



LJMU Research Online

Xia, F, Li, X, Li, X, Zheng, D, Sun, Q, Liu, J, Li, Y, Hua, J and Qi, B

Elevation of the Yields of Very Long Chain Polyunsaturated Fatty Acids via Minimal Codon Optimization of Two Key Biosynthetic Enzymes

<http://researchonline.ljmu.ac.uk/id/eprint/11457/>

Article

Citation (please note it is advisable to refer to the publisher's version if you intend to cite from this work)

Xia, F, Li, X, Li, X, Zheng, D, Sun, Q, Liu, J, Li, Y, Hua, J and Qi, B (2016) Elevation of the Yields of Very Long Chain Polyunsaturated Fatty Acids via Minimal Codon Optimization of Two Key Biosynthetic Enzymes. PLoS One, 11 (7). ISSN 1932-6203

LJMU has developed **LJMU Research Online** for users to access the research output of the University more effectively. Copyright © and Moral Rights for the papers on this site are retained by the individual authors and/or other copyright owners. Users may download and/or print one copy of any article(s) in LJMU Research Online to facilitate their private study or for non-commercial research. You may not engage in further distribution of the material or use it for any profit-making activities or any commercial gain.

The version presented here may differ from the published version or from the version of the record. Please see the repository URL above for details on accessing the published version and note that access may require a subscription.

For more information please contact researchonline@ljmu.ac.uk

<http://researchonline.ljmu.ac.uk/>

RESEARCH ARTICLE

Elevation of the Yields of Very Long Chain Polyunsaturated Fatty Acids via Minimal Codon Optimization of Two Key Biosynthetic Enzymes

Fei Xia¹*, Xueying Li¹*, Xinzheng Li¹*, Desong Zheng¹, Quanxi Sun¹, Jiang Liu¹, Yaxiao Li¹, Jinping Hua², Baoxiu Qi¹*

1 State Key Laboratory of Crop Biology, Shandong Agricultural University, Tai'an, 271000, China,

2 Department of Plant Genetics & Breeding, College of Agronomy and Biotechnology, China Agricultural University, No 2, Yuanmingyuan West Road, Haidian District, Beijing, 100193, China

* These authors contributed equally to this work.

* bssbq@bath.ac.uk



OPEN ACCESS

Citation: Xia F, Li X, Li X, Zheng D, Sun Q, Liu J, et al. (2016) Elevation of the Yields of Very Long Chain Polyunsaturated Fatty Acids via Minimal Codon Optimization of Two Key Biosynthetic Enzymes. PLoS ONE 11(7): e0158103. doi:10.1371/journal.pone.0158103

Editor: Alexander Chong Shu-Chien, Universiti Sains Malaysia, MALAYSIA

Received: February 27, 2016

Accepted: June 12, 2016

Published: July 19, 2016

Copyright: © 2016 Xia et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are within the paper and its Supporting Information files.

Funding: This work was supported to BQ by the National Natural Science Foundation of China (Grant No. 30970222) and genetically modified organisms breeding major projects of China (Grant No. 009ZX08005-024B).

Competing Interests: The authors have declared that no competing interests exist.

Abstract

Eicosapentaenoic acid (EPA, 20:5 $\Delta^{5,8,11,14,17}$) and Docosahexaenoic acid (DHA, 22:6 $\Delta^{4,7,10,13,16,19}$) are nutritionally beneficial to human health. Transgenic production of EPA and DHA in oilseed crops by transferring genes originating from lower eukaryotes, such as microalgae and fungi, has been attempted in recent years. However, the low yield of EPA and DHA produced in these transgenic crops is a major hurdle for the commercialization of these transgenics. Many factors can negatively affect transgene expression, leading to a low level of converted fatty acid products. Among these the codon bias between the transgene donor and the host crop is one of the major contributing factors. Therefore, we carried out codon optimization of a fatty acid delta-6 desaturase gene *PinD6* from the fungus *Phytophthora infestans*, and a delta-9 elongase gene, *IgASE1* from the microalga *Isochrysis galbana* for expression in *Saccharomyces cerevisiae* and *Arabidopsis* respectively. These are the two key genes encoding enzymes for driving the first catalytic steps in the $\Delta 6$ desaturation/ $\Delta 6$ elongation and the $\Delta 9$ elongation/ $\Delta 8$ desaturation pathways for EPA/DHA biosynthesis. Hence expression levels of these two genes are important in determining the final yield of EPA/DHA. Via PCR-based mutagenesis we optimized the least preferred codons within the first 16 codons at their N-termini, as well as the most biased CGC codons (coding for arginine) within the entire sequences of both genes. An expression study showed that transgenic *Arabidopsis* plants harbouring the codon-optimized *IgASE1* contained 64% more elongated fatty acid products than plants expressing the native *IgASE1* sequence, whilst *Saccharomyces cerevisiae* expressing the codon optimized *PinD6* yielded 20 times more desaturated products than yeast expressing wild-type (WT) *PinD6*. Thus the codon optimization strategy we developed here offers a simple, effective and low-cost alternative to whole gene synthesis for high expression of foreign genes in yeast and *Arabidopsis*.

Introduction

Very long chain polyunsaturated fatty acids (VLC-PUFAs), such as EPA and DHA have a wide range of physiological and biological functions in cells. A higher nutritional value is associated with a greater degree of unsaturation of VLC-PUFAs [1,2,3]. VLC-PUFAs are synthesized *de novo* via an aerobic desaturation/elongation pathway [4] and an anaerobic polyketide synthase pathway [5,6]. Desaturation and elongation of VLC-PUFAs is carried out via the classic $\Delta 6$ desaturation/ $\Delta 6$ elongation pathway and the alternative $\Delta 9$ elongation/ $\Delta 8$ desaturation pathways [7]. In the $\Delta 6$ desaturation/ $\Delta 6$ elongation pathway, the first step is catalyzed by the $\Delta 6$ desaturase that converts linolenic acid (LA, $18:2\Delta^{9,12}$) and α -linolenic acid (ALA, $18:3\Delta^{9,12,15}$) to γ -linolenic acid (GLA, $18:3\Delta^{6,9,12}$) and stearidonic acid (SDA, $18:4\Delta^{6,9,12,15}$), respectively. This is followed by a $\Delta 6$ elongation step and a $\Delta 5$ desaturation step to achieve the production of arachidonic acid EPA and (AA, $20:4\Delta^{5,8,11,14}$). In the alternative $\Delta 9$ elongation/ $\Delta 8$ desaturation pathway, however, LA and ALA are first elongated by a $\Delta 9$ elongase to yield eicosadienoic acid (EDA, $20:2\Delta^{11,14}$) and eicosatrienoic acid (EtrA, $20:2\Delta^{11,14,17}$), respectively. Subsequently, these products are desaturated by $\Delta 8$ and $\Delta 5$ desaturation to produce EPA and AA. DHA is then produced by $\Delta 5$ elongation and $\Delta 4$ desaturation of EPA.

The human body can synthesize small amounts of EPA and DHA mainly via the classic $\Delta 6$ desaturation/ $\Delta 6$ elongation pathway by consumption of the essential fatty acids LA and ALA in the diet. However, due to the very low efficiency of this conversion humans rely heavily on the consumption of deep sea fish, seafood products and commercially available fish oil capsules to obtain required amount of VLC-PUFAs directly [8]. Due to rapid decline of global fish stocks and contamination of the marine environment by hazardous chemicals these sources are becoming less sustainable and less preferred in recent years. Other sources of VLC-PUFAs are from commercially cultured algae and fungi which requires their cultivation, fermentation and oil extraction. However, the low yields obtained and the high costs involved in these processes are limiting large-scale production [9]. Therefore, developing a safe and sustainable alternative production route for these VLC-PUFAs is a high priority.

Following the initial success of producing appreciable amounts of EPA in transgenic *Arabidopsis* by our group [10], research has turned to genetic engineering of oilseed crops with multiple key genes of the VLC-PUFA biosynthetic pathways. As a result, a number of EPA and DHA producing transgenic crops have been generated [8,9,11,12,13,14]. However, a common problem with these transgenic crops is that the yield of EPA, and especially DHA, is low, thus preventing their commercialization.

Because the genes used in VLC-PUFA metabolic engineering come from microalgae, fungi and protists their expression levels in other organisms, such as yeast and higher plants, may be reduced due to codon-usage bias. This can then result in lower yields of VLC-PUFAs in these organisms. It is well established that synonymous codons are not used at equal frequencies and that the codon usage frequency varies widely in different species and even between genes expressed at high and low levels in the same species [15]. This codon bias can have a profound impact on the expression efficiency of heterologous genes [16,17,18,19,20,21,22]. The optimal codons are those that are used most frequently, and those used less frequently are termed rare or low-usage codons. If a transgene contains codon(s) rarely used by the host cell, its expression level in this host is expected to be very low [16,23,24]. If these rare codon(s) appear near the 5' end of the open reading frame, the expression of that gene will decrease even more dramatically compared to where they occur in the middle of the gene sequence [20,25,26]. Therefore, one of the strategies to improve heterologous expression of a foreign gene is to optimize its rare codons to match the ones utilized most frequently by the host whilst retaining the encoded amino acid sequences of the gene [19,27].

The effect of rare codon clusters on heterologous protein expression in *E. coli* has been reported [16,28]. For example, Kane [16] noted that the presence of individual AGG/AGA_{Arg}, CUA_{Leu}, AUA_{Ile}, CGA_{Arg} or CCC_{Pro} codons in the DNA sequence could cause translational problems in *E. coli*. If these codons form a cluster it can significantly reduce both the quantity and quality of the synthesized protein. Kim and Lee [20] also reported that clusters of rare codons at the 5'-end of a gene could lead to significantly reduced levels of heterologous gene expression in the host.

Codon usage has been studied in a number of different organisms. Initial studies on codon optimization were carried out in bacteria (*Escherichia coli*) where the heterologous production of mammalian proteins was tested, resulting in great improvement of gene expression levels. For example, the very first study was carried out by Itakura *et al* in 1977 who reported production of the first functional human polypeptide in *E. coli* by utilizing a codon-optimized 14-codon-long DNA molecule encoding human somatostatin. Expression of the codon-optimized DNA fragment yielded 1- to 40-fold more polypeptide than the non-optimized DNA molecule [29]. Later, Kink *et al* [30] successfully expressed the Ca²⁺-binding protein calmodulin from *Paramecium calmodulin*, again in *E. coli*, by changing four TAA codons to optimized CAA codons in its encoding gene. This optimization resulted in around 170 times more calmodulin production in *E. coli* than in *Paramecium* cells.

The effect of rare codons on expression of heterologous proteins in yeast has also been reported [27,31,32,33,34,35]. For example, Yadava *et al* [31] codon-optimized the F2 domain of the *Plasmodium falciparum* erythrocyte binding antigen (EBA-175), a strong candidate for a vaccine against malaria. Two synthetic genes were produced for its expression in *E. coli* and *Pichia pastoris*, with the aim of identifying the best heterologous host for high-level production of biologically active F2 domain of EBA-175 (EBA-F2). Their results showed that codon optimization significantly improved the expression of EBA-F2 in both systems compared to the native sequence. However, the protein produced in *Pichia* was superior in terms of its expression levels, solubility, and biological activity. Another example was the high efficiency expression of a β -1,3-1,4-glucanase from *Bacillus licheniformis* following codon optimization for *Pichia* expression. This led to a 10-fold increase in production of an active enzyme [34]. Thus, it is clear that codon optimization of target genes according to the codon bias of the host cell can frequently result in 10- to 50-fold increases of target protein production [27,33,35].

Here we report on codon optimization of two genes involved in VLC-PUFA biosynthesis for improved heterologous expression in the higher plant *Arabidopsis* and in yeast. The first gene is the $\Delta 9$ elongase gene, *IgASE1* which was isolated from a unicellular marine microalgae *I. galbana* [7], and the second a $\Delta 6$ desaturase gene, *PinD6*, from *P. infestans* [36]. Through PCR-based site-directed mutagenesis we optimized the first 15 codons of their N termini, and also all arginine codons in the coding sequences, as arginine displayed the highest degree of codon-usage bias. These modified variants together with their wild-type (WT) counterparts were subsequently expressed in *Arabidopsis* (*IgASE1*) and yeast (*PinD6*), respectively. Our results show that codon optimization of both genes resulted in marked increases in the production of VLC-PUFAs in transgenic hosts compared to the expression of the native genes.

Materials and Methods

Experimental Materials

Phytophthora infestans, isolate 88069 [37], was gifted by Dr. Ming Xue, Shandong Agricultural University, China and grown as described previously [36]. After 30 days, the mycelium was harvested by scraping it off the plate and then stored at -80°C for subsequent analysis. *Agrobacterium tumefaciens* strain GV3101 was a gift from Dr. Colin Lazarus, University of Bristol, UK.

Wild-type (WT) *Arabidopsis* ecotype ‘Columbia-0’ (Col-0) was obtained from the Nottingham Arabidopsis Stock Centre, UK. The delta-9 fatty acid elongase gene, *IgASE1* (GenBank accession number AAL37626) from *Isochrysis galbana* and the delta-6 desaturase gene *PinDes6* (GenBank Accession number XP_00290828.1) from *Phytophthora infestans* were isolated as described previously (7,36).

Codon Usage Frequency Analysis

Codon usage frequency was analyzed through codon analysis software “Graphical Codon Usage Analyser” (<http://gcu.schoedl.de/> and <http://www.kazusa.or.jp/codon/>).

Mutagenesis and Expression of *IgASE1* in *Arabidopsis*

Overlap extension PCR technology was employed according to Ho et al [37] and Qi et al [38], using primer pairs listed in Table 1. To optimize the first 16 codons of the N-terminus (N^{OP}) and to change CGC_{Arg}¹⁰ to AGA_{Arg}¹⁰ (R10^{OP}) of *IgASE1*, respective primer pairs of E9-N16/E9R and E9-1F/E9R (Table 1) were used. For optimization of all other CGC_{Arg} codons, and combined 2- or 3-codon changes 2 or more rounds of PCR reaction were usually involved. For example, to change CGC_{Arg} to AGA_{Arg} at the 35th position of *IgASE1* (R35^{OP}-*IgASE1*), two rounds of PCR were carried out. In the first round, primer pairs of E9F/E9-2R and E9-2F/E9R were used to generate two individual PCR fragments with change of CGC_{Arg} to AGA_{Arg} (Table 1, in red) plus 15–18 bases flanking AGA at the 3’ end of the E9F/E9-2R and the 5’ end of the E9-2F/E9R fragments, respectively. These two PCR fragments were purified and used as templates in a second round PCR reaction using a primer pair of E9F/E9R to assemble the full length R35^{OP}-*IgASE1*. In the same way, R84^{OP}-*IgASE1* was constructed with primer pairs E9F/E9-3R, E9-3F/E9R (first round PCR) and E9F/E9R (second round PCR) (Table 1). The WT-*IgASE1*, amplified with primer pair E9F and E9R, together with these 3 CGC codon optimized

Table 1. Primers used in this study.

Primer	Sequence (5’-3’)	gene	Restriction site
E9F	<u>GGATCC</u> ATG GCCCTCGCAAACGACGC	WT- <i>IgASE1</i>	<i>Bam</i> HI
E9R	GAGCTCCTAGAGCTGCTTGCCCGCCT	WT- <i>IgASE1</i>	<i>Sac</i> I
E9-1F	<u>GGATCC</u> ATGGCCCTCGCAAACGACGCGGGAGAG AGA AATCTGGGCGGCTGT	R10 ^{OP} - <i>IgASE1</i>	<i>Bam</i> HI
E9-2F	CAAACCGCTGCTC AGA AATCCGGGCTGGTGGATGAGAAG	R35 ^{OP} - <i>IgASE1</i>	-
E9-2R	CATCCACCAGCCCGGAATTTCTGAGCAGCGGTTTGAGTAG		-
E9-3F	GGCGCGTGGCTG AGA AAGGCAAACCGGCGACACACC	R84 ^{OP} - <i>IgASE1</i>	-
E9-3R	GTGTCGCCGGTTTGCCCTTCTCAGCCACGCGCCCGTACC		-
E9-N16	<u>GGATCC</u> ATGGCTCT TGCTAATGATGCTGGAGAAAGA ATTGGGCTGCTGTTACTGAT	N16 ^{OP} - <i>IgASE1</i>	<i>Bam</i> HI
D6F	<u>AAGCTT</u> ATG GTGGACGCCCCAAGACC	WT- <i>PinDes6</i>	<i>Hind</i> III
D6R	<u>GGATCC</u> TTACATGGCGGAAACTCGA	WT- <i>PinDes6</i>	<i>Bam</i> HI
D6-1F	AAGCTTATGGTGGACGCCCCAAGACCAAG AGA AAGATC	R9 ^{OP} - <i>PinDes6</i>	<i>Hind</i> III
D6-2F	TACTCTTCGCT AG ATTTCAGCTGGCTGTTGCAG	R276 ^{OP} - <i>PinDes6</i>	-
D6-2R	GCCAGCTGAATCTAGCGAAGAGTAGCA	WT- <i>PinDes6</i>	-
D6-3F	CACCACTACG AGA AACATTACGCCG	R376 ^{OP} - <i>PinDes6</i>	-
D6-3R	CGGCGTAATGTTTCTCGTAGTGGTGAC	WT- <i>PinDes6</i>	-
D6-4F	ATTCTAC AG AGGTCTCGTTGAGGT	R433 ^{OP} - <i>PinDes6</i>	-
D6-4R	CTCAACGAGACCTCTGTAGAATCCCGTC	WT- <i>PinDes6</i>	-
D6-N16	AAGCTTATGGT TGATGGTCCAAA ACT TAA AGAAAATTTCTTGGCA AGA AGTTAAGCAGCACG	N16 ^{OP} - <i>PinDes6</i>	<i>Hind</i> III

The translation start codon ATG is in bold, mutated bases are in bold and italic and restriction sites are underlined.

doi:10.1371/journal.pone.0158103.t001

DNA fragments were subsequently cloned into the pMD18-T vector (Takara) and sequenced to confirm the presence of the desired changes. These were used as templates to generate $R^{10,35op}$, $R^{10,84op}$, $R^{10,35,84op}$ and $N^{op}+3R^{op}$ *IgASE1* using different combinations of primer pairs listed in [Table 1](#). Individual DNA fragments were cloned in the pMD18-T vector and sequenced to verify changes as above.

For expression in *Arabidopsis*, the WT and all 8 mutated *IgASE1* DNA fragments were transferred via *Bam*HI and *Sac*I restriction sites to the plant expression vector pCambia 2300EC that contains a plant expression cassette that utilizes the CaMV 35S promoter and Nos terminator sequences [39]. *Agrobacterium tumefaciens* strain GV3101 was used to transform the WT *Arabidopsis* via the floral dipping method [40]. Transformants were selected on kanamycin containing ½ MS nutrient medium as described previously and PCR was carried out to check the presence of the transgene [41]. Ten transgenic plants from each line were randomly selected and total fatty acids from leaves were extracted and subjected to gas liquid chromatography as previously described [42]. Three higher C20 fatty acid producing lines that harbour a single copy of the transgene (segregating 3:1 for kanamycin resistant to kanamycin sensitive plants in the T2 seedlings) were taken to the T3 generation. Homozygous transgenic plants were isolated from these lines if they all survived on kanamycin-selective ½ MS agar plates. Again, total fatty acids were extracted and analysed from leaves of these plants.

Functional Characterization of PinD6 and Its Codon-Optimized Variants in Yeast

Individual fragments of the WT and mutated *PinD6* were generated as for *IgASE1* mutants, using primer pairs listed in [Table 1](#). They were cloned into the pMD18-T vector and sequenced to confirm the desired changes. Individual fragments were digested with *Bam*HI and *Hind*III and ligated into the corresponding restriction sites of the yeast expression vector pYES2 (Invitrogen), downstream of the GAL1 promoter, to generate pYES-WT-PinD6, pYES-N^{op}PinD6, pYES 4R^{op}-PinD6 and pYES-N^{op}+4R^{op}, respectively.

These were introduced into *S. cerevisiae* strain YPH500 (ura3-52, lys2-801^{amber}, ade2-101^{ochre}, trp1-Δ63, his3-Δ200, leu2-Δ1) (Stratagene) by the lithium acetate method and selected on minimal media without uracil [41]. Expression of these enzymes was induced by the addition of 2% (w/v) galactose to cultures grown in raffinose minimal liquid media as described previously [7]. After induction, the cultures were grown for a further 48 hours at 22°C in the same medium with or without individually exogenously supplied fatty acid substrates (LA and ALA, 250μM) and 1% Tergitol Type NP-40 (Sigma).

Fatty Acid Analysis

Total fatty acids were extracted from leaves of *Arabidopsis* and yeast cells and transmethylated with methanolic HCl according to Browse et al [42]. Fatty acid methyl esters (FAMES) were analyzed by gas chromatography (GC) on a 25-m×0.25-mm fused silica CP-Wax 52CB capillary column (Chrompack UK Ltd, London, UK), using a split/splitless injector (230°C, split ratio 50:1) and a flame ionization detector (300°C). After an initial temperature of 170°C for 3 min, the column was temperature-programmed at 4°C min⁻¹ to 220°C and then was held at 220°C for 15 min. Hydrogen carrier gas at an initial flow rate of 1ml min⁻¹ was used. FAMES were identified by comparing to retention time with known standards.

Results

Codon Optimization of the Δ9 Elongase Gene *IgASE1* from *I. galbana* for Expression in *Arabidopsis*

The Δ9 elongase is the first enzyme in the Δ9 elongation/Δ8 desaturation pathways; therefore, it plays a decisive role in the final yield of EPA/DHA. We first analyzed the codon preference of the microalgal Δ9 elongase gene, *IgASE1* [7] for expression in *Arabidopsis*. We found that the three arginine-encoding CGC_{Arg} codons at positions 10, 35 and 84 are the least preferred codons used by *Arabidopsis* having a predicted codon usage efficiency of only 20% (S1 Fig). This would be expected to lead to lower expression levels of the Δ9 elongase in this host plant and hence result in a reduced final yield of AA/EPA. After comparing the usage frequency of all six arginine-encoding synonymous codons by *Arabidopsis* we found the frequencies to be 100% for AGA, 57% for AGG, 49% for CGT, 34% for CGA, 26% for CGG, and 20% for CGC, respectively. Thus, we altered the 3 CGC codons to AGA codons individually in the first instance (single codon optimization). Next, we made two-site codon optimized combinations of R10^{OP}+R35^{OP}, R10^{OP}+R84^{OP}, R35^{OP}+R84^{OP}, and also triple codon optimization of R10^{OP}+R35^{OP}+R84^{OP} (Fig 1A). In addition, we optimized 11 codons within the first 15 codons (N^{OP}) at the N-terminus to their high usage synonymous codon equivalents where their predicted usage frequency by *Arabidopsis* is 100% (“ATG gcT ctT gcT aaT gaT gcT gga gaA AgA atT tgg gcT gct gtT” replacing “ATG GCC CTC GCA AAC GAC GCG GGA GAG CGC ATC TGG GCG GCT GTG” replacing “ATG GCC CTC GCA AAC GAC GCG GGA GAG CGC ATC TGG GCG GCT GTG”, Fig 1B).

All 8 codon-optimized *IgASE1* constructs and the WT sequence were cloned in the plant binary vector pCambia2300EC and transferred into WT *Arabidopsis* plants. Transgenic plants were selected on kanamycin-containing nutrient media agar plates. Total fatty acids from leaves of ten randomly selected mature plants carrying each construct were analyzed. *IgASE1*

(A) N-terminal optimisation of *IgASE1* (*IgASE1*-N^{OP})

AA position	+1	+2	+3	+4	+5	+6	+7	+8	+9	+10	+11	+12	+13	+14	+15
AA seq.	M	A	L	A	N	D	A	G	E	R	I	W	A	A	V
WT DNA seq.	ATG	GCC	CTC	GCA	AAC	GAC	GCG	GGA	GAG	CGC	ATC	TGG	GCG	GCT	GTG
		↓	↓	↓	↓	↓	↓		↓	↓	↓	↓	↓		↓
Optimized DNA seq.	ATG	GCT	CTT	GCT	AAT	GAT	GCT	GGA	GAA	AGA	ATT	TGG	GCT	GCT	GTT

(B) CGC_{Arg} optimisation for *IgASE1* at positions of 10, 35 and 84 (R10^{OP}, R35^{OP}, and R84^{OP})

AA position	+9	+10	+11	+35	+36	+37	+83	+84	+85	...
AA seq.	E	R	I	R	N	S	L	R	R
WT DNA seq.	GAG	CGC	ATC	CGC	AAT	TCC	CTG	CGC	AGG	...
			↓	↓		↓	↓				↓	↓	
Optimized DNA seq.	GAG	AGA	ATC	AGA	AAT	TCC	CTG	AGA	AGG	...

Fig 1. Positions of nucleotides that were changed for codon optimization of *IgASE1*. (a) Codon changes in the first 15 amino acids at the N-terminus; (b) Codon changes of CGC_{Arg} at positions 10, 35 and 84.

doi:10.1371/journal.pone.0158103.g001

catalyzes the conversion of LA to EDA ($\omega 6$ pathway) and ALA to ETrA ($\omega 3$ pathway) in yeast [7] and in *Arabidopsis* [10,43]. The conversion rate [conversion rate = $\text{product}/(\text{product}+\text{substrate})\times 100$] of LA to EDA ($\omega 6$ pathway), and ALA to ETrA ($\omega 3$ pathway) was analyzed. We found that the majority of WT-IgASE1-expressing transgenic plants had less than 10% conversion rate for $\omega 3$ and 30% for $\omega 6$ substrate fatty acids and none of them achieved more than 15% fatty acid conversion rate for $\omega 3$ or 40% for the $\omega 6$ pathway (Fig 2). However, all the codon-optimized variants had improved expression levels of *IgASE1*, and as a result increased production of EDA and ETrA in transgenic *Arabidopsis* (Fig 2). For example, expressing N^{OP}-*IgASE1* resulted in 6 out of the 10 transgenic lines converting $\geq 20\%$ of ALA to ETrA in the $\omega 3$ pathway, and 7/10 transgenic lines converting $\geq 45\%$ of LA to EDA in the $\omega 6$ pathway. Changing all 3 CGC_{Arg} to AGA_{Arg} codons also resulted in higher numbers of plants having fatty acid conversion rates of $\geq 20\%$ (5/10 for the $\omega 3$) and $\geq 45\%$ (4/10 for the $\omega 6$ pathway). Interestingly, optimizing 2 CGC and single CGC codons at different positions only increased the expression level of *IgASE1* slightly (Fig 2).

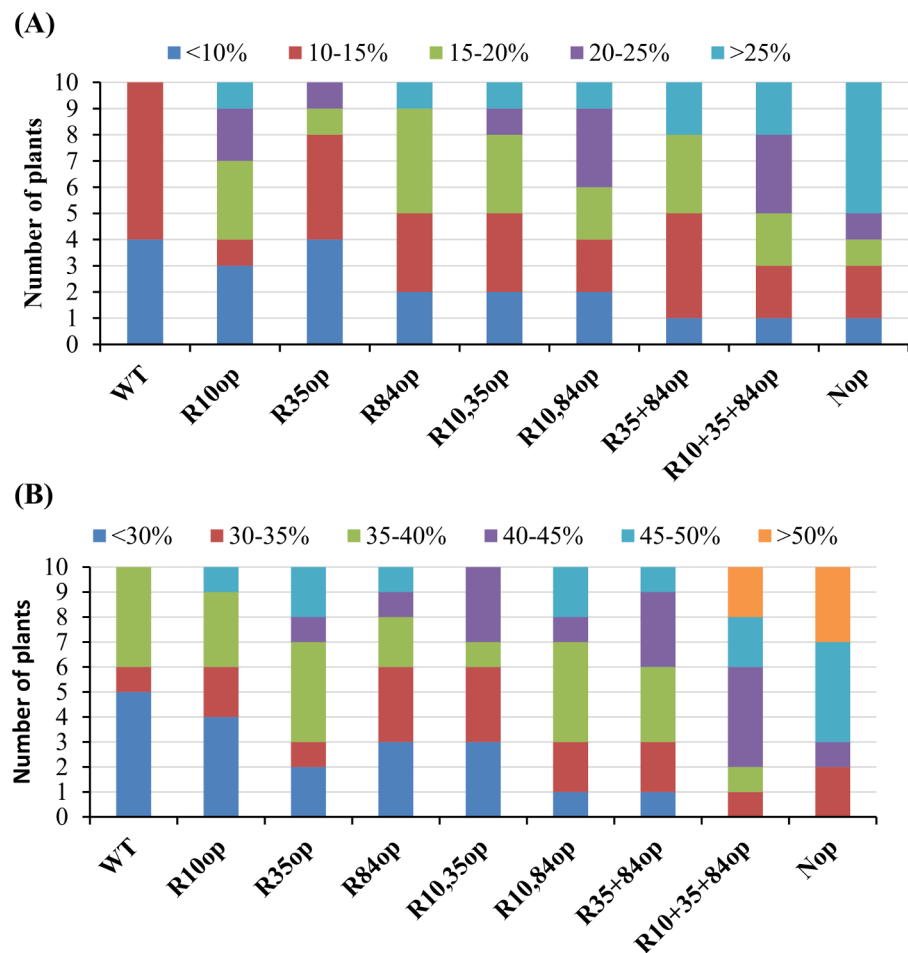


Fig 2. Effect of codon optimization of *IgASE1* on fatty acid conversion in transgenic *Arabidopsis*. (a) $\omega 3$ fatty acid conversion rates in transgenic *Arabidopsis* expressing WT and the codon optimized *IgASE1*; (b) $\omega 6$ fatty acid conversion rates in transgenic *Arabidopsis*. Number of transgenic plants per 10 transgenics having different fatty acid conversion rates are shown. Conversion rate of $\omega 3$ (ALA to ETrA) or $\omega 6$ (LA to EDA) fatty acids are calculated as: $(\text{product}/\text{product}+\text{substrate})\times 100$.

doi:10.1371/journal.pone.0158103.g002

Table 2. Effect of codon optimization of *IgASE1* on fatty acid composition in transgenic *Arabidopsis*.

Fatty Acid (mol% total FAs)	Plant source			
	WT <i>Arabidopsis</i>	Transgenics		
		WT- <i>IgASE1</i>	N ^{OP} - <i>IgASE1</i>	R ^{10,35,84OP} - <i>IgASE1</i>
16:0	15.9±0.12	16.4±0.13	16.2±0.12	15.5±0.14
16:1	3.9±0.03	3.3±0.04	3.1±0.03	3.2±0.04
16:3	13.8±0.20	13.5±0.19	14.1±0.13	14.4±0.15
18:0	1.0±0.03	1.1±0.04	1.5±0.02	1.2±0.03
18:1	3.7±0.15	2.5±0.14	3.5±0.11	2.8±0.13
18:2n-6 (LA)	15.9±0.16 ^a	8.3±0.17 ^b	5.5±0.08 ^c	6.0±0.05 ^c
18:3n-3 (ALA)	45.8±0.30 ^d	41.6±0.20 ^d	32.4±0.15 ^e	35.6±0.20 ^e
20:2n-6 (EDA)	-	5.1±0.20 ^f	9.9±0.10 ^g	8.9±0.11 ^g
20:3n-3 (ETrA)	-	8.2±0.20 ^h	13.8±0.14 ⁱ	12.4±0.12 ⁱ
Total C20 FAs	0	13.5	23.7	21.3
ω-6% FA conversion	-	-	64.3	59.7
ω-3% FA conversion	-	-	30.0	25.8

The values given are expressed as mol % of total fatty acid methyl esters identified by gas-liquid chromatography (GC). Total fatty acids were extracted from rosette stage leaves of WT *Arabidopsis* (Col-0 ecotype), or transgenic *Arabidopsis* expressing the WT, N-terminal optimised (N^{OP}), or all three CGC_{Arg} (R^{10,35,84OP}) optimized *IgASE1* variants. Three higher C20 fatty acid producing lines that harbour a single copy of the transgene were taken to T₃ generation and the homozygous plants were isolated. These were used for total fatty acid analysis. The % conversion for ω-6 fatty acids is calculated as: (EDA/EDA+LA)×100. Likewise, the % conversion for ω-3 fatty acids is calculated as: (ETrA/ETrA+ALA)×100.

Each value represents the mean ± standard deviation from measurements of three plants. Different letters indicate statistically different values after one-way ANOVA.

doi:10.1371/journal.pone.0158103.t002

A few transgenic plant lines harboring all 3 CGC_{Arg}, or the N-terminus-optimized *IgASE1* constructs are among the highest producers of the elongated C20 PUFAs (Table 2). For example, while the WT-*IgASE1* expressing transgenics produced 5.1 mol% total fatty acids of EDA (LA to EDA conversion rate was 38.1%), plants containing N^{OP}- and 3R^{OP}-*IgASE1* produced nearly twice as much EDA compared to WT-*IgASE1* expressing plants at about 9.9 and 8.9 mol% total FAs, and the conversion rate for LA to EDA was 64.3% and 59.7%, respectively. These plants also produced ETrA at 13.8 and 12.4 mol% total FAs compared to only 8.2 mol% total FAs in WT-*IgASE1* expressing plants (conversion rate of ALA to ETrA was 30.0% and 25.8% compared to 16.5% in WT) (Table 2). That is an increase of 1.7- and 1.5-fold of ETrA in N^{OP} and 3R^{OP} respectively, compared to WT-*IgASE1* expressing plants. The total C20 PUFAs (EDA+ETrA) accounted for 23.7 (N^{OP}-*IgASE1* transgenics) and 21.3% mol (3R^{OP}-*IgASE1* transgenics) compared to only 13.5% mol total FAs in WT-*IgASE1* expressing plants, an increase of 64% and 61%, respectively (Table 2).

Codon Optimization of *P. infestans* Δ6 desaturase Genes for Expression in Yeast

To test if the optimization methodology we developed for expression of *IgASE1* in plant also applies to other genes involved in the biosynthesis of VLC-PUFAs in a different host we next optimized the fungal *P. infestans* Δ6 desaturase PinD6, for expression in yeast. Delta6 desaturase is involved in the first step in the Δ6 desaturation pathway for the biosynthesis of VLC-PUFAs. It converts LA to GLA and ALA to SDA by adding a double bond at the Δ6 position of their hydrocarbon chains. Previously we reported that the activity of PinD6 was very low (with less than 5% substrate conversion rate) when expressed in yeast [36]. Codon usage analysis of

PinD6 revealed that there are four arginine-encoding CGC codons at positions of R9, R276, R376 and R433, respectively (S2 Fig). These codons are predicted to be the least preferred by yeast with the lowest expression efficiency being 13%. As for the AGA codon preference for arginine in *Arabidopsis*, yeast also has the strongest preference for this codon where 100% expression efficiency is predicted to be achieved (Fig 3A). Because the highest yield of EDA and ETrA was achieved for codon optimized IgASE1 in which all 3 CGC codons were changed to AGA codons in *Arabidopsis* we decided to mutate all four CGC codons of *PinD6* to AGA codons (4R^{OP} *PinD6*). Similarly, we also optimized the less preferred 14 codons within the first 16 codons at the N-terminus of *PinD6* (ATG GTG GAC GGC CCC AAG ACC AAG CGC AAG ATC TCG TGG CAG GAG GTC) to their synonymous codons (ATG GTT GAT GGT CCA AAA ACT AAA AGA AAA ATT TCT TGG CAA GAA GTT), where their predicted codon usage frequency in yeast is 100%. In addition, we generated a third DNA fragment containing all 4 AGA arginine encoding codons, plus the N-terminal 14 optimized codons to obtain 4R^{OP}+N^{OP} *PinD6* (Fig 3A and 3B).

The WT-*PinD6* and its codon-optimised variants (4 arginine optimised, 4R^{OP}; N-terminal 14 codon optimised, N^{OP}; combined 4 arginine optimised plus the N-terminal 14 codon optimised, 4R^{OP}+N^{OP}) were all cloned in the yeast expression vector pYES2 and transformed into yeast. The enzyme activities of the WT and the codon optimized *PinD6* were monitored by feeding the transgenic yeast cells with the substrate fatty acids LA and ALA in the presence of 2% galactose to induce protein expression. Consistent with our previous data [36] it was found that the WT-*PinD6* can convert LA and ALA to GLA and SDA, with each accounting for 0.8 mol% of total fatty acids with a conversion rate of 2.7% and 1.9% for LA and ALA respectively. However, expressing the N^{OP}-*PinD6* in yeast yielded GLA and SDA at 9 mol% and 16.9 mol% of the total fatty acids respectively, and the conversion rate was 31.5% and 38.3% for LA and ALA (Fig 3C; Table 3). This represents a 12-fold increase for GLA and a 19-fold increase for SDA production compared to the WT-*PinD6*-expressing yeast. The 4R^{OP}-*PinD6* transformed yeast cells contained the products GLA and SDA at 8.2 mol% and 14.6 mol% of the total fatty acids respectively and the conversion rate was 30.3% and 36.2% for LA and ALA (Fig 3C; Table 3). The fold change was 10.2 for GLA and 18.2 for SDA which was slightly lower than that of N^{OP}*PinD6*. Thus, optimization of all 4 CGC arginine codons significantly improved the enzyme activity of *PinD6*.

To see if the N^{OP} and 4R^{OP} mutations had an additive effect with respect to *PinD6* enzyme activity we next made a construct containing all these mutations. Analysis of the total fatty acid composition of the transgenic yeast cells showed that GLA and SDA accounted for 14.5 mol% and 19.0 mol% respectively and that the conversion rate was 36.5% for LA and 41% for ALA (Table 3; Fig 3C). This is equivalent to a 19.4-fold increase for GLA and a 23.8-fold increase for SDA in these transgenic yeast cells. This clearly demonstrates that the combined effect of N^{OP} and 4R^{OP} mutations on *PinD6* enzyme activity was much higher than the two individually optimized enzymes.

Discussion

The biosynthesis of VLC-PUFAs in a higher plant was first reported in 2004 by Qi et al [10] who introduced the $\Delta 9$ elongation and $\Delta 8$ desaturation pathway into *Arabidopsis*, utilizing one fatty acid elongase and 2 desaturase genes isolated from lower eukaryotes. This was followed by another report where the researchers demonstrated the possibility of producing VLC-PUFAs in seeds of tobacco and linseed via the D6 desaturation pathway [11]. Since then various attempts have been carried out to engineer both pathways into oilseed crops and significant advances have been made [9,13,14]. However, there still remain several challenges that must be

(A) Optimization of N-terminus of PinD6 (PinD6-N^{op})

AA position	+1	+2	+3	+4	+5	+6	+7	+8	+9	+10	+11	+12	+13	+14	+15	+16
AA seq.	M	V	D	G	P	K	T	K	R	K	I	S	W	Q	E	V
WT DNA seq.	ATG	GTG	GAC	GGC	CCC	AAG	ACC	AAG	CGC	AAG	ATC	TCG	TGG	CAG	GAG	GTC
Optimized DNA seq.	ATG	GTT	GAT	GGT	CCA	AAA	ACT	AAA	AGA	AAA	ATT	TCT	TGG	CAA	GAA	GTT

(B) CGC_{Arg} optimisation of PinD6 at positions of 9, 76, 376 and 433 (PinD6-4R^{op})

AA position	...	+8	+9	+10	+75	+76	+77	+375	+376	+377	+432	+433	+434
AA seq.	...	E	R	I	A	R	F	T	R	N	Y	R	G
WT DNA seq.	...	AAG	CGC	AAG	GCT	CGC	TCC	ACG	CGC	AAC	TAC	CGC	GGT
Optimized DNA seq.	...	AAG	AGA	AAG	GCT	AGA	TCC	ACG	AGA	AAC	TAC	AGA	GGT

(C) Comparison of ω6/3 fatty acid conversion rates in transgenic yeast expressing WT and the codon optimized PinD6

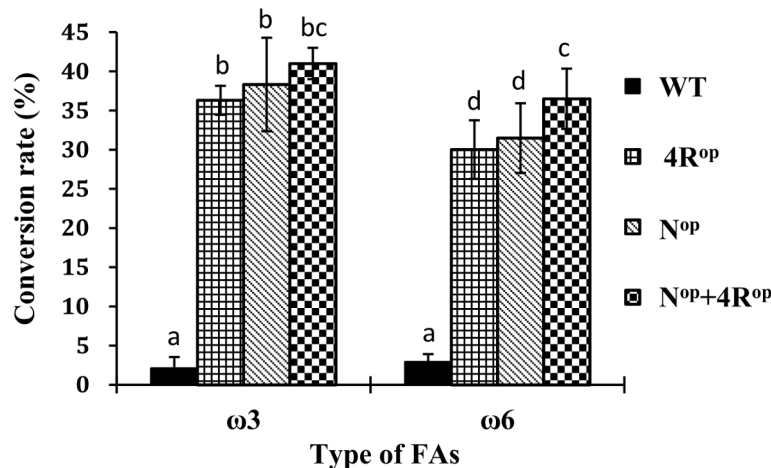


Fig 3. Effect of codon optimization of PinD6 on fatty acid conversion rate in transgenic yeast cells. (a) Codon changes in the first 16 amino acids at the N-terminus; (b) Codon changes of CGC_{Arg} to AGA_{Arg} at positions 9, 76, 376 and 433; (c) Fatty acid conversion rates of WT and codon-optimized variants of PinD6 in transgenic yeast. WT, 4R^{op}, N^{op} and N^{op}+4R^{op} represents transgenic yeast harboring the wild-type, four CGC to AGA-optimized, N-terminal 16 codons optimized and the N-terminal 16 codons plus the four CGC to AGA optimized PinD6, respectively. Each value represents the mean ± standard deviation from three independent repeats. Different letters indicate statistically different values after one-way ANOVA.

doi:10.1371/journal.pone.0158103.g003

met before engineered oilseed crops can be introduced for field production. Amongst these the accumulation of adequate levels of VLC-PUFA is the main hurdle.

The expression of multiple codon-optimized transgenes derived from lower eukaryotic organisms involved in the biosynthesis of EPA and DHA has resulted in high levels of these

Table 3. Fatty acid analysis of WT and transgenic yeast cells expressing codon-optimized variants of PinD6.

Fatty acid	Mol% of total fatty acids								
	WT yeast	+gal +ALA (18:3n-3)				+gal +LA (18:2n-6)			
		WT-D6	4R ^{OP}	N ^{OP}	N ^{OP} +4R ^{OP}	WT-D6	4R ^{OP}	N ^{OP}	N ^{OP} +4R ^{OP}
16:0	29.4±3.0	20.2±0.5	22.2±3.5	22.5±2.6	20.7±2.4	22.0±1.5	25.2±2.7	25.3±1.6	21.6±3.5
16:1	45.3±3.6	22.4±1.1	20.8±4.2	15.4±1.6	13.2±0.3	29.1±0.3	26.6±5.9	25.1±0.8	19.2±1.9
18:0	6.3±0.5	4.9±0.8	5.5±1.5	7.0±1.6	8.7±0.6	5.5±1.0	5.6±0.6	6.7±0.7	7.8±0.7
18:1	19.1±1.3	10.7±0.9	11.2±2.2	11.0±0.9	11.1±0.5	14.3±1.8	15.5±1.9	14.3±0.2	11.7±2.2
18:2n-6 (LA)	-	-	-	-	-	28.3±2.8 ^a	18.9±6.9 ^b	19.6±2.9 ^b	25.2±1.1 ^a
18:3n-6 (GLA)	-	-	-	-	-	0.8±0.1 ^c	8.2±2.4 ^d	9.0±1.2 ^d	14.5±2.4 ^b
18:3n-3 (ALA)	-	41.0±1.6 ^e	25.7±6.2 ^a	27.2±5.1 ^a	27.3±1.9 ^a	-	-	-	-
18:4n-3 (SDA)	-	0.8±0.2 ^c	14.6±2.9 ^b	16.9±2.0 ^b	19.0±1.3 ^b	-	-	-	-
% conversion	-	1.9	36.2	38.3	41	2.7	30.3	31.5	36.5

The values given are expressed as mol % of total fatty acid methyl esters identified by GC. In the case of desaturated substrates, this is also expressed as % conversion (product/(product + substrate) x100). All values are the means of triplicates from three separate experiments. WT-D6, 4R^{OP}, N^{OP} and N^{OP}+4R^{OP} represents transgenic *S. cerevisiae* expressing WT PinD6, all four CGC_{arg}-optimized, N-terminal 16 codons-optimized and combination of both N-terminal 16 codons and four CGC_{arg} optimized PinD6. Each value represents the mean ± standard deviation from three independent repeats. Different letters indicate statistically different values after one-way ANOVA.

doi:10.1371/journal.pone.0158103.t003

fatty acids being achieved in higher plants [44,45,46]. However, in these examples the transgenes were synthesized with the majority of their codons being changed to match the preferred codons used by the respective hosts. A minimum of two desaturation steps and an elongation step are needed for the production of EPA. In the case of DHA a further round of elongation and desaturation is required. The typical length of a fatty acid elongase is just under 1 kb and that of a desaturase is ~1.5 kb. Therefore, ~4,000–6,500 bps of nucleotides for host production of EPA or DHA have to be synthesized. This is a very costly option that most research labs cannot afford. In addition, the specific nucleotide arrangement in a gene sequence is important for its function, thus too many codon changes may alter this dynamic and result in a less active enzyme. Therefore, we aimed to find a codon optimization strategy where minimal changes to the nucleotide sequence of the transgene are made, but with the outcome that high enzyme activity is achieved, for the production of VLC-PUFAs.

In this study, we optimized two key genes, a Δ9 elongase gene from the microalga *I. galbana* [7] and a Δ6 desaturase gene from the fungus *P. infestans* [36]. We chose these two genes because they catalyze the first steps in the Δ9 elongation/Δ8 desaturation and the Δ6 desaturation/Δ6 elongation pathways, respectively. Therefore, their activities will have significant impact on the subsequent steps in both pathways for the biosynthesis of VLC-PUFAs.

We first optimized the Δ9 elongase for expression in *Arabidopsis* by changing the least preferred Arg-encoding CGC codon (CGC_{Arg}) that has a predicted usage frequency of only 20% in *Arabidopsis*. There are 3 such codons in the *IgASE1* ORF (S1 Fig). Through PCR-based site-directed mutagenesis, we changed all three to AGA_{Arg} codons which have a predicted usage frequency of 100% in *Arabidopsis*. As there were 3 CGC_{Arg} codon sites we optimized each one of these individually as well as producing different combinations of them in order to find the site(s) that were responsible for low *IgASE1* expression levels. We found that the expression level of the Δ9 elongase gene variants having just one CGC optimized codon (at either position +10, +35 and +84) or a pair of optimized codons (positions 10+35, 10+84, or 10+84) were only slightly higher than the WT Δ9 elongase control (Fig 2). However, when all these 3 CGC

codons were mutated to AGA simultaneously the enzyme activity of this variant was much higher than that of the WT IgASE1 and the other single and double AGA_{Arg} IgASE1 versions.

We also altered 11 out of the first 16 N-terminal codons to their frequently used *Arabidopsis* counterparts. Our data show that the N terminal codon-optimized IgASE1 had the highest enzyme activity compared to the WT- and all the AGA_{Arg}-IgASE1 variants (Fig 2; Table 2). Conveniently, this construct is also the easiest to produce as only one round of PCR is required to incorporate all the changes in this region. Thus, the best approach to achieve high levels of expression with minimal effort for this $\Delta 9$ elongase gene in *Arabidopsis* is to optimize all 11 N-terminal codons—this strategy resulted in lines producing up to 64% more C20 PUFAs than the WT-IgASE1 (Fig 2).

Codon optimization studies have predominately been carried out in *E. coli* where genes originating from humans and a range of other organisms have been heterologously expressed. For example, Burgess-Brown *et al* [47] achieved a higher expression efficiency of 30 human genes by optimizing the most biased codons. Similarly by adding rare codon tRNAs in cell lines high expression efficiency of these genes was achieved. Therefore, the availability of rare tRNAs seems to be the limiting factor in reduced protein expression [47]. Two of the six arginine-encoding codons, AGG and AGA, are amongst the rarest codons used by *E. coli*, and many studies demonstrated that they frequently lead to no, or low, protein expression of foreign genes. Therefore various codon alteration strategies for improving the yields of proteins expressed in *E. coli* involve the alteration of these two arginine-encoding codons. Consistent with this, we also found that one of the arginine-encoding codons, CGC, in IgASE1 is the most biased codon used by *Arabidopsis* where only 20% usage frequency is predicted (S1 Fig). By changing the 3 CGC_{Arg} codons in IgASE1 one-by-one to the favored *Arabidopsis* AGA_{Arg} codon we found higher levels of C20 FA production correlated with increasing numbers of optimised AGA_{Arg} codons. The positions of the CGC_{Arg} did not seem to play a significant role in this effect (Fig 2). Thus it appears that the total number of CGC_{Arg} codons, rather than their positions, contributes to lower IgASE1 activity (C20 FA yield) in transgenic *Arabidopsis* plants.

Changing all 11 out of the first 15 codons to their favorable *Arabidopsis* counterparts achieved even higher levels of C20 FAs than that obtained by changing all three CGC_{Arg} codons to AGA_{Arg} in transgenic *Arabidopsis* (Fig 2). Studies have shown that in *E. coli* if rare codons occur near the 5' end of an ORF, especially as clusters, the effect can be detrimental, resulting in very low to no expression of foreign genes [20,48,49,50,51,52]. By inserting the rarest codon AGA_{Arg} at different positions near the 5' end of a gene Kim and Lee [20] reported that positioning at the +2 and +3 positions had the most negative effect. As the position of an AGA_{Arg} codon moves further towards to the 3' end of the gene its effect becomes minimal. In IgASE1 the first arginine-encoding codon CGC is positioned at +10, however changing this codon to the favored AGA codon used by *Arabidopsis* resulted in only a very small enhancement in its activity (Fig 2). Therefore, the marked increase of enzyme activity achieved by optimizing all the 11 codons downstream of the ATG codon is very unlikely to be due to the mutation of CGC_{Arg} at the +10th amino acid position alone. There are also 3 low-usage alanine-encoding codons at the +2nd, +7th and +13th amino acid positions where only 37%, 33% and 33% usage frequencies are predicted to be used by *Arabidopsis* (S1 Fig). Therefore, we propose that these 3 alanine codons and the CGC_{Arg} at the +10th position could form a rare codon cluster at the N-terminus of IgASE1 which could result in lower expression of WT-IgASE1 and hence lower EDA and ETrA yields in *Arabidopsis* (Fig 2; Table 2).

In order to further verify the utility of the codon optimization strategy taken for the *I. galbana* $\Delta 9$ elongase gene, we codon-optimized a $\Delta 6$ desaturase gene we isolated from *P. infestans* (PinD6) which exhibited very low enzyme activity when expressed in yeast [36]. Based on the results obtained from the IgASE1, we made 3 constructs containing: i) 14 optimized N-terminal

codons, ii) 4 optimized highly biased CGC_{Arg} codons, and iii) a combination of all these mutations. We found that the enzyme activities of both N^{OP} and 4R^{OP} variants were significantly elevated compared to the WT, with the N-terminal-optimised $\Delta 6$ desaturase having yields higher than that of the variant containing all 4 CGC_{Arg} codons being optimized to AGA_{Arg}. This was consistent with the result obtained from codon optimization and expression of IgASE1 in *Arabidopsis*, hence further confirming the usefulness of our codon optimization strategies. Interestingly we also found that when both the N^{OP} and 4R^{OP} mutations were combined the enzyme activity of PinD6 was dramatically increased (Table 3; Fig 3C). Therefore, the N^{OP} and 4R^{OP} have an additive effect for PinD6 enzyme activity.

The yeast expression system has been used for verifying the functions of many eukaryotic genes. For example, most of the VLC-PUFA desaturase and elongase genes isolated from lower eukaryotic organisms were functionally characterized by exogenously expressing them in Baker's yeast. However, such genes have frequently shown very low [36, 53] or sometimes no (Qi and Lazarus, unpublished) activity when expressed in yeast, making the verification of their enzyme activities difficult or impossible. The fact that the desaturated fatty acid products of PinD6 could be increased more than 10-fold by simply optimizing 14 of its N-terminal codons for tailored expression in yeast highlights the importance of codon bias to protein expression. Of course, the lower activity of this delta-6 desaturase in yeast made the verification of our codon optimization strategy more convincing. Importantly, this N-terminal codon optimization strategy could be extended to the functional characterization of heterologous fatty acid desaturase and elongase genes in yeast.

Through codon optimization of two key enzymes for VLC-PUFA biosynthesis we provide further evidence that different genetic codon preference exists between organisms leading to lower enzyme activities following their heterologous expression. Both the N-terminal codons and rare codons across the whole ORF contribute to the low enzyme activities found for the two enzymes studied. Thus the best codon optimization strategy, based on our findings, is to optimize the first 16 codons at the N-terminus as well as all the most biased codons in the entire ORF of a given gene. Future studies will be focused on codon optimization of other desaturase and elongase genes involved in EPA and DHA biosynthetic pathways in order to enhance the VLC-PUFAs content in transgenic oilseed crops.

Supporting Information

S1 Fig. Prediction of frequency of codon usage of IgASE1 when expressed in *Arabidopsis thaliana*. The columns in red are the mostly rare codons used by *Arabidopsis thaliana*. The height of the column represents the frequency of codon usage. The 3 CGC codons at positions 10, 20 and 84 are marked in red arrows.
(PPTX)

S2 Fig. Prediction of frequency of codon usage of PinD6 when expressed in *S. cerevisiae*. The height of the column represents the frequency of codon usage. The rarest 4 CGC codons at positions 9, 276, 376 and 433 are marked in red arrows and cluster of 2 codons with less than 30% usage are marked with red stars.
(PPTX)

Acknowledgments

This work was supported by the National Natural Science Foundation of China (Grant No. 30970222 to BQ) and Genetically Modified Organisms Breeding Major Projects of China

(Grant No. 009ZX08005-024B to BQ). We thank Dr. James Doughty, University of Bath, for assisting with editing the manuscript.

Author Contributions

Conceived and designed the experiments: BQ Xueying Li. Performed the experiments: Xueying Li FX JL QS DZ YL. Analyzed the data: Xueying Li Xinzheng Li FX JL JH. Contributed reagents/materials/analysis tools: BQ. Wrote the paper: BQ FX.

References

1. Simopoulos AP. Importance of the ratio of omega-6/omega-3 essential fatty acids: evolutionary aspects of diet. *World Rev Nutr Diet.* 2004; 102:10–21.
2. Lauritzen L, Hansen HS, Jørgensen MH, Michaelsen KF. The essentiality of long chain n-3 fatty acids in relation to development and function of the brain and retina. *Prog Lipid Res.* 2001; 40:1–94. PMID: [11137568](#)
3. Trautwein E. n-3 Fatty acids-physiological and technical aspects for their use in food. *Eur J Lipid Sci Technol.* 2001; 103:45–55.
4. Sayanova OV, Napier JA. Eicosapentaenoic acid: biosynthetic routes and the potential for synthesis in transgenic plants. *Phytochemistry.* 2004; 65:147–58. PMID: [14732274](#)
5. Metz JG, Roessler P, Facciotti D, Levering C, Dittrich F, Lassner M, et al. Production of polyunsaturated fatty acids by polyketide synthases in both prokaryotes and eukaryotes. *Science.* 2001; 293(5528):290–3. PMID: [11452122](#)
6. Uttaro AD. Biosynthesis of polyunsaturated fatty acids in lower eukaryotes. *IUBMB life.* 2006; 58:563–71. PMID: [17028087](#)
7. Qi B, Beaudoin F, Fraser T, Stobart AK, Napier JA, Lazarus CM. Identification of a cDNA encoding a novel C18-Δ9 polyunsaturated fatty acid-specific elongating activity from the docosahexaenoic acid (DHA)-producing microalga, *Isochrysis galbana*. *FEBS Lett.* 2002; 510:159–65. PMID: [11801246](#)
8. Cheng B, Wu G, Vrinten P, Falk K, Bauer J, Qiu X. Towards the production of high levels of eicosapentaenoic acid in transgenic plants: the effects of different host species, genes and promoters. *Transgenic Res.* 2010; 19:221–9. doi: [10.1007/s11248-009-9302-z](#) PMID: [19582587](#)
9. Truksa M, Wu G, Vrinten P, Qiu X. Metabolic engineering of plants to produce very long-chain polyunsaturated fatty acids. *Transgenic Res.* 2006; 15:131–7. PMID: [16604455](#)
10. Qi B, Fraser T, Mugford S, Dobson G, Sayanova O, Butler J, et al. Production of very long chain polyunsaturated omega-3 and omega-6 fatty acids in plants. *Nat Biotechnol.* 2004; 22:739–45. PMID: [15146198](#)
11. Abbadi A, Domergue F, Bauer J, Napier JA, Welti R, Zähringer U, et al. Biosynthesis of very-long-chain polyunsaturated fatty acids in transgenic oilseeds: constraints on their accumulation. *The Plant Cell Online.* 2004; 16:2734–48.
12. Domergue F, Abbadi A, Heinz E. Relief for fish stocks: oceanic fatty acids in transgenic oilseeds. *Trends Plant Sci.* 2005; 10:112–6. PMID: [15749468](#)
13. Wu G, Truksa M, Datla N, Vrinten P, Bauer J, Zank T, et al. Stepwise engineering to produce high yields of very long-chain polyunsaturated fatty acids in plants. *Nat Biotechnol.* 2005; 23:1013–7. PMID: [15951804](#)
14. Petrie JR, Shrestha P, Zhou XR, Mansour MP, Liu Q, Belide S, et al. Metabolic engineering plant seeds with fish oil-like levels of DHA. *PLoS One.* 2012; 7:e49165. doi: [10.1371/journal.pone.0049165](#) PMID: [23145108](#)
15. Gouy M, Gautier C. Codon usage in bacteria: correlation with gene expressivity. *Nucleic Acids Res.* 1982; 10:7055–74. PMID: [6760125](#)
16. Kane JF. Effects of rare codon clusters on high-level expression of heterologous proteins in *Escherichia coli*. *Curr Opin Biotechnol.* 1995; 6:494–500. PMID: [7579660](#)
17. Deng T. Bacterial expression and purification of biologically active mouse c-Fos proteins by selective codon optimization. *FEBS Lett.* 1997; 409:269–72. PMID: [9202159](#)
18. Gustafsson C, Govindarajan S, Minshull J. Codon bias and heterologous protein expression. *Trends Biotechnol.* 2004; 22:346–53. PMID: [15245907](#)
19. Frelin L, Ahlen G, Alheim M, Weiland O, Barnfield C, Liljeström P, et al. Codon optimization and mRNA amplification effectively enhances the immunogenicity of the hepatitis C virus nonstructural 3/4A gene. *Gene Ther.* 2004; 11:522–33. PMID: [14999224](#)

20. Kim S, Lee SB. Rare codon clusters at 5'-end influence heterologous expression of archaeal gene in *Escherichia coli*. *Protein Expr Purif*. 2006; 50:49–57. PMID: [16962338](#)
21. Johansson A-S, Bolton-Grob R, Mannervik B. Use of silent mutations in cDNA encoding human glutathione transferase M2-2 for optimized expression in *Escherichia coli*. *Protein Expr Purif*. 1999; 17:105–12. PMID: [10497075](#)
22. Laguna-Becher M, Martín V, Kraemer M, Corigliano M, Yacono ML, Goldman A, et al. Effect of codon optimization and subcellular targeting on *Toxoplasma gondii* antigen SAG1 expression in tobacco leaves to use in subcutaneous and oral immunization in mice. *BMC Biotechnol*. 2010; 10:52. doi: [10.1186/1472-6750-10-52](#) PMID: [20633272](#)
23. Stenøien HK. Adaptive basis of codon usage in the haploid moss *Physcomitrella patens*. *Heredity (Edinb)*. 2004; 94:87–93.
24. Mirzahoseini H, Mafakheri S, Mohammadi NS, Enayati S, Mortazavidehkordi N. Heterologous proteins production in escherichia coli: an investigation on the effect of codon usage and expression host optimization. *Cell Journal(Yakhteh)*. 2011; 12:453–8.
25. Goldman E, Rosenberg AH, Zubay G, Studier WF. Consecutive low-usage leucine codons block translation only when near the 5' end of a message in *Escherichia coli*. *J Mol Biol*. 1995; 245:467–73. PMID: [7844820](#)
26. Griswold KE, Mahmood NA, Iverson BL, Georgiou G. Effects of codon usage versus putative 5'-mRNA structure on the expression of *Fusarium solani* cutinase in the *Escherichia coli* cytoplasm. *Protein Expr Purif*. 2003; 27:134–42. PMID: [12509995](#)
27. Sinclair G, Choy FY. Synonymous codon usage bias and the expression of human glucocerebrosidase in the methylotrophic yeast, *Pichia pastoris*. *Protein Expr Purif*. 2002; 26:96–105. PMID: [12356476](#)
28. Wang BQ, Lei L, Burton ZF. Importance of codon preference for production of human RAP74 and reconstitution of the RAP30/74 complex. *Protein Expr Purif*. 1994; 5:476–85. PMID: [7827505](#)
29. Itakura K, Hirose T, Crea R, Riggs AD, Heyneker HL, Bolivar F, Boyer HW Expression in *Escherichia coli* of a chemically synthesized gene for the hormone somatostatin. 1977; *Science* 198: 1056–1063. PMID: [412251](#)
30. Kink JA, Maley ME, Ling KY, Kanbrooki JA, Kung C. Efficient expression of the *Paramecium calmodulin* gene in *Escherichia coli* after four TAA-to-CAA changes through a series of polymerase chain reactions. *J Protozool*. 1991; 38:441–7. PMID: [1920142](#)
31. Yadava A, Ockenhouse CF. Effect of codon optimization on expression levels of a functionally folded malaria vaccine candidate in prokaryotic and eukaryotic expression systems. *Infect Immun*. 2003; 71:4961–9. PMID: [12933838](#)
32. Outchkourov NS, Stiekema WJ, Jongsma MA. Optimization of the expression of equistatin in *Pichia pastoris*. *Protein Expr Purif*. 2002; 24:18–24. PMID: [11812218](#)
33. Hu S, Li L, Qiao J, Guo Y, Cheng L, Liu J. Codon optimization, expression, and characterization of an internalizing anti-ErbB2 single-chain antibody in *Pichia pastoris*. *Protein Expr Purif*. 2006; 47:249–57. PMID: [16403645](#)
34. Teng D, Fan Y, Yang Y-I, Tian Z-g, Luo J, Wang J-h. Codon optimization of *Bacillus licheniformis* β -1, 3–1, 4-glucanase gene and its expression in *Pichia pastoris*. *Appl Microbiol Biotechnol*. 2007; 74:1074–83. PMID: [17216453](#)
35. Angov E, Hillier CJ, Kincaid RL, Lyon JA. Heterologous protein expression is enhanced by harmonizing the codon usage frequencies of the target gene with those of the expression host. *PLoS One*. 2008; 3: e2189. doi: [10.1371/journal.pone.0002189](#) PMID: [18478103](#)
36. Sun Q, Liu J, Zhang Q, Qing X, Dobson G, Li X, et al. Characterization of three novel desaturases involved in the delta-6 desaturation pathways for polyunsaturated fatty acid biosynthesis from *Phytophthora infestans*. *Appl Microbiol Biotechnol*. 2013; 97:7689–97. doi: [10.1007/s00253-012-4613-z](#) PMID: [23229570](#)
37. Ho SN, Hunt HD, Horton RM, Pullen JK, Pease LR. Site-directed mutagenesis by overlap extension using the polymerase chain reaction. *Gene*. 1989; 77:51–9. PMID: [2744487](#)
38. Qi B, Fraser T, Bleakley CL, Shaw EM, Stobart AK, Lazarus CM. The variant 'his-box' of the C18- Δ 9-PUFA-specific elongase IgASE1 from *Isochrysis galbana* is essential for optimum enzyme activity. *FEBS Lett*. 2003; 547:137–9. PMID: [12860401](#)
39. Sun Q, Liu J, Li Y, Zhang Q, Shan S, Li X, et al. Creation and validation of a widely applicable multiple gene transfer vector system for stable transformation in plant. *Plant Mol Biol*. 2013; 83:391–404. doi: [10.1007/s11103-013-0096-2](#) PMID: [23839253](#)
40. Clough SJ, Bent AF. Floral dip: a simplified method for *Agrobacterium*-mediated transformation of *Arabidopsis thaliana*. *The plant journal*. 1998; 16:735–43. PMID: [10069079](#)

41. Elble R. A simple and efficient procedure for transformation of yeasts. *BioTechniques*. 1992; 13(1):18. PMID: [1503765](#)
42. Browse J, McCourt PJ, Somerville CR. Fatty acid composition of leaf lipids determined after combined digestion and fatty acid methyl ester formation from fresh tissue. *Anal Biochem*. 1986; 152:141–5. PMID: [3954036](#)
43. Fraser TCM, Qi B, Elhoussein S, Chatrattanakunchai S, Stobart AK, Lazarus CM. Expression of the Isochrysis C18-Δ9 polyunsaturated fatty acid specific elongase component alters *Arabidopsis* glycerolipid profiles. *Plant Physiol*. 2004; 135:859–66. PMID: [15173563](#)
44. Wood CC, Petrie JR, Shrestha P, Mansour MP, Nichols PD, Green AG, et al. A leaf-based assay using interchangeable design principles to rapidly assemble multistep recombinant pathways. *Plant Biotechnol J*. 2009; 7:914–24. doi: [10.1111/j.1467-7652.2009.00453.x](#) PMID: [19843252](#)
45. Ruiz-López N, Haslam RP, Venegas-Calderón M, Li T, Bauer J, Napier JA, et al. Enhancing the accumulation of omega-3 long chain polyunsaturated fatty acids in transgenic *Arabidopsis thaliana* via iterative metabolic engineering and genetic crossing. *Transgenic Res*. 2012; 21:1233–43. doi: [10.1007/s11248-012-9596-0](#) PMID: [22350763](#)
46. Ruiz-Lopez N, Haslam RP, Napier JA, Sayanova O. Successful high-level accumulation of fish oil omega-3 long-chain polyunsaturated fatty acids in a transgenic oilseed crop. *The Plant Journal*. 2014; 77:198–208. doi: [10.1111/tbj.12378](#) PMID: [24308505](#)
47. Burgess-Brown NA, Sharma S, Sobott F, Loenarz C, Oppermann U, Gileadi O. Codon optimization can improve expression of human genes in *Escherichia coli*: A multi-gene study. *Protein Expr Purif*. 2008; 59:94–102. doi: [10.1016/j.pep.2008.01.008](#) PMID: [18289875](#)
48. Chen G, Inouye M. Role of the AGA/AGG codons, the rarest codons in global gene expression in *Escherichia coli*. *Genes Dev*. 1994; 8:2641–52. PMID: [7958922](#)
49. Olivares-Trejo JJ, Bueno-Martínez JG, Guamerros G, Hernández-Sánchez J. The pair of arginine codons AGA AGG close to the initiation codon of the lambda int gene inhibits cell growth and protein synthesis by accumulating peptidyl-tRNAArg4. *Mol Microbiol*. 2003; 49:1043–9. PMID: [12890027](#)
50. Thangadurai C, Suthakaran P, Barfal P, Anandaraj B, Pradhan SN, Ramalingam S, et al. Rare codon priority and its position specificity at the 5' of the gene modulates heterologous protein expression in *Escherichia coli*. *Biochem Biophys Res Commun*. 2008; 376:647–52. doi: [10.1016/j.bbrc.2008.09.024](#) PMID: [18801340](#)
51. Plotkin JB, Kudla G. Synonymous but not the same: the causes and consequences of codon bias. *Nature Reviews Genetics*. 2010; 12:32–42. doi: [10.1038/nrg2899](#) PMID: [21102527](#)
52. Kim MS, Jang JH, Kim YW. Overproduction of a thermostable 4-α-glucanotransferase by codon optimization at N-terminus region. *J Sci Food Agric*. 2013; 93:2683–90. doi: [10.1002/jsfa.6084](#) PMID: [23620355](#)
53. Eiamsa-ard P, Kanjana-Opas A, Cahoon EB, Chodok P, Kaewsuwan S. Two novel *Physcomitrella patens* fatty acid elongases (ELOs): identification and functional characterization. *Appl Microbiol Biotechnol*. 2013; 97:3485–97. doi: [10.1007/s00253-012-4556-4](#) PMID: [23138714](#)