## Supplementary Material

Defining and enhancing the biosynthesis of astaxanthin and Docosahexaenoic acid in Aurantiochytrium sp. SK4

Jingrun $\mathrm{Ye}^{\mathrm{a}, \mathrm{b}}$, Mengmeng Liu ${ }^{\mathrm{a}, \mathrm{b}}$, Mingxia He ${ }^{\mathrm{a}}$, Ying Ye ${ }^{\mathrm{a}, \mathrm{b}}$, Junchao Huang ${ }^{\mathrm{a} *}$
${ }^{\text {a Key Laboratory of Economic Plants and Biotechnology, Yunnan Key Laboratory }}$ for Wild Plant Resources, Kunming Institute of Botany, Chinese Academy of Sciences, Kunming 650201
${ }^{\text {b }}$ University of Chinese Academy of Sciences, Beijing 100049, People's Republic of China

## *Corresponding author

Email: huangjc@mail.kib.ac.cn.

Phone: +86-875-65228058


Figure S1. Uni-gene length distribution.


Figure S2. Functional annotation of assembled uni-genes in Aurantiochytrium sp. SK4. (A)
GO classification, (B) COG classification, (C) KEGG pathway.


Figure S3. KEGG pathway enrichment analysis of differentially expressed genes (DEGs).


Figure S4. Genes encoding enzymes of the polyketide synthase (PKS pathway) in Aurantiochytrium sp.SK4. Dark gray areas indicate proposed enzymatic domains. KS, 3ketoacyl synthase; KR, 3-ketoacyl-ACP reductase; ER, enoyl reducatase; DH, dehydrase/isomerase; PPTase, phosphopantetheine transferase; MAT, malonyl-CoA:ACP acyltransferase. aa, amino acid.


Figure S5. The growth curve of Aurantiochytrium sp.SK4 of Figure 1 and Figure 2.


| List of domain hits |  |  |  | $?$ |
| :---: | :---: | :---: | :---: | :---: |
| + Name | Accession | Description | Interval | E-value |
| [+] crti_fam | TIGR02734 | phytoene desaturase; Phytoene is converted to lycopene by desaturation at four (two ... | 55-551 | $1.36 \mathrm{e}-87$ |
| [+] COG1233 | COG1233 | Phytoene dehydrogenase-related protein [Secondary metabolites biosynthesis, transport and ... | 50-553 | 7.44e-61 |
| [H] SQS_PSY | pfam00494 | Squalene/phytoene synthase; | 597-856 | 2.57e-52 |
| [H] Trans_IPPS_HH | cd00683 | Trans-Isoprenyl Diphosphate Synthases, head-to-head; These trans-Isoprenyl Diphosphate ... | 590-857 | $8.87 \mathrm{e}-50$ |
| [ + ] ERG9 | COG1562 | Phytoene/squalene synthetase [Lipid transport and metabolism]; | 597-865 | $7.36 \mathrm{e}-45$ |
| $[+] \mathrm{HpnD}$ | TIGR03465 | squalene synthase HpnD; The genes of this family are often found in the same genetic locus ... | 610-860 | 1.12e-33 |
| [H] PLN02632 | PLN02632 | phytoene synthase | 587-861 | $1.68 \mathrm{e}-27$ |
| [ + ] Amino_oxidase | pfam01593 | Flavin containing amine oxidoreductase; This family consists of various amine oxidases, ... | 61-546 | $3.38 \mathrm{e}-23$ |
| [+] CarR_dom_SF | TIGR03462 | lycopene cyclase domain; This domain is often repeated twice within the same polypeptide, as ... | 910-1004 | $9.66 \mathrm{e}-17$ |
| [H] CarR_dom_SF | TIGR03462 | lycopene cyclase domain; This domain is often repeated twice within the same polypeptide, as ... | 1051-1141 | $2.98 \mathrm{e}-12$ |
| [H] PRK07233 | PRK07233 | hypothetical protein; Provisional | 52-97 | $2.05 \mathrm{e}-07$ |

B


Figure S6. The conserved domains of CrtIBY and alignment of amino acid sequences of different CrtIBY. The conserved domains of CrtIBY from Aurantiochytrium sp.SK4 were analyzed by CDD/SPARCLE (A) and alignment of amino acid sequences of possible trifunctional $\beta$-carotene synthases, $\mathrm{CrtIBY}(\mathbf{B})$. Sequences of CrtIBY of Aurantiochytrium sp. SK4 are compared with those of Aurantiochytrium sp. FCC1311, Aurantiochytrium sp. KH105, Schizochytrium sp. CCTCC M209059, Aurantiochytrium sp. T66, and Thraustochytrium sp. ATCC 26185.

## Table S1

Comparison of Aurantiochytrium sp. SK4 genome statistics to other five algae and the Arabidopsis thaliana genome.

| Organism | Aurantioc <br> hytrium <br> sp. SK4 | Arabidops <br> is <br> thaliana | Chlamydo <br> monas <br> reinhardti <br> $\boldsymbol{i}$ | Chlorella <br> sp. NC64 <br> A | Chromoc <br> hloris <br> zofingiens <br> is | Coccomyxa <br> subellipsoid <br> ea C-169 | Monorap <br> hidium <br> neglectum |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| Sequenced genome size <br> Percent G+C in | 49 Mbp | 119 Mbp | 107 Mbp | 42 Mbp | 58 Mbp | 49 Mbp | 67 Mbp |
| sequenced genome | $56.7 \%$ | $36 \%$ | $64 \%$ | $67 \%$ | $51 \%$ | $53 \%$ | $65 \%$ |
| Coding sequence in <br> sequenced genome | $63.0 \%$ | $28 \%$ | $37 \%$ | $32 \%$ | $39 \%$ | $25 \%$ | $26 \%$ |
| Percent G+C in coding <br> sequence | $56.9 \%$ | $44 \%$ | $70 \%$ | $69 \%$ | $53 \%$ | $61 \%$ | $70 \%$ |
| Average number of exons <br> Average exon length <br> Percentage transcript <br> with at least one intron | 2.4 | 903 nt | $55.5 \%$ | 237 nt | 261 nt | 166 nt | 291 nt |

## Table S2

The expression of genes associated with carotenoids and fatty acid biosynthesis in transcriptome.

| Pathway | Gene | RPKM |  |
| :---: | :---: | :---: | :---: |
|  |  | 24h | 96h |
| Astaxanthin pathway | HMGS | 293.68 | 2.67 |
|  | HMGR | 141.10 | 1.42 |
|  | MK | 194.72 | 10.42 |
|  | PMK | 21.56 | 12.61 |
|  | PPMD | 60.02 | 1.11 |
|  | IDI | 29.57 | 1.29 |
|  | Crtiby | 9.10 | 19.95 |
|  | Crtz | 3.87 | 12.66 |
| FAS pathway | Crto | 14.94 | 28.95 |
|  | Type 1 fatty acid synthase | 613.72 | 247.47 |
|  | $\Delta 12$ desaturase | 377.4 | 2.0 |
|  | $\triangle 5$ desaturase | 18.2 | 29.1 |
|  | $\omega-3$ desaturase | 0 | 0.6 |
|  | 44 desaturase | 114.4 | 19.7 |
|  | $\triangle 6$ desaturase | 0.7 | 0 |
|  | $\triangle 9$ desaturase | 1.08 | 0 |
| PKS pathway | PKS pfaA | 754.56 | 1.95 |
|  | PKS pfaB | 193.79 | 9.46 |
|  | PKS pfaC | 381.54 | 4.38 |

## Table S3

Contents of squalene in wild-type SK4 and the transformant AT26 at different stages.

|  | Sampling time | SK4 | AT26 |
| :---: | :---: | :---: | :---: |
|  | 48 h | $10.98 \pm 0.13$ | $0.34 \pm 0.02$ |
| Squalene content $\left(\mathbf{m g} \times \mathbf{g}^{\mathbf{- 1}} \mathbf{D W}\right)$ | 72 h | $21.08 \pm 0.06$ | $0.41 \pm 0.06$ |
|  | 96 h | $13.18 \pm 0.05$ | $0.64 \pm 0.07$ |

Data are shown as mean $\pm \mathrm{SD}, \mathrm{n}=3$.

## Table S4

Primers Used for qRT-PCR and the detection of the $\mathrm{p}^{--V H b-b l e-2 A-I D I-2 A-G P S ~(V B I G) . ~}$

| Primer | Sequence |
| :---: | :---: |
| qRT-pcr-Actin-F | GAGGCCATGTTTCAGACCAT |
| qRT-pcr-Actin-R | ACGAGAGCCGTCATTTCTGT |
| qRT-pcr-HMGS--- | CGCCGGCGTCGACAGCAT |
| qRT-pcr-HMGS-R | GGGCACGGCGGGCAAGAC |
| qRT-pcr-HMGR-F | CCGGCGCAAAATGTCGAGTCT |
| qRT-pcr-HMGR-R | CCGCCGACAGTGCCAACCTC |
| qRT-pcr-MK-F | CCGCAACCACGAAATCCTCCAAAA |
| qRT-pcr-MK-R | CGAGAGCGCCGGCAGACTTG |
| qRT-pcr-PMK-F | GCCGTCTTTGCAGTTGTTGTTGATTC |
| qRT-pcr-PMK-R | GCCGCCGATCTTCACTCAGCAA |
| qRT-pcr-crtIBY-F | TGGTGACCTCGATCATGTGT |
| qRT-pcr-crtIBY-R | CGGCTCTACAGGTAATGAGT |
| qRT-pcr-FASI-R | GAGAACGTCAGCACCTTTGC |
| qRT-pcr--FASI--F | AGGCTCGAGAGAGCCTTGAC |
| qRT-pcr-pfaA-F | TGATCCCTTCGTGAATGACC |
| qRT-pcr-pfaA-R | GCTCGTTGTGGAACTGAAGG |
| qRT-pcr-pfaB-F | GTCATTCTGCCCCTCATCATCAACC |
| qRT-pcr-pfaB-R | GACTGCTTGGCGACCTGGTTCAC |
| qRT-pCr-pfaC-F | CCGCCCCATCCACGTCATCCTC |
| qRT-pcr-pfaC-R | CCGGACTGCTTGGCGACCTGGTT |
| VBIGF | GGCTTTGGCGATGACGGTATTG |
| VBIGR | CCCCTCCTCATCTCGTCCCTGT |

