loss of phototrophy and the light sensitivity typical of PSI-deficient mutants (Figure 4c).

The *cw15* + pPsaA\* cell line was transformed with pWUCA2, plating on minimal medium in the light under aerobic conditions. Although the intact *psbH* gene in pWUCA2 can be used for selection in a *psbH* mutant line such as TN72, the recipient cell line used here already has an intact *psbH* so this gene was merely being used as a homologous flanking region for integration of *trnW}\_{\text{UCA}} into the chloroplast genome. Instead, selection was directly for  $\text{trnW}_{\text{UCA}}$  to translate full-length, W693-containing PsaA protein for restored photosynthesis, effectively using the tRNA gene as a positive selectable marker. Four out of seven *cw15* + pPsaA\* + pWUCA2 colonies checked by PCR were homoplasmic for  $\text{trnW}_{\text{UCA}}$  after a single round of streaking out on minimal medium, demonstrating that selection was successful. We continued with transformants 1 and 2 (Figure 4b and Table S2); DNA sequencing of part of *psaA3* confirmed that they were not W693 TGA  $\rightarrow$  TGG revertants (Figure S4). These cell lines can grow on minimal medium in the presence of oxygen*

(Figure 4c), contrasting with the oxygen-sensitive phenotype of PsaA mutants that have other amino acids at position 693 (Purton *et al.*, 2001).

The ability of  $\text{trnW}_{\text{UCA}}$  to complement the *psaA*\* mutation indicates that the single nucleotide change in the anticodon loop from the natural tryptophan tRNA,  $\text{trnW}_{\text{CCA}}$ , to the synthetic tRNA,  $\text{trnW}_{\text{UCA}}$ , does not prevent the tryptophanyl tRNA synthetase from recognizing this as a tRNA to be charged with tryptophan.

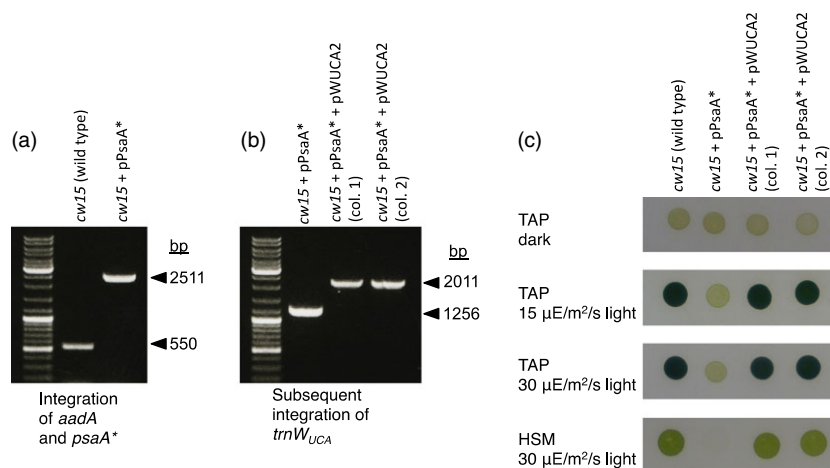
### Genes whose products are toxic to *Escherichia coli* can be cloned using the synthetic tRNA system

The  $\text{trnW}_{\text{UCA}}$  cloning scheme using the combined vector (Figure 1) was tested using two genes whose products are known to be toxic to *E. coli*. The SPN9CC endolysin, from a *Salmonella* Typhimurium bacteriophage, has previously been shown to lyse *E. coli* (Lim *et al.*, 2014). A codon-optimized version of this endolysin gene was designed in which a single TGG codon was altered to TGA. This was cloned into pWUCA2 to make plasmid pSty. The second gene was a *Shewanella denitrificans* ORF identified from the PanDaTox database as unclonable in *E. coli* during genome sequencing. Further work by the compilers of the database showed that the ORF can be cloned under an inducible promoter in *E. coli* but that cells die upon induction of expression (Kimelman *et al.*, 2012). A codon-optimized version of this ORF was designed with two TGG to TGA mutations, and cloned into pWUCA2 to give plasmid pSde. DNA sequences for the two genes are given in Appendix S1.

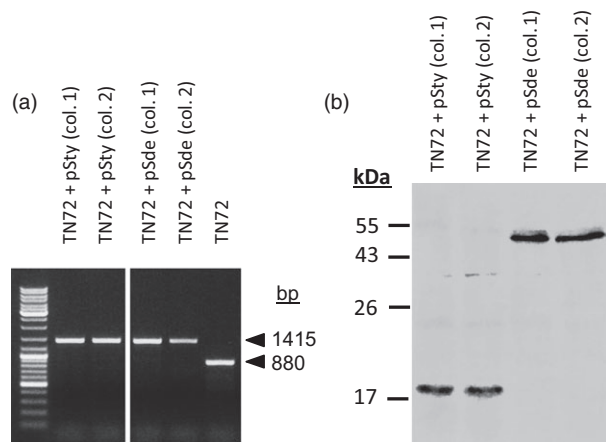
*Chlamydomonas reinhardtii* TN72 was transformed with pSty and pSde separately; homoplasmic integration of the transgenes and  $\text{trnW}_{\text{UCA}}$  was demonstrated by PCR (Figure 5a and Table S2). The accumulation of the SPN9CC endolysin and the *S. denitrificans* protein was demonstrated by immunoblotting with an anti-HA antibody (Figure 5b).

### Stop codon usage in other microalgae

The selection of microalgal species for industrial biotechnology depends on many factors including the ease of genetic manip-



**Figure 4** The mutation of a tryptophan codon that is essential for PsaA function in *Chlamydomonas reinhardtii* can be complemented by  $\text{trnW}_{\text{UCA}}$ . (a) PCR on *C. reinhardtii* cell lines showing the homoplasmic integration of *aadA* downstream of *psaA* in the chloroplast genome, with the concurrent mutation of the *psaA* W693 codon to TGA (confirmed by sequencing). (b) PCR showing the homoplasmic integration of pWUCA2 into the *C. reinhardtii* pPsaA\* cell line. Selection was directly for  $\text{trnW}_{\text{UCA}}$  to restore phototrophic growth by allowing the translation of *psaA*\*. (c) Growth properties of the cell lines on media containing acetate (TAP) or minimal medium (HSM). *Chlamydomonas reinhardtii* cultures were adjusted to equal optical densities, spotted onto agar and incubated at 25 °C in the presence of oxygen.



**Figure 5** Use of the *trnW<sub>UCA</sub>* system to clone and express genes in *C. reinhardtii* whose products are toxic to *Escherichia coli*. (a) PCR demonstrating homoplasmic integration of the transgenes into *C. reinhardtii* cell line TN72. (b) Western blot with anti-HA antibody, demonstrating accumulation of the foreign proteins.

ulation, growth rates, media costs, harvesting costs and, for some products, the lipid content. We surveyed the stop codon usage of 12 microalgal species for which chloroplast genome sequences were available to assess whether a similar approach to genetic code manipulation might be possible in species other than *C. reinhardtii* (Table S3). The chloroplast genomes ranged from 72 to 269 kb in size and were predicted to contain between 61 and 224 protein-coding genes according to the annotations in the NCBI database. For each of the 12 genomes, UAA was the most frequently used stop codon and UGA was the least frequently used. As noted by Robbins *et al.* (2007), the chloroplast genome of *Ostreococcus tauri* does not use the UGA codon. *Lobosphaera (Parietochloris) incisa* also lacks UGA stop codons, and only a single putative gene (encoding a 45 amino acid hypothetical protein) has this stop codon in *Chlorella sorokiniana* (Table S3). However, the chloroplast transformation of these three species has yet to be reported. *Dunaliella salina*, *Euglena gracilis* and *Phaeodactylum tricorutum* each use the UGA stop codon three to four times in the chloroplast genome, so may be amenable to emancipation or dual use of this codon. Species in each of these three genera have been shown to have transformable chloroplast genomes (Doetsch *et al.*, 2001; Georgianna *et al.*, 2013; Xie *et al.*, 2014).

## Discussion

The UGA codon is one of three triplet nucleotide codons used as stop signals in the standard genetic code. However, some species are known to have reassigned UGA to a sense codon. In the *C. reinhardtii* nucleus, UGA encodes stop or selenocysteine depending on the RNA context (Novoselov *et al.*, 2002; Rao *et al.*, 2003). Gracilibacteria translate this codon as glycine (Rinke *et al.*, 2013), whilst UGA reassignments to tryptophan (known as Genetic Code 4) have been observed in mycoplasmas and their phages, in the bacterium *Candidatus* Hodgkinia cicadicola, and in some mitochondria (Ivanova *et al.*, 2014; McCutcheon *et al.*, 2009). The mitochondrion of the green alga *Pedinomonas minor* decodes both UGG and UGA as tryptophan using a single *trnW* (Turmel *et al.*, 1999). These examples set a precedent for the reassignment of UGA in the *C. reinhardtii* chloroplast.

The recognition of more than one codon sequence by an anticodon ('wobble') involves post-transcriptional modifications of the tRNA by one or more enzymes (Crick, 1966; El Yacoubi *et al.*, 2012). Unmodified uridine in the wobble position of an anticodon ( $U_{34}$ ) binds only to adenosine in the third position of the codon (Agris *et al.*, 2007); in the case of the synthetic *C. reinhardtii* chloroplast *trnW<sub>UCA</sub>* this would mean that only the codon UGA would be recognized, as desired.  $U_{34}$  modifications in naturally occurring *trnW<sub>UCA</sub>* include 5-taurinomethyluridine in *Bos taurus* mitochondria and 5-carboxymethylaminomethyluridine in *Tetrahymena thermophila* mitochondria (see Modomics database at <http://modomics.genesilico.pl>).  $U_{34}$  modifications can allow the anticodon to recognize certain other bases in the codon's wobble position, but the relationship is rather complex and not fully understood (Agris *et al.*, 2007). We have not determined whether  $U_{34}$  in the synthetic tRNA is modified, but two findings indirectly suggest that wobble may not occur. First, the heteroplasmic nature of *C. reinhardtii* mutants that suppress a TGG→TGA mutation in *rbcL* (Spreitzer *et al.*, 1985) suggests that a homoplasmic CCA→UCA mutation in the endogenous *trnW* anticodon would be lethal due to the prevention of UGG translation. Second, pleiotropic effects might be expected if our synthetic tRNA could insert tryptophan in native proteins in response to UGU or UGC cysteine codons, but no change in growth rate was observed in *C. reinhardtii* cell lines containing the tRNA.

Peptide chain release factors specifically recognize stop codons and can antagonize attempts to reassign these to sense codons. However, we found no observable difference in the level of CrCD protein accumulation between *C. reinhardtii* cell lines with an intact *crCD* gene or *crCD\*\* + trnW<sub>UCA</sub>* (Figure S2). This suggests that either there is no release factor that recognizes UGA codons in the *C. reinhardtii* chloroplast or such a factor is outcompeted successfully by *trnW<sub>UCA</sub>*. In eubacteria and the *Arabidopsis thaliana* chloroplast, release factor PrfA (RF1) recognizes UAA and UAG codons whilst PrfB (RF2) recognizes UAA and UGA; both trigger peptidyl-tRNA hydrolysis. Since neither the chloroplast nor mitochondria in *C. reinhardtii* use UGA as a stop (or sense) codon, PrfB should not be required, but nevertheless a *prfB* orthologue (locus Cre01.g010864) is present in the nuclear genome. It is not clear from signal peptide analysis whether the resulting PrfB protein would be targeted to the chloroplast, mitochondria or both; the cytosol has its own eukaryotic release factor system. The motifs required for PrfB function and specificity are well studied in other organisms (Frolova *et al.*, 1999; Ito *et al.*, 2000; Johnson *et al.*, 2011; Wilson *et al.*, 2000). The stop codon recognition motif SPF, peptide release motif GGQ and essential Ser246 residue are all intact in the *C. reinhardtii* PrfB orthologue, although the predicted N-terminus is dissimilar to those of the *E. coli* and *A. thaliana* PrfB proteins. *Chlamydomonas reinhardtii* may eventually lose the *prfB* gene, as has happened in *Candidatus* Hodgkinia cicadicola (McCutcheon *et al.*, 2009). Alternatively it may be retained if it is required for efficient UAA termination or has been adapted and recruited to stabilize particular RNA transcripts, as is the case for PrfB3 in *A. thaliana* (Stoppel *et al.*, 2011).

The synthetic tRNA did not allow detectable readthrough of UGA codons in *E. coli*. This observation allows the gene of interest and *trnW<sub>UCA</sub>* to be combined into a single vector and also reduces the risk that transgenes would be translatable if they spread to other organisms by horizontal gene transfer. The two main factors likely to contribute to this lack of readthrough in

*E. coli* are competition with the release factor PrfB and differences in the mechanism of specific tRNA function between *E. coli* and the *C. reinhardtii* chloroplast. Regarding the latter factor, the strongest recognition elements for the aminoacylation of *E. coli* trnW with tryptophan are the discriminator base G73, which is conserved across prokaryotic (but not eukaryotic) trnW, and the anticodon CCA (Himeno *et al.*, 1991; Hughes and Ellington, 2010). The *C. reinhardtii* chloroplast trnW and its synthetic trnW<sub>UCA</sub> counterpart do contain the G73 base. However, since the synthetic trnW<sub>UCA</sub> necessarily carries a mutated anticodon, the *E. coli* tryptophanyl-tRNA synthetase may not recognize it to be charged with tryptophan. Differences in the synthesis of the tRNA 3' acceptor stem may also contribute to a lack of tRNA transferability between chloroplasts and *E. coli*: this trinucleotide sequence (also CCA) is included in tRNA genes in *E. coli*, whereas in the *C. reinhardtii* chloroplast and cyanobacteria it is added post-transcriptionally by a nucleotidyltransferase enzyme (Schmidt and Subramanian, 1993; Xiong and Steitz, 2006).

An alternative strategy for cloning genes that are toxic to *E. coli* into chloroplast expression vectors was demonstrated by Oey *et al.* (2009), who inserted bacterial transcription termination signals between a selectable marker gene and a downstream endolysin gene. After cloning in *E. coli* and transformation of tobacco, the selectable marker gene and termination signals were subsequently removed by Cre-*loxP* recombination to enhance endolysin expression. Whilst this strategy worked well for the gene tested and is a good compromise for genetic systems that lack a spare codon, there was a low level of leaky expression in *E. coli* so it would not be suitable for proteins that are highly toxic to this bacterium. In contrast, with the trnW<sub>UCA</sub> strategy we did not detect any CrCD protein in the *E. coli* pWUCA2-CD\*\* cell line. If necessary, any leaky translation could be reduced further by increasing the number of TGA codons in the gene, with the maximum being the number of tryptophan residues in the protein.

Another approach to reduce the expression of bactericidal proteins during cloning is to culture the *E. coli* at a lower temperature, which increases plasmid supercoiling and reduces transcription of the transgene, at least when plant *rrn* and *psbA* promoters are used (Madesis *et al.*, 2010). Due to the differential sensitivity of plastid promoters to topology (Stirdivant *et al.*, 1985), the efficacy of this strategy is likely to vary between chloroplast expression vectors. Antimicrobial peptides, which would be lethal to *E. coli* if expressed on their own, are often expressed as fusion proteins to temporarily mask their function when this bacterium is used as an expression platform or cloning host (Lee *et al.*, 2011; Li, 2009). This is an effective strategy but requires the extra processing step of protease cleavage, adding to the time and cost of protein production and making it inappropriate for the manipulation of metabolic pathways.

Genetic code manipulation can be used to introduce non-canonical amino acids into certain positions within proteins *in vivo*, with applications including altering enzyme properties, enabling chemical modifications and providing trophic biocontainment by making an organism dependent on unnatural amino acids (Ravikumar and Liu, 2015). For example Zhang *et al.* (2004b) adapted the *Bacillus subtilis* tryptophanyl tRNA and its cognate synthetase to incorporate 5-hydroxytryptophan in response to the opal stop codon UGA in mammalian cells. This required altering the tRNA anticodon sequence and a single amino acid in the active site of the synthetase.

5-hydroxytryptophan has unique spectral properties and can be used to study protein structure and function. As long as non-canonical amino acids can be taken up into the chloroplast from the growth medium, the spare UGA codon in the *C. reinhardtii* chloroplast genetic code could be used to encode such an amino acid. According to the Codon Usage Database ([www.kazusa.or.jp](http://www.kazusa.or.jp)) there are no spare codons in the plastomes of the model higher plants *Arabidopsis thaliana*, *Nicotiana tabacum* or *Zea mays*, but the reassignment of existing codons such as UGA or the use of quadruplet codons may be possible, if less efficient. Indeed, non-canonical amino acids have been engineered into bacteria, yeast and mammalian cells despite the lack of spare codons in these organisms (Niu *et al.*, 2013; Wang and Wang, 2012).

A typical *E. coli* genome contains 2765 TAA (ochre), 321 TAG (amber) and 1249 TGA (opal) stop codons ([www.kazusa.or.jp](http://www.kazusa.or.jp), Codon Usage Database). Amber is the rarest codon in *E. coli* and has been successfully assigned to encode non-canonical amino acids using transgenic orthogonal aminoacyl-tRNA synthetase/tRNA pairs. In standard synthetic *E. coli* amber suppressor lines, UAG is still used for translation termination in many endogenous genes, and recent studies show that the cells evolve to counteract amber suppression by inserting transposons into the new aminoacyl-tRNA synthetase gene and decreasing plasmid copy number (Wang *et al.*, 2014). Growth rates are also decreased. These issues could hinder the development of stable, fast-growing *E. coli* amber suppressor lines for industrial use. The generation of an *E. coli* strain with a truly emancipated amber codon for genetic engineering purposes involved the replacement of all 321 TAG stop codons in the genome with TAA, then the deletion of *prfA* encoding release factor 1, to prevent competition with the transgenic tRNA; unfortunately this strain has a 60% increased doubling time (Johnson *et al.*, 2012; Lajoie *et al.*, 2013).

In contrast, the *C. reinhardtii* chloroplast genome uses ochre, amber and opal stop codons 65, four and zero times respectively. This reflects the considerably smaller gene content of the chloroplast genome: most proteins in *C. reinhardtii* are encoded by the nuclear genome, including many chloroplast-targeted proteins such as some components of the photosynthetic machinery (Harris *et al.*, 2009). The genetic isolation of the plastidial translation system should enable its manipulation independently of the nuclear and mitochondrial genomes. Although UGA is the only codon that is completely absent from protein-coding genes in the *C. reinhardtii* chloroplast genome, several other codons are rarely used and are potential targets for reassignment if several different non-canonical amino acids were required in a single cell line. As well as the low number of amber (UAG) codons mentioned above, there are only six instances of CGG and 14 of CUC. The finding that the *C. reinhardtii* chloroplast UGA codon can be efficiently assigned simply with the addition of a modified tRNA gene provides a starting point for more advanced genetic code manipulation in microalgae.

## Experimental procedures

### *Chlamydomonas reinhardtii* strains and growth conditions

For experiments using pPsaA\*, the cell wall-deficient strain *C. reinhardtii* cw15 was used as the recipient. For all other experiments, the recipient strain was *C. reinhardtii* TN72 (Young

and Purton, 2014), which is a *cw15 psbH*-deletion mutant. Cell lines were maintained on Tris-acetate phosphate (TAP) plates with 2% agar (Harris *et al.*, 2009) and were cultured for growth tests and Western analysis in flasks containing 20 mL TAP, shaking at 120 rpm and 25 °C. Where required, optical density was measured at 750 nm using a spectrophotometer.

### Plasmid construction

A synthetic *trnW<sub>UCA</sub>* gene was designed by taking the *C. reinhardtii* chloroplast *trnW* gene sequence with 100 nt flanking DNA on each side (i.e. Genbank accession number BK000554.2, position 17481 to 17207; Appendix S1), altering the anticodon from CCA to TCA, and adding MluI restriction sites at both ends of the fragment for cloning. The DNA was synthesized as a linear fragment by Integrated DNA Technologies (Coralville, IA, USA) and cloned into the MluI site in pBev1 (Hallahan *et al.*, 1995) to make pWUCA1, or into the MluI site in pSRSapI (Young and Purton, 2014) to make pWUCA2. The sequences of pWUCA1 and pWUCA2 are given in Appendix S1. Restriction enzymes were purchased from New England Biolabs (Ipswich, MA). Plasmids were cloned by the transformation of chemically competent *E. coli* DH5 $\alpha$  using ampicillin selection (Sambrook and Russell, 2001) and extracted by alkaline lysis (Sambrook and Russell, 2001) or with a QIAfilter Plasmid Midi kit (Qiagen, Venlo, The Netherlands).

To make the pCD\*\* construct, two TGG→TGA mutations were introduced into plasmid pCD/pRY127d (Young and Purton, 2014) by one-step isothermal assembly using three PCR products (Gibson *et al.*, 2009); primers are listed in Table S1 and DNA was amplified using Phusion High-Fidelity DNA Polymerase (Thermo Scientific, Waltham, MA) according to the manufacturer's instructions. The mutations correspond to tryptophans W21 and W147 of the 436 aa CrCD protein. This is the first and sixth of the seven tryptophan residues in CrCD.

To make the pPsaA\* construct, a TGG→TGA mutation at amino acid position W693 was introduced into the *psaA* exon 3 ORF in the plasmid pBev1 (Hallahan *et al.*, 1995) by one-step isothermal assembly of a single PCR product whose ends overlapped by 22 bp; primer sequences are given in Table S1. The DNA sequence of pPsaA\* is given in Appendix S1.

Modified versions of the *Salmonella* Typhimurium bacteriophage SPN9CC endolysin and *Shewanella denitrificans* Sden\_1266 genes were synthesized by Eurofins Genomics (Ebersberg, Germany) and Integrated DNA Technologies, respectively; see Appendix S1. They were each cloned into pWUCA2 using SapI and SphI restriction enzymes, placing them under a *psaA* exon 1 promoter. The resulting plasmids (pSty and pSde) are described in Table 1.

### *Chlamydomonas reinhardtii* transformation

Transformation was carried out using the glass bead vortex method as described in Young and Purton (2014), with selection on high-salt minimal medium for constructs that restore *psbH* or PsaA accumulation and selection on TAP + 100  $\mu$ g/mL spectinomycin for constructs containing an *aadA* spectinomycin resistance cassette. Colonies were checked for homoplasmy of the insertion by PCR using the primers shown in Table S2 and Phusion Polymerase (see above). PCR products were analysed on 1% agarose gels alongside GeneRuler DNA Ladder Mix (Thermo Scientific).

### Western blot analysis

10 mL mid-log phase *C. reinhardtii* cultures grown under 90  $\mu$ E/m<sup>2</sup>/s light were harvested for Western blot analysis of proteins. Preparation of lysates, SDS-PAGE gels, blotting onto a nitrocellulose membrane and incubation in the primary  $\alpha$ HA or  $\alpha$ RbcL antibody were performed as described previously (Young and Purton, 2014) except that  $\alpha$ HA was prepared in TBS with 0.1% Tween (TBS-T) and 0.5% milk. Blots were then incubated for 1 h in the secondary antibody, goat  $\alpha$ rabbit Dylight 800 (Thermo Scientific) diluted 1:25 000 in TBS-T and 0.5% milk, washed in TBS-T and imaged with an Odyssey Fc Imaging System (LI-COR, Lincoln, NE) at 800 nm.

For the *E. coli* Western blot, strains were grown in LB with 100  $\mu$ g/mL ampicillin overnight at 37 °C. Optical densities were measured at 600 nm then cultures were pelleted and resuspended in sample buffer to equal densities as described in the Mini-PROTEAN Tetra Cell manual (Bio-Rad, Hercules, CA). The protocol was then continued exactly as for the *C. reinhardtii* Western blots described in Young and Purton (2014). Briefly, samples were boiled, loaded onto a 15% acrylamide gel, blotted onto a nitrocellulose membrane and probed with  $\alpha$ HA primary antibody and ECL  $\alpha$ rabbit IgG HRP-linked secondary antibody, with detection via chemiluminescence.

### 5-fluorocytosine sensitivity test

Liquid *C. reinhardtii* cultures grown in TAP medium for 48 h were adjusted to an optical density of 0.4 at 750 nm. 5  $\mu$ L was spotted onto TAP plates containing 2% agar and either no drug or 2 mg/mL 5-fluorocytosine (Sigma-Aldrich, St. Louis, MO, USA). Plates were incubated under 50  $\mu$ E/m<sup>2</sup>/s light at 25 °C for 10 days.

### *Escherichia coli* growth curve

5 mL LB broths containing 100  $\mu$ g/mL ampicillin were inoculated with overnight *E. coli* DH5 $\alpha$  cultures containing each plasmid so that the starting absorbance at 600 nm was 0.1. Cultures were incubated at 37 °C with shaking, and the absorbance at 600 nm was measured every 90 min using a spectrophotometer.

### Acknowledgements

This work was funded by the UK Biotechnology and Biological Sciences Research Council, grant BB/I007660/1. We thank UCL for covering the cost of open access publication. The authors have no conflicts of interest to declare.

### References

- Agris, P.F., Vendeix, F.A. and Graham, W.D. (2007) tRNA's wobble decoding of the genome: 40 years of modification. *J. Mol. Biol.* **366**, 1–13.
- Amitai, G. and Sorek, R. (2012) PanDaTox: a tool for accelerated metabolic engineering. *Bioengineered* **3**, 218–221.
- Bock, R. (2015) Engineering plastid genomes: methods, tools, and applications in basic research and biotechnology. *Annu. Rev. Plant Biol.* **66**, 211–241.
- Boudreaux, B., MacMillan, F., Teutloff, C., Agalarov, R., Gu, F., Grimaldi, S., Bittl, R. *et al.* (2001) Mutations in both sides of the photosystem I reaction center identify the phyloquinone observed by electron paramagnetic resonance spectroscopy. *J. Biol. Chem.* **276**, 37299–37306.
- Cognat, V., Pawlak, G., Duchene, A.M., Daujat, M., Gigant, A., Salinas, T., Michaud, M. *et al.* (2013) PlantRNA, a database for tRNAs of photosynthetic eukaryotes. *Nucleic Acids Res.* **41**, D273–D279.
- Crick, F.H. (1966) Codon-anticodon pairing: the wobble hypothesis. *J. Mol. Biol.* **19**, 548–555.



- Doetsch, N.A., Favreau, M.R., Kuscuoglu, N., Thompson, M.D. and Hallick, R.B. (2001) Chloroplast transformation in *Euglena gracilis*: splicing of a group III twintron transcribed from a transgenic *psbK* operon. *Curr. Genet.* **39**, 49–60.
- El Yacoubi, B., Bailly, M. and de Crécy-Lagard, V. (2012) Biosynthesis and function of posttranscriptional modifications of transfer RNAs. *Annu. Rev. Genet.* **46**, 69–95.
- Frolova, L.Y., Tsvikovskii, R.Y., Sivolobova, G.F., Oparina, N.Y., Serpinsky, O.I., Blinov, V.M., Tatkov, S.I. et al. (1999) Mutations in the highly conserved GGQ motif of class 1 polypeptide release factors abolish ability of human eRF1 to trigger peptidyl-tRNA hydrolysis. *RNA*, **5**, 1014–1020.
- Georgianna, D.R., Hannon, M.J., Marcuschi, M., Wu, S., Botsch, K., Lewis, A.J., Hyun, J. et al. (2013) Production of recombinant enzymes in the marine alga *Dunaliella tertiolecta*. *Algal Res.* **2**, 2–9.
- Gibson, D.G., Young, L., Chuang, R.Y., Venter, J.C., Hutchison, C.A. 3rd and Smith, H.O. (2009) Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat. Methods* **6**, 343–345.
- Hallahan, B.J., Purton, S., Ivson, A., Wright, D. and Evans, M.C. (1995) Analysis of the proposed Fe-SX binding region of Photosystem 1 by site directed mutation of PsaA in *Chlamydomonas reinhardtii*. *Photosynth. Res.* **46**, 257–264.
- Harris, E.H., Stern, D.B. and Witman, G.B. (2009) *The Chlamydomonas Sourcebook*, 2nd edn. Oxford: Elsevier Inc.
- Himeno, H., Hasegawa, T., Asahara, H., Tamura, K. and Shimizu, M. (1991) Identity determinants of *E. coli* tryptophan tRNA. *Nucleic Acids Res.* **19**, 6379–6382.
- Hughes, R.A. and Ellington, A.D. (2010) Rational design of an orthogonal tryptophanyl nonsense suppressor tRNA. *Nucleic Acids Res.* **38**, 6813–6830.
- Ito, K., Uno, M. and Nakamura, Y. (2000) A tripeptide 'anticodon' deciphers stop codons in messenger RNA. *Nature*, **403**, 680–684.
- Ivanova, N.N., Schwientek, P., Tripp, H.J., Rinke, C., Pati, A., Huntemann, M., Visel, A. et al. (2014) Stop codon reassignments in the wild. *Science*, **344**, 909–913.
- Johnson, D.B., Xu, J., Shen, Z., Takimoto, J.K., Schultz, M.D., Schmitz, R.J., Xiang, Z. et al. (2011) RF1 knockout allows ribosomal incorporation of unnatural amino acids at multiple sites. *Nat. Chem. Biol.* **7**, 779–786.
- Johnson, D.B., Wang, C., Xu, J., Schultz, M.D., Schmitz, R.J., Ecker, J.R. and Wang, L. (2012) Release factor one is nonessential in *Escherichia coli*. *ACS Chem. Biol.* **7**, 1337–1344.
- Jordan, P., Fromme, P., Witt, H.T., Klukas, O., Saenger, W. and Krauss, N. (2001) Three-dimensional structure of cyanobacterial photosystem I at 2.5 Å resolution. *Nature*, **411**, 909–917.
- Kimelman, A., Levy, A., Sberro, H., Kidron, S., Leavitt, A., Amitai, G., Yoder-Himes, D.R. et al. (2012) A vast collection of microbial genes that are toxic to bacteria. *Genome Res.* **22**, 802–809.
- Lajoie, M.J., Rovner, A.J., Goodman, D.B., Aerni, H.R., Haimovich, A.D., Kuznetsov, G., Mercer, J.A. et al. (2013) Genomically recoded organisms expand biological functions. *Science*, **342**, 357–360.
- Lee, S.B., Li, B., Jin, S. and Daniell, H. (2011) Expression and characterization of antimicrobial peptides Retrocyclin-101 and Protegrin-1 in chloroplasts to control viral and bacterial infections. *Plant Biotech J.* **9**, 100–115.
- Li, Y. (2009) Carrier proteins for fusion expression of antimicrobial peptides in *Escherichia coli*. *Biotechnol. Appl. Biochem.* **54**, 1–9.
- Lim, J.A., Shin, H., Heu, S. and Ryu, S. (2014) Exogenous lytic activity of SPN9CC endolysin against gram-negative bacteria. *J. Microbiol. Biotechnol.* **24**, 803–811.
- Madesis, P., Osathanunkul, M., Georgopoulou, U., Gisby, M.F., Mudd, E.A., Nianiou, I., Tsitoura, P. et al. (2010) A hepatitis C virus core polypeptide expressed in chloroplasts detects anti-core antibodies in infected human sera. *J. Biotech.* **145**, 377–386.
- Maul, J.E., Lilly, J.W., Cui, L., dePamphilis, C.W., Miller, W., Harris, E.H. and Stern, D.B. (2002) The *Chlamydomonas reinhardtii* plastid chromosome: islands of genes in a sea of repeats. *Plant Cell* **14**, 2659–2679.
- McCutcheon, J.P., McDonald, B.R. and Moran, N.A. (2009) Origin of an alternative genetic code in the extremely small and GC-rich genome of a bacterial symbiont. *PLoS Genet.* **5**, e1000565.
- Michelet, L., Lefebvre-Legendre, L., Burr, S.E., Rochaix, J.D. and Goldschmidt-Clermont, M. (2011) Enhanced chloroplast transgene expression in a nuclear mutant of *Chlamydomonas*. *Plant Biotechnol. J.* **9**, 565–574.
- Nakamura, A., Tsukada, T., Auer, S., Furuta, T., Wada, M., Koivula, A., Igarashi, K. et al. (2013) The tryptophan residue at the active site tunnel entrance of *Trichoderma reesei* cellobiohydrolase Cel7A is important for initiation of degradation of crystalline cellulose. *J. Biol. Chem.* **288**, 13503–13510.
- Niu, W., Schultz, P.G. and Guo, J. (2013) An expanded genetic code in mammalian cells with a functional quadruplet codon. *ACS Chem. Biol.* **8**, 1640–1645.
- Novoselov, S.V., Rao, M., Onoshko, N.V., Zhi, H., Kryukov, G.V., Xiang, Y., Weeks, D.P. et al. (2002) Selenoproteins and selenocysteine insertion system in the model plant cell system, *Chlamydomonas reinhardtii*. *EMBO J.* **21**, 3681–3693.
- Oey, M., Lohse, M., Scharff, L.B., Kreikemeyer, B. and Bock, R. (2009) Plastid production of protein antibiotics against pneumonia via a new strategy for high-level expression of antimicrobial proteins. *Proc. Natl Acad. Sci. USA* **106**, 6579–6584.
- Purton, S., Stevens, D.R., Muhiuddin, I.P., Evans, M.C., Carter, S., Rigby, S.E. and Heathcote, P. (2001) Site-directed mutagenesis of PsaA residue W693 affects phyloquinone binding and function in the photosystem I reaction center of *Chlamydomonas reinhardtii*. *Biochemistry*, **40**, 2167–2175.
- Rao, M., Carlson, B.A., Novoselov, S.V., Weeks, D.P., Gladyshev, V.N. and Hatfield, D.L. (2003) *Chlamydomonas reinhardtii* selenocysteine tRNA[Ser] Sec. *RNA*, **9**, 923–930.
- Rasala, B.A. and Mayfield, S.P. (2015) Photosynthetic biomanufacturing in green algae; production of recombinant proteins for industrial, nutritional, and medical uses. *Photosynth. Res.* **123**, 227–239.
- Ravikumar, A. and Liu, C.C. (2015) Biocontainment through reengineered codes. *ChemBioChem* **16**, 1149–1151.
- Rinke, C., Schwientek, P., Sczyrba, A., Ivanova, N.N., Anderson, I.J., Cheng, J.F., Darling, A. et al. (2013) Insights into the phylogeny and coding potential of microbial dark matter. *Nature*, **499**, 431–437.
- Robbins, S., Derelle, E., Ferraz, C., Wuyls, J., Moreau, H. and Van de Peer, Y. (2007) The complete chloroplast and mitochondrial DNA sequence of *Ostreococcus tauri*: organelle genomes of the smallest eukaryote are examples of compaction. *Mol. Biol. Evol.* **24**, 956–968.
- Sambrook, J. and Russell, D.W. (2001) *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor: CSHL Press.
- Scaife, M.A., Nguyen, G.T., Rico, J., Lambert, D., Helliwell, K.E. and Smith, A.G. (2015) Establishing *Chlamydomonas reinhardtii* as an industrial biotechnology host. *Plant J.* **82**, 532–546.
- Schmidt, J. and Subramanian, A.R. (1993) Sequence of the cyanobacterial tRNA (w) gene in *Synechocystis* PCC 6803: requirement of enzymatic 3' CCA attachment to the acceptor stem. *Nucleic Acids Res.* **21**, 2519.
- Specht, E.A. and Mayfield, S.P. (2013) Synthetic oligonucleotide libraries reveal novel regulatory elements in *Chlamydomonas* chloroplast mRNAs. *ACS Synth. Biol.* **2**, 34–46.
- Spreitzer, R.J., Chastain, C.J. and Ogren, W.L. (1984) Chloroplast gene suppression of defective ribulosebiphosphate carboxylase/oxygenase in *Chlamydomonas reinhardtii*: evidence for stable heteroplasmic genes. *Curr. Genet.* **9**, 83–89.
- Spreitzer, R.J., Goldschmidt-Clermont, M., Rahire, M. and Rochaix, J.D. (1985) Nonsense mutations in the *Chlamydomonas* chloroplast gene that codes for the large subunit of ribulosebiphosphate carboxylase/oxygenase. *Proc. Natl Acad. Sci. USA* **82**, 5460–5464.
- Stirdivant, S.M., Crossland, L.D. and Bogorad, L. (1985) DNA supercoiling affects in vitro transcription of two maize chloroplast genes differently. *Proc. Natl Acad. Sci. USA* **82**, 4886–4890.
- Stoppel, R., Lezhneva, L., Schwenkert, S., Torabi, S., Felder, S., Meierhoff, K., Westhoff, P. et al. (2011) Recruitment of a ribosomal release factor for light- and stress-dependent regulation of *petB* transcript stability in *Arabidopsis* chloroplasts. *Plant Cell* **23**, 2680–2695.
- Timmis, J.N., Ayliffe, M.A., Huang, C.Y. and Martin, W. (2004) Endosymbiotic gene transfer: organelle genomes forge eukaryotic chromosomes. *Nat. Rev. Genet.* **5**, 123–135.
- Turlum, M., Lemieux, C., Burger, G., Lang, B.F., Otis, C., Plante, I. and Gray, M.W. (1999) The complete mitochondrial DNA sequences of *Nephroselmis olivacea* and *Pedinomonas minor*. Two radically different evolutionary patterns within green algae. *Plant Cell* **11**, 1717–1730.

- Wang, Q. and Wang, L. (2012) Genetic incorporation of unnatural amino acids into proteins in yeast. *Methods Mol. Biol.* **794**, 199–213.
- Wang, Q., Sun, T., Xu, J., Shen, Z., Briggs, S.P., Zhou, D. and Wang, L. (2014) Response and adaptation of *Escherichia coli* to suppression of the amber stop codon. *ChemBioChem* **15**, 1744–1749.
- Wilson, D.N., Guevremont, D. and Tate, W.P. (2000) The ribosomal binding and peptidyl-tRNA hydrolysis functions of *Escherichia coli* release factor 2 are linked through residue 246. *RNA*, **6**, 1704–1713.
- Xie, W.H., Zhu, C.C., Zhang, N.S., Li, D.W., Yang, W.D., Liu, J.S., Sathishkumar, R. *et al.* (2014) Construction of novel chloroplast expression vector and development of an efficient transformation system for the diatom *Phaeodactylum tricorutum*. *Mar. Biotechnol. (NY)* **16**, 538–546.
- Xiong, Y. and Steitz, T.A. (2006) A story with a good ending: tRNA 3'-end maturation by CCA-adding enzymes. *Curr. Opin. Struct. Biol.* **16**, 12–17.
- Young, R.E. and Purton, S. (2014) Cytosine deaminase as a negative selectable marker for the microalgal chloroplast: a strategy for the isolation of nuclear mutations that affect chloroplast gene expression. *Plant J.* **80**, 915–925.
- Yu, W. and Spreitzer, R.J. (1992) Chloroplast heteroplasmy is stabilized by an amber-suppressor tryptophan tRNA(CUA). *Proc. Natl Acad. Sci. USA* **89**, 3904–3907.
- Zhang, Y., Deshpande, A., Xie, Z., Natesh, R., Acharya, K.R. and Brew, K. (2004a) Roles of active site tryptophans in substrate binding and catalysis by alpha-1,3 galactosyltransferase. *Glycobiology*, **14**, 1295–1302.
- Zhang, Z., Alfonta, L., Tian, F., Bursulaya, B., Uryu, S., King, D.S. and Schultz, P.G. (2004b) Selective incorporation of 5-hydroxytryptophan into proteins in mammalian cells. *Proc. Natl Acad. Sci. USA* **101**, 8882–8887.

## Supporting information

Additional Supporting information may be found in the online version of this article:

**Figure S1** Homoplasmy PCR for *crCD* strains.

**Figure S2** Western blot of CrCD protein levels in *Chlamydomonas reinhardtii* cell lines.

**Figure S3** *Chlamydomonas reinhardtii* growth curves with and without  $\text{trnW}_{\text{UCA}}$ .

**Figure S4** Sequence confirmation of *psaA* codon alteration.

**Table S1** Primers used in the construction of plasmids.

**Table S2** Primers used to confirm homoplasmic integration of transgenes.

**Table S3** Stop codon distribution in microalgae.

**Appendix S1** DNA and amino acid sequences.