

# SCIENTIFIC REPORTS



OPEN

## Specific Upregulation of a Cotton Phytoene Synthase Gene Produces Golden Cottonseeds with Enhanced Provitamin A

Dan Yao, Yi Wang, Qian Li, Xufen Ouyang, Yaohua Li, Chuannan Wang, Lingli Ding, Lei Hou, Ming Luo & Yuehua Xiao

Provitamin A (PVA) bio-fortification of crops offers a sustainable strategy to prevent the prevalence of vitamin A deficiency (VAD), one of the world's major public health problems. The present work aimed to enhance PVA accumulation in cottonseed, the main by-product in the production of cotton fibers and the third largest source of edible plant oil in the world. On the basis of comprehensive identification of carotenoid synthase genes and their expression levels in various cotton tissues, we selected phytoene synthase as the target for manipulating carotenoid biosynthesis in the developing cottonseeds. After functional verification in transgenic tobacco, a cotton phytoene synthase gene (*GhPSY2D*) driven by a seed-specific promoter was transformed into cotton. The transgenic cottonseeds showed golden appearance and contained over 6-fold higher carotenoid contents in the extracted oil than the non-transgenic control. Thin layer chromatograph analysis indicated that the main PVA carotenoid  $\beta$ -carotene was predominant in the transgenic cottonseeds, but undetectable in the wild-type control. By simultaneously providing economically valuable fibers and edible oils, the transgenic cottons bio-fortified with  $\beta$ -carotene in seeds may be a new powerful tool against VAD in low-income regions.

Vitamin A (retinol) is an essential nutrient for vertebrates, including humans. Vitamin A deficiency (VAD) may result in a series of disorders in animals, including impaired growth, reproduction, epithelial integrity, and disease resistance<sup>1</sup>. In human, this nutrient deficiency causes xerophthalmia, including night blindness, and increases the risk of infant morbidity and mortality from measles and diarrhoea in children<sup>2</sup>. VAD has been one of major human health problems for a long time. Although the overall prevalence of worldwide VAD has significantly decreased in the last two decades, this deficiency is still unacceptably serious in some regions, such as south Asia and sub-Saharan Africa<sup>3,4</sup>. This may be largely attributed to insufficient dietary diversification, unsuccessful food fortification, and the restricted vitamin A capsule delivery in these regions<sup>3</sup>. Compared to the supply of vitamin A capsule and vitamin A-rich animal-source foods, growing and consuming crops with high PVA level (PVA biofortified crops) is more sustainable and effective to alleviate VAD prevalence, especially in the low-income regions<sup>5,6</sup>. Several PVA biofortified crops have been reported to prevent the VAD prevalence effectively in various populations<sup>7,8</sup>.

PVA include a group of carotenoids containing at least one non-substituted  $\beta$ -ring, such as  $\beta$ -carotene,  $\alpha$ -carotene, and  $\beta$ -cryptoxanthin, which are synthesized via a complex carotenoid pathway (Supplementary Fig. S1)<sup>6,9–11</sup>. In higher plants, carotenoids are synthesized in the plastids from geranylgeranyl pyrophosphate (GGPP), which is the precursor for multiple pathways and contains four molecules of isopentenyl pyrophosphate originating from glyceraldehyde phosphate and pyruvate (Supplementary Fig. S1). Phytoene synthase (PSY) catalyzes the first committed reaction of carotenoid pathway to synthesize phytoene from two GGPP molecules. Accumulating evidence indicates that PSY is the key regulatory enzyme in the biosynthesis of carotenoids<sup>6,11–13</sup>. Biotechnological strategies have been successfully employed to increase PVA level (especially  $\beta$ -carotene) in a wealth of crops, resulting in golden canola<sup>14</sup>, rice<sup>15–17</sup>, wheat<sup>18</sup>, sorghum<sup>19</sup>, corn<sup>20,21</sup>, cassava<sup>22</sup>, potato<sup>23</sup>, sweet potato<sup>24</sup>, tomato<sup>25</sup>, soybean<sup>26</sup> and banana<sup>5</sup>. These works mainly adopted two strategies, i.e. to up-regulate PSY

Biotechnology Research Center, Chongqing Key Laboratory of Application and Safety Control of Genetically Modified Crops, Southwest University, Beibei, Chongqing, 400715, China. Dan Yao and Yi Wang contributed equally to this work. Correspondence and requests for materials should be addressed to Y.X. (email: [xiaoyuehua@swu.edu.cn](mailto:xiaoyuehua@swu.edu.cn))

solely<sup>5,14</sup> or with other synthases<sup>5,15–20</sup> to promote carotenoid biosynthesis, and to express an *Orange* gene to enhance carotenoid accumulation<sup>15,21,24,27</sup>. Recently, it was reported that the Orange protein might promote carotenoid biosynthesis via stabilizing the key synthase PSY<sup>28,29</sup>.

Cotton (*Gossypium*) is the leading natural fiber crop in the world, and is one of the major economic drivers in developing countries. In addition to fiber, cottonseed is an important source of edible oil (ranking 3rd in the world) and high-quality proteins<sup>30,31</sup>. Therefore, cotton is a potential supporting crop for poor regions to improve the economic and nutrient status simultaneously. PVA bio-fortified cotton and cottonseed oil may be a powerful tool against VAD prevalence in low-income cotton-growing regions, for example in south Asia and sub-Saharan Africa<sup>3,32,33</sup>. Our work aimed to increase PVA content of cottonseed and the resultant cottonseed oil. To this end, a functional PSY gene predominantly expressed in cotton was cloned, and upregulated specifically in the developing seeds. The transgenic cottons, with normal growth and development, produced golden cottonseeds and cottonseed oil fortified with  $\beta$ -carotene. This work significantly improved the nutritional value of cottonseeds, which provided this cash crop with the potential to prevent VAD prevalence.

## Results

**Identification and expression analysis of carotenoid synthase genes in *Gossypium*.** To characterize intrinsic carotenoid biosynthesis in various cotton tissues, especially in the developing seeds, we compared transcript levels of carotenoid related genes using transcriptomic data<sup>34</sup>. Firstly, we identified 36, 29 and 54 carotenoid synthase genes from the assembled cotton genomes of *G. raimondii*, *G. arboreum* and *G. hirsutum*, respectively (Table 1)<sup>34–36</sup>. These genes encoded all the enzymes catalyzing violaxanthin synthesis from GGPP and 3 key enzymes upstream to GGPP. As shown in Fig. 1, the transcript levels of the investigated carotenoid synthase genes varied with tissues and development stages, indicating that the carotenoid biosynthesis in cotton was developmentally regulated at transcription level. Generally, the transcript levels of carotenoid synthase genes were lower in the developing fibers and ovules compared to those in the roots, stems, leaves and petals. In the developing ovules, the PSY and LYC- $\epsilon$  genes had only a trace of transcription, while genes coding the rest carotenoid synthases had low-to-moderate expression. The transcript profile of carotenoid synthase genes suggested that the transcription of PSY genes may be a limiting factor of carotenoid biosynthesis in cottonseeds.

**Cloning and functional analysis of the *GhPSY2* gene.** To manipulate carotenoid biosynthesis in developing cottonseeds, we firstly cloned the coding sequences of the predominantly-expressed cotton PSY2 genes (Fig. 1) and analyzed their biological functions in transgenic tobacco. The coding regions of *GhPSY2A* and *GhPSY2D* were amplified from the leaf cDNA of upland cotton line T586, and their sequences were identical to those in the assembled TM-1 genome. Both *GhPSY2A* and *GhPSY2D* encoded proteins of 398 aa. *GhPSY2A/2D* were highly similar to plant group I PSYs, and distantly related to group II and III (Supplementary Fig. S2A). Multiple sequence alignment indicated that both *GhPSY2* proteins had conserved DxxxD motifs, substrate binding pocket, catalytic residues and active site lid motifs (Supplementary Fig. S2B). Along with the expression data, these results suggested that *GhPSY2A/2D* might encode biologically functional PSYs.

To further elucidate the biological functions of the cloned PSY genes, *GhPSY2D* was overexpressed in tobacco. Compared with the wild-type control, the *GhPSY2D* over-expressers accumulated higher levels of carotenoids, and had golden appearance in several organs, such as leaves, stems, filaments, and developing seeds (Fig. 2A–E). Meanwhile, the over-expressers showed retarded growth and dwarf phenotype (Supplementary Fig. S3) as reported in tomatoes<sup>37</sup>. These results indicated that *GhPSY2D* was biologically functional to promote carotenoid synthesis in plants.

**Golden cottonseed from specific upregulation of *GhPSY2D*.** To increase the carotenoid content of cottonseeds, *GhPSY2D* was constructed downstream to a seed-specific promoter *pV*<sup>38,39</sup>, and transformed into upland cotton. We finally obtained 3 transformants (#1, #2 and #3) with *GhPSY2D* highly expressed in developing seeds (Fig. 3). As expected, *GhPSY2D* transcript (Fig. 3B) and carotenoid (Fig. 3C) levels in the transgenic embryos dramatically increased at the mid-to-late stage (after 25 DPA) and the embryos turned golden accordingly (Fig. 3A). The mature *pV:GhPSY2D* cottonseeds had golden kernel and the carotenoid contents in the extracted oil of transgenic cottonseeds were significantly higher (over 6-fold) than that from the wild type (Fig. 4A–D). Importantly, thin layer chromatography (TLC) analysis indicated that the major carotenoid increased in the transgenic cottonseeds was  $\beta$ -carotene (Fig. 4E), which was the main plant PVA carotenoid. Moreover, the *pV:GhPSY2D* cottons showed no obvious defect in plant growth and fiber development compared to the wild-type control (Supplementary Fig. S4), although the germination rates of transgenic cottonseeds decreased as reported in Arabidopsis (Supplementary Fig. S5)<sup>40</sup>. These results demonstrated that the transgenic cottonseeds were successfully bio-fortified for PVA, rendering the transgenic cotton a potential tool against VAD in cotton-growing regions including south Asia and sub-Saharan Africa<sup>32,33</sup>.

## Discussion

The objective of the present work is to generate PVA bio-fortified cottonseeds, which may be useful tool against VAD prevalence in low-income regions. On the basis of comprehensive identification of carotenoid synthase genes in the assembled cotton genomes, we compared the transcript levels of these genes in various cotton tissues, including developing ovules, and demonstrated that the transcription of PSY genes might be the limiting factor for carotenoid biosynthesis and accumulation in cottonseeds. Next, we cloned the *GhPSY2* genes, and confirmed their biological functions by over-expression in tobacco. Finally, the *GhPSY2D* gene driven by a seed-specific promoter *pV* was transformed into upland cotton, and significantly expressed in the mid-to-late embryos. The resultant transgenic cottonseeds had golden kernel, and the extracted oil contained significantly higher level of carotenoids, especially  $\beta$ -carotene (the major active PVA). Given cottons are widely grown as cash crop in south

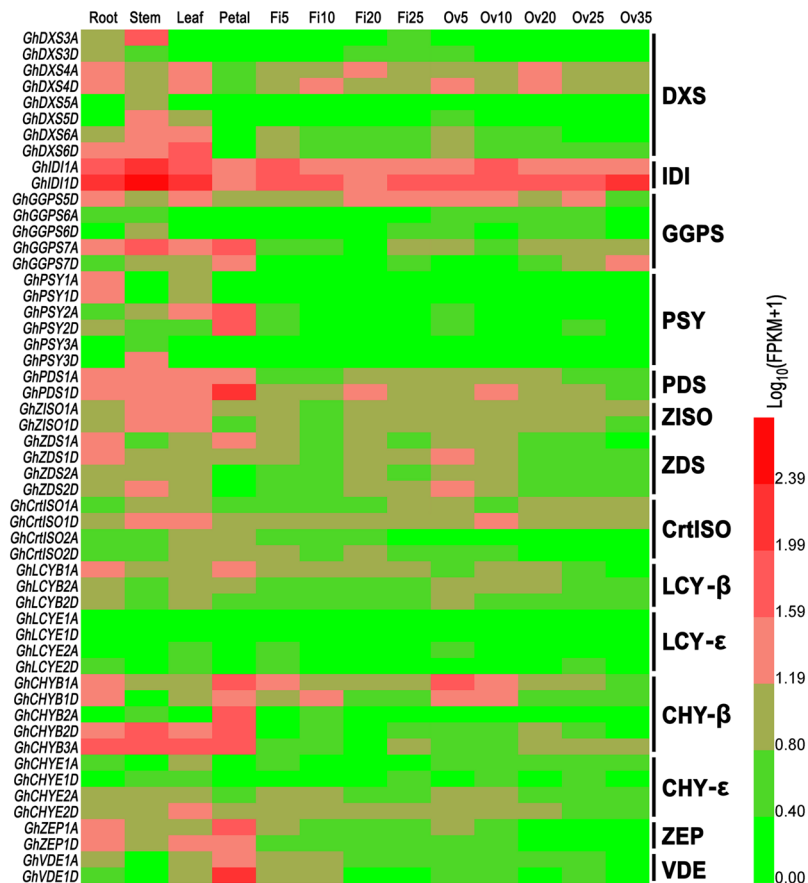
Enzymes	Genes	<i>G. raimondii</i> (D5)	<i>G. arboreum</i> (A2)	<i>G. hirsutum</i> -Dt1	<i>G. hirsutum</i> -At1
DXS	<i>GoDXS1</i>	Gorai.004G030400	Cotton_A_06764	—	—
	<i>GoDXS2</i>	Gorai.004G030800	Cotton_A_06768	—	—
	<i>GoDXS3</i>	Gorai.004G030900	Cotton_A_06769	Gh_D08G0270	Gh_A08G0193
	<i>GoDXS4</i>	Gorai.004G111100	Cotton_A_21777	Gh_D08G1005	Gh_A08G2461
	<i>GoDXS5</i>	Gorai.008G093600	Cotton_A_31168	Gh_D12G2793	Gh_A12G0784
	<i>GoDXS6</i>	Gorai.011G184200	Cotton_A_39687	Gh_D10G1640	Gh_A10G2292
IDI	<i>GoIDI1</i>	Gorai.004G262800	Cotton_A_05787	Gh_D08G2391	Gh_A08G1997
GGPS	<i>GoGGPS1</i>	Gorai.002G251300	Cotton_A_00433	Gh_D01G2110	Gh_A01G1866
	<i>GoGGPS2</i>	Gorai.002G251400	—	—	—
	<i>GoGGPS3</i>	Gorai.005G054800	—	—	—
	<i>GoGGPS4</i>	Gorai.005G120800	—	—	—
	<i>GoGGPS5</i>	Gorai.007G051300	Cotton_A_02023	Gh_D11G0475	—
	<i>GoGGPS6</i>	Gorai.007G206700	Cotton_A_32196	Gh_D11G1883	Gh_A11G1725
	<i>GoGGPS7</i>	Gorai.011G102700	Cotton_A_17013	Gh_D10G0907	Gh_A10G0844
	<i>GoGGPS8</i>	Gorai.011G102800	—	—	—
	<i>GoGGPS9</i>	Gorai.011G103000	—	—	—
	<i>GoGGPS10</i>	Gorai.011G133100	—	—	—
	PSY	<i>GoPSY1</i>	Gorai.001G083700	Cotton_A_12202	Gh_D07G0746
<i>GoPSY2</i>		Gorai.006G009400	Cotton_A_14731	Gh_D09G0078	Gh_A09G0080
<i>GoPSY3</i>		Gorai.010G126900	Cotton_A_36110	Gh_D06G1167	Gh_A06G0932
<i>GoPSY4</i>		Gorai.012G039100	Cotton_A_29856	Gh_D04G0308	Gh_A05G3296
PDS	<i>GoPDS1</i>	Gorai.011G149900	Cotton_A_36228	Gh_D10G1327	Gh_A10G1167
ZISO	<i>GoZISO1</i>	Gorai.005G080900	Cotton_A_34612	Gh_D02G0721	Gh_A02G0673
ZDS	<i>GoZDS1</i>	Gorai.007G284900	Cotton_A_14988	Gh_D11G2616	Gh_A11G2306
	<i>GoZDS2</i>	Gorai.013G201600	Cotton_A_12871	Gh_D13G1826	Gh_A13G1497
CrtISO	<i>GoCrtISO1</i>	Gorai.002G153800	Cotton_A_34115	Gh_D01G1197	Gh_A01G1122
	<i>GoCrtISO2</i>	Gorai.002G218000	Cotton_A_27012	Gh_D01G1808	Gh_A01G1557
LCY-β	<i>GoLCYB1</i>	Gorai.006G113300	Cotton_A_26267	Gh_D09G0946	Gh_A09G0916
	<i>GoLCYB2</i>	Gorai.013G002400	Cotton_A_20533	Gh_D13G0026	Gh_A13G0010
LCY-ε	<i>GoLCYE1</i>	Gorai.001G012900	Cotton_A_05719	Gh_D07G2366	Gh_A07G0117
	<i>GoLCYE2</i>	Gorai.009G032900	Cotton_A_09351	Gh_D05G0313	Gh_A05G0229
CHY-β	<i>GoCHYB1</i>	Gorai.006G199400	Cotton_A_27296	Gh_D09G1728	Gh_A09G1634
	<i>GoCHYB2</i>	Gorai.008G249600	Cotton_A_32712	Gh_D12G2438	Gh_A12G2678
	<i>GoCHYB3</i>	Gorai.008G274800	Cotton_A_01286	—	Gh_A12G2302
CHY-ε	<i>GoCHYE1</i>	Gorai.007G372200	Cotton_A_05513	Gh_D11G3263	Gh_A11G2878
	<i>GoCHYE2</i>	Gorai.013G180200	Cotton_A_26341	Gh_D13G1644	Gh_A13G1336
ZEP	<i>GoZEP1</i>	Gorai.007G213500	Cotton_A_24363	Gh_D11G3469	Gh_A11G1774
VDE	<i>GoVDE1</i>	Gorai.007G268500	Cotton_A_18131	Gh_D11G2473	Gh_A11G2173

**Table 1.** The coding genes of carotenoid synthases identified in cottons. Enzymes are abbreviated as in Supplementary Fig. S1. Genes are named alphabetically according to their IDs in *G. raimondii*, and homeologs annotated in different genomes (D5, A2, Dt1 and At1) are list in the same line.

Asia and sub-Sahara Africa<sup>32,33</sup>, the PVA bio-fortified transgenic cottonseeds may be quite potential to prevent VAD prevalence in these regions.

To prevent VAD prevalence, a wealth of crops were bred or engineered to contain high levels of carotenoids or PVAs<sup>6,24</sup>. Compared with the previously reported PVA bio-fortified crops, cotton has significant advantages in the battle against VAD prevalence. Firstly, cottonseeds are lipid-rich and PVAs are easily extracted with cottonseed oil (Fig. 4)<sup>30,31,41</sup>. As reported, fat in diet significantly increase the β-carotene bioavailability<sup>42,43</sup>, implying that enhanced PVA in cottonseed and cottonseed oil should be more easily utilized by human compared with these in starch-rich crops. Secondly, human malnutrition, including VAD, generally occurs in developing regions, where the demands on economic development and nutrient improvement, somewhat mutually dependent and contradictory, are both pressing. As one of the most important cash crops in developing regions including south Asia and sub-Sahara Africa<sup>32,33</sup>, cotton plays a crucial role in economic development and poverty reduction. PVA bio-fortified transgenic cottons can simultaneously meet the demands on PVA supply and income increments, therefore may become a powerful tool in the battle against poverty and VAD prevalence. Notably, our method is easy to combine with other tactics, such as specific inhibition of gossypol synthesis via RNA interfering<sup>31</sup>, to further enhance the nutrient value and utilization of cottonseeds.

Carotenoid biosynthesis in plants involves a multi-step complex pathway (Fig. 1)<sup>6,9–11</sup>. The final carotenoid level and profile in a certain tissue are collectively determined by the substrate accessibility and enzyme activities



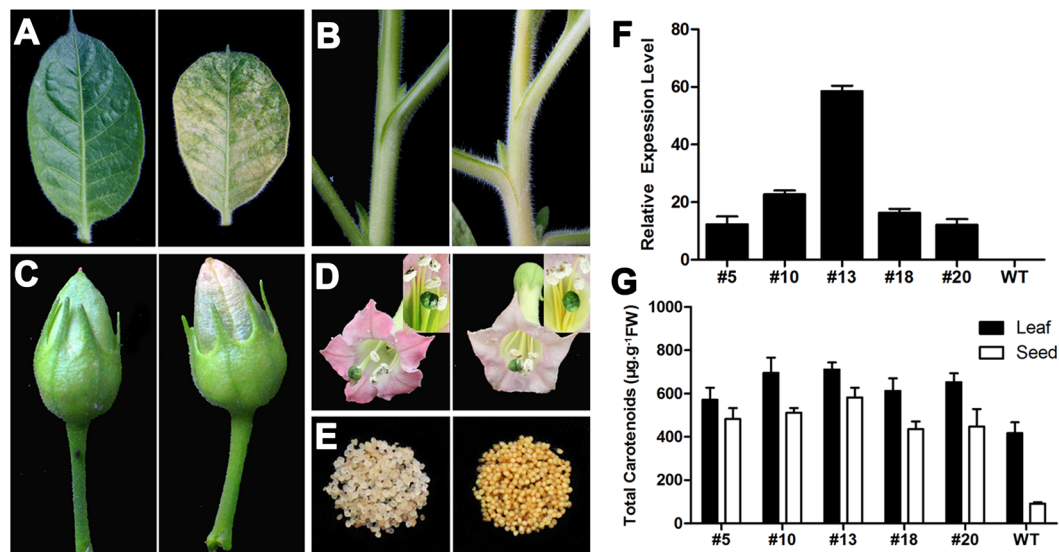
**Figure 1.** Transcript levels of carotenoid synthase genes in various upland cotton tissues. The gene expression levels (FPKM) in root, stem, leaf, petal, and fibers (Fi5–25) and ovules (Ov5–35) of various days post anthesis (DPA) are converted to  $\text{Log}_{10}(\text{FPKM} + 1)$  and illustrated as seven classes in the heat map. Genes are named as the abbreviation of species and enzyme plus code plus A or D to indicate the subgenome origin. Enzymes are abbreviated as in Supplementary Fig. S1. Transcript levels are inferred from the transcriptomic data<sup>34</sup>. The genes with very low transcript level (total FPKM < 2) are omitted.

catalyzing these synthesis steps. For example, the golden rice from upregulation of PSY and CrtI mainly accumulates  $\beta$ -carotene, instead of lycopene, attributing to the constitutively expressed intrinsic carotenoid synthases<sup>44</sup>. Before the sequenced genomes and comprehensive expression data available, designing strategies to engineer carotenoid and other secondary metabolites largely depended on experience and generally multiple enzymes were simultaneously targeted, which substantially added difficulty in gene manipulation<sup>5,15,26,45</sup>. In this work, we designed the transgenic strategy for PVA bio-fortification on the basis of comprehensive identification and expression analysis of carotenoid synthase genes using assembled genomes and transcriptomic data (Table 1 and Fig. 1), and obtained the golden cottonseed by specifically upregulating a single gene (*GhPSY2D*). Consistent with the very low transcript levels of the encoding genes of LCY- $\epsilon$ , ZEP and VDE, and moderate expressions of PDS, ZISO, ZDS, crtISO, LCY- $\beta$  and CHY- $\beta$  genes in the late-stage cottonseeds (35 DPA, Fig. 1), *GhPSY2D* upregulation in cottonseeds mainly promoted the accumulation of  $\beta$ -carotene and another carotenoid, probably zeaxanthin (Fig. 4E). These results indicated that comprehensive evaluation of all the genes involved in a certain pathway may be a useful foundation for manipulation of secondary metabolites in plants.

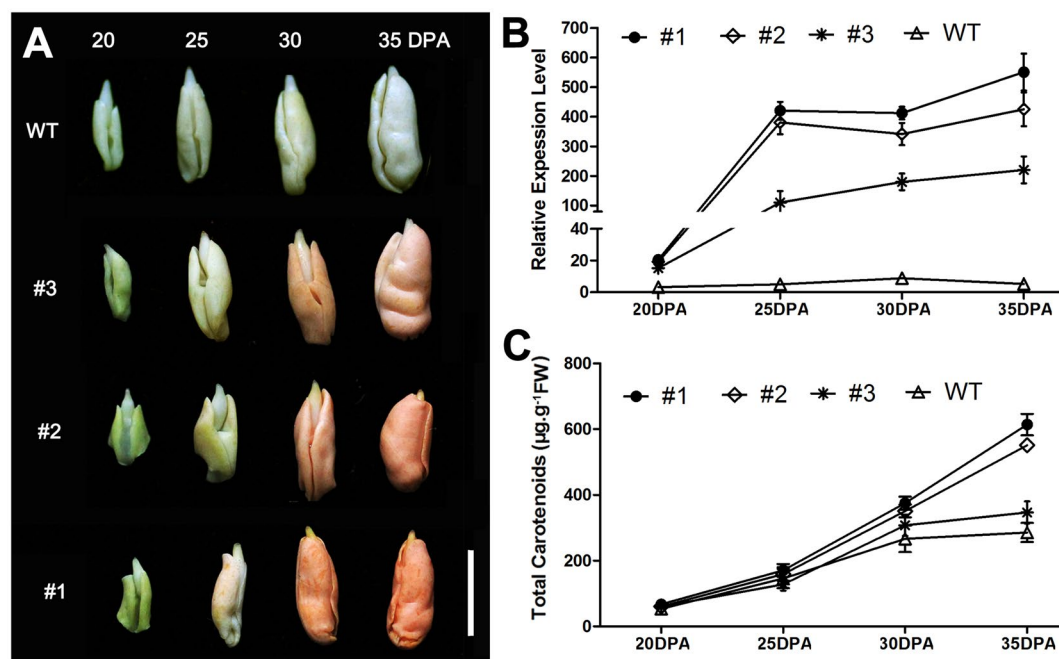
## Methods

**Identification and expression analysis of carotenoid synthase genes in cotton.** To identify the carotenoid synthase genes in cottons, Arabidopsis proteins were employed as probe to query homologous sequences from *G. raimondii* genome in phytozome (<https://phytozome.jgi.doe.gov/>)<sup>46</sup>. The resultant *G. raimondii* sequences were aligned with all the annotated proteins of certain synthase, and subjected to construct a neighbor-joining tree with 1000 replicates of bootstrap test in MEGA6.0<sup>47</sup>. The homologs clustered with certain Arabidopsis carotenoid synthase were regarded as carotenoid synthases in *G. raimondii*. The corresponding orthologous genes in *G. arboreum* and *G. hirsutum* were identified with a standalone BLAST software using *G. raimondii* genes as probe<sup>48</sup>.

The transcript levels of all predicted *G. hirsutum* genes were evaluated routinely using public available transcriptomic data<sup>34</sup> released by Zhang's lab (<http://mascotton.njau.edu.cn>). The transcript levels (in FPKM,



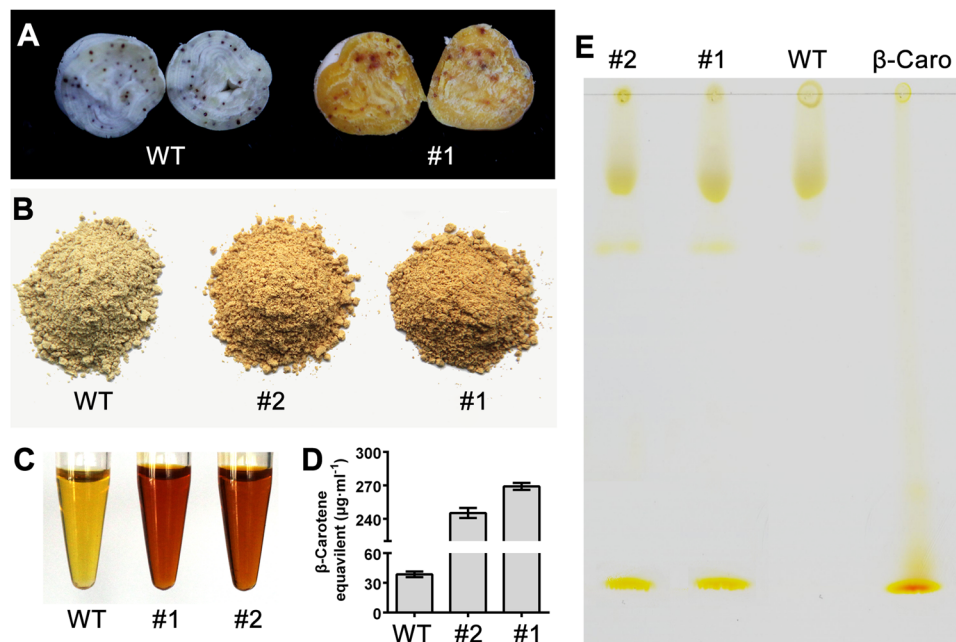
**Figure 2.** Overexpression of *GhPSY2D* promotes carotenoid biosynthesis in tobacco. (A–E) Phenotypic comparison of *GhPSY2D* over-expressor (transformant #13, right) and the wild-type control (a null segregant of transformant #13, left); (A) Leaves; (B) Stems; (C) Capsules; (D) Flower and stamens; (E), 12-DPA seeds; (F) Transcript levels of *GhPSY2D* in wild-type control (WT) and *GhPSY2D* over-expressers (#5, #10, #13, #18 and #20); (G) Carotenoid quantification in leaves and 12-DPA seeds. Error bars indicate standard deviations of 3 biological replicates.



**Figure 3.** *GhPSY2D* expression and carotenoid accumulation in the developing cottonseeds. Colors (A), *GhPSY2D* transcript levels (B) and carotenoid contents (C) in the developing embryos (20–35 DPA) of the transformants #1, #2 and #3, and the wild-type control (WT, a null segregant of transformant #1) are indicated.

fragment per kilobase per million) of identified carotenoid synthase genes in various *G. hirsutum* tissues were depicted as heat map using the program HemI 1.0<sup>49</sup>.

**RNA Extraction and qRT-PCR.** Total RNAs were extracted from approximately 100 mg of plant tissues using a rapid plant RNA extraction kit (Aidlab, Beijing, China). The first-stranded cDNAs were synthesized from 1 µg total RNA using a reverse transcriptase kit (TaKaRa, Dalian, China). Quantitative PCR was performed on a CFX96 real-time PCR detection system using SYBR Green Supermix (Bio-Rad, CA, USA) and gene-specific primers



**Figure 4.** Carotenoids in mature cottonseeds. (A) Transverse view of mature seed kernels; (B) Seed kernel powder; (C) Cottonseed oils; (D) Total carotenoid content in cottonseed oils; (E) TLC analysis of carotenoid components in cottonseed oils. Standards β-carotene is separated along with samples.

(Table S1). The thermocycling parameters were as follows: 95 °C for 3 min, followed by 40 cycles of 95 °C for 5 s, 57 °C for 20 s, and a standard melting curve to monitor PCR specificity. The actin<sup>50</sup> and histone3 (AF024716)<sup>51</sup> gene were amplified as internal control in tobacco and cotton, respectively. The analyses included three biological replicates and data were analyzed using the software Bio-Rad CFX Manager 2.0 provided by the manufacturer.

**Cloning and bioinformatics analysis of *GhPSY2*.** A pair of primers (*GhPSY2*-U and *GhPSY2*-D, Supplementary Table S2) encompassing the full-length ORF, designed according to the *GhPSY2* sequences identified in *G. hirsutum*, was employed to amplify the cDNA coding sequences from leaves. The reaction included 1 μl first-stranded cDNA, 0.2 μM each primer, and 1 × PrimeSTAR max Premix (TaKaRa, Dalian, China), and amplified for 35 cycles of 98 °C for 10 s, 56 °C for 15 s and 72 °C for 40 s. The PCR products were cloned into pGEM-T easy vector (Promega, Shanghai, China), sequenced in BGI (Shenzhen, China), and further compared to the *GhPSY2* genes identified in the assembled *G. hirsutum* genome. *GhPSY2* proteins were aligned with Arabidopsis, rice and maize PSYs from Phytozome (<https://phytozome.jgi.doe.gov/>)<sup>46</sup> using ClustalW, and the NJ tree was constructed with 1000 replicates of bootstrap test in MEGA6.0<sup>47</sup>.

**Vector construction and plant transformation.** A modified pBI121 vector p5 (pBI121-GN), containing selection marker *NPTII* and *GUS* genes, was used to construct the plant expression vectors<sup>52</sup>. The cDNA sequences of *GhPSY2D* ORF was excised from the cloning vector pGEM-T easy using *Bam*HI and *Eco*RI, and inserted downstream to a CaMV35S promoter in the p5 vector restricted by the same enzymes, resulting in the overexpression vector. To construct seed-specific expression vectors, the promoter of *Phaseolus vulgaris* β-type phaseolin storage protein gene (*pV*, GenBank accession no. J01263.1)<sup>39</sup>, were amplified with restriction sites of *Hind*III and *Bam*HI (Table S1). The promoter was constructed upstream to *GhPSY2D* by replacing the CaMV35S promoter in the overexpression vector using *Hind*III and *Bam*HI sites. All these expression vectors were transferred into *Agrobacterium tumefaciens* strains (LBA4404), and the resulting *Agrobacterium* strain was used for tobacco and cotton (Jimian No. 14) transformation as previously described<sup>50,52</sup>.

**Carotenoid extraction and analysis.** Carotenoids in fresh tobacco tissues and developing cottonseeds were extracted according Fuentes' method<sup>53</sup> with some modification. In brief, approximately 500 mg of fresh tissues were ground to fine powder in liquid nitrogen, extracted for 15 min in 3 ml hexane/acetone/ethanol (2:1:1 v/v/v) with shaking. Two to three successive extractions were performed to remove carotenoids until the tissues were colorless. The extracts were combined, dried with nitrogen, and re-suspended in 1 ml acetone. Total carotenoids were measured spectrophotometrically at 474 nm, and quantified according to a standard curve of β-carotene. The extractions were repeated in three biological replicates.

Carotenoids in dry mature cottonseeds were extracted along with oil with Soxhlet extractor (Buchi B-811, Switzerland) with ether. The resultant cottonseed oil were 20-fold diluted in acetone and subjected to spectrophotometric quantification of total carotenoids as mentioned above and TLC analysis<sup>54</sup>. Ten microliter of diluted cottonseed oil were loaded and separated on Silica gel plate (0.2 mm-thick, Jiangyou Silica Gel Co., Yantai, Shandong, China) along with 5 μg β-carotene standards. The plate was developed in hexane:ether:acetone (60:30:20, v/v) and photographed directly.

## References

- Underwood, B. A. Vitamin A deficiency disorders: international efforts to control a preventable “pox”. *The Journal of nutrition* **134**, 231S–236S (2004).
- Harjes, C. E. *et al.* Natural genetic variation in lycopene epsilon cyclase tapped for maize biofortification. *Science (New York, N.Y.)* **319**, 330–333, <https://doi.org/10.1126/science.1150255> (2008).
- Stevens, G. A. *et al.* Trends and mortality effects of vitamin A deficiency in children in 138 low-income and middle-income countries between 1991 and 2013: a pooled analysis of population-based surveys. *The Lancet Global Health* **3**, e528–e536, [https://doi.org/10.1016/s2214-109x\(15\)00039-x](https://doi.org/10.1016/s2214-109x(15)00039-x) (2015).
- Hamer, D. H. & Keusch, G. T. Vitamin A deficiency: slow progress towards elimination. *The Lancet Global Health* **3**, e502–e503, [https://doi.org/10.1016/s2214-109x\(15\)00096-0](https://doi.org/10.1016/s2214-109x(15)00096-0) (2015).
- Paul, J. Y. *et al.* Golden bananas in the field: elevated fruit pro-vitamin A from the expression of a single banana transgene. *Plant Biotechnol J* **15**, 520–532, <https://doi.org/10.1111/pbi.12650> (2017).
- Giuliano, G. P. A biofortification of crop plants: a gold rush with many miners. *Current opinion in biotechnology* **44**, 169–180, <https://doi.org/10.1016/j.copbio.2017.02.001> (2017).
- Girard, A. W. *et al.* Promotion of Orange-Fleshed Sweet Potato Increased Vitamin A Intakes and Reduced the Odds of Low Retinol-Binding Protein among Postpartum Kenyan Women. *The Journal of nutrition* **147**, 955–963, <https://doi.org/10.3945/jn.116.236406> (2017).
- De Moura, F. F. *et al.* Biofortified beta-carotene rice improves vitamin A intake and reduces the prevalence of inadequacy among women and young children in a simulated analysis in Bangladesh, Indonesia, and the Philippines. *The American journal of clinical nutrition* **104**, 769–775, <https://doi.org/10.3945/ajcn.115.129270> (2016).
- Enfissi, E. M., Nogueira, M., Bramley, P. M. & Fraser, P. D. The regulation of carotenoid formation in tomato fruit. *The Plant journal: for cell and molecular biology* **89**, 774–788, <https://doi.org/10.1111/tjp.13428> (2017).
- Ruiz-Sola, M. Á. & Rodríguez-Concepción, M. Carotenoid Biosynthesis in Arabidopsis: A Colorful Pathway. *The Arabidopsis Book*, e0158, doi:10.1199/tab.0158 (2012).
- Nisar, N., Li, L., Lu, S., Khin, N. C. & Pogson, B. J. Carotenoid metabolism in plants. *Molecular plant* **8**, 68–82, <https://doi.org/10.1016/j.molp.2014.12.007> (2015).
- Toledo-Ortiz, G., Huq, E. & Rodríguez-Concepción, M. Direct regulation of phytoene synthase gene expression and carotenoid biosynthesis by phytochrome-interacting factors. *Proceedings of the National Academy of Sciences of the United States of America* **107**, 11626–11631, <https://doi.org/10.1073/pnas.0914428107> (2010).
- Welsch, R. *et al.* Provitamin A accumulation in cassava (*Manihot esculenta*) roots driven by a single nucleotide polymorphism in a phytoene synthase gene. *Plant Cell* **22**, 3348–3356, <https://doi.org/10.1105/tpc.110.077560> (2010).
- Shewmaker, C. K., Sheehy, J. A., Daley, M., Colburn, S. & Ke, D. Y. Seed-specific overexpression of phytoene synthase: increase in carotenoids and other metabolic effects. *The Plant journal: for cell and molecular biology* **20**, 401–412X (1999).
- Bai, C. *et al.* Bottlenecks in carotenoid biosynthesis and accumulation in rice endosperm are influenced by the precursor-product balance. *Plant Biotechnol J* **14**, 195–205, <https://doi.org/10.1111/pbi.12373> (2016).
- Ye, X. *et al.* Engineering the provitamin A (beta-carotene) biosynthetic pathway into (carotenoid-free) rice endosperm. *Science (New York, N.Y.)* **287**, 303–305 (2000).
- Paine, J. A. *et al.* Improving the nutritional value of Golden Rice through increased pro-vitamin A content. *Nat Biotech* **23**, 482–487, [http://www.nature.com/nbt/journal/v23/n4/supinfo/nbt1082\\_S1.html](http://www.nature.com/nbt/journal/v23/n4/supinfo/nbt1082_S1.html) (2005).
- Zeng, J. *et al.* Metabolic Engineering of Wheat Provitamin A by Simultaneously Overexpressing CrtB and Silencing Carotenoid Hydroxylase (TaHYD). *Journal of agricultural and food chemistry* **63**, 9083–9092, <https://doi.org/10.1021/acs.jafc.5b04279> (2015).
- Che, P. *et al.* Elevated vitamin E content improves all-trans beta-carotene accumulation and stability in biofortified sorghum. *Proceedings of the National Academy of Sciences of the United States of America* **113**, 11040–11045, <https://doi.org/10.1073/pnas.1605689113> (2016).
- Naqvi, S. *et al.* Transgenic multivitamin corn through biofortification of endosperm with three vitamins representing three distinct metabolic pathways. *Proceedings of the National Academy of Sciences of the United States of America* **106**, 7762–7767, <https://doi.org/10.1073/pnas.0901412106> (2009).
- Berman, J. *et al.* The Arabidopsis ORANGE (AtOR) gene promotes carotenoid accumulation in transgenic corn hybrids derived from parental lines with limited carotenoid pools. *Plant cell reports* **36**, 933–945, <https://doi.org/10.1007/s00299-017-2126-z> (2017).
- Sayre, R. *et al.* The BioCassava plus program: biofortification of cassava for sub-Saharan Africa. *Annual review of plant biology* **62**, 251–272, <https://doi.org/10.1146/annurev-arplant-042110-103751> (2011).
- Diretto, G. *et al.* Metabolic engineering of potato tuber carotenoids through tuber-specific silencing of lycopene epsilon cyclase. *BMC plant biology* **6**, 13, <https://doi.org/10.1186/1471-2229-6-13> (2006).
- Kang, L. *et al.* Metabolic engineering of carotenoids in transgenic sweetpotato. *Breeding science* **67**, 27–34, <https://doi.org/10.1270/jsbbs.16118> (2017).
- Rosati, C. *et al.* Metabolic engineering of beta-carotene and lycopene content in tomato fruit. *The Plant journal: for cell and molecular biology* **24**, 413–419 (2000).
- Kim, M. J. *et al.* Genetic modification of the soybean to enhance the beta-carotene content through seed-specific expression. *PLoS One* **7**, e48287, <https://doi.org/10.1371/journal.pone.0048287> (2012).
- Diretto, G. *et al.* Metabolic Engineering of Potato Carotenoid Content through Tuber-Specific Overexpression of a Bacterial Mini-Pathway. *PLoS ONE* **2**, e350 (2007).
- Park, S. *et al.* Orange protein has a role in phytoene synthase stabilization in sweetpotato. *Scientific reports* **6**, 33563, <https://doi.org/10.1038/srep33563> (2016).
- Zhou, X. *et al.* Arabidopsis OR proteins are the major posttranscriptional regulators of phytoene synthase in controlling carotenoid biosynthesis. *Proceedings of the National Academy of Sciences of the United States of America* **112**, 3558–3563, <https://doi.org/10.1073/pnas.1420831112> (2015).
- Yuan, D. *et al.* The genome sequence of Sea-Island cotton (*Gossypium barbadense*) provides insights into the allopolyploidization and development of superior spinnable fibres. *Scientific reports* **5**, 17662, <https://doi.org/10.1038/srep17662> (2015).
- Sunilkumar, G., Campbell, L. M., Puckhaber, L., Stipanovic, R. D. & Rathore, K. S. Engineering cottonseed for use in human nutrition by tissue-specific reduction of toxic gossypol. *Proceedings of the National Academy of Sciences of the United States of America* **103**, 18054–18059, <https://doi.org/10.1073/pnas.0605389103> (2006).
- Statista. *World cotton production by country 2016*, <https://www.statista.com/statistics/263055/cotton-production-worldwide-by-top-countries/> (2017).
- Staritz, C. & Tröster, B. Cotton-based development in Sub-Saharan Africa? Global commodity chains, national market structure and development outcomes in Burkina Faso, Mozambique and Tanzania. *Working Papers* (2015).
- Zhang, T. *et al.* Sequencing of allotetraploid cotton (*Gossypium hirsutum* L. acc. TM-1) provides a resource for fiber improvement. *Nature biotechnology* **33**, 531–537, <https://doi.org/10.1038/nbt.3207> (2015).
- Li, F. *et al.* Genome sequence of the cultivated cotton *Gossypium arboreum*. *Nature genetics* **46**, 567–572, <https://doi.org/10.1038/ng.2987> (2014).
- Paterson, A. H. *et al.* Repeated polyploidization of *Gossypium* genomes and the evolution of spinnable cotton fibres. *Nature* **492**, 423–427, <https://doi.org/10.1038/nature11798> (2012).

37. Fray, R. G. *et al.* Constitutive expression of a fruit phytoene synthase gene in transgenic tomatoes causes dwarfism by redirecting metabolites from the gibberellin pathway. *The Plant Journal* **8**, 693–701, <https://doi.org/10.1046/j.1365-313x.1995.08050693.x> (1995).
38. Chen, Z. L., Pan, N. S. & Beachy, R. N. A DNA sequence element that confers seed-specific enhancement to a constitutive promoter. *The EMBO journal* **7**, 297–302 (1988).
39. Slightom, J. L., Sun, S. M. & Hall, T. C. Complete nucleotide sequence of a French bean storage protein gene: Phaseolin. *Proceedings of the National Academy of Sciences of the United States of America* **80**, 1897–1901 (1983).
40. Lindgren, L. O., Stalberg, K. G. & Hoglund, A. S. Seed-specific overexpression of an endogenous Arabidopsis phytoene synthase gene results in delayed germination and increased levels of carotenoids, chlorophyll, and abscisic acid. *Plant physiology* **132**, 779–785, <https://doi.org/10.1104/pp.102.017053> (2003).
41. Lukonge, E., Labuschagne, M. T. & Hugo, A. The evaluation of oil and fatty acid composition in seed of cotton accessions from various countries. *Journal of the Science of Food and Agriculture* **87**, 340–347, <https://doi.org/10.1002/jsfa.2731> (2007).
42. La Frano, M. R., de Moura, F. F., Boy, E., Lonnerdal, B. & Burri, B. J. Bioavailability of iron, zinc, and provitamin A carotenoids in biofortified staple crops. *Nutrition reviews* **72**, 289–307, <https://doi.org/10.1111/nure.12108> (2014).
43. Lipkie, T. E. *et al.* Bioaccessibility of carotenoids from transgenic provitamin A biofortified sorghum. *Journal of agricultural and food chemistry* **61**, 5764–5771, <https://doi.org/10.1021/jf305361s> (2013).
44. Schaub, P., Al-Babili, S., Drake, R. & Beyer, P. Why Is Golden Rice Golden (Yellow) Instead of Red? *Plant Physiology* **138**, 441–450, <https://doi.org/10.1104/pp.104.057927> (2005).
45. Tanaka, Y. & Ohmiya, A. Seeing is believing: engineering anthocyanin and carotenoid biosynthetic pathways. *Current opinion in biotechnology* **19**, 190–197, <https://doi.org/10.1016/j.copbio.2008.02.015> (2008).
46. Goodstein, D. M. *et al.* Phytozome: a comparative platform for green plant genomics. *Nucleic acids research* **40**, D1178–D1186, <https://doi.org/10.1093/nar/gkr944> (2012).
47. Tamura, K., Stecher, G., Peterson, D., Filipski, A. & Kumar, S. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Molecular biology and evolution* **30**, 2725–2729, <https://doi.org/10.1093/molbev/mst197> (2013).
48. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *Journal of molecular biology* **215**, 403–410, [https://doi.org/10.1016/s0022-2836\(05\)80360-2](https://doi.org/10.1016/s0022-2836(05)80360-2) (1990).
49. Deng, W., Wang, Y., Liu, Z., Cheng, H. & Xue, Y. HemI: a toolkit for illustrating heatmaps. *PLoS One* **9**, e111988, <https://doi.org/10.1371/journal.pone.0111988> (2014).
50. Luo, K. *et al.* Leaf senescence is delayed in tobacco plants expressing the maize knotted1 gene under the control of a wound-inducible promoter. *Plant cell reports* **25**, 1246–1254, <https://doi.org/10.1007/s00299-006-0178-6> (2006).
51. Zhu, Y. Q., Xu, K. X., Luo, B., Wang, J. W. & Chen, X. Y. An ATP-binding cassette transporter GhWBC1 from elongating cotton fibers. *Plant physiology* **133**, 580–588, <https://doi.org/10.1104/pp.103.027052> (2003).
52. Luo, M. *et al.* GhDET2, a steroid 5 $\alpha$ -reductase, plays an important role in cotton fiber cell initiation and elongation. *The Plant Journal* **51**, 419–430, <https://doi.org/10.1111/j.1365-313x.2007.03144.x> (2007).
53. Fuentes, P. *et al.* Light-dependent changes in plastid differentiation influence carotenoid gene expression and accumulation in carrot roots. *Plant molecular biology* **79**, 47–59, <https://doi.org/10.1007/s11103-012-9893-2> (2012).
54. Angaman, D. *et al.* Precursor uptake assays and metabolic analyses in isolated tomato fruit chromoplasts. *Plant Methods* **8**, 1, <https://doi.org/10.1186/1746-4811-8-1> (2012).

## Acknowledgements

This work was partially supported by the National Natural Science Foundation of China (30971713 and 31571582 to Y.X.) and by the Genetically Modified Organisms Breeding Major Project of China (2016ZX08005005–001 to Y.X.).

## Author Contributions

D.Y. and Y.X. conceived the experiments and wrote the paper, D.Y., Y.W., Q.L. and X.O. conducted the experiments, D.Y., Y.W., Q.L., Y.L. and Y.X. analyzed the data, C.W., L.D., L.H. and M.L. contributed reagents/materials/analysis tools. All authors reviewed manuscript.

## Additional Information

**Supplementary information** accompanies this paper at <https://doi.org/10.1038/s41598-018-19866-1>.

**Competing Interests:** The authors declare that they have no competing interests.

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2018