



OPEN

## Multiple gene substitution by Target-AID base-editing technology in tomato

Johan Hunziker<sup>1</sup>, Keiji Nishida<sup>2</sup>, Akihiko Kondo<sup>2</sup>, Sanae Kishimoto<sup>3</sup>, Tohru Ariizumi<sup>4</sup> & Hiroshi Ezura<sup>4,5</sup>✉

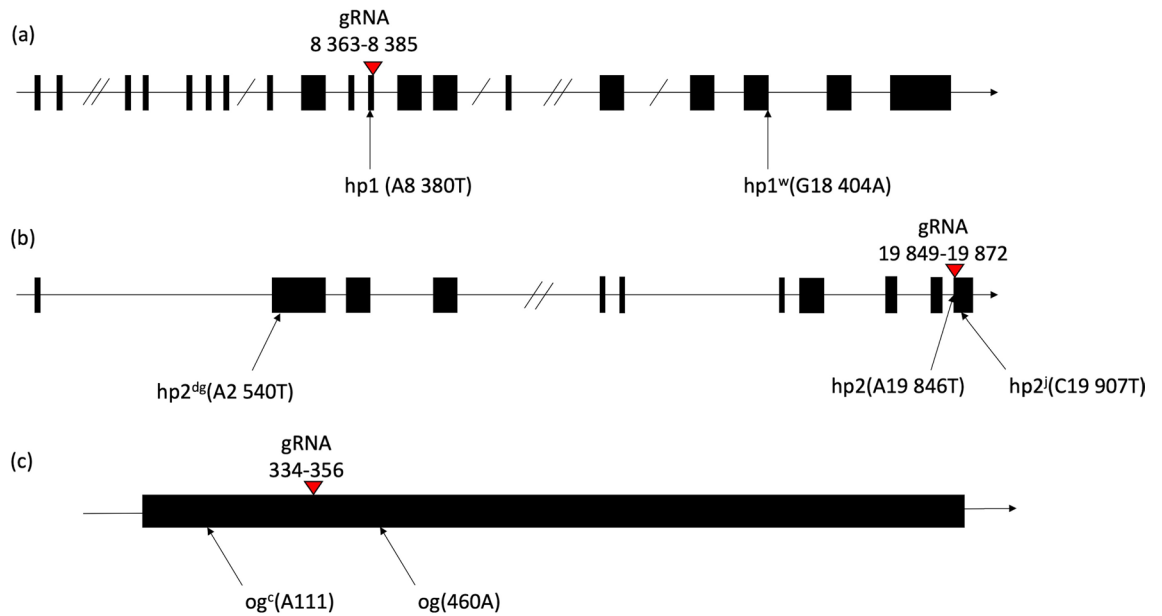
The use of Target activation-induced cytidine deaminase (Target-AID) base-editing technology with the CRISPR-Cas 9 system fused with activation-induced cytidine deaminase (AID) resulted in the substitution of a cytidine with a thymine. In previous experiments focusing on a single target gene, this system has been reported to work in several plant species, including tomato (*Solanum lycopersicum* L.). In this research, we used Target-AID technology to target multiple genes related to carotenoid accumulation in tomato. We selected 3 genes, *SIDDB1*, *SIDET1* and *SICYC-B*, for their roles in carotenoid accumulation. Among 12 edited T<sub>0</sub> lines, we obtained 10 independent T<sub>0</sub> lines carrying nucleotide substitutions in the three targeted genes, with several allelic versions for each targeted gene. The two edited lines showed significant differences in carotenoid accumulation. These results demonstrate that Target-AID technology is a highly efficient tool for targeting multiple genes with several allelic versions.

Gene editing is a technology based on the capacity to modify a target gene by removing, inserting or substituting DNA. Base-editing (BE) technology has been developed with the aim of inducing local targeted nucleotide substitutions. The clustered regularly interspaced short palindromic repeats (CRISPR)/CRISPR-associated protein 9 (CRISPR/Cas9) system is the gene editing tool that currently appears to be the most popular tool due to its efficiency and application to any living organism<sup>1</sup>. CRISPR/Cas9 acts as molecular scissors with two nuclease domains inducing DNA double-strand breaks (DSBs). A DSB can be fixed by a nonhomologous end joining (NHEJ) during mitosis after the duplication of DNA. Additionally, in the presence of a homologous DNA template containing the desired DNA sequence, such cleavage can be fixed through homology-directed repair (HDR) during the resorption of DSBs, which involves the introduction of the desired sequence through homologous recombination. Both processes result in the formation of abundant indels, rather than substitutions alone, inducing gene knockout, which can be deleterious depending on the gene or the position within it. To improve the efficiency of substitution, a number of BE technologies have been developed.

Some major BE technologies are based on an engineered Cas9 protein fused with a single strand DNA-specific cytidine deaminase (CDA)<sup>2–4</sup>. Another BE technique has been developed based adenosine deaminase (ADA)<sup>5</sup> but has only been tested in bacteria. Recently, the CDA and ADA technologies were fused in a single plasmid to form a protein complex that is able to specifically modify nucleotides<sup>6</sup>. Such technologies have been studied in organisms such as bacteria<sup>3,6</sup>, animals<sup>7–9</sup> and plants<sup>5,10–12</sup>, with a variable range of efficiency. Some of these technologies have been reviewed but are still being improved<sup>13</sup>. Target-AID technology, in which an *Arabidopsis thaliana*-optimized codon nickase Cas9 (nCas9) fused with a lamprey CDA is used, has been demonstrated as a powerful tool for precise BE in bacteria for multiplexing applications<sup>14</sup>. However, in plants, only single gene targets in rice and tomato have been reported thus far<sup>12</sup>.

Tomato is one of the most important commercial horticulture crops grown in the world. The improvement of yield and quality in tomato is a major challenge for breeders, as fruit yield and quality are controlled by several genes acting either independently or in concert. The identification of genes and alleles and their rapid integration in elite cultivars are necessary but time-consuming and expensive processes, as the insertion of a new allelic version in an elite line requires several years and crossing between generations, as observed for the introgression of mutations isolated among mutant collections of the Micro-Tom cultivar<sup>15</sup>. The recent development of gene

<sup>1</sup>Graduate School of Life and Environmental Sciences, University of Tsukuba, Tsukuba, Japan. <sup>2</sup>Graduate School of Science, Technology and Innovation, Kobe University, Kobe, Japan. <sup>3</sup>Institute of Vegetable and Floricultural Science, NARO, Tsukuba, Ibaraki, Japan. <sup>4</sup>Faculty of Life and Environmental Sciences, University of Tsukuba, Tsukuba, Japan. <sup>5</sup>Tsukuba Plant Innovation Research Center, University of Tsukuba, Tsukuba, Japan. ✉email: ezura.hiroshi.fa@u.tsukuba.ac.jp



**Figure 1.** Representation of *SIDDB1*, *SIDE1* and *SICYC-B* target sites. Schematic representation of each candidate gene, (a) *SIDDB1*, (b) *SIDE1* and (c) *SICYC-B*, with the previously described mutations and their corresponding annotations with mutation position compared to the first nucleotide of start codon on gene sequence. Black boxes represent exons regions and lines represent introns or UTR regions. Double slashes in introns represent reduction of the intron size. The red triangles represent the site targeted by the Target-AID with their corresponding nucleotides localization on gene.

editing technologies using programmable nucleases appears to be a promising and efficient approach for both breeding and basic research involving forward and reverse genetics in tomato.

Tomato fruit is known as a good source of lycopene, vitamin C,  $\beta$ -carotene, folate, and potassium<sup>16</sup>. Carotenoids play an important role in human nutrition because of their pro-vitamin A activity. However, there are other significant health benefits that are attributed to carotenoids, which are thought to be associated with their activity as antioxidants, such as their anticancer potential. Several genes involved in the improvement of carotenoid accumulation have been identified in tomato during the last two decades, such as tomato DNA Damage UV Binding protein 1 (*SIDDB1*)<sup>17,18</sup>, deetiolated1 (*SIDE1*)<sup>19</sup> and Lycopene beta cyclase (*SICYC-B*)<sup>20</sup> for lycopene accumulation or with *SICYC-B*<sup>20–22</sup> and *SICrtISO*<sup>23</sup> for  $\beta$ -carotene accumulation. *Slhp1* and *Slhp2* are mutations corresponding to *SIDDB1* and *SIDE1* genes respectively. These mutants confer the unique characteristics including purple roots resulting from anthocyanin accumulation<sup>19,24</sup>, dwarfism, dark leaves and dark green fruits<sup>17–19</sup>.

In this study, we tested Target-AID technology for generation of substitution mutations in multiple genes. As a model experiment, we targeted three genes, *SIDDB1*, *SIDE1* and *SICYC-B*, which are responsible for carotenoid accumulation, especially that of lycopene in tomato. For this purpose, we used a multiplexing target plasmid carrying three sgRNAs targeting the *SIDDB1*, *SIDE1* and *SICYC-B* genes with an nCas9-CDA-UGI (Uracil-DNA glycosylase inhibitor) protein complex. This approach resulted in the generation of several lines showing substitutions in each target gene simultaneously, demonstrating the usefulness of the Target-AID technology for multiple gene substitutions in tomato.

## Results

**Target site selection in *SIDDB1*, *SIDE1* and *SICYC-B*.** As reported previously in tomato, individual mutations in *SIDDB1*<sup>17,18</sup>, *SIDE1*<sup>19</sup> and *SICYC-B*<sup>20</sup>, particularly the *Slhp1*, *Slhp2* and *Slog* mutations, respectively, result in lines presenting high accumulation of lycopene. Regions close to each original mutation were selected as target site to obtain a new allelic version of each gene (Fig. 1). For *SIDDB1*, the original *Slhp1* mutation is located at nucleotide 8380, where a guanine is substituted by an adenine, resulting in the substitution of an Asp by a Tyr. Because an sgRNA containing this guanine could not be designed, the nucleotide targeted in our construct by using a reverse sgRNA was the guanine located at nucleotide 8383, resulting in the expected substitution of Ala by Pro. For *SIDE1*, the original *Slhp2* mutation, consisting of an AG-to-TG substitution, is located at the junction of intron 10 and exon 11 inducing alternative splicing of the mRNA leading to the deletion of 3 amino acids (Gly, Pro and Glu) of exon 11 within the second Nuclear localization signal (NLS). As the construct could not reproduce the exact substitution, the target region in the exon is expected to produce a nonviable NLS domain, inducing a loss of function resulting in an *hp* phenotype. For *SICYC-B*, the reported mutations involve the deletion of an adenine at nucleotide 111 for the *Slog<sup>c</sup>* allele and an insertion around nucleotide 460 for the *Slog* allele, resulting in a frameshift of the protein in both cases and the production of a nonfunctional enzyme. An expected target site presenting the possibility of introducing a stop codon by cytidine substitution has been selected, resulting in a truncated protein. Based on these assumptions, we expected that a construct inducing only a substitution instead of an indel could produce nonlethal new allelic versions of the gene, with a potential

| Target                   | Coordinates (Soylc ID)                         | Strand | No. of mismatches | Target Sequence              | PAM | Distance to the closest exon | Location of the target site | Gene name                      |
|--------------------------|--|--------|-------------------|------------------------------|-----|------------------------------|-----------------------------|--------------------------------|
| <i>DDB1</i> on-target    | SL2.50ch02:23374003-233747025 (Soylc02g021650) | +      | 0                 | AGCATTGCTAGGTATGATA          | TGG | 0                            | E                           | DNA damage-binding protein 1   |
| <i>DDB1</i> off-target1  | SL2.50ch02:8359160-8359182 (Soylc02g005390)    | -      | 4                 | <b>CCCTA</b> TGCTAGGTATGATA  | GGG | 24247                        | -                           | Unknown Protein                |
| <i>DDB1</i> off-target2  | SL2.50ch12:21742100-21742122                   | +      | 4                 | <b>GGACTTT</b> TCTAGGTATGATA | GGG | NA                           | -                           | NA                             |
| <i>DET1</i> on-target    | SL2.50ch01:53615357-53615379                   | +      | 0                 | CCTGAAGCTGGCAGCACAGA         | TGG | 0                            | E                           | DET1-like protein              |
| <i>DET1</i> off-target1  | SL2.50ch04:29300186-29300208 (Soylc04g024750)  | +      | 4                 | <b>ACAGAAGT</b> TGGAAGCACAGA | TGG | 11753                        | -                           | Unknown Protein                |
| <i>DET1</i> off-target2  | SL2.50ch09:13564134-13564156 (Soylc09g018210)  | -      | 4                 | <b>CCAGAACTAGA</b> AGCACAGA  | TGG | 10218                        | -                           | FAR1 Related sequence 5        |
| <i>CYC-B</i> on-target   | SL2.50ch06:45899368-45899390 (soylc06g074240)  | +      | 0                 | GGCCAACATGGAGAGTGGTGA        | AGG | 0                            | E                           | Ethylene receptor              |
| <i>CYC-B</i> off-target1 | SL2.50ch03:64496503-64496525 (Soylc03g114520)  | -      | 4                 | <b>AGCGAGAT</b> TGAGAGTGGTGA | GGG | 1043                         | -                           | Kinetochore protein            |
| <i>CYC-B</i> off-target2 | SL2.50ch02:50124240-50124262 (Soylc02g087860)  | -      | 4                 | <b>GGCGAAT</b> TGATAGTGGTGA  | GGG | 1805                         | -                           | Transcription factor style 2.1 |

**Table 1.** Prediction of *SIDDB1*, *SIDET1* and *SICYC-B* off-target sites. Distance to the closest exon: 0 indicates that the target site and exon coordinates overlap; NA indicates that the target sites are farther than 100 kb from the next exon. Location of the target site: E, - indicate exonic, intronic and intergenic, respectively.

combination of mutated *SIDDB1* and *SIDET1*, inducing a complementary improvement of lycopene accumulation, as synergistic phenotype could be observed on hypocotyl phenotype<sup>18</sup>. All targets are summarized in Table 1 and schematically represented within the genes in Fig. 1.

**Efficiency of base-editing at the targeted site in the T<sub>0</sub> lines.** Approximately 13 T<sub>0</sub> lines from 9 independent calli resulting from 151 explants infected were regenerated. It represents about 8% of diploid plants that were transformed via the infection of cotyledons. One regenerated line was identified as a positive control because it did not contain the selection marker and was excluded from further analysis, remaining 12 plants as transgenic. The remaining 12 T<sub>0</sub> plants (#8, #16, #16-1, #16-2, #16-3, #18, #26, #27, #30, #36-2, #43-1, #43-2) showed a typical *hp* mutant phenotype involving dwarfism, dark leaves and dark green fruits, observable in segregating generation (data not show). The sequencing of each target gene was performed to confirm the presence of a substitution rather than an indel. Results show 11 lines among 12 lines showed simultaneous substitution on 3 targets at the same time (Table 2a), resulting in a ratio of about 91.6% of efficiency of edition. Seven regenerated lines were isolated (#16, #16-1, #16-2, #16-3, #18, #43-1, #43-2), which provided the T<sub>1</sub> seeds. As plants in the T<sub>0</sub> generation show heterozygotic or chimeric status according to Sanger sequencing, these lines were used to perform amplicon sequencing by NGS to confirm the base-editing efficiency of our construct. Most of the lines showed a high efficiency of cytosine to thymine substitution at the 3' site of the gRNA for all 3 target genes, with the same pattern of mutation being observed in all independent lines (Fig. 2, Supplementary Tables S1a, S2). Regarding off-target effects, none of the lines presented any significant mutation obtained from the construct, as no substitutions could be observed among the few identified PCR errors (Supplementary Table S1b). Interestingly, no indels for *SIDET1* were present in the T<sub>0</sub> lines, whereas for *SIDDB1* and *SICYC-B*, the proportion of total reads showing indels was highly variable, ranging from approximately 12% to more than 90% of the total number of reads (Supplementary Table S1a). The NGS data and the T<sub>0</sub> Sanger data are summarized in Supplementary Tables S1, S2, respectively. All of these lines were generated to segregate the mutation pattern in the T<sub>1</sub> generation.

**Characterization of the triple mutations induced in the T<sub>1</sub> lines.** Seven lines descendances providing fruits and seeds are investigated. The seeds of the 7 T<sub>1</sub> lines (#16, #16-1, #16-2, #16-3, #18, #43-1, #43-2) were grown and observed. A minimum of 6 plant per line were observed, and line #16-1 showed few viable seeds. As expected, all lines show substitutions on all targets, resulting from segregation (Table 2b). In line #18, as the T<sub>0</sub> generation showed a very high percentage of indels, above 90%, more T<sub>1</sub> lines were analyzed. Interestingly, none of the lines showed a homozygous indel in *SIDDB1*, suggesting a potential lethal effect of the knockout of this gene. This was also confirmed in heterozygous T<sub>2</sub> plants from line #18, in which no plants carried a homozygous indel. Only plants showing homozygous or biallelic patterns were used for analysis, which were limited to 2 lines obtained from independent calli. The obtained plants showed the same phenotype as the *hp* mutant as expected, with purple roots, a dwarf morphology and darker green fruit compared to the WT (Fig. 3a,b, Table 3c, Supplementary Fig. S1a). To analyze the impact on photomorphogenesis, T<sub>2</sub> generation was used.

| Plant lines | SIDDB1 |              | SIDET1 |              | SICYC-B |              |
|-------------|--------|--------------|--------|--------------|---------|--------------|
|             | Indel  | substitution | Indel  | substitution | Indel   | substitution |
| <b>(a)</b>  |        |              |        |              |         |              |
| #8          | +      | +            | -      | +            | -       | +            |
| #16         | +      | +            | -      | +            | -       | +            |
| #16-1       | +      | +            | -      | +            | +       | +            |
| #16-2       | +      | +            | -      | +            | -       | +            |
| #16-3       | +      | +            | -      | +            | +       | +            |
| #18         | +      | +            | -      | +            | -       | +            |
| #26         | +      | +            | +      | +            | -       | +            |
| #27         | -      | +            | -      | +            | +       | +            |
| #30         | -      | +            | +      | +            | -       | +            |
| #36-2       | -      | -            | -      | +            | -       | -            |
| #43-1       | +      | +            | -      | +            | -       | +            |
| #43-2       | -      | +            | -      | +            | -       | +            |
| <b>(b)</b>  |        |              |        |              |         |              |
| #16         | +      | +            | +      | +            | -       | +            |
| #16-1       | -      | +            | +      | +            | +       | +            |
| #16-2       | +      | +            | -      | +            | -       | +            |
| #16-3       | -      | +            | -      | +            | -       | +            |
| #18         | +      | +            | -      | +            | -       | +            |
| #43-1       | -      | +            | -      | +            | -       | +            |
| #43-2       | +      | +            | -      | +            | +       | +            |

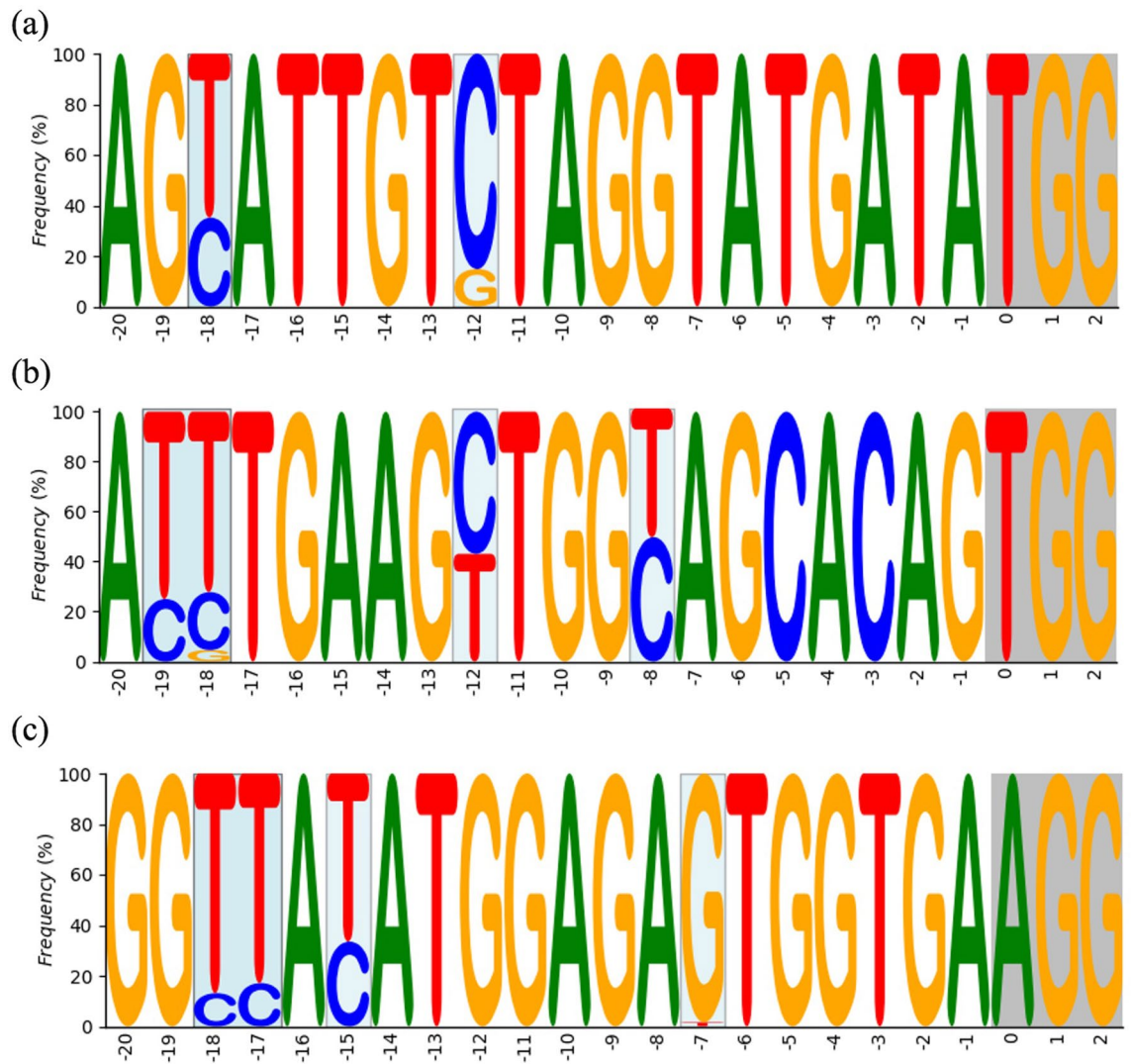
**Table 2.** Mutation type obtained in  $T_0$  and  $T_1$  generations of edited tomato lines. (a) Mutation type observed in all  $T_0$  diploid plants obtained and (b) in all  $T_1$  segregated lines. For each target, the presence or absence of sub mentioned type of mutation is indicated with a '+' for presence or '-' for absence, respectively.

We could observe an impact on photomorphogenesis with smaller hypocotyl on  $T_2$  mutant lines than WT line, under light and dark conditions (Supplementary Fig. S1b,c), as reported previously on double mutant grown under darkness<sup>18</sup>. The fruits did not show a significantly darker red epidermal color but tended to show a darker red color of the placenta, especially in line #43-2, more noticeable under light (Fig. 3c). As somatic mutation induced by the presence of the construct remained possible, only null segregant plants that did not show the presence of transgene markers were selected for further experiments. The patterns of mutation observed in all  $T_1$  lines analyzed are summarized in Supplementary Table S3. For *SIDDB1*, Sanger sequencing showed a common pattern of substitution in a homozygous state, where guanine was replaced by adenine at nucleotide 8 383 in lines #16, #16-2, #16-3, #18 and #43-1, resulting in the substitution of Ala by Thr, while in lines #16 and #16-2, a guanine was replaced by a cytidine or thymine at nucleotide 8 377, resulting in the substitution of Asp by His or Tyr, respectively. Interestingly, several plants showed indel in  $T_1$  which was not noticed in parental plants. Line #43-2 showed one plant with an indel in *SIDDB1* and *SICYC-B*, not visible in the parental plant by NGS analysis, and a biallelic state, carrying both substitutions described above in several other plants, not noticed in the parental plant (Table 2b, Fig. 4a). For *SIDET1*, lines #16 and #16-1 present an indel which was not noticed by NGS sequencing too (Table 2b). It will probably result from a remaining activity of the nCAS9 in plants carrying such substitutions.

In the case of *SIDET1*, 3 patterns of substitutions in a homozygous state could be observed, all involving the replacement of cytidine with thymine. The double substitution of nucleotides 19,850 and 19,851 resulted in the substitution of Pro by Phe in line 18; a triple substitution at nucleotide 19,857 resulting in the substitution of Asp by Val as the 3rd substitution in lines #16, #16-3, #43-1 and #43-2; and a quadruple substitution resulting in a silent mutation in a codon for Gly at nucleotide 19,861 for line #16-1. Other patterns of mutations were found, but because they were present in a heterozygous form in a single plant, they were excluded from the analysis (Fig. 4b).

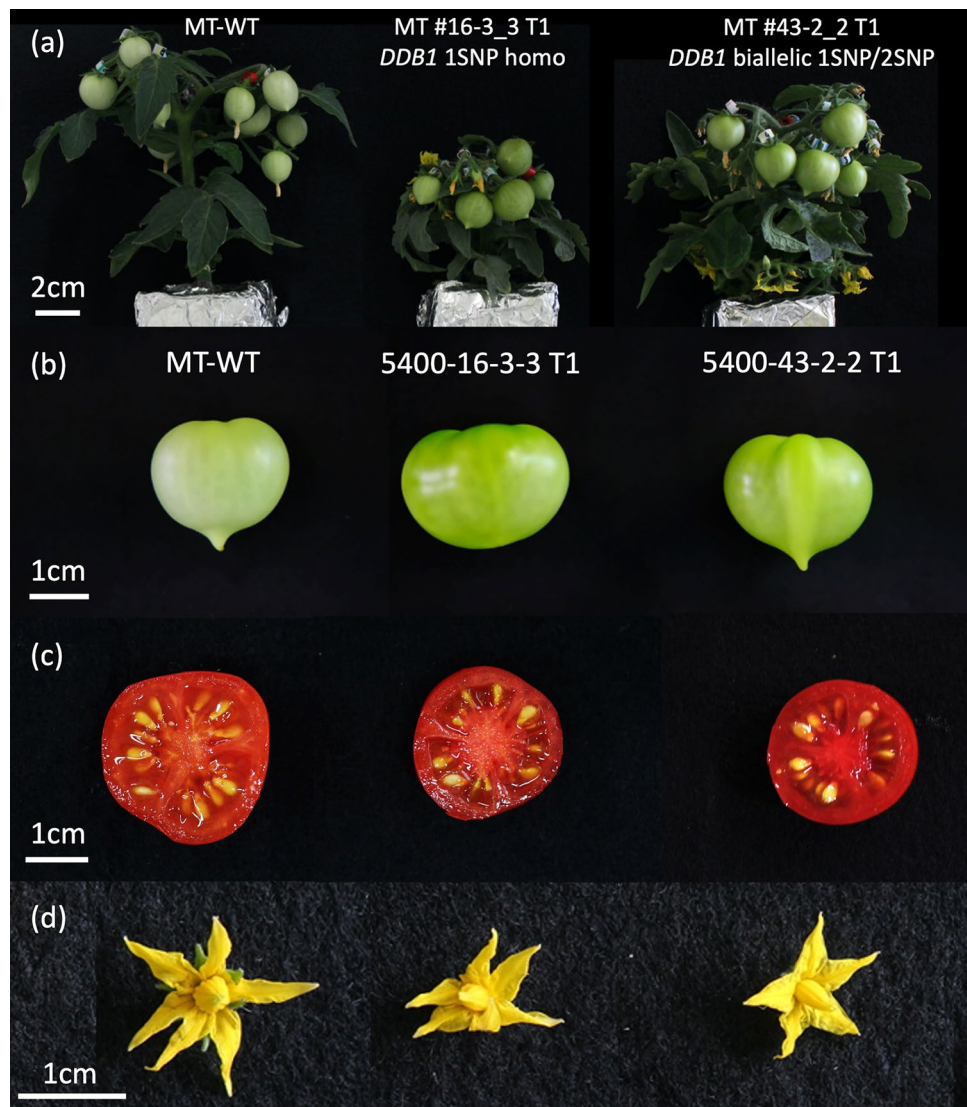
Finally, for *SICYC-B*, only 2 patterns involving the substitution of guanine by adenine could be observed in a homozygous state. A double substitution led to the formation of a stop codon at nucleotides 353 and 354 and a triple substitution at nucleotides 351, 353 and 354 resulted in the substitution of Met by Ile and the formation of a stop codon immediately thereafter (Fig. 4c). The interesting phenotype reported for mutations in *SICYC-B*<sup>20</sup> of darker petals could not be observed (Fig. 3d).

**Evaluation of carotenoid accumulation in the  $T_1$  lines.** After selection, only 2 independent callus lines were selected for carotenoid accumulation analysis in the  $T_1$  generation on the basis of being triple-substituted mutant lines. Five fruits per plant was harvested. In each independent  $T_1$  line selected, only 1 plant showed null segregation and was used for segregant analysis to avoid the somatic mutation effect due to the presence of the transgene in  $T_1$  plants (Supplementary Table S4). Other lines, such as line #43-2\_7', were not used due to a premature death before obtaining seeds. Fruits were harvested at 12 days post-breaker stage (DPB), and only the pericarp was retained to prevent bias resulting from the whiter columella color in the WT. The total



**Figure 2.** Logo Fig. of the deep sequencing analysis of on-target sites following Target-AID induction in tomato. Representation of the pattern of mutation induced by Target-AID with the corresponding sgRNA sequence in (a) *SIDD1*, (b) *S1DET1* and (c) *S1CYC-B*. Light blue background represents substitution areas, grey background represents PAM sequences. Numbers under the sequence represent the position of nucleotide from the PAM sequence, zero starting from the nucleotide N of NGG PAM sequence.

carotenoid content was measured by HPLC and adjusted to the lycopene standard for calculation. As each of the null segregant lines #16-3\_3  $T_1$  and #43-2\_2  $T_1$  presented a different mutation in *DDBI*, with a biallelic version of the gene occurring in #43-2\_2 (Fig. 4a), differentiation of the carotenoid composition and accumulation was expected. Line #43-2\_2 showed a 2nd SNP at nucleotide 928 only in the heterozygous state, and as plants with no mutation at this position showed a similar phenotype to the heterozygotes (Fig. 3c), this mutation was considered recessive and did not affect the results under biallelic conditions. The results of the analysis in the  $T_1$  lines showed an almost two-fold increase in lycopene in lines #43-2\_2, whereas #16-3\_3 showed an increase of approximately 30% (Table 3a). The accumulation of  $\beta$ -carotene was expected to be reduced compared to that in the WT, resulting in an expected disruption of the *CYC-B* enzyme. Interestingly, we observed an increase in  $\beta$ -carotene in #16-3\_3 of more than two-fold in the  $T_1$  generation. In the #43-2\_2 line, the  $\beta$ -carotene content remained stable compared to that in the WT. Same tendency of accumulation could be observed in  $T_2$  generation, with a total carotenoid, lycopene and  $\beta$ -carotene higher accumulation in mutated lines compare to the WT (Table 3b). However, at the green stage, fruits from line #16-3\_3 showed a darker color (Fig. 3c), resulting from a higher content of chlorophyll, suggesting a greater increase in carotenoids compared with line #43-2\_2 (Table 3c). Regarding the upstream carotenoid pathway, both mutated lines showed greater accumulation of phytofluene in the  $T_1$  line. In the downstream portion of the lycopene pathway, only line #16-3\_3 showed an increase in the accumulation of lutein and neoxanthin.



**Figure 3.** Plant phenotype of segregant generation. (a) Six-week-old plants in the  $T_1$  generation. (b) Fruit size and color at the green mature stage (28 days post-anthesis). (c) Fruit color at the red ripe stage (12 days post-breaker stage). (d) Flower shape and color at the anthesis stage. We noted a similar degree of darkness between the red fruits of the WT and both mutant lines.

## Discussion

Our results confirm that Target-AID is an efficient tool for inducing multiple-gene cytidine substitutions in tomato, with a substitution ratio higher than 40% of total number of reads without indel in all three targets simultaneously for all lines analyzed in the  $T_0$  generation, and no off-targets were detected (visualized in Fig. 2, Supplementary Table S1b). As reported previously, Target-AID tends to produce C to T and G to A substitutions<sup>3,12</sup>. As tomato is routinely subjected to transformation with standard protocols developed over decades<sup>25,26</sup>, the ratio of regeneration and transformation is known to follow standard patterns close to 10%. Among the 12 regenerated plants, 11 presented substitutions at all three targets simultaneously, representing about 91.6% of efficiency in regeneration. This ratio was much higher than that obtained in previous studies for multiplex targeting by non-modified Cas9 system. A study on the carotenoid pathway in tomato, showed more variable results depending on the target in the pathway (0 to 95.83% per target, with only 8.33% carrying more than 2 target mutations)<sup>27</sup>. Same observation was also reported on parthenocarpity trait, using non modified Cas9 system (11 to 65%)<sup>28</sup>. However, the promoter of the Cas protein or gRNA influences the efficiency of edition<sup>28</sup>, as observed for codon-optimized versions such as CDA *Arabidopsis thaliana*-optimized versions<sup>12</sup>.

The nCas9 protein generates single-strand breaks, and the UGI protein decreases indel formation after replication<sup>3</sup>, resulting in an indel rate lower than that coming from the use of the conventional Cas9 system. The Cas9 protein is mainly used to generate indels after the fixation of a lesion, and the most high-efficiency edition results in frameshift mutations<sup>27,29,30</sup>. In the case of Target-AID, the segregation of the generated mutations tends to produce lines carrying substitutions instead of indels in either plant or animal cells<sup>9,12</sup>. In our study, the ratio of indels varied from more than 90% of total reads resulting from NGS sequencing for *SIDDB1* in line #18 to 0%

| Plant lines | Phytoene          | Phytofluene       | Lycopene           | $\beta$ -Carotene | Lutein           | Neoxanthin       | Relative total carotenoid |
|-------------|-------------------|-------------------|--------------------|-------------------|------------------|------------------|---------------------------|
| <b>(a)</b>  |                   |                   |                    |                   |                  |                  |                           |
| WT          | 14.29 $\pm$ 3.45  | 3.91 $\pm$ 0.97   | 30.75 $\pm$ 9.50   | 2.02 $\pm$ 0.65   | 0.38 $\pm$ 0.16  | 0.06 $\pm$ 0.03  | 68.24 $\pm$ 17.39         |
| #16-3_3     | 10.72 $\pm$ 0.83* | 6.00 $\pm$ 0.51*  | 40.51 $\pm$ 7.40   | 4.19 $\pm$ 0.96*  | 1.06 $\pm$ 0.21* | 0.20 $\pm$ 0.04* | 100.03 $\pm$ 13.58*       |
| #43-2_2     | 11.61 $\pm$ 0.88* | 6.52 $\pm$ 0.43*  | 44.63 $\pm$ 3.28*  | 4.03 $\pm$ 0.74*  | 1.02 $\pm$ 0.18* | 0.19 $\pm$ 0.06* | 107.85 $\pm$ 4.08*        |
| <b>(b)</b>  |                   |                   |                    |                   |                  |                  |                           |
| WT          | 15.11 $\pm$ 3.33  | 3.60 $\pm$ 0.73   | 31.28 $\pm$ 9.96   | 1.80 $\pm$ 0.47   | 0.31 $\pm$ 0.06  | 0.05 $\pm$ 0.02  | 66.15 $\pm$ 17.75         |
| #16-3_3     | 9.52 $\pm$ 1.26*  | 2.48 $\pm$ 0.36*  | 38.12 $\pm$ 11.50  | 4.86 $\pm$ 1.42*  | 1.17 $\pm$ 0.26* | 0.25 $\pm$ 0.08* | 83.64 $\pm$ 24.26         |
| #43-2_2     | 17.86 $\pm$ 4.23  | 2.73 $\pm$ 0.92*  | 43.39 $\pm$ 13.78* | 2.34 $\pm$ 0.38*  | 0.65 $\pm$ 0.08* | 0.08 $\pm$ 0.01* | 90.37 $\pm$ 23.47*        |
| Plant lines | Chlorophyll 'a'   | Chlorophyll 'b'   | Total chlorophyll  |                   |                  |                  |                           |
| <b>(c)</b>  |                   |                   |                    |                   |                  |                  |                           |
| WT          | 18.94 $\pm$ 2.08  | 7.46 $\pm$ 1.03   | 26.40 $\pm$ 3.04   |                   |                  |                  |                           |
| #16-3_3     | 50.81 $\pm$ 4.25* | 18.53 $\pm$ 1.31* | 69.34 $\pm$ 5.52*  |                   |                  |                  |                           |
| #43-2_2     | 37.67 $\pm$ 3.38* | 14.14 $\pm$ 1.63* | 51.81 $\pm$ 5.01*  |                   |                  |                  |                           |

**Table 3.** Carotenoid content in  $\mu\text{g g}^{-1}\text{FW}$  of  $T_1$  and  $T_2$  segregant null at red ripe fruits and chlorophyll contents of  $T_2$  segregants at green mature fruits. (a) and (b) The carotenoid content was measured in the fruit pericarp by HPLC at the breaker-stage + 12 days (12 DPB) in  $T_1$  null plants and  $T_2$  segregants, respectively. Concentrations are expressed in  $\mu\text{g g}^{-1}\text{FW}$ . (c) The chlorophyll content was measured in the fruit pericarp by spectrophotometry at the green mature stage (28 DPA) in  $T_2$  segregants. Concentrations are expressed in  $\mu\text{g g}^{-1}\text{FW}$ . Statistical analysis with T-test independent variable was performed between the mutated lines to the control line for  $T_1$  and  $T_2$  for carotenoid content and for chlorophyll content, with significant differences were considered to exist at  $p < 0.05$ . \*Means significantly different at 0.05.

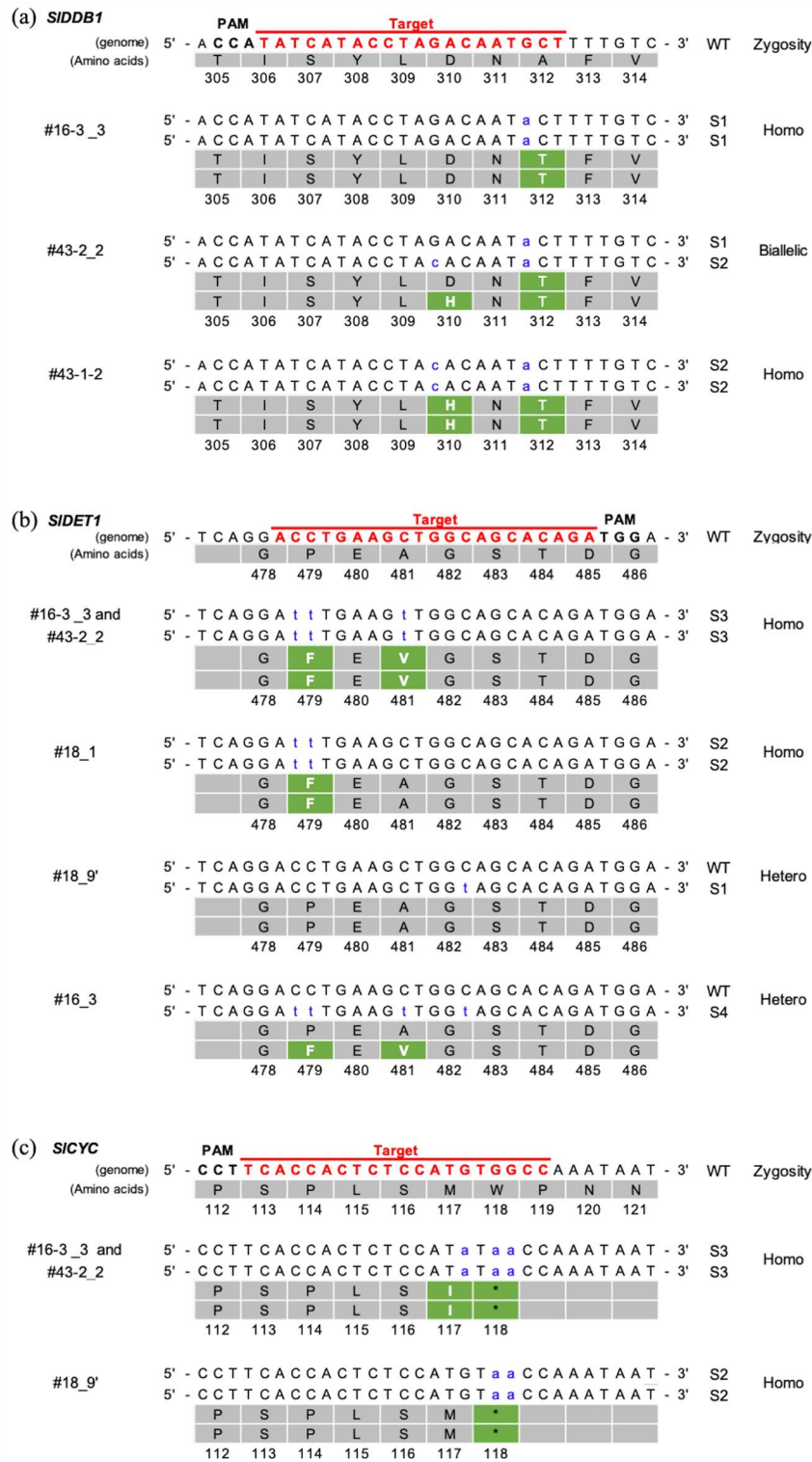
for *SIDET1* in all lines, with an average of 15% to 25% (Supplementary Table S1). This was observed upon direct PCR product sequencing in the  $T_0$  generation, with very low histogram peaks mix noticed downstream of the indel site on Sanger reads (Supplementary Fig. S2), coming from the presence of few amplicons of the DNA strand carrying indel, resulting in a potential error of interpretation as a noise background. This low amount of indel can also be noticed by the few amounts of heterozygote plants carrying an indel in the next generation (Supplementary Table S3).

The mutations in *SIDDB1* and *SIDET1* originally described in previous publications regarding *hp* tomato mutants<sup>17–19</sup>, resulted in SNPs that do not induce total knockout of the gene, probably due to lethal effects in the homozygous condition. As indels were present only in the heterozygous condition for *SIDDB1* and *SIDET1* (Supplementary Table S3), this could have resulted from a potential deleterious effect of the knockout of these genes. In the case of *SIDET1*, the original *Slhp2* mutation led to alternative splicing, inducing a deletion of three amino acids in exon 11, resulting in a truncated NLS<sup>19</sup>. However, this mutation was described as an incomplete knock-out mutation, with approximately 10% of the mRNA being correctly spliced. No indel allele mRNA of *SIDDB1* has been reported previously. In our case, the absence of homozygous plants for an indel in *SIDDB1* or *SIDET1* could be explained by the necessity for a small amount of functional protein, as no  $T_2$  homozygous indel lines could be obtained from line #18, which showed the highest segregation of indels.

Two alleles of *SIDDB1* could be obtained, with greater availability of the single SNP version. For *SIDDB1*, one mutation was located 4 bp from the original *Slhp1* mutation, resulting in the substitution of the subsequent amino acid compared to the original mutation<sup>17</sup>. The substitution of alanine by threonine was located at a position that is not strongly conserved compared to the previous asparagine to tyrosine substitution in *Slhp1* (Supplementary Fig. S3). As each amino acid belong to a different group, exhibiting a hydrophobic side chain or an uncharged polar chain, respectively, these substitutions might induce a change of conformation of the protein. However, the substitution of aspartic acid by histidine at amino acid 315 in line #42-2-2 occurred in a strongly conserved domain, potentially resulting in a strong effect on plant growth, explaining the difficulty of obtaining a homozygote for this position (Supplementary Fig. S3). A potentially stronger effect on carotenoid accumulation compared to line #16-3-3, with a codominant effect, may be considered. The impact of this mutation should be observed in segregating homozygotes in the  $T_2$  population.

Impact on photomorphogenesis could be observed with smaller hypocotyl of double mutant grown under light and dar conditions (Supplementary Fig. S1b,c) as reported previously in tomato<sup>18</sup>. But analysis of impact of each mutation on segregation generation should be performed to confirm if we have an additional effect of both mutations on the plant growth, as reported<sup>18</sup>.

The generation of a truncated protein is possible by the introduction of a stop codon, as in our example involving *SICYC-B*. The original *Slog<sup>c</sup>* mutant has been reported to exhibit an insertion resulting in a frameshift, showing the nonlethal effect of knocking out the gene. As the *SICYC-B* target resulted in a stop codon and, thus, a predicted truncated protein, the mutant should show a darker orange flower phenotype compared to the WT, as described in *Slog* mutants<sup>20</sup>. Interestingly, we could not clearly observe such a phenotype of tawny orange flowers (Fig. 3). The *SICYC-B* protein also functions as a neoxanthin synthetase in addition to a chromoplasmic lycopene- $\beta$ -cyclase, which results in the transformation of violaxanthin to neoxanthin<sup>31</sup>. In our loss of function of *SICYC-B* lines, we do not observe a decrease in  $\beta$ -carotene or neoxanthin contents as expected. Interestingly,



**Figure 4.** Target AID vectors induce nucleotide and amino acid substitutions in gene-edited tomato plants. Representative substitution spectra and changes in amino acid residues in T<sub>1</sub> plants. All figures show target sequences, PAM sequences and DNA modifications are in bold red, bold black, and bold blue font, respectively. Numbers under amino acid sequence represent the residue number position in the protein sequence. (a) Changes in amino acid residues in DDB1-targeted T<sub>1</sub> plants; (b) changes in amino acid residues in DET1-targeted T<sub>1</sub> plants. (c) Changes in amino acid residues in CYC-B-targeted T<sub>1</sub> plants. ‘S’ followed by a number correspond to substitution, with the number of nucleotides substituted. The substitutions in amino acid residues were due to the corresponding DNA substitutions in T<sub>1</sub> plants. The marker gene-free plants are highlighted.



we can observe an increase in both lutein content,  $\beta$ -carotene and neoxanthin (Table 3a,b). Such observation of no decrease of  $\beta$ -carotene in *SICYC-B* defective lines has been already reported in the Micro-Tom background<sup>32</sup>. This can probably be explained by the Micro-Tom genotype, as 2 different populations, MT Brazil<sup>32</sup> and MT Japan, in our case, present a mutation in *SICYC-B* without a phenotype of modification in lycopene accumulation.

In conclusion, Target-AID is an efficient tool for inducing multiple-gene substitutions in tomato, which are identified in a mutant population. The conventional Cas9 technique can cause a potential problem regarding the lethal effect of gene knock-out caused by indel mutations, as observed for *SIDDB1* and *SIDET1*. Although the generation of a single target mutation has already been demonstrated in the *Sletr1* and *Slidella*<sup>12</sup> genes, this is the first report of the introduction of substitutions among multiple genes in tomato. This technique will be a useful tool for the engineering of breeding traits controlled by multiple genes in elite lines of tomato. It can also be used to create new alleles from spontaneous mutations discovered in other mutant populations, including those obtained through spontaneous, gamma- or X-ray or Ethyl methanesulfonate (EMS) mutagenesis, such as the original *Sletr1*<sup>33</sup> and *Slidella*<sup>34</sup> mutants. New technologies for targeted replacement in the genomic DNA space and the introduction of new codons are still under development using CRISPR-Cas technology such as Prime Editing<sup>35</sup>. However, as such technology still shows a very low efficiency, the Target-AID is currently one alternative whereby breeders can introduce allelic variations in multiple targets in a single line and single generation.

## Materials and methods

**Vector construction.** The vector construct used in this study was derived from a previous study<sup>12</sup> but was modified by the insertion of the UGI fragment and amplified by PCR from pScl\_dCas9-CDA-UL (Addgene #108551)<sup>14</sup> using the following primers: kn2052, CTAAGAAGAAACGTAAAGTAgggcCcATGACCAACCTT TCCGAC and kn2053, agctgggaggcctggatcAgGgctatgcaaccagtcTAGCATC. The fragment was then inserted into *Apal*-digested pDicAID\_nCas9-PmCDA\_NptII by Gibson assembly to obtain pDicAID\_nCas9-CDA-UGI-LVA\_NptII.

The target guide cassette was replaced as described below. Each of the target sgRNA sequences for *SIDDB1*, *SIDET1* and *SICYC-B* was individually inserted into pDicAID between the *AtU6* promoter and the chimeric sgRNA scaffold (presenting an *AscI* site in the 5'-terminal region and an *MluI* site in the 3'-terminal region of the gRNA) by a PCR method and then circularized by Gibson assembly. Each resulting target sgRNA expression unit consisted of the *AtU6* promoter with an *I-SceI-AscI* site in the 5'-terminal region and a target-specific sgRNA with an *MluI-SpeI-I-SceI* site in the 3'-terminal region. New multiplexing plasmids were generated by double digestion as follows: *MluI* and *SpeI* double digestion for target 1 (*SIDDB1*) to open the vector and *AscI* and *SpeI* double digestion for target 2 (*SIDET1*) to release the fragment for extraction by gel purification. The products were ligated by recombination at the *AscI*-digested site with *MluI* and then transferred into highly competent *E. coli*. The same process was repeated for target 3 (*SICYC-B*) using the previous plasmid construct obtained from the ligation of the target 1 and target 2 products. Primer list used for the construction of the plasmid can be found in supplementary data (Supplementary Table S5).

**Plant transformation and culture.** The tomato (*Solanum lycopersicum*) cv. Micro-Tom Japan was used in this study, which was provided by the NBRP tomato program at the University of Tsukuba. The vector construct was introduced by *Agrobacterium* transformation with slight modification<sup>25,26</sup>. Transgenic diploid plants were selected by kanamycin resistance. The plants and their descendants ( $T_1$  and  $T_2$  lines) were grown in a culture chamber under an LED photosynthetic photon flux of 270  $\mu\text{mol m}^{-2} \text{s}^{-1}$  light (16 h/8 h) at a constant temperature of 25 °C. About 151 explants were infected and only 13 diploid plants were successfully regenerated and analyzed in further experimentations.

Photomorphogenesis experimentation has been conducted on  $T_2$  generation under same light condition as growing plants, or under total darkness, for 7 days. Seeds were treated with a bleach for 15 min followed by 4 washes with autoclaved tap water and transferred on ½ MS medium in Magenta box. Box were transferred for 2 days under darkness, then for 5 days under light for light experiment, or 6 consecutive days in dark followed by one day in light condition. Hypocotyl length were measured using digital caliper (Mitutoyo, Japan).

**PCR and sequencing analysis.** For PCR amplification, the primers used in this study are listed in Supplementary Table S5. For NGS sequencing, each target locus was amplified using PrimeSTAR GXL (Takara, Japan) following the manufacturer's protocol, and the second-round PCR products were extracted from 0.8% agarose gels under blue LED light to prevent PCR product degradation. For Sanger sequencing, targets were amplified using KOD FX Neo (Takara, Japan) according to the manufacturer's protocol, and the PCR product was directly purified using the FastGene Gel/PCR Extraction kit (Nippon Genetic, Japan).

**NGS analysis.** The design of the target was conducted as reported in our previous study on the Target-AID system<sup>12</sup> using CCTop (<https://crispr.cos.uni-heidelberg.de/index.html>), and selection was performed according to the proximity of the original mutation described for *SIDDB1*<sup>17,18</sup> and *SIDET1*<sup>19</sup> and the potential for the generation of a stop codon in *SICYC-B*<sup>20</sup>. Off-target associations were selected based on their higher scores. Both target and off-target effects were analyzed in 7 independent  $T_0$  lines, which were then used for the production of  $T_1$  seeds. The fragment containing the target region (1 kb) was firstly isolated by PCR reaction from genomic DNA, using the first primer set. After gel verification, to obtain an adaptor-containing amplicon (~0.5 kb), a second round of nested PCR was conducted using the first product as a template and a second pair of primers containing adaptor sequences. Expected products were extracted after separation in 0.8% agarose gels and band extraction using a FastGene Gel/PCR Extraction kit (Nippon Genetic, Japan). The concentration and quality of the samples were determined using the KAPA Library Quantification kit for Illumina (Kapa Biosystems, Roche,

CH) on a BioRad CFX96 qPCR system (BioRad, CA, USA). Approximately 50 ng of the product per sample was bulked for analysis. The primers used in this experiment are listed in Supplementary Table S5. The amplicons were labelled using NEBNext Multiplex Oligos for Illumina (Index Primer Set 1 and Dual Index Primer Set 1) (New England BioLabs, MA, USA). Deep sequencing was performed on a MiSeq sequencing system (Illumina, CA, USA), and more than 200 000 paired 300 bp reads per sample were obtained on average. Data analysis was performed using Geneious Prime 2019 (Biomatters, Auckland, New Zealand (<https://www.geneious.com>)). The sequence reads were paired and trimmed using the BBDuk Adapter/quality Trimming Version 38.37 plug-in from Brian Bushnell, based on the following settings: Trim Adapter with all libraries, Kmer length: 37; maximum substitutions: 1; maximum substitutions + INDELS: 0; trim low quality: both ends; minimum quality: 30; trim adapters based on paired read overhangs: minimum overlap: 20; discard short reads, minimum length: 200 bp. Paired and trimmed samples were mapped to the reference using Geneious mapper with the following settings: sensitivity: highest sensitivity; find structural variants, short insertions, and deletion of any size; minimum mapping quality: 30; trim paired read overhangs; only map paired reads that map nearby; minimum support for structural variant discovery: 2 read, include insertions in structural variants. Variant calling was performed with the following settings: minimum coverage: 10; minimum variant frequency: 0.05; maximum variant p-value: 10<sup>-3</sup>: 3 (0.1% to see variant by chance); p-value calculation method: approximate; advanced: merge adjacent variations. Both in- and off-target mutation frequencies were calculated using the numbers of reads with DNA modifications per total reads sequenced, directly obtained by variant calling. The logo Fig. 1 was produced using the Logomaker plug-in, under Python software<sup>36</sup> (<https://github.com/jbkinney/logomaker>).

### Analysis of the mutation spectrum in T<sub>0</sub> and descendant transgenic plants by Sanger sequencing.

To analyze the mutation spectrum of DNA modifications, in combination with NGS sequencing, target regions were amplified by PCR using specific primers (Supplementary Table S5) and genomic DNA extracted from the leaves of WT and transgenic plants (T<sub>0</sub>, T<sub>1</sub>) as templates. For the T<sub>0</sub> lines, in the presence of chimeric plants at a high ratio, the amplified PCR fragments were cloned into the pGEM-T easy plasmid (Promega, USA) after Taq treatment at 72 °C for 30 min, which was then cloned into the DH5 $\alpha$  strain (Toyobo, Japan) for sequencing on the ABI 3130xl system (Applied Biosystems, CA, USA). The bacteria were plated on LB agar plates containing 50  $\mu$ g ml<sup>-1</sup> ampicillin. Approximately 10 colonies for each target were picked using blue/white selection and confirmed by the PCR amplification of the appropriate fragment from each colony. Their insert sequences were determined by the Sanger method using target-specific primers. The T<sub>1</sub> and T<sub>2</sub> generation were sequenced directly after purification using the FastGene Gel/PCR Extraction kit (Nippon Genetic, Japan).

The conservation scores of the amino acid sequences of *SIDDB1* and *SIDET1* were determined on the ConSurf website server<sup>37</sup>, using the PBD ID of the protein and maintaining the standard settings.

### Carotenoid and chlorophyll analysis.

Carotenoid analysis was performed as follows. Five red fruits per plant were harvested at 12 DPB, and the pericarp was frozen in liquid nitrogen and ground to a fine powder using a Multi Beads Shocker (Yasui Kikai, Japan). The powder was stored at -80 °C for multiple purposes. For T<sub>1</sub> experimentation, only 1 plant of edited line could be analyzed as null segregant, divided in 3 individuals technical (n = 3) replicates and 5 plants for WT line (n = 15). For T<sub>2</sub> generation, 4 plants for WT line, line #16-3\_3 (n = 12) and 3 plants for line #43-2\_2 (n = 9) have been used, divided in 3 technical replicates for each plant. To perform carotenoid extraction, 100 mg of powder was used, and extraction was conducted following the petal carotenoid protocol<sup>38</sup> with some modifications. Before extraction, the samples were placed in a 2 ml tube with 2 zirconium beads and shaken for 30 s to break up any large powder agglomerates. Then, 30  $\mu$ l of 1 M Tris-HCl (pH 7.6–8) and 200  $\mu$ l of acetone was added, followed by mixing by hand. Then, 500  $\mu$ l of diethyl ether was added, followed by mixing by hand and centrifugation for 1 min at 10,000 rpm using a table centrifuge. The upper organic phase was then transferred to a new tube, and the previous steps were repeated until the total discoloration of the organic phase. Thereafter, 500  $\mu$ l of 5 mM Tris-HCl (pH 7.6–8) and a few drops of saturated saline solution were added to the collected organic phase, followed by mixing by hand and brief centrifugation. The lower aqueous phase was removed by pipetting and discarded, and the previous step was repeated 3 times to remove the remaining acetone and water present in the organic phase. The upper phase was then transferred to a new tube, and half of the volume of 10% KOH-MeOH (W/V) is added, followed by mixing by hand. The mixture was kept for 1 to 2 h in the dark to saponify the ester groups. Then, 1 ml of distilled water was added to the tube, followed by mixing and centrifugation. The lower aqueous phase was then removed, and the previous step was repeated 3 times. The upper phase was transferred to a new tube, and 300  $\mu$ l of diethyl ether was added to recover as much of the carotenoid extract as possible. In the new tube, 100% EtOH was added to the total organic volume at a rate of 20 to 30%, followed by mixing. The samples were dried under a vacuum desiccator for 1 h at RT, and the pellet was checked to ensure that it was correctly dried. The samples were stored at -80 °C until analysis. For analysis, each extract was analyzed by means of high-performance liquid chromatography (HPLC) (X-LC, JASCO, Tokyo, Japan)<sup>39</sup> with a photodiode array detector (X-LC 3110MD, JASCO) under the following conditions: column, YMC Carotenoid (100 mm  $\times$  4.6 mm i.d., 3  $\mu$ m; YMC, Kyoto, Japan); solvent A, methanol (MeOH)/methyl tert-butyl ether (MTBE)/H<sub>2</sub>O = 95:1:4 (v/v/v); solvent B, MeOH/MTBE/H<sub>2</sub>O = 25:71:4; gradient, 0/100, 5/100, 38/0 (min/% A); flow rate, 1.0 ml/min; column temperature, 35 °C; UV-Vis monitoring range, 200–600 nm. The contents were calculated according to the total peak area of the HPLC chromatograms at a wavelength of 450 nm for lycopene,  $\beta$ -carotene, lutein and neoxanthin. Phytoene and phytofluene were measured at 285 nm and 345 nm, respectively. Total carotenoids were expressed as the lycopene equivalents and neoxanthin as lutein equivalent. Lycopene,  $\beta$ -carotene, lutein, phytoene and phytofluene were measured with the corresponding standards. All other carotenoid compounds are expressed as lutein equivalents of the tissue.

The chlorophyll content was analyzed as described by Nagata and Yamashita<sup>40</sup>. Five green fruits per plant were harvested at 28 DPA, from 3 plants per line, and the pericarp was frozen in liquid nitrogen and ground to a fine powder using a Multi Beads Shocker (Yasui Kikai, Japan). The powder was stored at  $-80^{\circ}\text{C}$  for multiple purposes. For chlorophyll extraction, 200 mg of powder was used per biological sample, with 3 technical replicates ( $n = 9$ ), and extraction was performed using 2 ml of acetone-hexane (4:6 v/v) in a 5 ml tube, which was vortexed and then sonicated for 20 s. All samples were kept on ice and protected from light to prevent degradation. The samples were briefly centrifuged, 700  $\mu\text{l}$  of the organic phase was transferred to a quartz cuvette, from which the absorbance at 645 ( $A_{645}$ ), 663 ( $A_{663}$ ) and 750 ( $A_{750}$ ) nm was read on a spectrophotometer (SpectraMax M2, Molecular Devices, USA). The contents of chlorophyll 'a' and chlorophyll 'b' were calculated using the following equations in  $\mu\text{g g}^{-1}\text{FW}$ :

$$\text{Chlorophyll 'a'} = (0.999A_{663} - 0.0989A_{645}) \times 100$$

$$\text{Chlorophyll 'b'} = (-0.328A_{663} + 1.77A_{645}) \times 100$$

**Statistical analysis.** Statistical analysis was performed with IBM SPSS Statistics 25 (SPSS Inc., Chicago, USA) via independent-samples T test and a significant difference was considered to occur at  $p < 0.05$ .

Hypocotyl size figures (Supplementary Fig. S1b,c) were produced on R software, using R package *ggplot2* version 3.3.2<sup>41</sup> (<https://ggplot2.tidyverse.org/>).

### Data availability

Short read sequence data have been deposited at the NCBI Sequence Read Archive (SRA) under the accession codes PRJNA633920. The remaining data that support the findings of this study are available from the corresponding author upon reasonable request.

Received: 6 July 2020; Accepted: 9 November 2020

Published online: 24 November 2020

### References

- Hsu, P. D., Lander, E. S. & Zhang, F. Development and applications of CRISPR-Cas9 for genome engineering. *Cell* **157**, 1262–1278 (2014).
- Komor, A. C., Kim, Y. B., Packer, M. S., Zuris, J. A. & Liu, D. R. Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. *Nature* **533**, 420–424 (2016).
- Nishida, K. *et al.* Targeted nucleotide editing using hybrid prokaryotic and vertebrate adaptive immune systems. *Science* **353**, 6305 (2016).
- Nishimasu, H. *et al.* Engineered CRISPR-Cas9 nuclease with expanded targeting space. *Science* **9**, 1259–1262 (2018).
- Zong, Y. *et al.* Precise base editing in rice, wheat and maize with a Cas9-cytidine deaminase fusion. *Nat. Biotechnol.* **35**, 438–440 (2017).
- Sakata, R. C. *et al.* A single CRISPR base editor to induce simultaneous C-to-T and A-to-G mutations. *bioRxiv*. <https://doi.org/10.1101/729269> (2019).
- Lu, X. *et al.* Optimized Target-AID system efficiently induces single base changes in zebrafish. *J. Genet. Genomics* **45**, 215–217 (2018).
- Ma, Y. *et al.* Highly efficient and precise base editing by engineered dCas9-guide tRNA adenosine deaminase in rats. *Cell Discov.* **4**, 39 (2018).
- Sasaguri, H. *et al.* Introduction of pathogenic mutations into the mouse *Psen1* gene by Base Editor and Target-AID. *Nat. Commun.* **9**, 2892 (2018).
- Li, C. *et al.* Expanded base editing in rice and wheat using a Cas9-adenosine deaminase fusion. *Genome Biol.* **19**, 59 (2018).
- Negishi, K. *et al.* An adenine base editor with expanded targeting scope using SpCas9-NGv1 in rice. *Plant Biotechnol. J.* **17**, 1476–1478 (2019).
- Shimatani, Z. *et al.* Targeted base editing in rice and tomato using a CRISPR-Cas9 cytidine deaminase fusion. *Nat. Biotechnol.* **35**, 441–443 (2017).
- Rees, H. A. & Liu, D. R. Base editing: precision chemistry on the genome and transcriptome of living cells. *Nat. Rev. Genet.* **19**, 770–788 (2018).
- Banno, S., Nishida, K., Arazoe, T., Mitsunobu, H. & Kondo, A. Deaminase-mediated multiplex genome editing in *Escherichia coli*. *Nat. Microbiol.* **3**, 423–429 (2018).
- Carvalho, R. F. *et al.* Convergence of developmental mutants into a single tomato model system: 'Micro-Tom' as an effective toolkit for plant development research. *Plant Methods* **7**, 18 (2011).
- Willcox, J. K., Catignani, G. L. & Lazarus, S. Tomatoes and cardiovascular health. *Crit. Rev. Food Sci. Nutr.* **43**, 1–18 (2003).
- Lieberman, M., Segev, O., Gilboa, N., Lalazar, A. & Levin, I. The tomato homolog of the gene encoding UV-damaged DNA binding protein 1 (*DDB1*) underlined as the gene that causes the *high pigment-1* mutant phenotype. *Theor. Appl. Genet.* **108**, 1574–1581 (2004).
- Liu, Y. *et al.* Manipulation of light signal transduction as a means of modifying fruit nutritional quality in tomato. *Proc. Natl. Acad. Sci. U.S.A.* **101**, 9897–9902 (2004).
- Mustilli, A. C., Fenzi, F., Ciliento, R., Alfano, F. & Bowler, C. Phenotype of the tomato *high pigment-2* mutant is caused by a mutation in the tomato homolog of *DEETIOLATED1*. *Plant Cell* **11**, 145–157 (1999).
- Ronen, G., Carmel-Goren, L., Zamir, D. & Hirschberg, J. An alternative pathway to  $\beta$ -carotene formation in plant chromoplasts discovered by map-based cloning of *Beta* and *old-gold* color mutations in tomato. *Proc. Natl. Acad. Sci.* **97**, 11102–11107 (2000).
- Hwang, I., Kim, Y., Han, J. & Nou, I. S. Orange color is associated with *CYC-B* expression in tomato fleshy fruit. *Mol. Breed.* **36**, 42 (2016).
- Mohan, V., Pandey, A., Sreelakshmi, Y. & Sharma, R. Neofunctionalization of chromoplast specific lycopene beta cyclase gene (*CYC-B*) in tomato clade. *PLoS ONE* **11**, e0153333 (2016).
- Isaacson, T., Ronen, G., Zamir, D. & Hirschberg, J. Cloning of *tangerine* from tomato reveals a carotenoid isomerase essential for the production of  $\beta$ -carotene and xanthophylls in plants. *Plant Cell* **14**, 333–342 (2002).

24. Peters, J. L., Schreuder, M. E. L., Verduin, S. J. W. & Kendrick, R. E. Physiological characterization of a high-pigment mutant of tomato. *TGC Rep.* **51**, 23–26 (1992).
25. Gupta, S. & Van Eck, J. Modification of plant regeneration medium decreases the time for recovery of *Solanum lycopersicum* cultivar M82 stable transgenic lines. *Plant Cell. Tissue Organ Cult.* **127**, 417–423 (2016).
26. Shikata, M. & Ezura, H. Micro-Tom tomato as an alternative plant model system: mutant system: mutant collection and efficient transformation. *Plant Signal Transduct. Methods Protoc.* **1363**, 47–55 (2016).
27. Li, X. *et al.* Lycopene is enriched in tomato fruit by CRISPR/Cas9-mediated multiplex genome editing. *Front. Plant Sci.* **9**, 599 (2018).
28. Hashimoto, R., Ueta, R., Abe, C., Osakabe, Y. & Osakabe, K. Efficient multiplex genome editing induces precise, and self-ligated type mutations in tomato plants. *Front. Plant Sci.* **9**, 916 (2018).
29. Nonaka, S., Arai, C., Takayama, M., Matsukura, C. & Ezura, H. Efficient increase of  $\gamma$ -aminobutyric acid (GABA) content in tomato fruits by targeted mutagenesis. *Sci. Rep.* **7**, 7057 (2017).
30. Li, R. *et al.* Multiplexed CRISPR/Cas9-mediated metabolic engineering of  $\gamma$ -aminobutyric acid levels in *Solanum lycopersicum*. *Plant Biotechnol. J.* **16**, 415–427 (2018).
31. Bouvier, F., D'Harlingue, A., Backhaus, R. A., Kumagai, M. H. & Camara, B. Identification of neoxanthin synthase as a carotenoid cyclase paralog. *Eur. J. Biochem.* **267**, 6346–6352 (2000).
32. Sestari, I. *et al.* Near-isogenic lines enhancing ascorbic acid, anthocyanin and carotenoid content in tomato (*Solanum lycopersicum* L. cv Micro-Tom) as a tool to produce nutrient-rich fruits. *Sci. Hortic.* **175**, 111–120 (2014).
33. Okabe, Y. *et al.* Tomato TILLING technology: development of a reverse genetics tool for the efficient isolation of mutants from micro-tom mutant libraries. *Plant Cell Physiol.* **52**, 1994–2005 (2011).
34. Shinozaki, Y. *et al.* Identification and functional study of a mild allele of *SIDELLA* gene conferring the potential for improved yield in tomato. *Sci. Rep.* **8**, 12043 (2018).
35. Anzalone, A. V. *et al.* Search-and-replace genome editing without double-strand breaks or donor DNA. *Nature* **576**, 149–157 (2019).
36. Tareen, A. & Kinney, J. B. Logomaker: beautiful sequence logos in Python. *Bioinformatics* **36**, 2272–2274 (2019).
37. Ashkenazy, H., Erez, E., Martz, E., Pupko, T. & Ben-Tal, N. ConSurf 2010: calculating evolutionary conservation in sequence and structure of proteins and nucleic acids. *Nucleic Acids Res.* **38**, 529–533 (2010).
38. Kishimoto, S., Sumitomo, K., Yagi, M., Nakayama, M. & Ohmiya, A. Three routes to orange petal color via carotenoid components in 9 compositae species. *J. Jpn. Soc. Hortic. Sci.* **76**, 250–257 (2007).
39. Liu, H., Kishimoto, S., Yamamizo, C., Fukuta, N. & Ohmiya, A. Carotenoid accumulations and carotenogenic gene expressions in the petals of *Eustoma grandiflorum*. *Plant Breed.* **132**, 417–422 (2013).
40. Nagata, M. & Yamashita, I. Simple method for simultaneous determination of chlorophyll and carotenoids in tomato fruit. *Nippon Shokuhin Kogyo Gakkaishi* **39**, 925–928 (1992).
41. Wickham, H. *ggplot2: Elegant Graphics for Data Analysis* (Springer, New York, 2016).

## Acknowledgements

We would like to thank the Prof. PERES Lazaro from the University of Sao Paulo for his kind suggestion to our question on flower phenotype on their Micro-Tom NILS line carrying *og* allele. We thank P. Gramazio, Assistant Professor F. Lombardo and Assistant Professor M. Takayama for assistance and advices in short read sequence preparation and analysis. We also thank S. Kashojiya for her support in tables construction. This work was supported by the Cabinet Office, Government of Japan, Cross-ministerial Strategic Innovation Promotion Program (SIP), “Technologies for creating next-generation agriculture, forestry and fisheries” (funding agency: Bio-oriented Technology Research Advancement Institution, NARO). J. Hunziker was supported by the Monbukagakusho: MEXT scholarship for doctoral program (April 2017–March 2021).

## Author contributions

J.H., T.A. and H.E. wrote the main manuscript text, prepared figures and supplemental information. J.H., N.K., A.K. and S.K. did experiments. All authors reviewed the manuscript.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary information** is available for this paper at <https://doi.org/10.1038/s41598-020-77379-2>.

**Correspondence** and requests for materials should be addressed to H.E.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**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 licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence 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 licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2020