PNAS
www.pnas.org

Supplementary Information for

Green diatom mutants reveal an intricate biosynthetic pathway of fucoxanthin

Yu Bai[1], Tianjun Cao[1], Oliver Dautermann[1], Paul Buschbeck, Michael B. Cantrell, Yinjuan Chen, Christopher D. Lein, Xiaohuo Shi, Maxwell A. Ware, Fenghua Yang, Huan Zhang, Lihan Zhang, Graham Peers[2], Xiaobo Li[2], and Martin Lohr[2]

[1]Y.B., T.C. and O.D. contributed equally to this work.

[2]Correspondence to: Graham.Peers@colostate.edu, lixiaobo@westlake.edu.cn or lohr@uni-mainz.de

This PDF file includes:

Supplementary text
Figures S1 to S18
Table S1
Legends for Datasets S1 to S5
SI References

Other supplementary materials for this manuscript include the following:

Datasets S1 to S5

Supplementary Information Text

Text S1. Sequencing of PCR genotyping products from genomic DNA of wild type, *vdl2*- and *zep1*- knockout mutants of *P. tricornutum*.

All transformants obtained on selective agar plates should harbor the *Ble* transgene, but only a fraction of them can be expected to be VDL2- or ZEP1-deficient. The integration of *Ble* may be on-target in the event of homology-directed repair (HDR), or off-target through other mechanisms such as non-homologous end joining (NHEJ). In addition, because of the diploid nature of the *P. tricornutum* genome, if the HDR event only occurred on a single chromosome, a monoallelic mutant would be yielded, with one wild-type allele retained. This potential genotypic diversity of transformants required us to carefully examine the relationship between their genotypes and pigment phenotypes. Therefore, Sanger sequencing was performed on the PCR products shown in **Fig. 1** and **Fig. S1** from both ends, using the original PCR primers and additional primers (see **Datasets S1A** and **S1C**).

For *VDL2*, when the primer pair was designed to bind upstream and downstream of the protospacer region, respectively (blue primer pair in **Fig. 1*A***), wild type DNA yielded a band corresponding to the expected size of 1054 bp, whereas DNA from the green mutants yielded a larger band with an expected size of 2342 bp. The sequencing results for the wild-type PCR product were consistent with amplification of fragments from both *VDL2* alleles in the genome with equal efficiency, as indicated by diagnostic single nucleotide polymorphisms (SNPs) in the sequence reads (**Fig. S2**). The biallelic nature of the SNPs was confirmed by BLAST searches in publicly available whole genome shotgun data of *P. tricornutum* CCAP 1055/1 from project SRX8974960 in the NCBI sequence read archive (SRA). These searches also confirmed that all our PCR primers bound to target regions devoid of naturally occurring biallelic SNPs. When the primer pair was designed to amplify the junction between the *Ble* insertion and *VDL2* (red primer pair in **Fig. 1*A***), only genomic DNA template from the green mutant lines generated a band in the expected size range of 1847 bp (**Fig. 1** and **Fig. S1**). Sequencing of the PCR products from the five mutants confirmed the specific integration of the *Ble* construct at the target site of the *VDL2* gene (**Fig. S2**). However, none of these PCR products contained any of the SNPs discerning the two wild type alleles, indicating a loss of heterozygosity around the integration site.

The genotyping results on *ZEP1* wild type and the five *zep1* mutants were similar to the observations for *VDL2* (**Fig. 1** and **Fig. S1**), with the notable exception that the PCR products from genomic DNA of the mutants zep1-1 and zep1-5 showed the same SNP pattern as the PCR product amplified from wild-type genomic DNA (**Fig. S3**). This observation indicates that in these mutants both alleles have successfully been interrupted by the *Ble* cassette and were retained in the genome.

Text S2. PCR genotyping of brown colonies isolated from the selective plates after transformation with *VDL2*- or *ZEP1*-knockout constructs.

We further validated our screening and genotyping approach by genotyping randomly chosen brown colonies from the selective plates after transformation with the *VDL2*- or the *ZEP1*-knockout constructs. None of these brown colonies were biallelic knockouts. For *ZEP1*, 14 brown colonies were examined (**Fig. S4*B***). PCR-genotyping showed that all transformant lines contained at least 1 intact allele of the target gene. For 9 of the 14 lines, we detected an additional product indicating genomic integration of our *Ble* construct, and for 2 of these 9 lines a PCR product showing target-specific *Ble* insertion by HDR (see **Fig. S4*B***). This indicates that these 2 lines were monoallelic knockouts of the target gene, whereas the *Ble* construct likely was randomly inserted in the other lines. In the case of *VDL2*, again all 13 brown colonies examined contained at least one intact target allele (note weak bands of expected product sizes for transformant 7), the *VDL2*-knockout construct was detected in 7 of these lines, and none of the lines showed target-site specific integration of the transgene (see **Fig. S4*A***).

Text S3. Implications of the mutant genotyping and sequence data.

We did not systematically investigate the efficiency of HDR for our target genes, but during initial screening of the first transformation plates we observed growth of 3 green colonies and 137 brown colonies (= 2.1 % green colonies) for *vdl2* and growth of 1 green and 27 brown colonies (= 3.6% green colonies) for *zep1*. While the frequency of biallelic knockouts of *VDL2* and *ZEP1* by HDR appears to be lower than in previous reports (1, 2), this may have been caused by the strongly reduced light use efficiency of the green mutants that yielded only very small colonies even after prolonged illumination which may have prevented identification of some green colonies.

We were initially surprised to not observe many more monoallelic HDR mutants compared to biallelic HDR ones, considering that the latter mutants may require HDR to occur on both alleles. This observation, however, and the lack of SNP signals in the sequencing data from the five *vdl2* mutants and three of the five *zep1* mutants are consistent with previous reports that genome editing in *P. tricornutum* often leads to biallelic mutants with identical indels on both chromosomes, i.e., loss of heterozygosity, when no HDR template is provided (3). This can be explained by recombination between the two homologous chromosomes without meiosis by a double-strand break induced gene conversion mechanism. Indeed, it has been demonstrated that mitotic interhomolog recombination leading to loss of heterozygosity occurs in *P. tricornutum* at a frequency ten times higher than in the budding yeast *Saccharomyces cerevisiae* (4). In conclusion, PCR genotyping and HPLC data of all investigated green mutants give strong support for successful biallelic knockout of the target genes even in those mutants whose sequencing results did not indicate the presence of heterozygous alleles. This conclusion is further supported by the observation that the phenotype of all knockout mutant lines has been stable since their isolation more than two years ago.

Text S4. NMR data and structural determination of haptoxanthin.

Haptoxanthin was obtained as orange-yellow solid and its molecular formula was determined to be $C_{42}H_{56}O_4$ by positive-ion HR-MS data (molecular mass for $[M+H]^+$ calculated as 625.4251 and measured as 625.4241) (**Fig. S6**). The structure of haptoxanthin was elucidated by thorough analysis of ^{1}H NMR, ^{13}C NMR, DEPT spectroscopic data (**Figs. S7** to **S9**; **Table S1**) as well as 2D-NMR of COSY, ROESY, HSQC and HMBC spectra (**Figs. S10** to **S13**). Briefly, DEPT and HSQC correlations suggested that haptoxanthin contained 11 CH_3 carbons, 4 CH_2 carbons, 13 CH carbons and 14 fully substituted carbons and had a carbon skeleton similar to fucoxanthin. The analyses of COSY, ROESY and HMBC spectra showed that haptoxanthin contained the same allene and ring B skeleton as that of fucoxanthin, while the ring A region of haptoxanthin was significantly different from fucoxanthin. Two *sp* carbons C7 and C8 (δC 89.0, 98.6) were assigned as a triple bond, and the HMBC correlations between H18 and C7, and H19 and C8 indicated the conjugated alkyne on the cyclohexenyl ring A. The remaining part of the structure was also determined by NMR comparison to fucoxanthin.

The stereochemistry and conformation of ring A was further determined by ROESY and *J*-coupling constants. As shown in **Fig. S14**, H3 had ROESY correlations with H2a, H4a and Me17 but no correlations with H2b or H4b, indicating that H3 and Me17 were both axial bonds, H2a and H4a were equatorial bonds, and H2b and H4b were axial bonds in the opposite of H3. These results were also consistent with the coupling constant values *J*H3-H2b = 12.2 Hz and *J*H3-H4b = 9.8 Hz. The stereochemistry and conformation of ring B was identified by comparison to fucoxanthin as they showed almost identical NMR spectra.

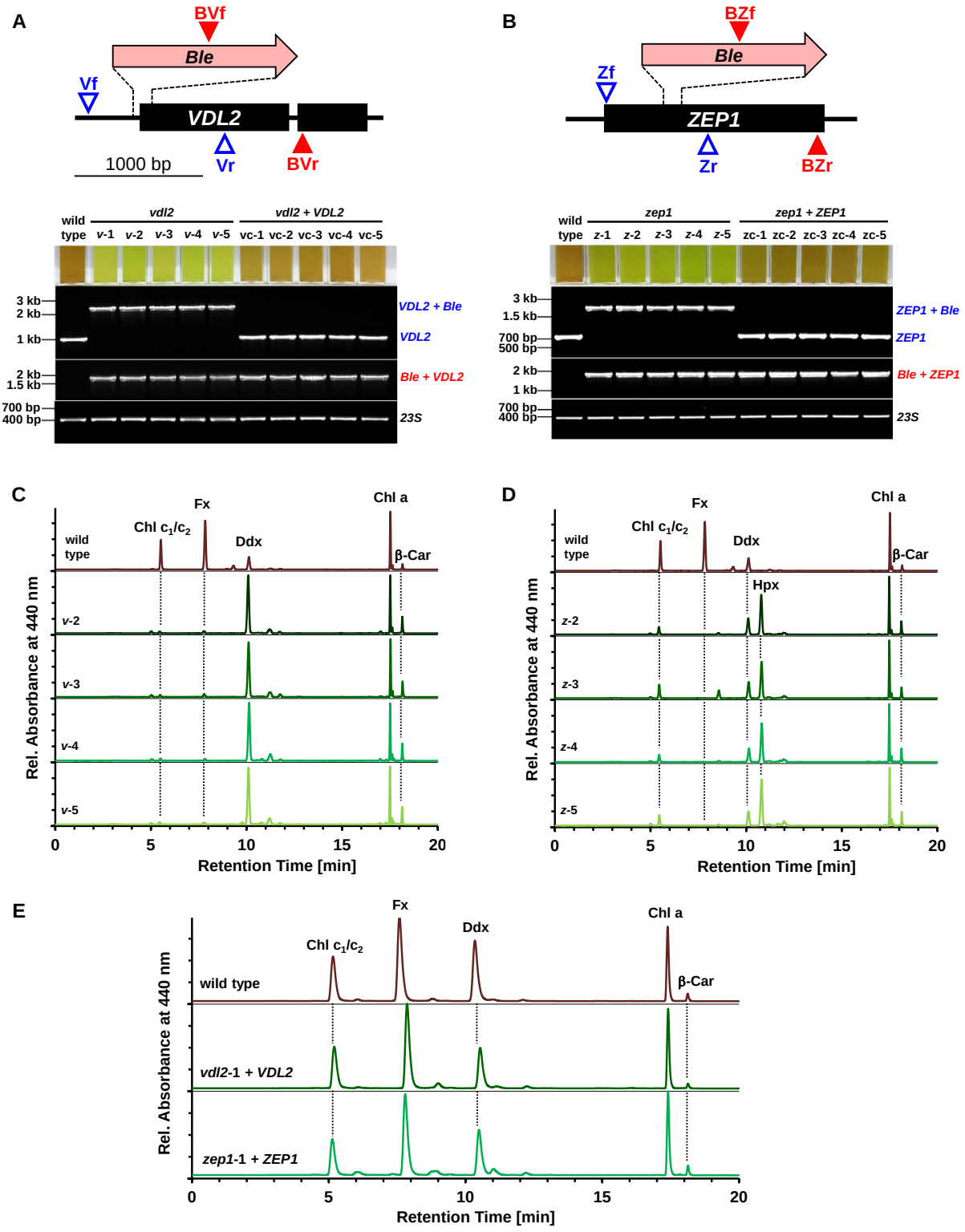


Fig. S1. Molecular characterization and pigment phenotypes of additional *VDL2*- and *ZEP1*-knockout mutants from *P. tricornutum* and of mutants after complementation with the corresponding native genes. (*A* to *B*) Schemes showing insertion sites of the 1.4 kB zeocin

resistance cassette (*Ble*) in the target genes; PCR primer binding sites for differentiation of wild type and mutants indicated by blue triangles, for detection of the inserted Ble genes by red triangles; photographs show cuvettes with cultures of wild type, five independently generated green-colored knockout lines and five independently complemented lines of the respective first knockout line for *VDL2* (*v*-1 in A = *vdl2*-1 in **Fig. 1*A***) and *ZEP1* (*z*-1 in B = *zep1*-1 in **Fig. 1*B***) that have regained the brown color of the wild type. The PCR primers used for genotyping of all lines and the resulting product sizes are the same as in **Fig. 1**, confirming integration of the 1.4 kB zeocin resistance cassette (Ble) into the respective target gene for both the green-colored mutant lines and complemented lines as well as the additional presence of native gene copies in the latter (in the complemented lines, amplification of fragments of re-introduced randomly integrated native genes was strongly favored over genes with Ble insertion). (*C* to *D*) HPLC analyses (system II) of pigment extracts from wild type and the four additional knockout mutants of VDL2 (*C*) and ZEP1 (*D*) that were not included in **Fig. 1**. (*E*) HPLC analyses (system II) of pigment extracts from wild type and the first complemented knockout line for *VDL2* (*vc*-1 in *A*) and *ZEP1* (*zc*-1 in *B*) confirming that the complemented strains have regained the ability to synthesize fucoxanthin. Car, carotene; Chl, chlorophyll; Ddx, diadinoxanthin; Fx, fucoxanthin.

Figure S2

A

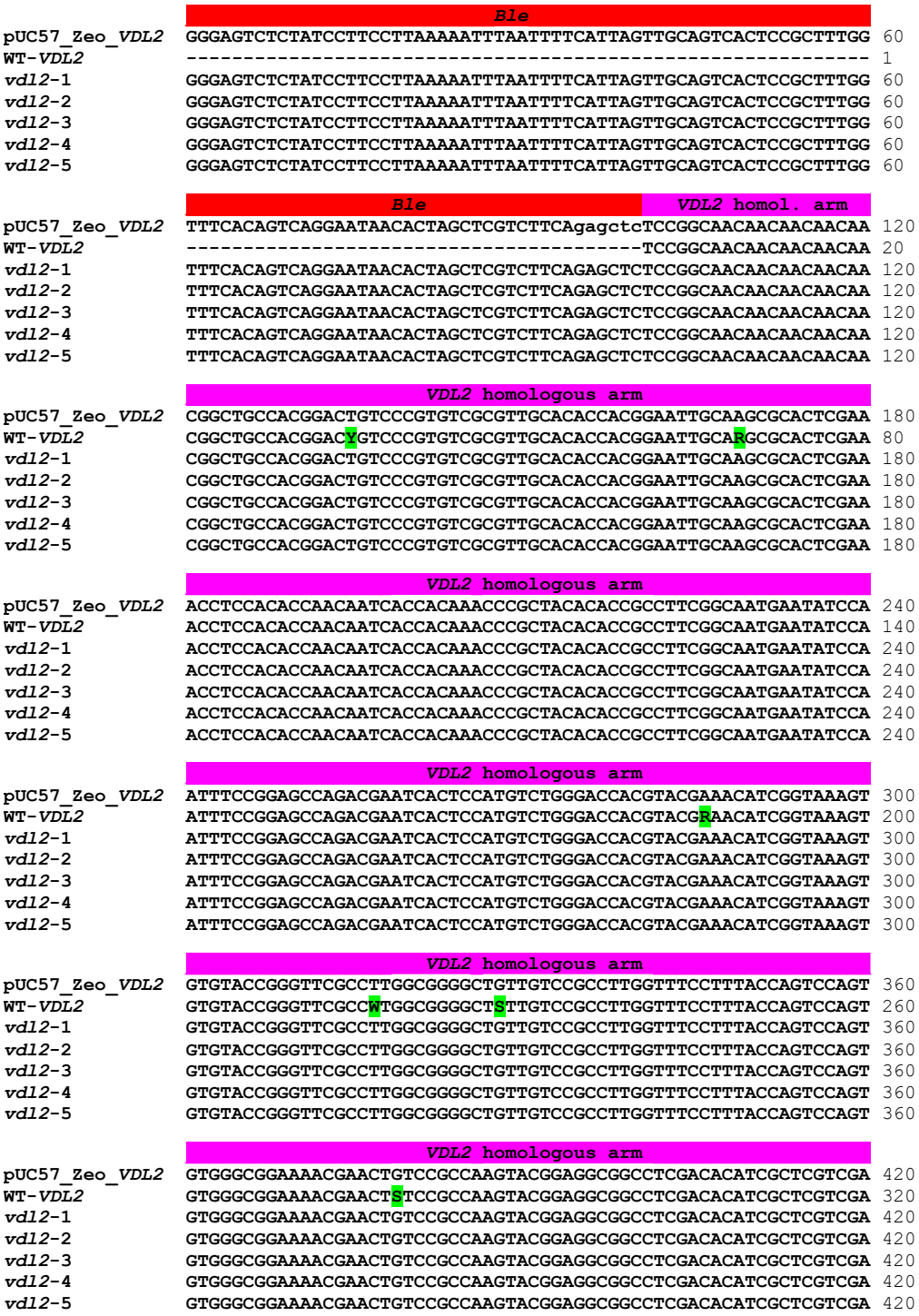

Ble
pUC57_Zeo_VDL2 GGGAGTCTCTATCCTTCCTTAAAAATTTAATTTTCATTAGTTGCAGTCACTCCGCTTTGG 60
WT-VDL2 -- 1
vdl2-1 GGGAGTCTCTATCCTTCCTTAAAAATTTAATTTTCATTAGTTGCAGTCACTCCGCTTTGG 60
vdl2-2 GGGAGTCTCTATCCTTCCTTAAAAATTTAATTTTCATTAGTTGCAGTCACTCCGCTTTGG 60
vdl2-3 GGGAGTCTCTATCCTTCCTTAAAAATTTAATTTTCATTAGTTGCAGTCACTCCGCTTTGG 60
vdl2-4 GGGAGTCTCTATCCTTCCTTAAAAATTTAATTTTCATTAGTTGCAGTCACTCCGCTTTGG 60
vdl2-5 GGGAGTCTCTATCCTTCCTTAAAAATTTAATTTTCATTAGTTGCAGTCACTCCGCTTTGG 60
Ble
VDL2 homol. arm
pUC57_Zeo_VDL2 TTTCACAGTCAGGAATAACACTAGCTCGTCTTCAgagctcTCCGGCAACAACAACAACAA 120
WT-VDL2 --TCCGGCAACAACAACAACAA 20
vdl2-1 TTTCACAGTCAGGAATAACACTAGCTCGTCTTCAGAGCTCTCCGGCAACAACAACAACAA 120
vdl2-2 TTTCACAGTCAGGAATAACACTAGCTCGTCTTCAGAGCTCTCCGGCAACAACAACAACAA 120
vdl2-3 TTTCACAGTCAGGAATAACACTAGCTCGTCTTCAGAGCTCTCCGGCAACAACAACAACAA 120
vdl2-4 TTTCACAGTCAGGAATAACACTAGCTCGTCTTCAGAGCTCTCCGGCAACAACAACAACAA 120
vdl2-5 TTTCACAGTCAGGAATAACACTAGCTCGTCTTCAGAGCTCTCCGGCAACAACAACAACAA 120
VDL2 homologous arm
pUC57_Zeo_VDL2 CGGCTGCCACGGACTGTCCCGTGTCGCGTTGCACACCACGGAATTGCAAGCGCACTCGAA 180
WT-VDL2 CGGCTGCCACGGACYGTCCCGTGTCGCGTTGCACACCACGGAATTGCARGCGCACTCGAA 80
vdl2-1 CGGCTGCCACGGACTGTCCCGTGTCGCGTTGCACACCACGGAATTGCAAGCGCACTCGAA 180
vdl2-2 CGGCTGCCACGGACTGTCCCGTGTCGCGTTGCACACCACGGAATTGCAAGCGCACTCGAA 180
vdl2-3 CGGCTGCCACGGACTGTCCCGTGTCGCGTTGCACACCACGGAATTGCAAGCGCACTCGAA 180
vdl2-4 CGGCTGCCACGGACTGTCCCGTGTCGCGTTGCACACCACGGAATTGCAAGCGCACTCGAA 180
vdl2-5 CGGCTGCCACGGACTGTCCCGTGTCGCGTTGCACACCACGGAATTGCAAGCGCACTCGAA 180
VDL2 homologous arm
pUC57_Zeo_VDL2 ACCTCCACACCAACAATCACCACAAACCCGCTACACACCGCCTTCGGCAATGAATATCCA 240
WT-VDL2 ACCTCCACACCAACAATCACCACAAACCCGCTACACACCGCCTTCGGCAATGAATATCCA 140
vdl2-1 ACCTCCACACCAACAATCACCACAAACCCGCTACACACCGCCTTCGGCAATGAATATCCA 240
vdl2-2 ACCTCCACACCAACAATCACCACAAACCCGCTACACACCGCCTTCGGCAATGAATATCCA 240
vdl2-3 ACCTCCACACCAACAATCACCACAAACCCGCTACACACCGCCTTCGGCAATGAATATCCA 240
vdl2-4 ACCTCCACACCAACAATCACCACAAACCCGCTACACACCGCCTTCGGCAATGAATATCCA 240
vdl2-5 ACCTCCACACCAACAATCACCACAAACCCGCTACACACCGCCTTCGGCAATGAATATCCA 240
VDL2 homologous arm
pUC57_Zeo_VDL2 ATTTCCGGAGCCAGACGAATCACTCCATGTCTGGGACCACGTACGAAACATCGGTAAAGT 300
WT-VDL2 ATTTCCGGAGCCAGACGAATCACTCCATGTCTGGGACCACGTACGRAACATCGGTAAAGT 200
vdl2-1 ATTTCCGGAGCCAGACGAATCACTCCATGTCTGGGACCACGTACGAAACATCGGTAAAGT 300
vdl2-2 ATTTCCGGAGCCAGACGAATCACTCCATGTCTGGGACCACGTACGAAACATCGGTAAAGT 300
vdl2-3 ATTTCCGGAGCCAGACGAATCACTCCATGTCTGGGACCACGTACGAAACATCGGTAAAGT 300
vdl2-4 ATTTCCGGAGCCAGACGAATCACTCCATGTCTGGGACCACGTACGAAACATCGGTAAAGT 300
vdl2-5 ATTTCCGGAGCCAGACGAATCACTCCATGTCTGGGACCACGTACGAAACATCGGTAAAGT 300
VDL2 homologous arm
pUC57_Zeo_VDL2 GTGTACCGGGTTCGCCTTGGCGGGGCTGTTGTCCGCCTTGGTTTCCTTTACCAGTCCAGT 360
WT-VDL2 GTGTACCGGGTTCGCCWTGGCGGGGCTSTTGTCCGCCTTGGTTTCCTTTACCAGTCCAGT 260
vdl2-1 GTGTACCGGGTTCGCCTTGGCGGGGCTGTTGTCCGCCTTGGTTTCCTTTACCAGTCCAGT 360
vdl2-2 GTGTACCGGGTTCGCCTTGGCGGGGCTGTTGTCCGCCTTGGTTTCCTTTACCAGTCCAGT 360
vdl2-3 GTGTACCGGGTTCGCCTTGGCGGGGCTGTTGTCCGCCTTGGTTTCCTTTACCAGTCCAGT 360
vdl2-4 GTGTACCGGGTTCGCCTTGGCGGGGCTGTTGTCCGCCTTGGTTTCCTTTACCAGTCCAGT 360
vdl2-5 GTGTACCGGGTTCGCCTTGGCGGGGCTGTTGTCCGCCTTGGTTTCCTTTACCAGTCCAGT 360
VDL2 homologous arm
pUC57_Zeo_VDL2 GTGGGCGGAAAACGAACTGTCCGCCAAGTACGGAGGCGGCCTCGACACATCGCTCGTCGA 420
WT-VDL2 GTGGGCGGAAAACGAACTSTCCGCCAAGTACGGAGGCGGCCTCGACACATCGCTCGTCGA 320
vdl2-1 GTGGGCGGAAAACGAACTGTCCGCCAAGTACGGAGGCGGCCTCGACACATCGCTCGTCGA 420
vdl2-2 GTGGGCGGAAAACGAACTGTCCGCCAAGTACGGAGGCGGCCTCGACACATCGCTCGTCGA 420
vdl2-3 GTGGGCGGAAAACGAACTGTCCGCCAAGTACGGAGGCGGCCTCGACACATCGCTCGTCGA 420
vdl2-4 GTGGGCGGAAAACGAACTGTCCGCCAAGTACGGAGGCGGCCTCGACACATCGCTCGTCGA 420
vdl2-5 GTGGGCGGAAAACGAACTGTCCGCCAAGTACGGAGGCGGCCTCGACACATCGCTCGTCGA 420

Figure S2

A *continued*

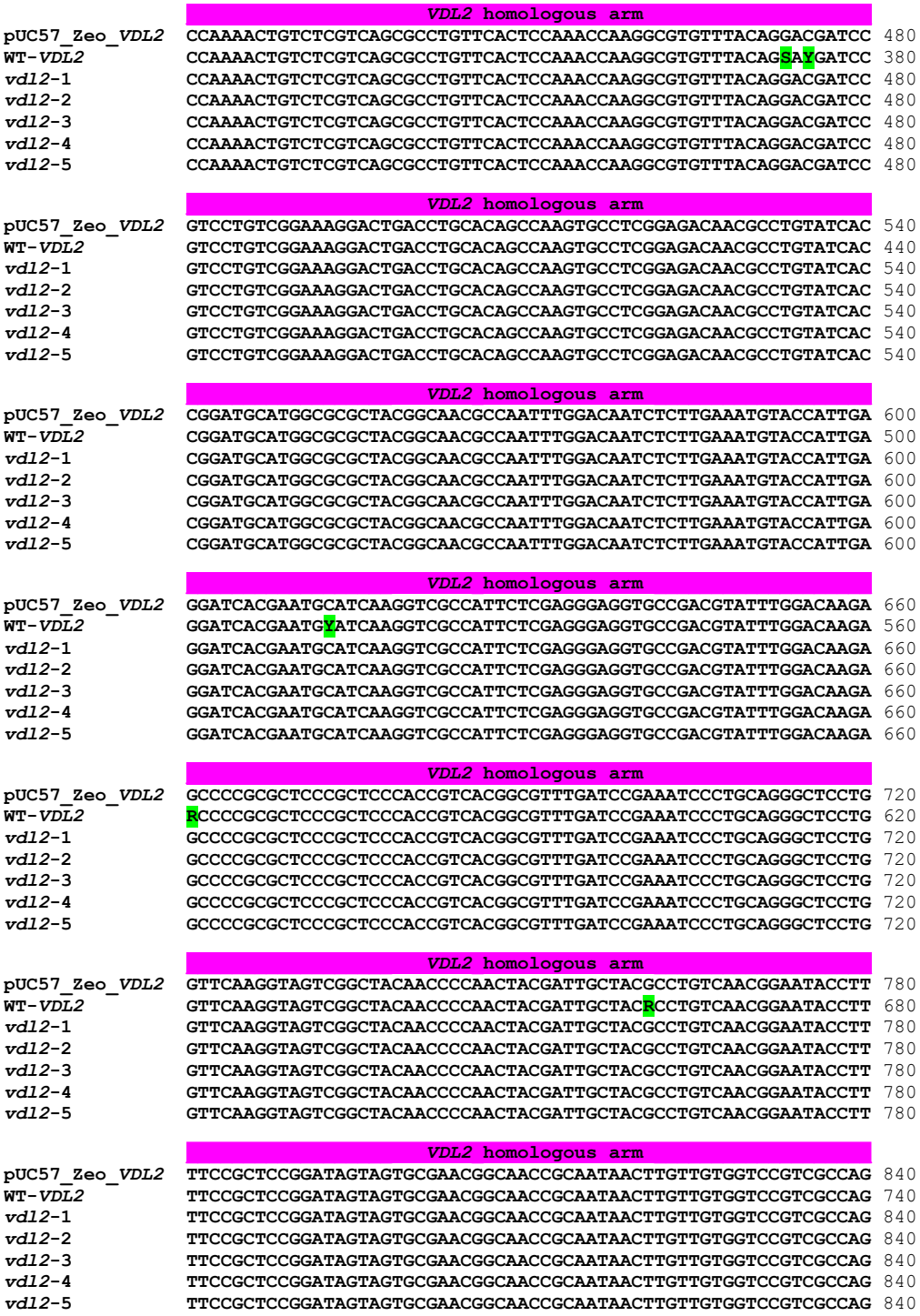


Figure S2

A *continued*

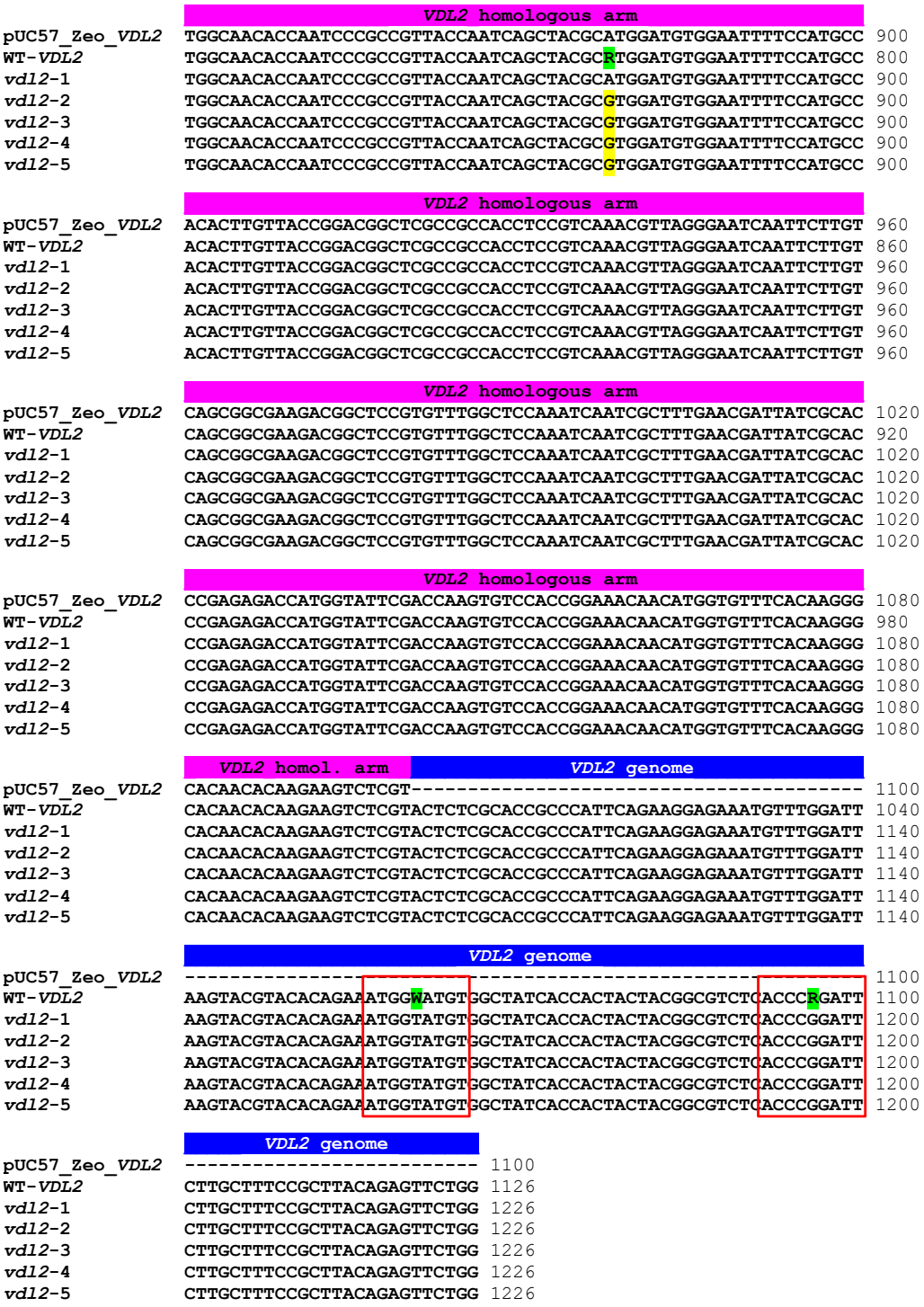


Figure S2

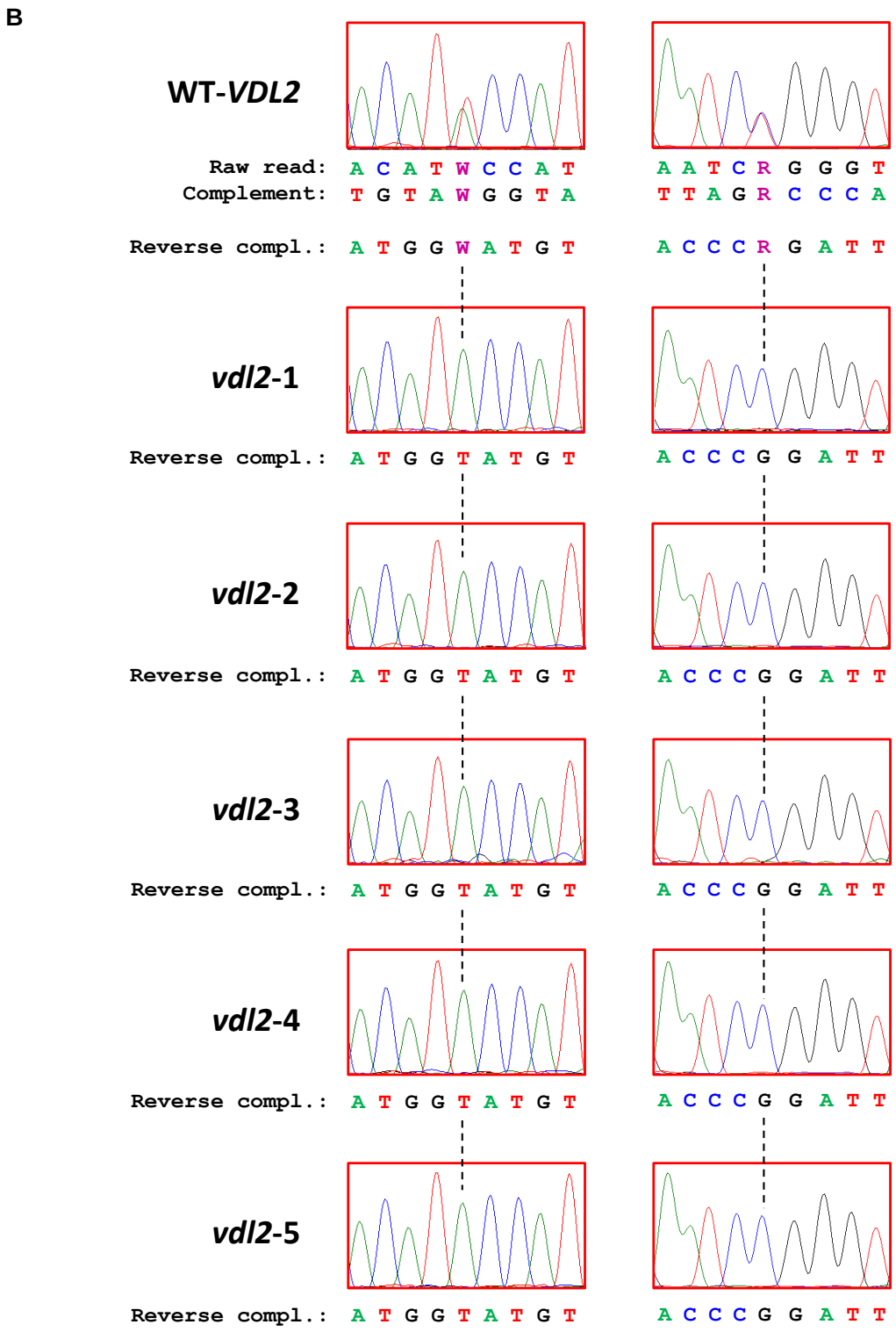


Fig. S2. Sequencing results of PCR products from *VDL2*-wild type and *vdl2*-knockout mutant lines of *P. tricornutum*. (*A*) Alignment of partial sequences of the *VDL2*-knockout construct

(pUC57_Zeo_*VDL2*) used for transformation, the PCR product from WT cells (WT-*VDL2*), and the PCR products from the five green knockout mutant lines shown in **Fig. S1** (*vdl2*-1 to *vdl2*-5) resulting from the red primer pair confirming specific integration of the *Ble* construct into the target gene. Colored bars above the alignment indicate partial sequence of the knockout construct consisting of the *Ble* gene (red) and the 1000 bp *VDL2*-homologous arm (purple) used for HDR, and the VDL2-genomic region downstream of the integration site (*VDL2* genome; blue). Diagnostic single nucleotide polymorphisms (SNPs) between the two *VDL2* alleles detected in the sequence of the WT-*VDL2* product are highlighted by green background and indicated by IUPAC nucleotide codes R (A or G), S (G or C), W (A or T), and Y (C or T). Yellow background highlights bases in the mutant products differing from the corresponding base in the *VDL2*-homologous arm, indicating integration of only a part of the homology arm into the target region. (*B*) Exemplary reverse sequence reads of the two regions labeled by red boxes in *A* that cover diagnostic biallelic SNPs and indicate amplification of both alleles from the wild type genome with equal efficiency. The SNPs were not detected in the PCR products from the five mutants, suggesting that the genome editing led to biallelic mutants with identical alleles, consistent with previous observations on CRISPR/Cas9 mutants in *P. tricornutum* (3) (see **Texts S1** and **S3** for further discussion).

Figure S3

A

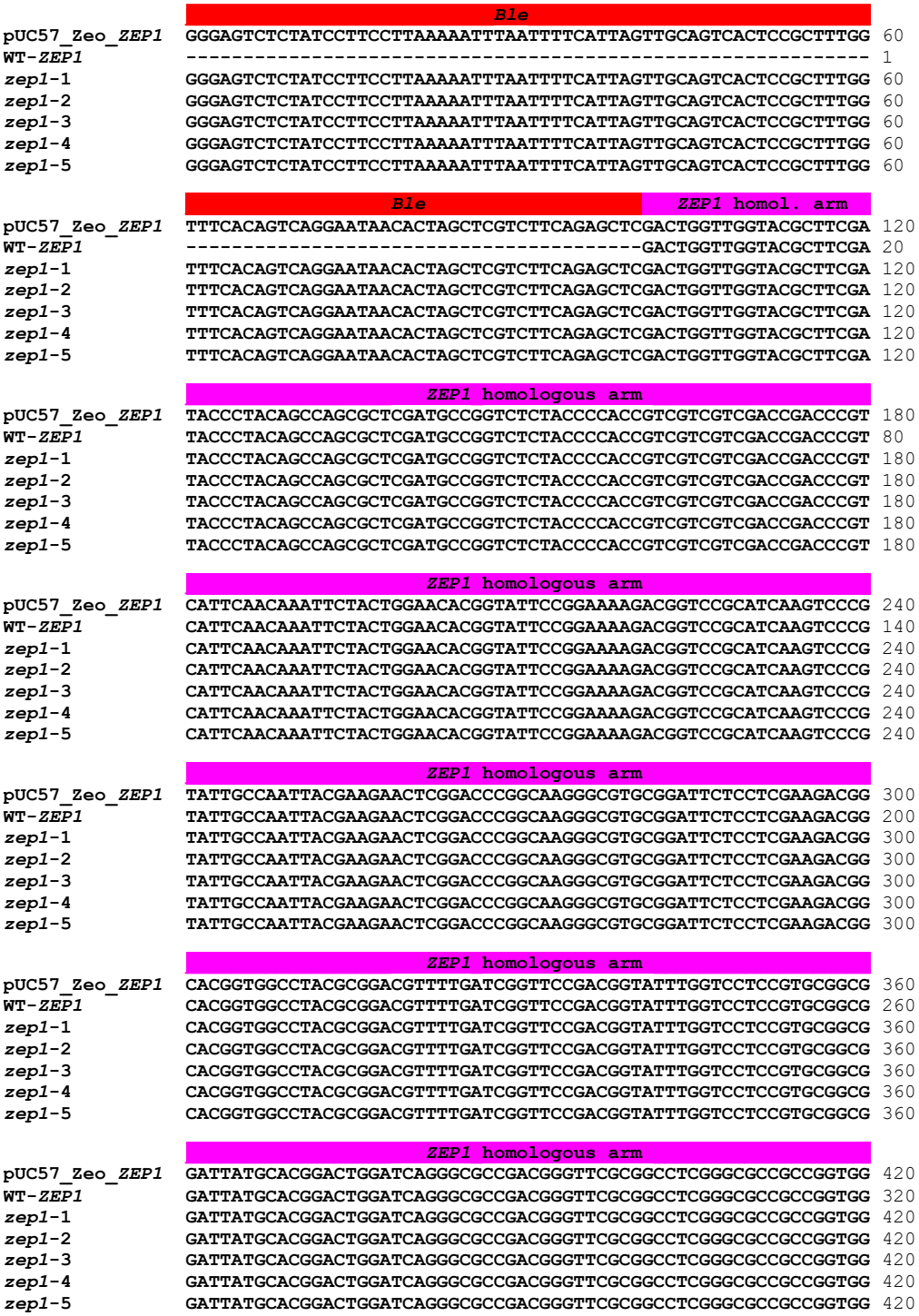

Ble
pUC57_Zeo_ZEP1 GGGAGTCTCTATCCTTCCTTAAAAATTTAATTTTCATTAGTTGCAGTCACTCCGCTTTGG 60
WT-ZEP1 -- 1
zep1-1 GGGAGTCTCTATCCTTCCTTAAAAATTTAATTTTCATTAGTTGCAGTCACTCCGCTTTGG 60
zep1-2 GGGAGTCTCTATCCTTCCTTAAAAATTTAATTTTCATTAGTTGCAGTCACTCCGCTTTGG 60
zep1-3 GGGAGTCTCTATCCTTCCTTAAAAATTTAATTTTCATTAGTTGCAGTCACTCCGCTTTGG 60
zep1-4 GGGAGTCTCTATCCTTCCTTAAAAATTTAATTTTCATTAGTTGCAGTCACTCCGCTTTGG 60
zep1-5 GGGAGTCTCTATCCTTCCTTAAAAATTTAATTTTCATTAGTTGCAGTCACTCCGCTTTGG 60
Ble
ZEP1 homol. arm
pUC57_Zeo_ZEP1 TTTCACAGTCAGGAATAACACTAGCTCGTCTTCAGAGCTCGACTGGTTGGTACGCTTCGA 120
WT-ZEP1 --GACTGGTTGGTACGCTTCGA 20
zep1-1 TTTCACAGTCAGGAATAACACTAGCTCGTCTTCAGAGCTCGACTGGTTGGTACGCTTCGA 120
zep1-2 TTTCACAGTCAGGAATAACACTAGCTCGTCTTCAGAGCTCGACTGGTTGGTACGCTTCGA 120
zep1-3 TTTCACAGTCAGGAATAACACTAGCTCGTCTTCAGAGCTCGACTGGTTGGTACGCTTCGA 120
zep1-4 TTTCACAGTCAGGAATAACACTAGCTCGTCTTCAGAGCTCGACTGGTTGGTACGCTTCGA 120
zep1-5 TTTCACAGTCAGGAATAACACTAGCTCGTCTTCAGAGCTCGACTGGTTGGTACGCTTCGA 120
ZEP1 homologous arm
pUC57_Zeo_ZEP1 TACCCTACAGCCAGCGCTCGATGCCGGTCTCTACCCCACCGTCGTCGTCGACCGACCCGT 180
WT-ZEP1 TACCCTACAGCCAGCGCTCGATGCCGGTCTCTACCCCACCGTCGTCGTCGACCGACCCGT 80
zep1-1 TACCCTACAGCCAGCGCTCGATGCCGGTCTCTACCCCACCGTCGTCGTCGACCGACCCGT 180
zep1-2 TACCCTACAGCCAGCGCTCGATGCCGGTCTCTACCCCACCGTCGTCGTCGACCGACCCGT 180
zep1-3 TACCCTACAGCCAGCGCTCGATGCCGGTCTCTACCCCACCGTCGTCGTCGACCGACCCGT 180
zep1-4 TACCCTACAGCCAGCGCTCGATGCCGGTCTCTACCCCACCGTCGTCGTCGACCGACCCGT 180
zep1-5 TACCCTACAGCCAGCGCTCGATGCCGGTCTCTACCCCACCGTCGTCGTCGACCGACCCGT 180
ZEP1 homologous arm
pUC57_Zeo_ZEP1 CATTCAACAAATTCTACTGGAACACGGTATTCCGGAAAAGACGGTCCGCATCAAGTCCCG 240
WT-ZEP1 CATTCAACAAATTCTACTGGAACACGGTATTCCGGAAAAGACGGTCCGCATCAAGTCCCG 140
zep1-1 CATTCAACAAATTCTACTGGAACACGGTATTCCGGAAAAGACGGTCCGCATCAAGTCCCG 240
zep1-2 CATTCAACAAATTCTACTGGAACACGGTATTCCGGAAAAGACGGTCCGCATCAAGTCCCG 240
zep1-3 CATTCAACAAATTCTACTGGAACACGGTATTCCGGAAAAGACGGTCCGCATCAAGTCCCG 240
zep1-4 CATTCAACAAATTCTACTGGAACACGGTATTCCGGAAAAGACGGTCCGCATCAAGTCCCG 240
zep1-5 CATTCAACAAATTCTACTGGAACACGGTATTCCGGAAAAGACGGTCCGCATCAAGTCCCG 240
ZEP1 homologous arm
pUC57_Zeo_ZEP1 TATTGCCAATTACGAAGAACTCGGACCCGGCAAGGGCGTGCGGATTCTCCTCGAAGACGG 300
WT-ZEP1 TATTGCCAATTACGAAGAACTCGGACCCGGCAAGGGCGTGCGGATTCTCCTCGAAGACGG 200
zep1-1 TATTGCCAATTACGAAGAACTCGGACCCGGCAAGGGCGTGCGGATTCTCCTCGAAGACGG 300
zep1-2 TATTGCCAATTACGAAGAACTCGGACCCGGCAAGGGCGTGCGGATTCTCCTCGAAGACGG 300
zep1-3 TATTGCCAATTACGAAGAACTCGGACCCGGCAAGGGCGTGCGGATTCTCCTCGAAGACGG 300
zep1-4 TATTGCCAATTACGAAGAACTCGGACCCGGCAAGGGCGTGCGGATTCTCCTCGAAGACGG 300
zep1-5 TATTGCCAATTACGAAGAACTCGGACCCGGCAAGGGCGTGCGGATTCTCCTCGAAGACGG 300
ZEP1 homologous arm
pUC57_Zeo_ZEP1 CACGGTGGCCTACGCGGACGTTTTGATCGGTTCCGACGGTATTTGGTCCTCCGTGCGGCG 360
WT-ZEP1 CACGGTGGCCTACGCGGACGTTTTGATCGGTTCCGACGGTATTTGGTCCTCCGTGCGGCG 260
zep1-1 CACGGTGGCCTACGCGGACGTTTTGATCGGTTCCGACGGTATTTGGTCCTCCGTGCGGCG 360
zep1-2 CACGGTGGCCTACGCGGACGTTTTGATCGGTTCCGACGGTATTTGGTCCTCCGTGCGGCG 360
zep1-3 CACGGTGGCCTACGCGGACGTTTTGATCGGTTCCGACGGTATTTGGTCCTCCGTGCGGCG 360
zep1-4 CACGGTGGCCTACGCGGACGTTTTGATCGGTTCCGACGGTATTTGGTCCTCCGTGCGGCG 360
zep1-5 CACGGTGGCCTACGCGGACGTTTTGATCGGTTCCGACGGTATTTGGTCCTCCGTGCGGCG 360
ZEP1 homologous arm
pUC57_Zeo_ZEP1 GATTATGCACGGACTGGATCAGGGCGCCGACGGGTTCGCGGCCTCGGGCGCCGCCGGTGG 420
WT-ZEP1 GATTATGCACGGACTGGATCAGGGCGCCGACGGGTTCGCGGCCTCGGGCGCCGCCGGTGG 320
zep1-1 GATTATGCACGGACTGGATCAGGGCGCCGACGGGTTCGCGGCCTCGGGCGCCGCCGGTGG 420
zep1-2 GATTATGCACGGACTGGATCAGGGCGCCGACGGGTTCGCGGCCTCGGGCGCCGCCGGTGG 420
zep1-3 GATTATGCACGGACTGGATCAGGGCGCCGACGGGTTCGCGGCCTCGGGCGCCGCCGGTGG 420
zep1-4 GATTATGCACGGACTGGATCAGGGCGCCGACGGGTTCGCGGCCTCGGGCGCCGCCGGTGG 420
zep1-5 GATTATGCACGGACTGGATCAGGGCGCCGACGGGTTCGCGGCCTCGGGCGCCGCCGGTGG 420
ZEP1 homologous arm

Figure S3

A *continued*

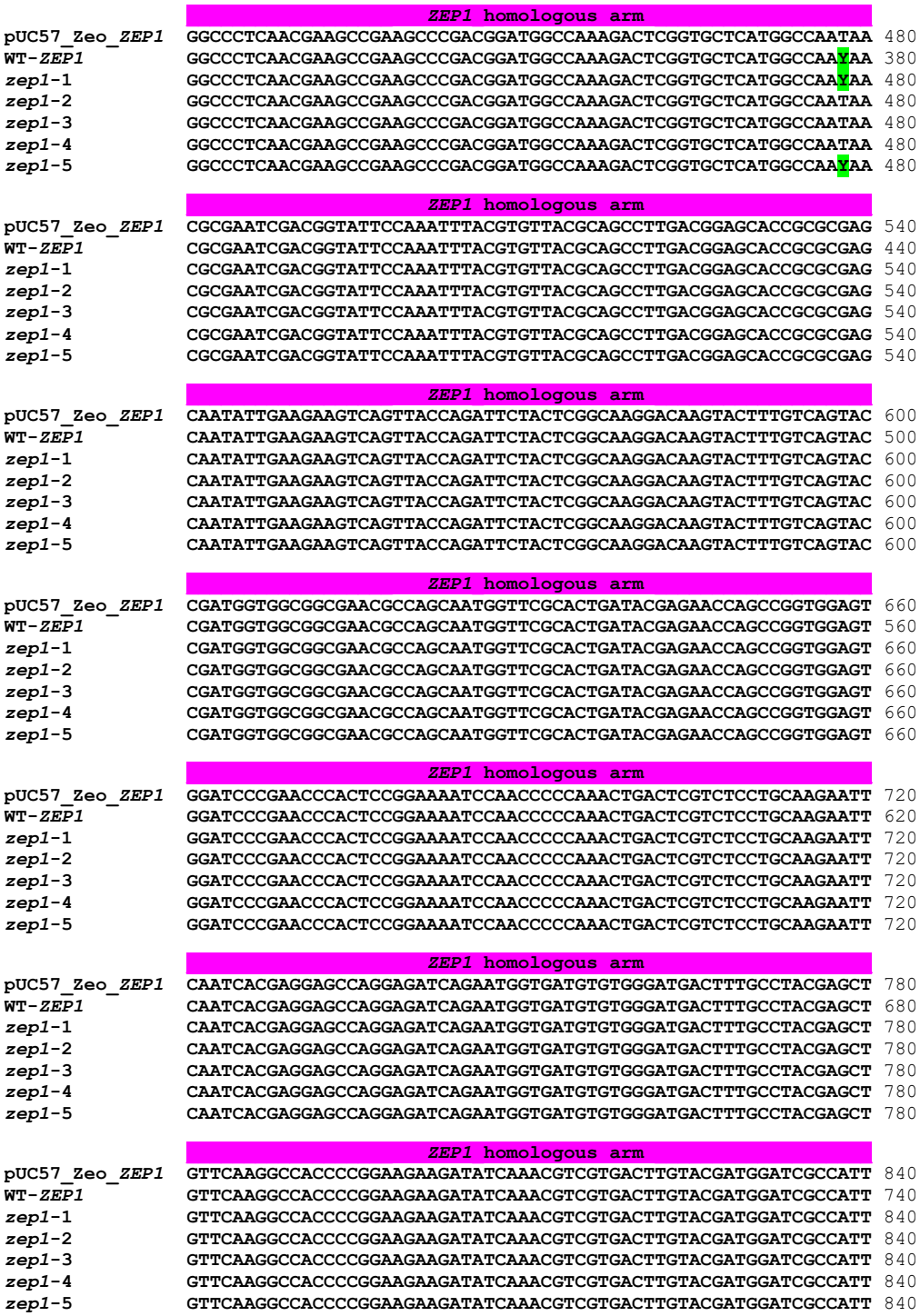


Figure S3

A *continued*

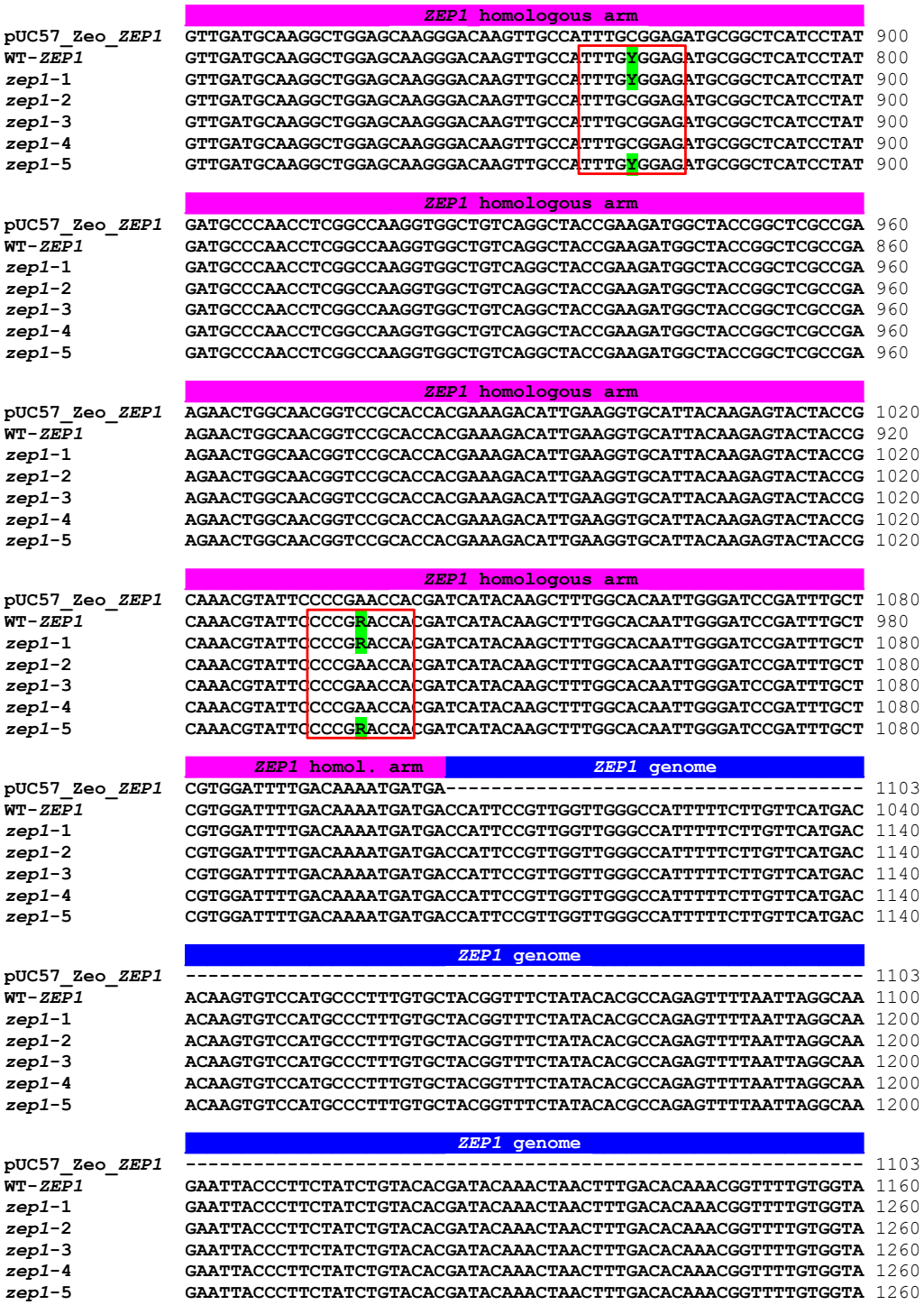


Figure S3

A *continued*

```
                                  ZEP1 genome
pUC57_Zeo_ZEP1    ---------------------------------------- 1103
WT-ZEP1           GGCTGTCGACGTWGACCTTCATCATTCTGTCGGTTCCAAAAG 1202
zep1-1            GGCTGTCGACGTWGACCTTCATCATTCTGTCGGTTCCAAAAG 1302
zep1-2            GGCTGTCGACGTTGACCTTCATCATTCTGTCGGTTCCAAAAG 1302
zep1-3            GGCTGTCGACGTTGACCTTCATCATTCTGTCGGTTCCAAAAG 1302
zep1-4            GGCTGTCGACGTTGACCTTCATCATTCTGTCGGTTCCAAAAG 1302
zep1-5            GGCTGTCGACGTWGACCTTCATCATTCTGTCGGTTCCAAAAG 1302
```

B

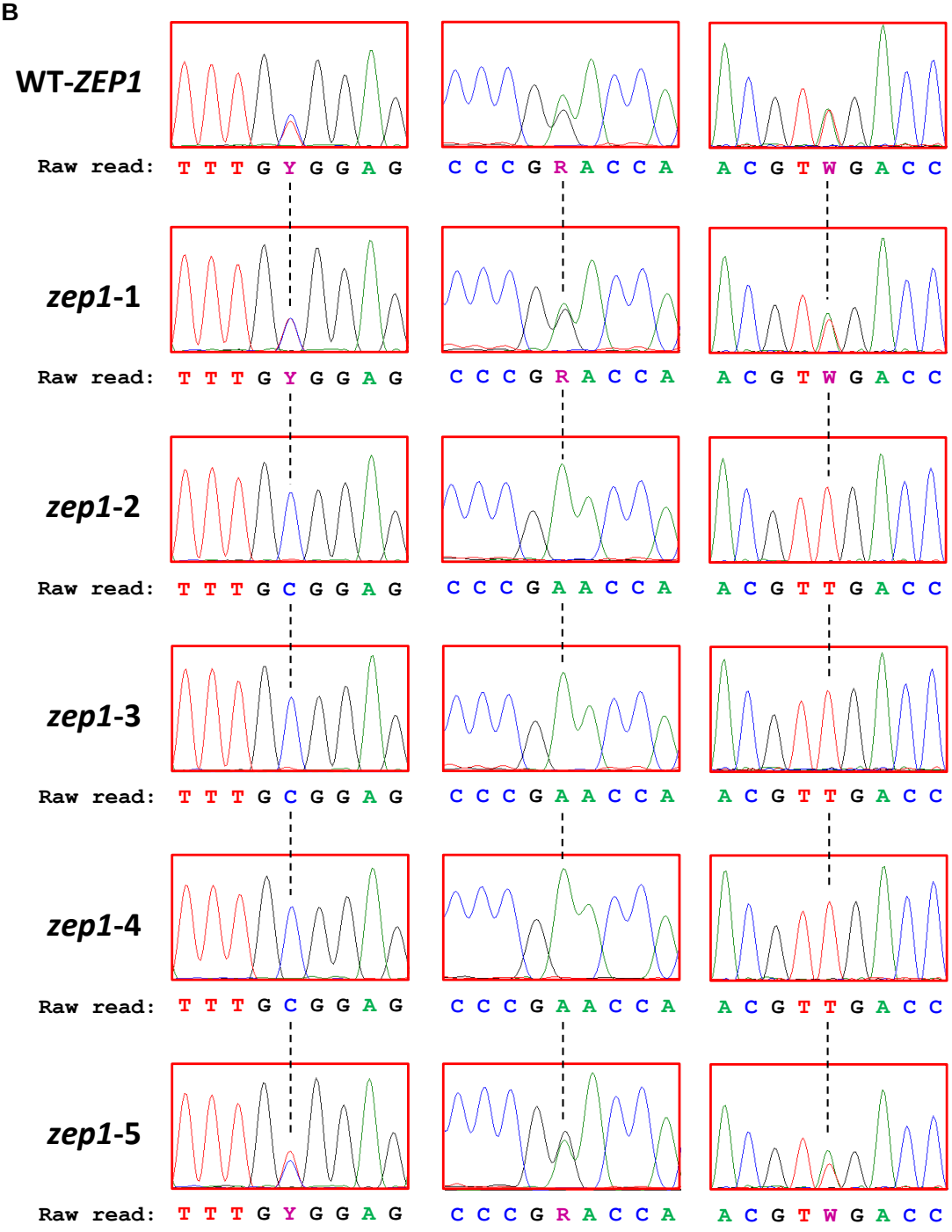


Fig. S3. Sequencing results of PCR products from *ZEP1*-wild type and *zep1*-knockout mutant lines of *P. tricornutum*. (*A*) Alignment of partial sequences of the *ZEP1*-knockout construct (pUC57_Zeo_ZEP1) used for transformation, the PCR product from WT cells (WT-ZEP1), and the PCR products from the five green knockout mutant lines shown in **Fig. S1** (zep1-1 to zep1-5) resulting from the red primer pair confirming specific integration of the *Ble* construct into the target gene. Colored bars above the alignment indicate partial sequence of the knockout construct consisting of the *Ble* gene (red) and the 1000 bp *ZEP1*-homologous arm (purple) used for HDR, and the *ZEP1*-genomic region downstream of the integration site (*ZEP1* genome; blue). Diagnostic single nucleotide polymorphisms (SNPs) between the two *ZEP1* alleles detected in the sequence of the WT-ZEP1 product and in the sequences of the *zep1*-1 and *zep1*-5 PCR products are highlighted by green background and indicated by IUPAC nucleotide codes R (A or G), S (G or C), W (A or T), and Y (C or T). (*B*) Exemplary sequence reads of the three regions labeled by red boxes in *A* covering diagnostic biallelic SNPs and indicating amplification of both alleles from the genomic DNA with equal efficiency. The SNPs were detected in the PCR products from 2 of the 5 mutants, indicating insertion in both alleles in the knockout mutants *zep1*-1 and *zep1*-5. No SNPs were detected in the PCR products from the other 3 mutants, suggesting that here the genome editing again led to biallelic mutants with identical alleles (3) (see **Texts S1** and **S3** for further discussion).

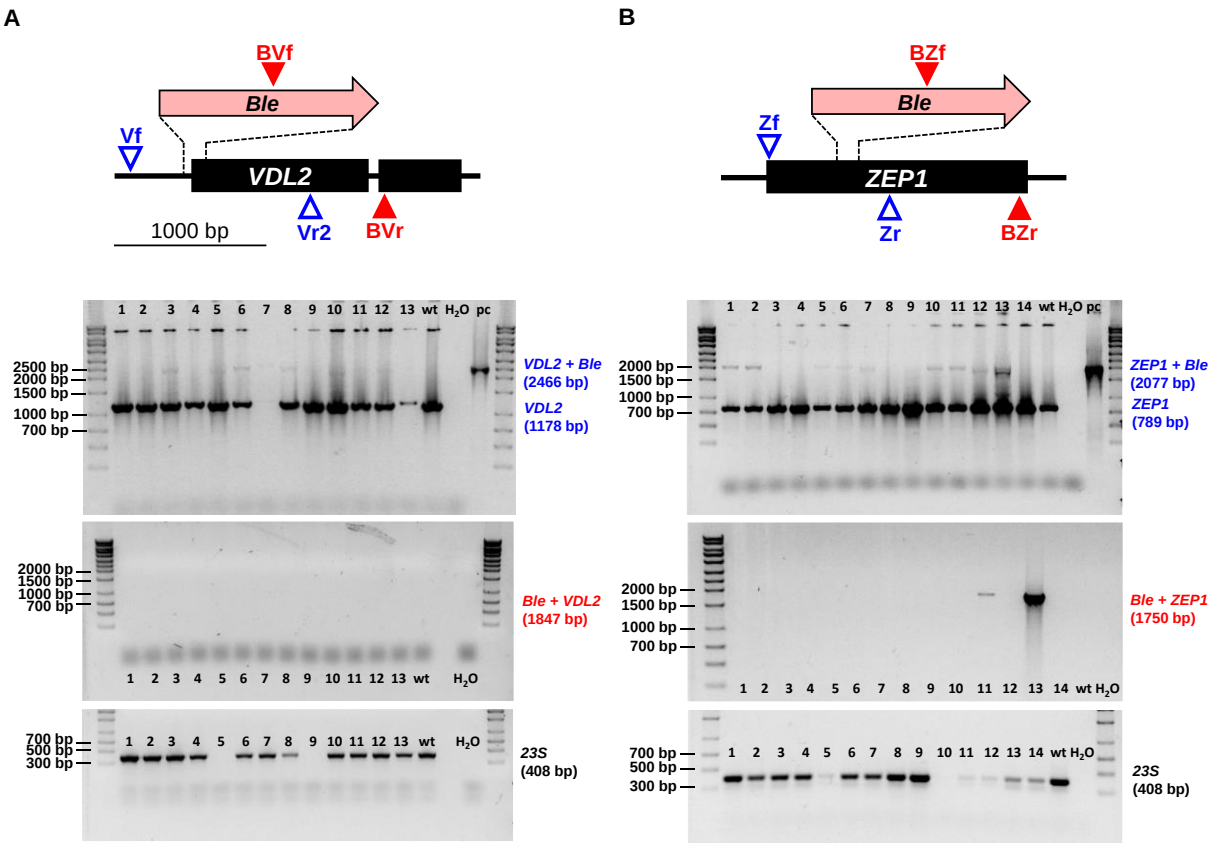


Fig. S4. Molecular characterization of brown colonies of *P. tricornutum* randomly isolated from the initial transformant screening plates. Agarose gels of PCR products from genomic DNA of 13 colonies from transformation with the *VDL2*-knockout construct (*A*) and 14 colonies from transformation with the *ZEP1*-knockout construct (*B*). Products labelled in blue result from combination of the primers indicated by blue open triangles in the top schemes that detect wt alleles of the target genes (*VDL2* / *ZEP1*) and to some extent also integration of the knockout construct into the genome (*VDL2* / *ZEP1* + *Ble*; either targeted or random). In samples with only a wt product band, absence of the larger product band detecting knockout construct integration most likely is due to the amplification of the shorter wt gene outcompeting that of the longer fragment. Note weak bands of expected product sizes for transformant 7 in *A*. Products labelled in red (*Ble* + *VDL2* / *ZEP1*) result from combination of the primers indicated by red filled triangles in the top schemes that detect specific integration of the knockout constructs by HDR in at least one allele of the respective target gene. DNA controls amplified a fragment of the plastome-encoded 23S rRNA gene. H_2O, water as PCR template; pc (positive control), *Ble* cassette with homology arms as PCR template.

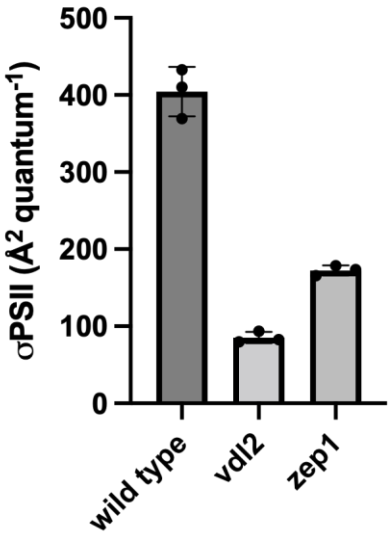


Fig. S5. Comparison of the photosystem II functional absorption cross section (σPSII) of *P. tricornutum* wild type and the green knockout strains. In the knockout strains devoid of fucoxanthin, σPSII is between 55% and 80% lower than in the wild type. Strains were grown in liquid culture under constant illumination with 60±5 µmol photons m^{-2} s^{-1} on a shaker and the measurements performed with cells in exponential growth. σPSII was calculated from chlorophyll fluorescence induction curves measured with a FIRe fluorometer (arithmetic means of n=3 for each strain, error bars are 1 standard deviation, points show independent biological replicates).

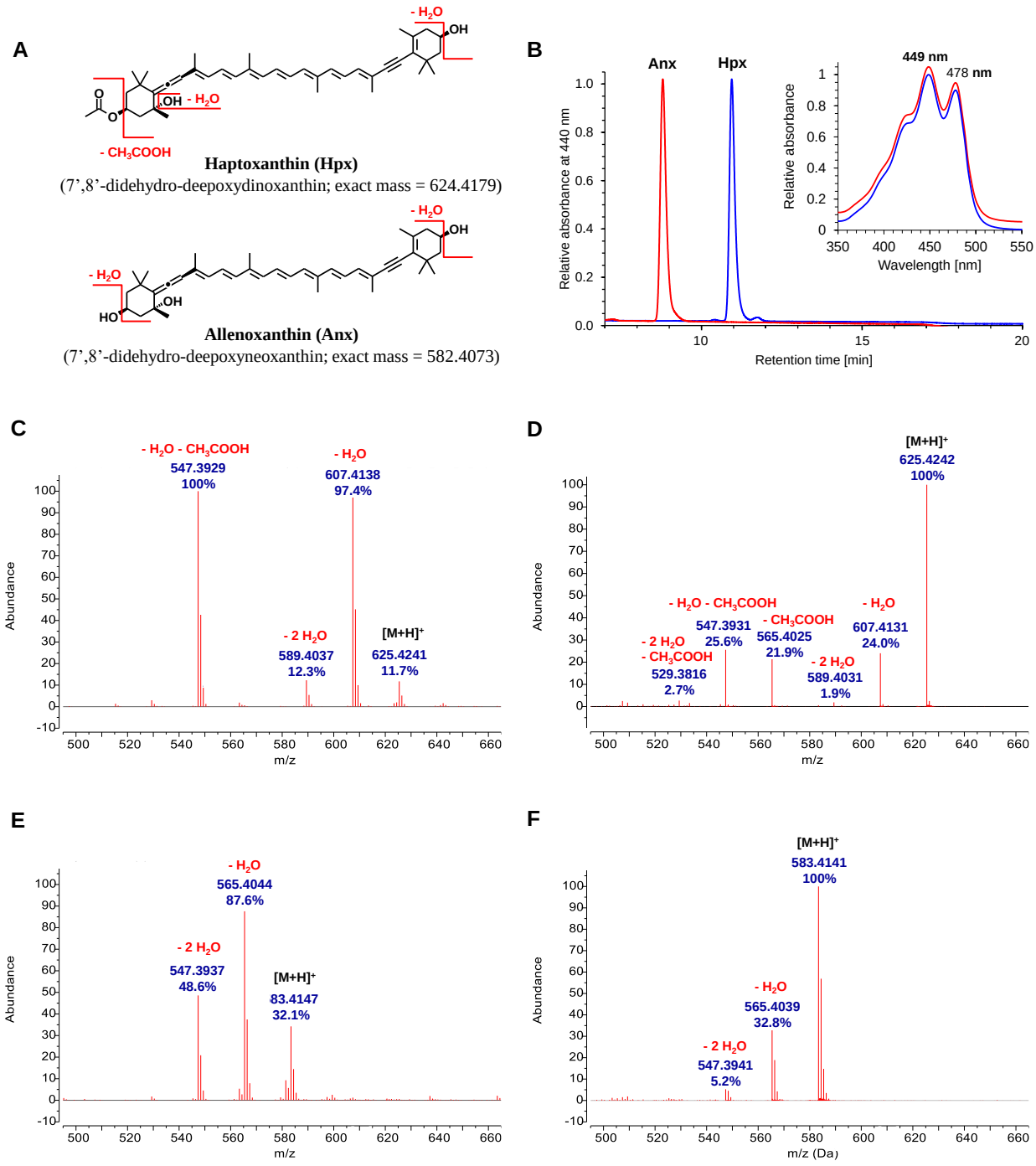


Fig. S6. Mass-spectrometric identification of the novel pigment haptoxanthin accumulating in the ZEP1 knockout mutant of *P. tricornutum*. (*A*) Proposed molecular structure of haptoxanthin (Hpx), its saponification product allenoxanthin (Anx), and some of their characteristic MS fragments. (*B*) Normalized HPLC scans (system II) and online-absorbance spectra (inset) of Hpx purified by consecutive separation on two different HPLC systems (blue lines) and of its saponification product Anx (red lines); while the absorbance spectra of Hpx and Anx (offset by 0.05 absorbance units for clarity) do not differ, Axn has a shorter retention time consistent with cleavage

of the acetyl ester resulting in a free hydroxyl group. APCI-MS scans (positive ion mode) of (*C*) Hpx (8 spectra averaged) with a calculated mass of 625.4251 for $[M+H]^+$, and (*E*) Anx (7 spectra averaged) with a calculated mass of 583.4146 for $[M+H]^+$. APCI-MS-MS scans (positive ion mode) of (*D*) the $[M+H]^+$ peak of Hpx in *C* at m/z of 625.4241 (4 spectra averaged) and (*F*) the $[M+H]^+$ peak of Anx in *E* at m/z of 583.4147 (4 spectra averaged); m/z and relative abundance of major mass peaks in blue. Masses of the observed peaks were in perfect agreement with the fragmentation pattern predicted from the proposed molecular structures in *A* and *B*, and the mass difference between Hpx and Anx further confirmed the loss of an acetyl group on saponification; the difference between expected and measured masses was below 2 ppm (max. 0.001 u) for all fragment peaks analyzed.

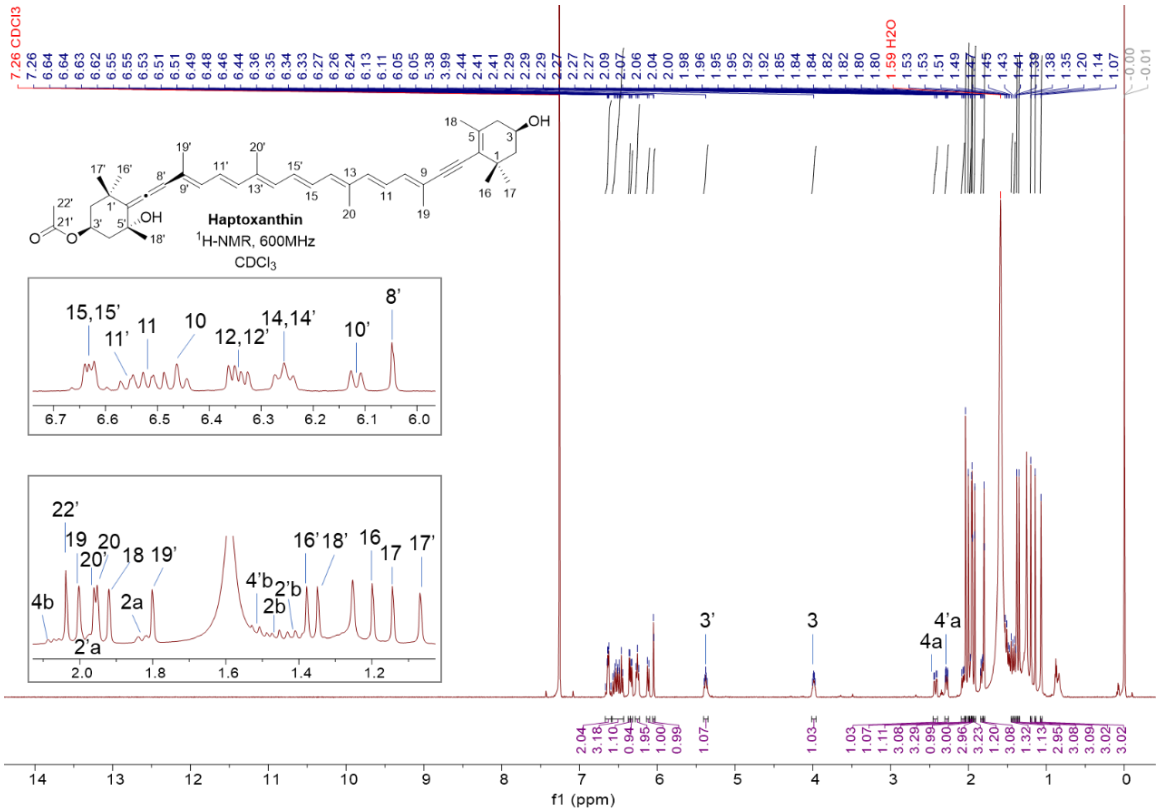


Fig. S7. ^{1}H NMR spectra (600 MHz, $CDCl_3$, 298 K) of haptoxanthin.

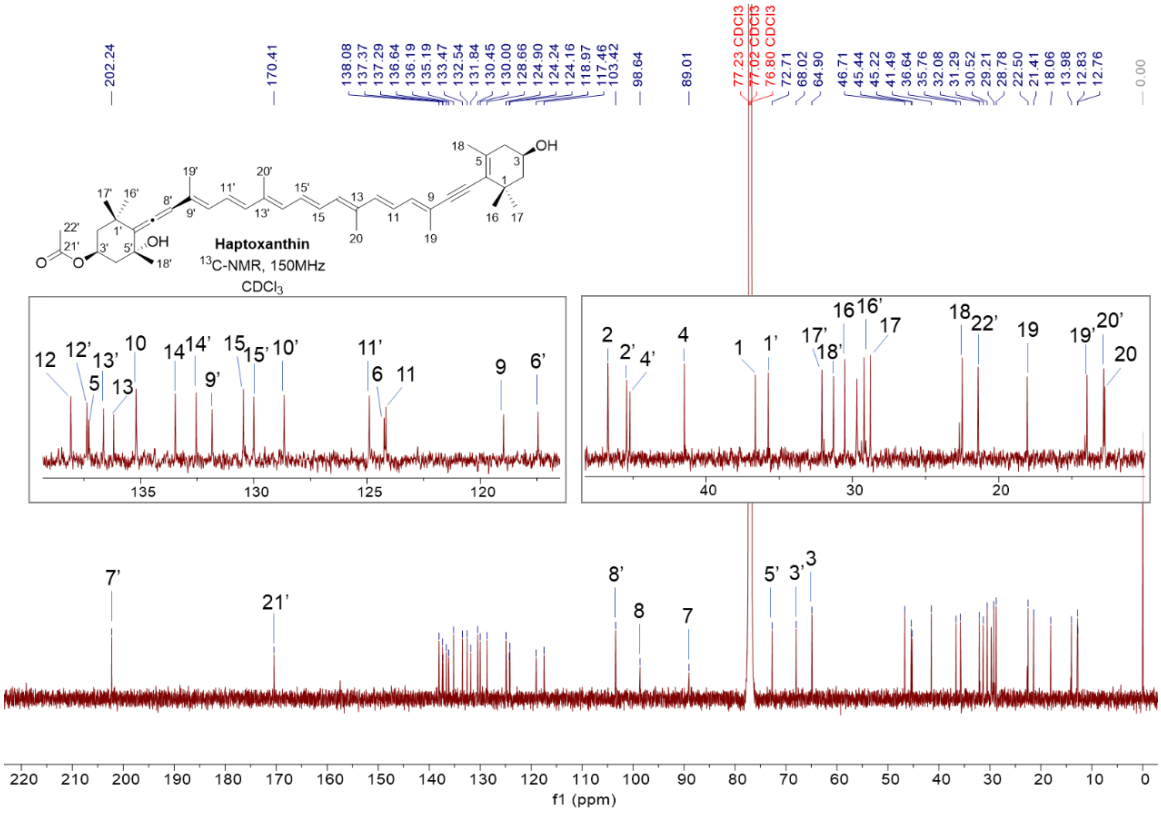


Fig. S8. ^{13}C NMR spectra (150 MHz, $CDCl_3$, 298 K) of haptoxanthin.

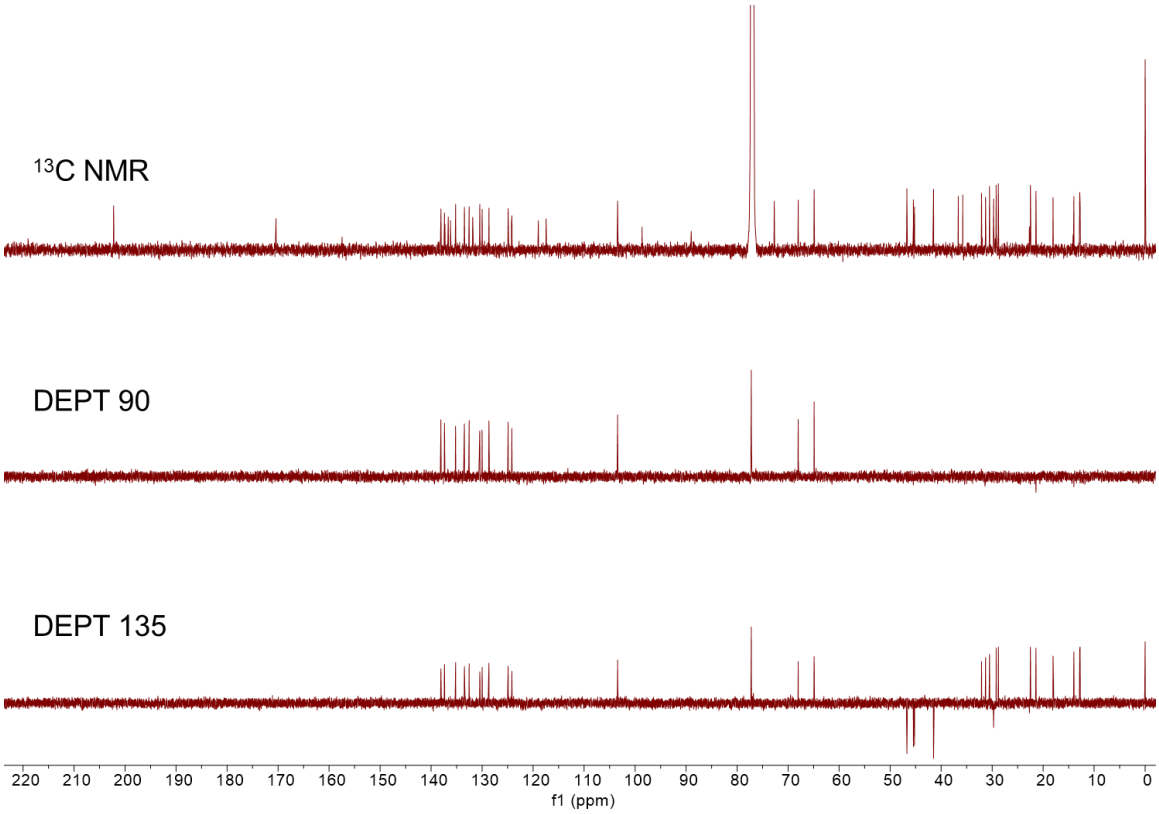


Fig. S9. DEPT spectra (150 MHz, $CDCl_3$, 298 K) of haptoxanthin.

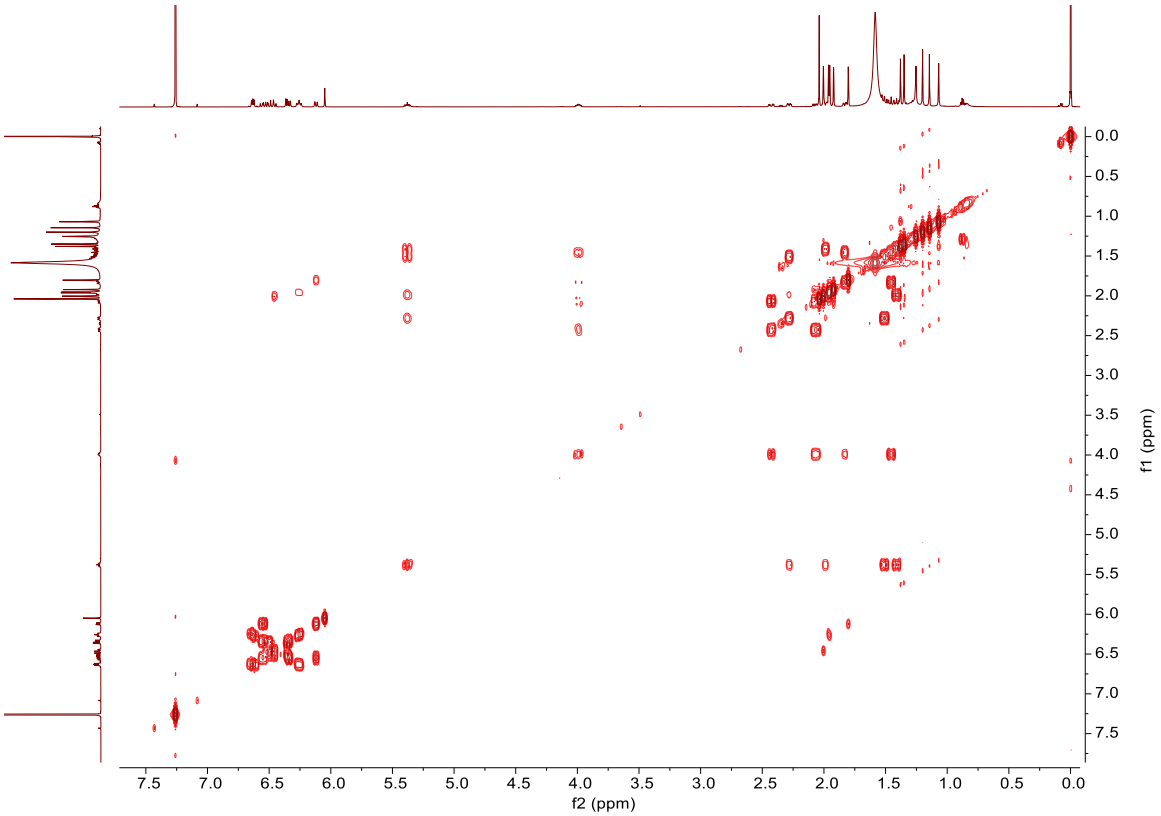


Fig. S10. ^{1}H-^{1}H COSY spectrum (600 MHz, $CDCl_3$, 298 K) of haptoxanthin.

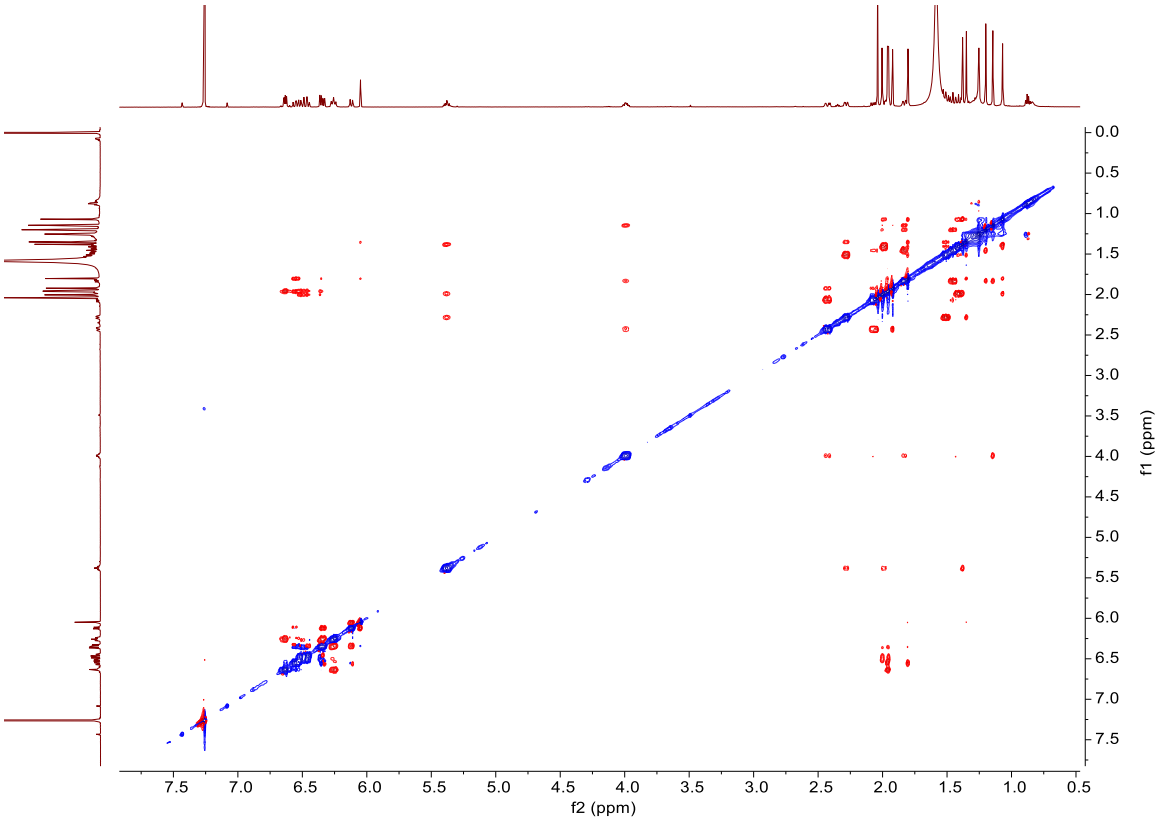


Fig. S11. ^{1}H-^{1}H ROESY spectrum (600 MHz, $CDCl_3$, 298 K) of haptoxanthin.

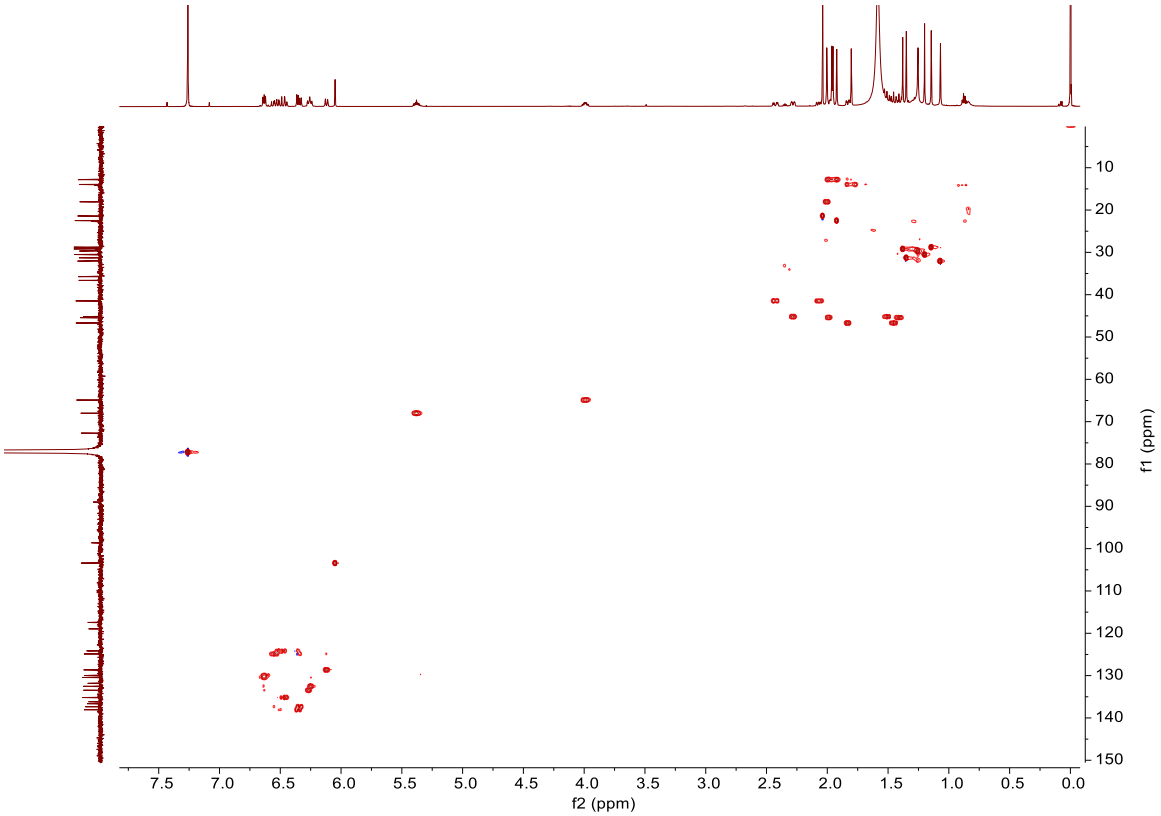


Fig. S12. HSQC spectrum (600 MHz, $CDCl_3$, 298 K) of haptoxanthin.

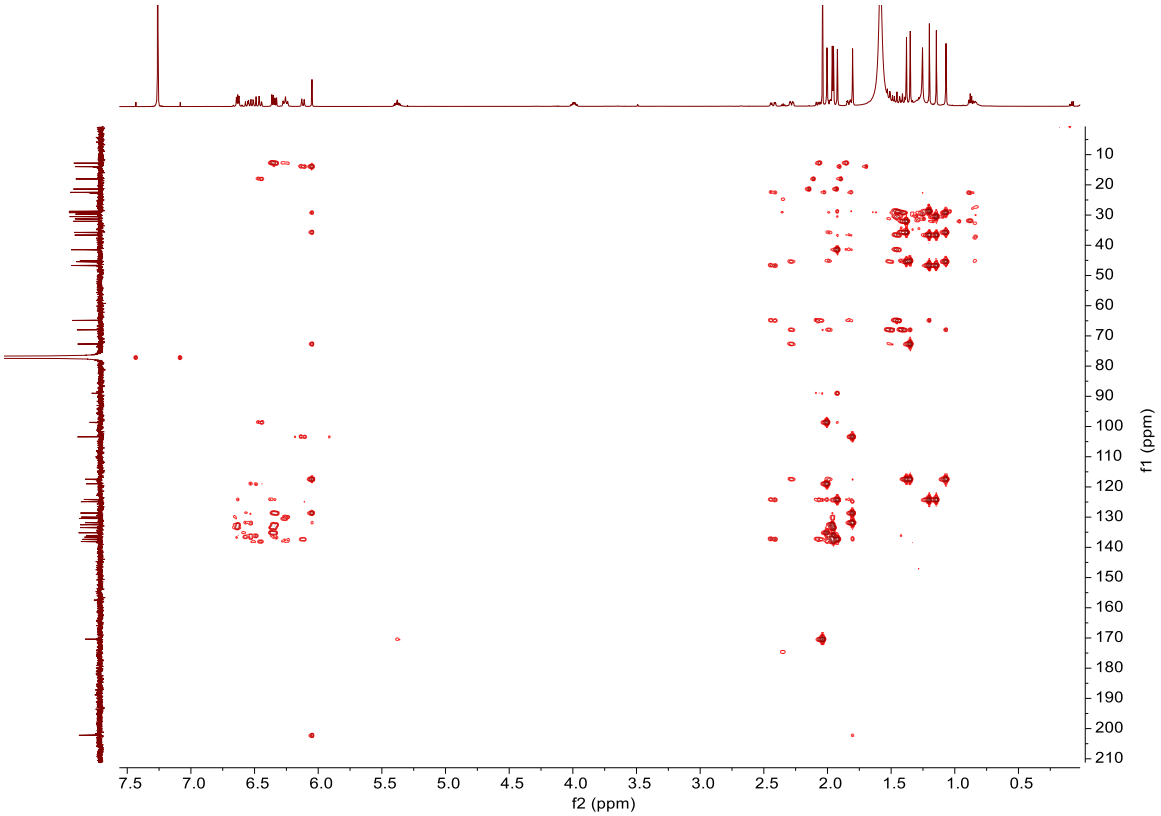


Fig. S13. HMBC spectrum (600 MHz, $CDCl_3$, 298 K) of haptoxanthin.

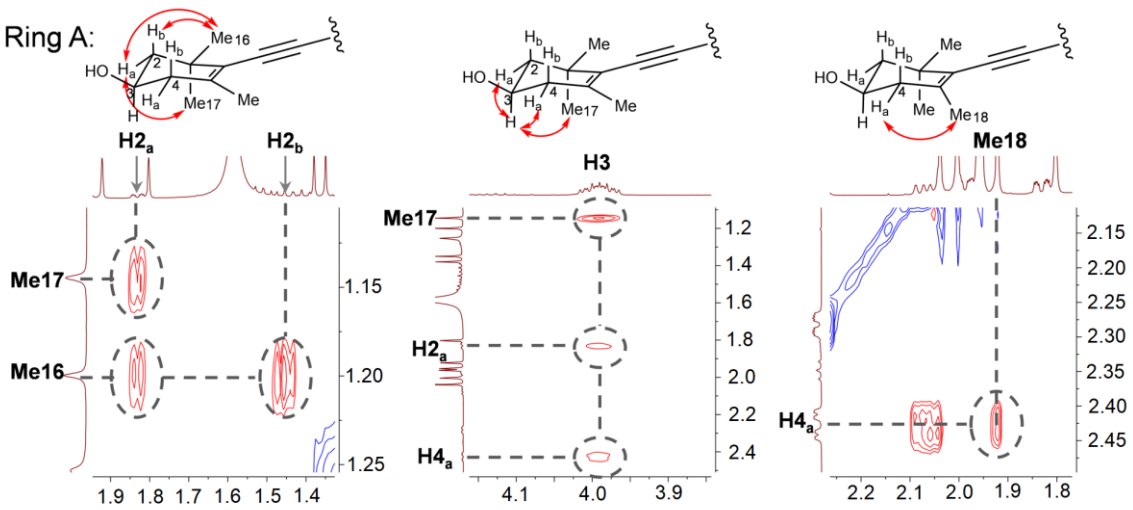


Fig. S14. The ^{1}H-^{1}H ROESY correlations in ring A of haptoxanthin.

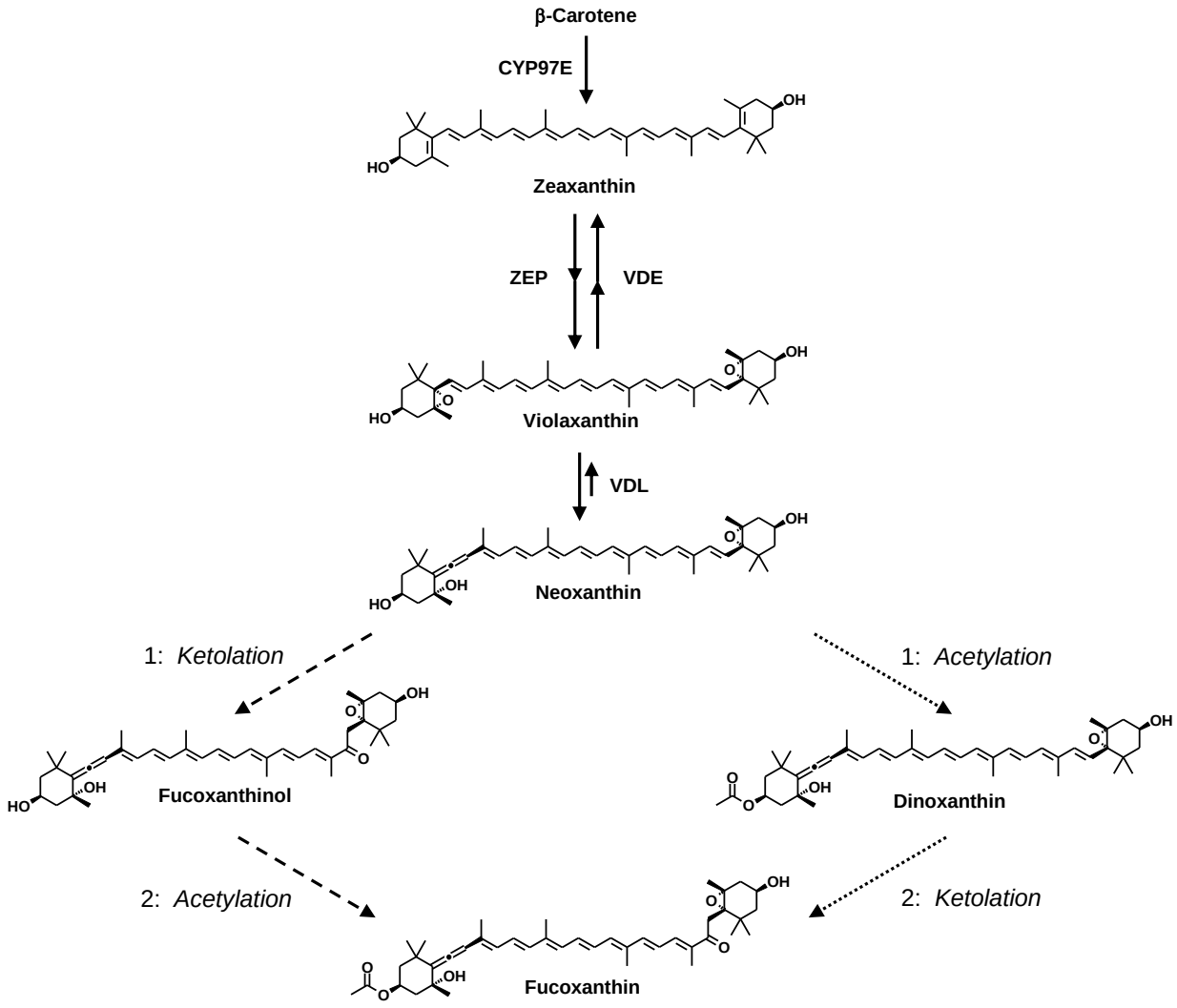


Fig. S15. Hypothetical alternative pathways of fucoxanthin biosynthesis from neoxanthin omitting diadinoxanthin as an intermediate. The pathway on the left denoted by dashed arrows and proceeding via fucoxanthinol was postulated for diatoms (5) and brown algae (6, 7). As shown in **Fig. 2**, we found the pathway in diatoms to be more complex, involving diadinoxanthin as obligate precursor and excluding fucoxanthinol as an intermediate. For reasons discussed in the main manuscript, we propose that brown algae may instead use the pathway on the right denoted by dotted arrows and proceeding via dinoxanthin for fucoxanthin biosynthesis (identical to the pathway in **Fig. 2** denoted by arrows in khaki). VDE, violaxanthin de-epoxidase; VDL, violaxanthin de-epoxidase-like protein; ZEP, zeaxanthin epoxidase.

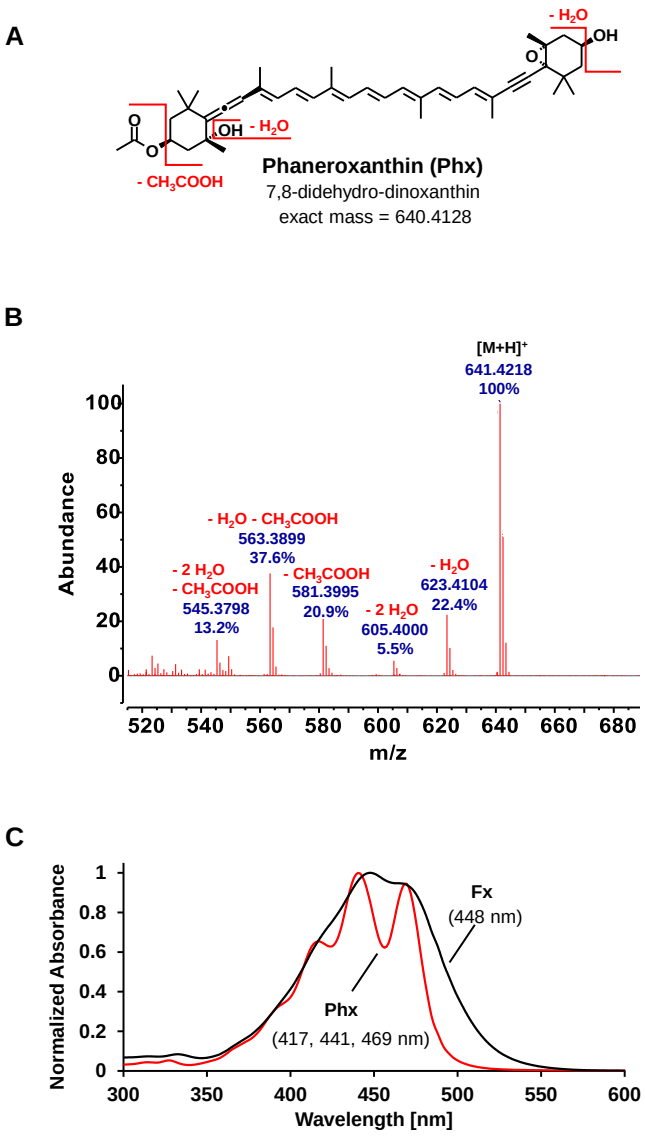


Fig. S16. MS-MS analysis of phaneroxanthin resulting from *in vitro* epoxidation of haptoxanthin by ZEP1 from *P. tricornutum*. (*A*) Chemical structure of phaneroxanthin and some of its characteristic MS fragments. (*B*) APCI-MS-MS scan (positive ion mode; 51 spectra averaged) of the $[M+H]^+$ peak of phaneroxanthin at m/z of 641.4208 in **Fig. 4*C***; m/z and relative abundance of major mass peaks in blue. Masses of the observed peaks were in perfect agreement with the fragmentation pattern predicted from the proposed molecular structures in (A); the difference between expected and measured masses was below 4 ppm (max. 0.002 u) for all fragment peaks analyzed. (*C*) Comparison of online absorbance spectra of phaneroxanthin and fucoxanthin (Fx).

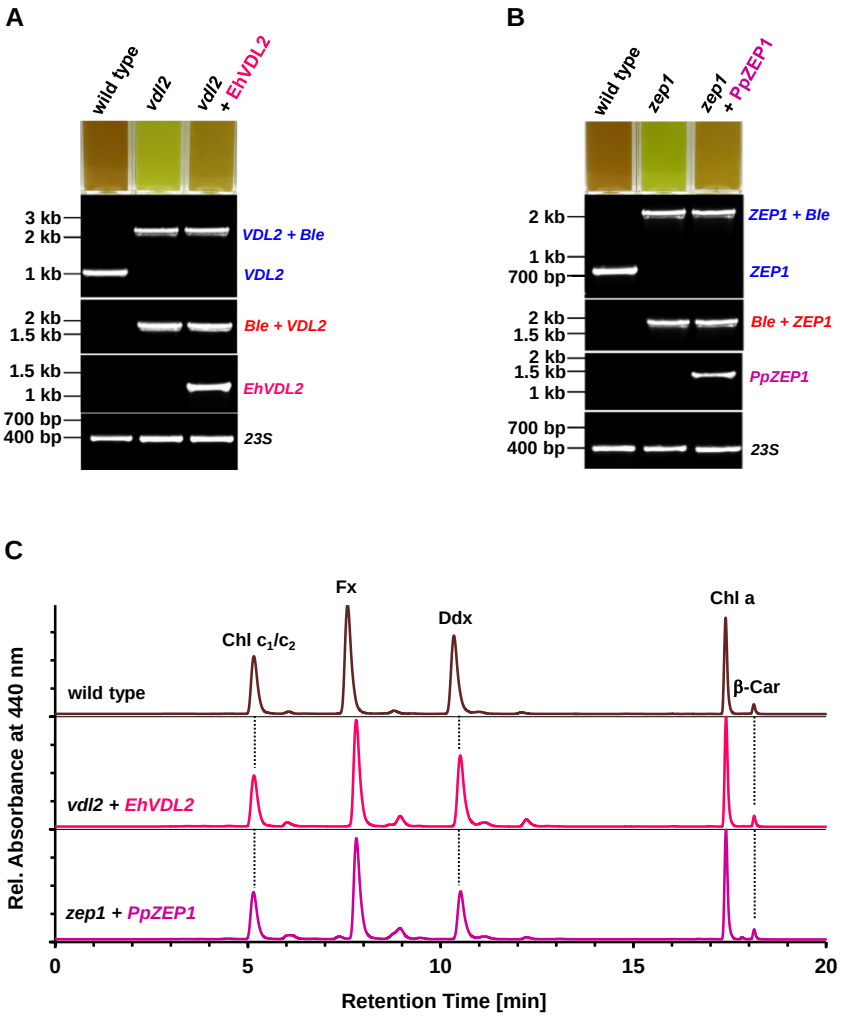


Fig. S17. Molecular characterization and pigment phenotypes of the *VDL2*- and *ZEP1*-knockout mutants from *P. tricornutum* after complementation with the corresponding orthologs from haptophyte algae. (*A* to *B*) Photographs show cuvettes with cultures of wild type, green-colored knockout lines and of knockout lines complemented with the corresponding orthologs from haptophyte algae that have regained the brown color of the wild type. *EhVDL2* was cloned from *Emiliania huxleyi* and *PpZEP1* from *Prymnesium parvum*. The PCR primers used for genotyping of the knockout lines and the resulting product sizes are the same as in **Fig. 1**. The additional haptophyte genes in the complemented lines were detected by PCR primers for specific amplification of fragments with expected sizes of 1174 bp for *EhVDL2* (*A*) and 1414 bp for *PpZEP1* (*B*). (*C*) HPLC analyses (system II) of pigment extracts from wild type and the respective knockout lines complemented with *EhVDL2* (*A*) or *PpZEP1* (*B*) confirming that the complemented strains regained the ability to synthesize fucoxanthin. Car, carotene; Chl, chlorophyll; Ddx, diadinoxanthin; Fx, fucoxanthin.

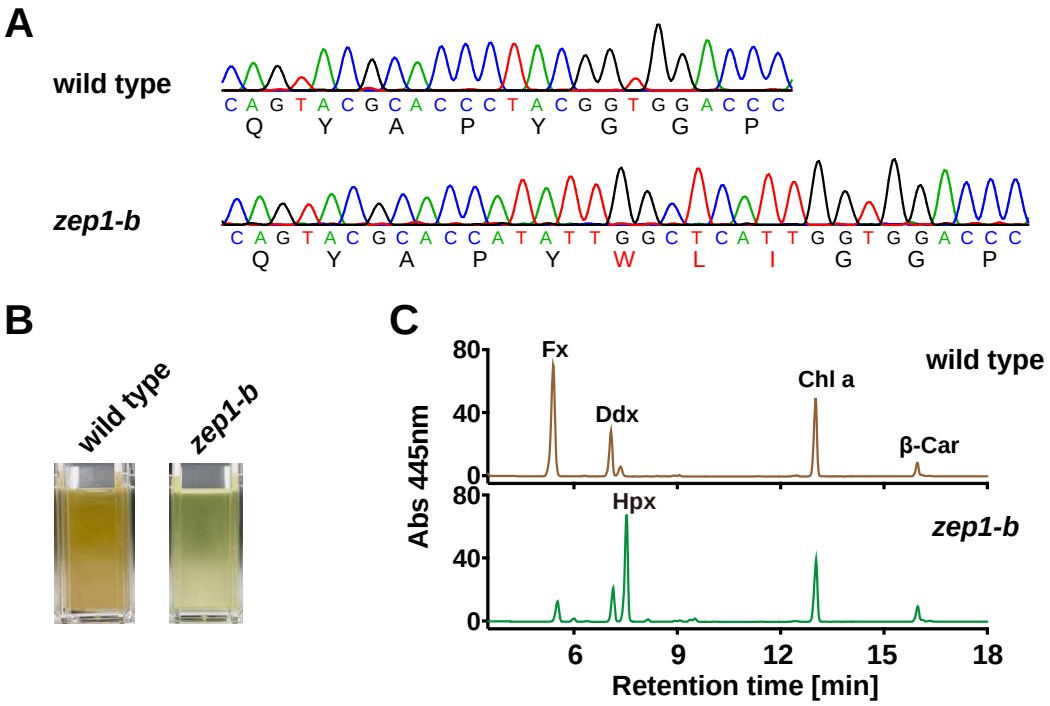
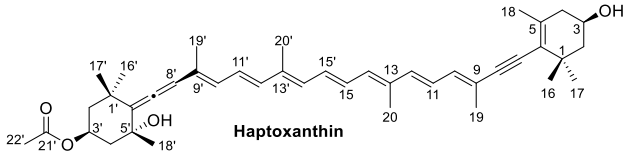


Fig. S18. Phenotypic and genotypic characterization of the ZEP1-targeted mutant line used for pigment purification for NMR and *in vitro* enzyme assays. (*A*) Part of sequencing results from *zep1-b* PCR product over the guide RNA site, which shows a 9-bp insertion in *ZEP1*. Predicted amino acid sequences are shown below the sequencing reads. (*B*) Appearance of *zep1-b* in comparison to wild type. A total of 10^6 cells in 1 mL were imaged. (*C*) HPLC analyses of pigments extracts from wild type and the *zep1-b* mutant. Phenotypes are similar to that of the ZEP1-knockout mutant shown in **Fig. 1**. "wild type" refers to the background strain XLF5 of the *zep1-b* mutant. Car, carotene; Chl, chlorophyll; Ddx, diadinoxanthin; Fx, fucoxanthin; Hpx, haptoxanthin.

Table S1. ^{13}C NMR and ^{1}H NMR data for haptoxanthin in $CDCl_3$ (δ in ppm).



Position	δ_C (150 MHz)	δ_H, multiplicity (*J* in Hz, 600 MHz)	DEPT	HMBC
1	36.7		C	
2	46.7	H_a, 1.83, ddd (*J* = 12.2 Hz, 3.6 Hz, 2.0 Hz) H_b, 1.44, dd (*J* = 12.2 Hz, 12.2 Hz)	CH2	1, 3, 16, 17
3	64.8	3.99, m	CH	
4	41.5	H_a, 2.42, ddd (*J* = 17.6 Hz, 5.5 Hz, 1.2 Hz) H_b, 2.07, ddd (*J* = 17.6 Hz, 9.8 Hz, 1.8 Hz)	CH2	3, 5, 6
5	137.2		C	
6	124.2		C	
7	89.0		C	
8	98.6		C	
9	119.0		C	
10	135.2	6.45, d (*J* = 11.4 Hz)	CH	8, 9, 12, 19
11	124.2	6.50, dd (*J* = 11.4 Hz, 14.5 Hz)	CH	10, 12
12	138.0	6.35, d (*J* = 14.5 Hz)	CH	10, 13, 14
13	136.2		C	
14	133.4	6.27, d (*J* = 10.8 Hz)	CH	12, 15', 20
15	130.5	6.64, dd (*J* = 10.8 Hz, 15.6 Hz)	CH	13, 14', 15'
16	30.5	1.19, s	CH3	1, 2, 6, 17
17	28.7	1.14, s	CH3	1, 2, 6, 16
18	22.5	1.92, s	CH3	4, 5, 6, 7
19	18.0	2.00, s	CH3	8, 9, 10
20	12.8	1.95, s	CH3	12, 13, 14
1'	35.7		C	
2'	45.4	H_a, 1.99, dd (*J* = 12.1 Hz, 4.1 Hz, 2.1 Hz) H_b, 1.41, dd (*J* = 12.1 Hz, 12.1 Hz)	CH2	1', 3', 16'
3'	68.0	5.38, m	CH	21'
4'	45.2	H_a, 2.28, ddd (*J* = 12.8 Hz, 4.2 Hz, 2.0 Hz) H_b, 1.51, dd (*J* = 12.8 Hz, 11.1 Hz)	CH2	3', 5'
5'	72.7		C	
6'	117.4		C	
7'	202.2		C	
8'	103.4	6.05, s	CH	1', 5', 6', 7', 10', 19'
9'	131.8		C	
10'	128.6	6.12, d (*J* = 11.3 Hz)	CH	8', 12',19'
11'	124.9	6.55, dd (*J* = 11.3 Hz, 14.9 Hz)	CH	9', 10', 13'
12'	137.4	6.34, d (*J* = 14.9 Hz)	CH	10', 13', 14'
13'	136.6		C	
14'	132.5	6.25, d (*J* = 10.8 Hz)	CH	12', 15'
15'	130.0	6.63, dd (*J* = 10.8 Hz, 15.6 Hz)	CH	13', 14, 15
16'	29.2	1.38, s	CH3	1', 2', 3', 17'
17'	32.1	1.07, s	CH3	1', 2', 3', 16'
18'	31.3	1.35, s	CH3	3', 4', 5', 6'
19'	14.0	1.80, s	CH3	8', 9', 10'
20'	12.8	1.96, s	CH3	12', 13', 14'
21'	170.4		C	
22'	21.4	2.04, s	CH3	3', 21'

Dataset S1 (separate file). Primers used for construct assemblies, colony screening, genotyping and sequencing.

Dataset S2 (separate file). Accessions of the VDE family sequences from algae and plants included in the phylogenetic tree in Fig. 3.

Dataset S3 (separate file). Occurrence and sequence accessions of VDL2 and ZEP1 in taxa analyzed in Fig. 5*A*.

Dataset S4 (separate file). Accessions of the ZEP family sequences from algae and plants included in the phylogenetic tree in Fig. 5*B*.

Dataset S5 (separate file). Accessions of genomic clusters of VDE/ZEP3 and of VDL2/ZEP1 in diatoms.

SI References

1. M. A. Moosburner *et al.*, Multiplexed Knockouts in the Model Diatom *Phaeodactylum* by Episomal Delivery of a Selectable Cas9. *Front Microbiol* **11** (2020).
2. D. Jallet *et al.*, Mitochondrial fatty acid β-oxidation is required for storage-lipid catabolism in a marine diatom. *New Phytol.* **228**, 946-958 (2020).
3. M. Nymark, A. K. Sharma, T. Sparstad, A. M. Bones, P. Winge, A CRISPR/Cas9 system adapted for gene editing in marine algae. *Sci. Rep.* **6**, 24951 (2016).
4. P. Bulankova *et al.*, Mitotic recombination between homologous chromosomes drives genomic diversity in diatoms. *Curr. Biol.* **31**, 3221-3232. e3229 (2021).
5. P. Kuczynska, M. Jemiola-Rzeminska, K. Strzalka, Photosynthetic pigments in diatoms. *Mar. Drugs* **13**, 5847-5881 (2015).
6. J. A. Haugan, S. Liaaen-Jensen, Algal carotenoids 54. Carotenoids of brown algae (Phaeophyceae). *Biochem. Syst. Ecol.* **22**, 31-41 (1994).
7. D. Nègre *et al.*, Genome–Scale Metabolic Networks Shed Light on the Carotenoid Biosynthesis Pathway in the Brown Algae *Saccharina japonica* and *Cladosiphon okamuranus*. *Antioxidants* **8**, 564 (2019).