



OPEN

# Biofortified tomatoes provide a new route to vitamin D sufficiency

Jie Li<sup>1</sup>, Aurelia Scarano<sup>2</sup>, Nestor Mora Gonzalez<sup>3</sup>, Fabio D'Orso<sup>1,4</sup>, Yajuan Yue<sup>1</sup>, Krisztian Nemeth<sup>5</sup>, Gerhard Saalbach<sup>1</sup>, Lionel Hill<sup>1</sup>, Carlo de Oliveira Martins<sup>1</sup>, Rolando Moran<sup>6</sup>, Angelo Santino<sup>2</sup> and Cathie Martin<sup>1</sup>✉

**Poor vitamin D status is a global health problem; insufficiency underpins higher risk of cancer, neurocognitive decline and all-cause mortality. Most foods contain little vitamin D and plants are very poor sources. We have engineered the accumulation of provitamin D<sub>3</sub> in tomato by genome editing, modifying a duplicated section of phytosterol biosynthesis in Solanaceous plants, to provide a biofortified food with the added possibility of supplement production from waste material.**

Vitamin D prevents deficiency diseases affecting skeletal development<sup>1</sup> and is converted to products with steroid hormone bioactivities, which function in signalling in many organs including the brain (Supplementary Fig. 1)<sup>2</sup>. Consequently, deficiencies in vitamin D impact immune function and inflammation and are associated with increased risk of micronutrient deficiencies<sup>3</sup>, cancer<sup>3</sup>, Parkinson's disease<sup>4</sup>, depression<sup>5</sup>, neurocognitive decline<sup>6</sup>, dementia<sup>7</sup> and the severity of coronavirus disease 2019 (COVID-19) infection<sup>8</sup>. Vitamin D can be synthesized by humans from 7-dehydrocholesterol (7-DHC), also known as provitamin D<sub>3</sub>, following exposure of skin to ultraviolet B (UVB) light<sup>9</sup>, but the major source is dietary<sup>10</sup>. Approximately one billion people worldwide suffer from vitamin D insufficiency<sup>11</sup>, and numbers are increasing largely because of inadequate dietary availability (Supplementary Fig. 1a). We have developed a new dietary source of vitamin D in plants to meet the increasing demand for ways to address vitamin D insufficiency, which is of particular relevance to those adopting plant-rich, vegetarian or vegan diets<sup>10</sup>.

The original identification of vitamin D<sub>2</sub> in plants was eventually shown to be due to fungal infection<sup>12</sup>. 7-DHC is synthesized by some plants such as tomato, on route to cholesterol and steroidal glycoalkaloid (SGA) synthesis, predominantly in leaves. UVB exposure of leaves of tomato produces vitamin D<sub>3</sub> (ref. <sup>13</sup>). However, plants are generally poor dietary sources of vitamin D<sub>3</sub>, and the best are fish and dairy products. Mushrooms and yeast can be used as sources of vitamin D<sub>3</sub>, following exposure to UVB light, but vitamin D<sub>2</sub> has been reported to be substantially less bioeffective than vitamin D<sub>3</sub> in several epidemiological studies<sup>14</sup>.

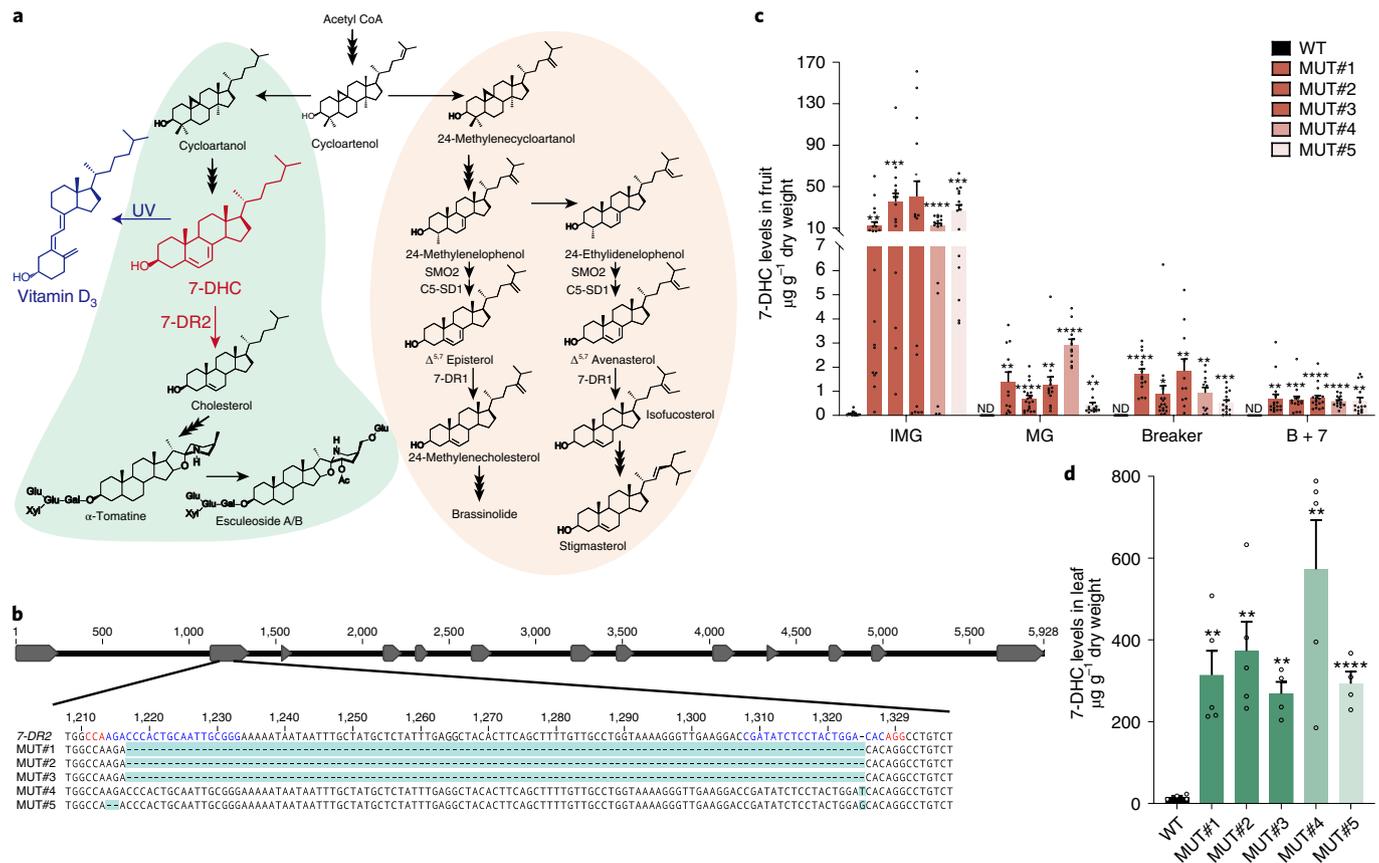
Although 7-DHC has been identified in tomato leaves, it does not normally accumulate in fruit, where it serves as an intermediate in the formation of SGAs: tomatines in green fruit and esculosides in ripe fruit<sup>13,15</sup>. Recently it has been shown that a duplicate pathway operates in Solanaceous species, including tomato, where specific isoforms of some enzymes, that are generally responsible for phytosterol and brassinosteroid biosynthesis, produce cholesterol for the

formation of SGAs<sup>16</sup> (Fig. 1a). This partial separation of phytosterol and cholesterol biosynthesis allows metabolic flexibility for the synthesis of important hormones (brassinosteroids) and more specialized stress chemicals, such as SGAs, with fungicidal, antimicrobial and insecticidal properties<sup>16</sup>. The existence of a 'duplicate' pathway for SGA biosynthesis in tomato makes the engineering of 7-DHC accumulation relatively straightforward. A specific isoform of 7-dehydrocholesterol reductase (SI7-DR2) converts 7-DHC to cholesterol for the synthesis of  $\alpha$ -tomatine in leaves and fruit<sup>16</sup> (Fig. 1a). Consequently, knocking out the activity of SI7-DR2 should result in the accumulation of 7-DHC with minimal impact on phytosterol and brassinosteroid biosynthesis. With the aim of biofortifying tomato in provitamin D<sub>3</sub>, we tested the efficacy of blocking SI7-DR2 activity using clustered regularly interspaced short palindromic repeats–CRISPR-associated protein 9 (CRISPR–Cas9) genome editing to increase 7-DHC levels. Two single-guide RNAs (sgRNAs) were designed to sequences within the second exon of the *SI7-DR2* gene (Fig. 1b), with minimal homology between the sgRNAs and the only other sterol  $\Delta^7$  reductase gene in tomato (*SI7-DR1*). We recovered five independent knockout alleles of the *SI7-DR2* gene within the T1 generation, three of which carried identical deletions of 108 bp of the exon 2 sequence between the two sgRNAs (MUT#1, MUT#2 and MUT#3). Two other knockout alleles were created by deletion of 2 bp with insertion of 1 bp (MUT#5) or by insertion of 1 bp only in the second exon (MUT#4), both of which caused frame shifts and predicted premature termination of the SI7-DR2 protein (Fig. 1b). Homozygous-knockout alleles were recovered in the T1 generation, and homozygous-knockout lines lacking the transfer DNA (T-DNA) carrying the *Cas9* gene and the sgRNA sequences were recovered for four of the five lines within the T2 generation.

Five homozygous-knockout alleles of *SI7-DR2* were selected, following segregation, in the T2 generation. No off-target edits of the *SI7-DR1* gene were detected in these mutant lines (Supplementary Fig. 3). Fruit and leaves were analysed for 7-DHC content as well as levels of other phytosterols, cholesterol and SGAs using liquid chromatography–mass spectrometry (LC–MS)<sup>13,17</sup>.

Loss of SI7-DR2 activity had no effect on the growth, development or yield of the tomato lines (Supplementary Fig. 1c,d). This contrasts with the phenotype of the loss-of-function mutation of the gene encoding sterol  $\Delta^7$ -reductase involved in phytosterol biosynthesis in *Arabidopsis* (*DWARF5*), which is dwarfed because of an inhibition of brassinosteroid biosynthesis<sup>18</sup>. The lack of effect of mutations in *SI7-DR2* on phytosterol metabolism was confirmed by

<sup>1</sup>John Innes Centre, Norwich Research Park, Norwich, UK. <sup>2</sup>Institute of Sciences of Food Production, C.N.R., Unit of Lecce, Lecce, Italy. <sup>3</sup>Recombinant Biopharmaceutical Laboratory, Department of Pharmacology, Biological Sciences Faculty, University of Concepción, Concepción, Chile. <sup>4</sup>CREA—Research Centre for Genomics and Bioinformatics, Rome, Italy. <sup>5</sup>University of Glasgow, Glasgow, UK. <sup>6</sup>University of Camagüey, Camagüey, Cuba. ✉e-mail: [cathie.martin@jic.ac.uk](mailto:cathie.martin@jic.ac.uk)



**Fig. 1 | 7-DHC accumulated in *SI7-DR2*-homozygous-knockout lines.** **a**, The cholesterologenesis pathway (depicted in light green) and phytosterol biosynthesis pathway (depicted in light orange) in tomato, redrawn from Sonawane et al.<sup>16</sup>. 7-DHC is converted by 7-DR2 to cholesterol, which can be converted to vitamin D<sub>3</sub> by exposure to UVB light. SMO, C-4 sterol methyl oxidase; C5-SD1, sterol C-5(6) desaturase 1. **b**, Five independent *SI7-DR2*-knockout lines were generated by genome editing. Top: schematic structure of *SI7-DR2* gene, with exons indicated as grey arrows. Bottom: recovered mutations in each line are highlighted in light blue. The CRISPR-Cas9-targeted sequences and the protospacer-adjacent motif sequences are shown in blue and red, respectively. **c**, 7-DHC contents in wild-type (WT) and *SI7-DR2*-knockout tomato fruit at different stages of ripening (IMG, immature green; MG, mature green; Breaker, fruit turning ripe; B + 7, 7 days after breaker-ripe fruit). Data are presented as the mean  $\pm$  s.e.m. From left to right:  $n = 14, 19, 16, 15, 16, 14, 13, 11, 18, 13, 10, 15, 15, 14, 17, 11, 11, 15, 9, 17, 14, 17, 15$  and 15 biologically independent fruit samples. ND, not detected. **d**, 7-DHC content of leaves of wild-type and *SI7-DR2*-knockout lines. Data are presented as mean  $\pm$  s.e.m. From left to right:  $n = 4, 5, 5, 4, 5$  and 4 biologically independent leaf samples. Statistical significance between WT and mutants at each fruit ripening stage (**c**) or in leaves (**d**) was assessed using two-tailed *t*-tests (\* $P \leq 0.05$ , \*\* $P \leq 0.01$ , \*\*\* $P \leq 0.001$ , \*\*\*\* $P \leq 0.0001$ ).

comparing the levels of stigmasterol, an end product of the phytosterol pathway in tomato, in leaves of wild-type and edited lines (Supplementary Fig. 1e) and other phytosterols in fruit and leaves (Extended Data Fig. 1). In wild-type plants, 7-DHC was detected only in immature green fruit and was undetectable in ripening and ripe fruit. In contrast, loss of *SI7-DR2* activity resulted in substantial increases in 7-DHC levels in leaves and green fruit (Fig. 1c,d). Levels of 7-DHC were lower in ripe fruit of the *SI7-DR2* mutants (Fig. 1c) but remained high enough that, if converted to vitamin D<sub>3</sub> by treatment with UVB, the amount in one tomato would be equivalent to that in two medium-sized eggs or 28 g of tuna, which are both recommended dietary sources of vitamin D (FoodData Central USDA, <https://fdc.nal.usda.gov/>). For the elderly with declining levels of 7-DHC, consuming fruit biofortified with 7-DHC might address their deficiencies directly<sup>19</sup>.

Matrix-assisted laser desorption/ionization (MALDI) imaging showed that the increases in 7-DHC were distributed in both the flesh and peel of tomatoes (Fig. 2a and Supplementary Fig. 1f).  $\alpha$ -Tomatine and dehydrotomatine are converted to esculeoside A and dehydroesculeoside A, respectively, during fruit ripening,

meaning that tomatines are reduced to low levels in ripe fruit<sup>20</sup>. MALDI imaging of mutant and wild-type green fruit showed that  $\alpha$ -tomatine was lower in the *SI7-DR2* mutants than in controls (Fig. 2a and Supplementary Fig. 1f,g), and leaves showed substantially lower levels in the mutants, although  $\alpha$ -tomatine was not eliminated (Fig. 2b). A strong reduction in the levels of the SGA, esculeoside A, was also observed in ripe fruit of the mutants compared with the control (Fig. 2c). The reduction in  $\alpha$ -tomatine might be considered beneficial because of its reported toxicant/antinutritional activity. Interestingly, cholesterol levels were generally higher than in wild-type controls in both fruit (Fig. 2a and Supplementary Fig. 1h) and leaves (Fig. 2d). This suggested that the block in flux along the SGA biosynthetic pathway may be compensated by increased flux of intermediates, perhaps catalysed by the enzymes of the phytosterol pathway (or at least *SI7-DR1*), which supplement cholesterol production and limit reductions in SGA accumulation. However, this does not involve compensatory changes in expression of the genes encoding enzymes in either pathway as shown by reverse transcription quantitative PCR (RT-qPCR) analysis of these genes in leaves of wild-type and mutant lines (Supplementary

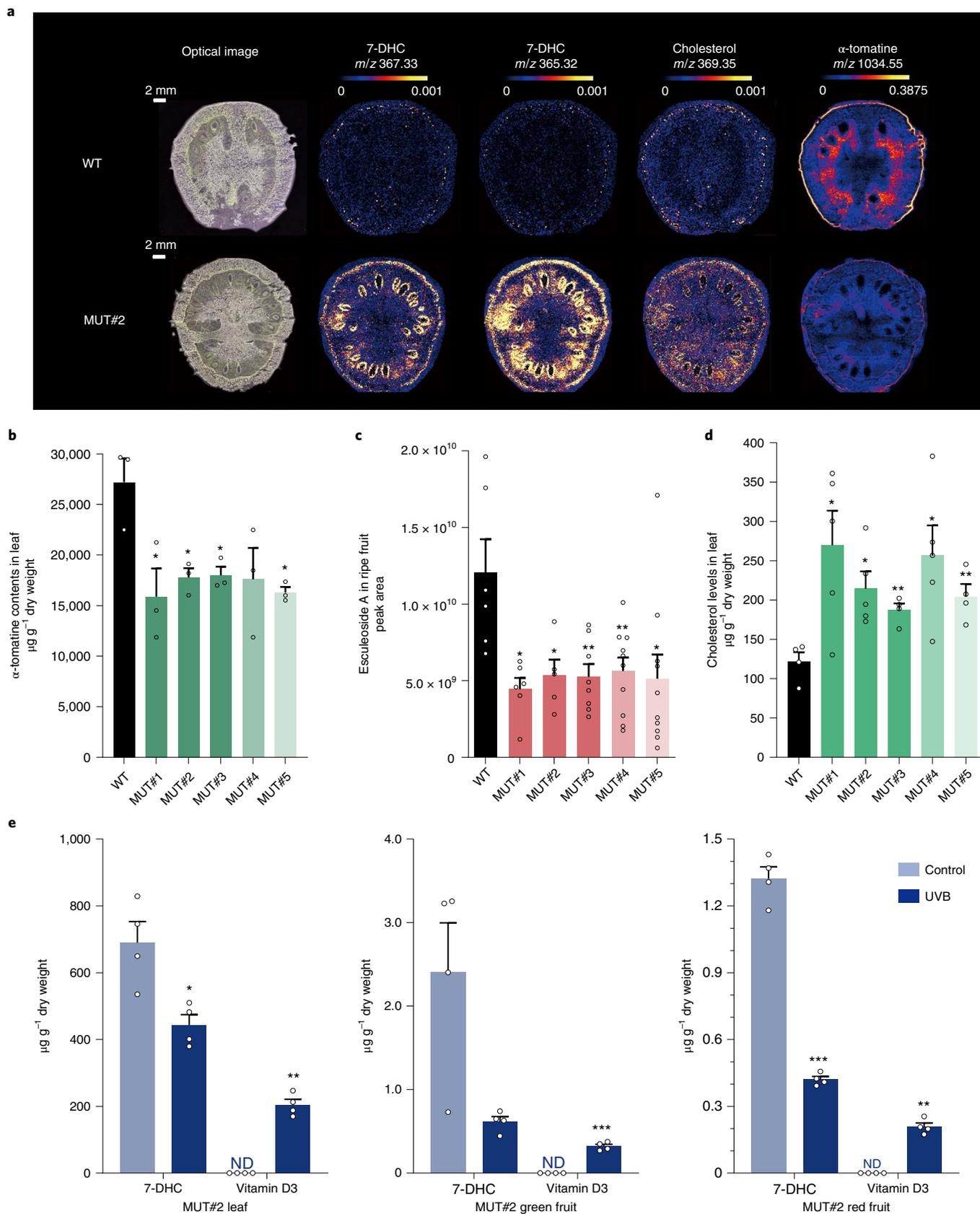


Fig. 2a). Only *SIC5-SD1* in the phytosterol pathway showed consistently lower (~30%) transcript levels than controls (Fig. 1a and Supplementary Fig. 2b).

As an important confirmation of our biofortification strategy, we tested whether the elevated levels of 7-DHC in *Sl7-DR2*-mutant plants could be converted to vitamin D<sub>3</sub> by irradiating leaves and

**Fig. 2 | Localization and quantitative comparison of 7-DHC, SGAs and cholesterol in wild-type and 7-DR2-knockout lines and conversion of 7-DHC in *Sl7-DR2* knockouts to vitamin D<sub>3</sub> by UVB irradiation.** **a**, MALDI images of 7-DHC ( $m/z$  367.33) and its laser-induced derivative ion ( $m/z$  365.32), cholesterol ( $m/z$  369.35) and  $\alpha$ -tomatine ( $m/z$  1,034.55). Scale bar, 2 mm. The HotMetal2 colour scale indicates the range of total ion current-normalized intensity. The same metabolite is shown with identical scale intensity for wild-type and mutant samples. It is not straightforward to compare the relative abundance of different metabolites using MALDI images due to potentially different ionization efficiencies. **b**,  $\alpha$ -Tomatine contents of leaves of wild-type and *Sl7-DR2*-knockout lines (mean  $\pm$  s.e.m,  $n = 3$  biologically independent leaf samples for each line). **c**, Relative esculeoside A content of red-ripe (seven days after breaker) fruit of wild-type and *Sl7-DR2*-knockout lines (mean  $\pm$  s.e.m). From left to right:  $n = 6, 6, 5, 8, 10$  and 10 biologically independent fruit samples. **d**, Cholesterol content of leaves of wild-type and *Sl7-DR2*-knockout lines (mean  $\pm$  s.e.m). From left to right,  $n = 4, 5, 5, 4, 5$  and 4 biologically independent leaf samples. **e**, Contents of 7-DHC and vitamin D<sub>3</sub> in control and UVB-treated leaves or fruit (mean  $\pm$  s.e.m,  $n = 4$  biologically independent leaf or fruit samples at each stage for control and MUT#2). Tissues of Mut#2 were irradiated by UVB light for 1 h. The experiment was repeated three times. ND, not detected. Statistical significance between WT and mutant values (**b–d**) and between control and UVB-treated tissue (**e**) was assessed using two-tailed *t*-tests (\* $P \leq 0.05$ , \*\* $P \leq 0.01$ , \*\*\* $P \leq 0.001$ , \*\*\*\* $P \leq 0.0001$ ).

sliced fruit with UVB light for 1 h as described by Jäpelt et al.<sup>17</sup>. Treatment of leaves was very effective, resulting in yields of vitamin D<sub>3</sub> of nearly 200  $\mu\text{g g}^{-1}$  dry weight (Fig. 2e). Yields from fruit were lower, reaching 0.3  $\mu\text{g g}^{-1}$  dry weight in green fruit and 0.2  $\mu\text{g g}^{-1}$  dry weight in red fruit, reflecting the declining content of 7-DHC in green fruit and red-ripe fruit compared with leaves (Fig. 2e). A medium-size tomato has a dry weight of about 8–10 g, and the levels of vitamin D<sub>3</sub> that could be achieved in one *Sl7-DR2*-mutant tomato approach 30% in green fruit and 20% in red fruit of the recommended daily allowance (RDA) calculated on baseline dietary intakes of 2.8  $\mu\text{g}$  per day (United Kingdom) and 4.3  $\mu\text{g}$  per day (United States)<sup>21,22</sup>. Vitamin D<sub>3</sub> in ripe fruit might be increased further by extended exposure to UVB, for example, during sun-drying.

The duplicate pathway for cholesterol/SGA biosynthesis exists in other food crops of the Solanaceae family, including eggplant (*Solanum melongena*), potato (*Solanum tuberosum*) and pepper (*Capsicum annuum*)<sup>16</sup>. The close association between cholesterol/SGA biosynthesis, 7-DHC accumulation and photosynthesis in leaves and green fruit of tomato (Figs. 1c,d and 2b–d and Supplementary Fig. 1g,h) suggests that knockouts of 7-DR2 activity in pepper, where fruit may be green when eaten, might also provide a vitamin D<sub>3</sub>-biofortified, plant-based food. Mutations that increase UVB penetration into fresh fruit, such as *y* in tomato that causes the loss of UV-protecting chalcones from the skin of ‘pink tomatoes’, might offer increased UVB conversion of provitamin D<sub>3</sub> to vitamin D<sub>3</sub>. Such stacking could be achieved by further gene editing or by introgression<sup>23</sup>. The leaves of the *Sl7-DR2* mutants are rich sources of provitamin D<sub>3</sub> and consequently could provide an important new feed stock using the waste vegetative material from tomato cultivation for the manufacture of vitamin D<sub>3</sub> supplements from plants that would be suitable for vegans. Editing of *Sl7-DR2* could generate similar alterations in any elite tomato variety, meaning that tomato could be developed as a plant-based, sustainable source of vitamin D<sub>3</sub>.

## Methods

**Plant materials.** Tomato (*Solanum lycopersicum*) cv. Money Maker and knockout mutants of *Sl7-DR2* were grown in the greenhouse at an average ambient temperature of 20–22 °C. Supplemental lighting was available to maintain 16 h of light per day when necessary.

**Plasmid construction.** Two specific target sequences (Fig. 1b) in exon 2 of the *Sl7-DR2* (*Solyc06g074090*) gene were selected to generate *Sl7-DR2*-knockout mutants. These were introduced into the sgRNA scaffold by PCR. To make the sgRNA-expression cassette, each sgRNA amplicon and a synthesized U6-III promoter (pICSL90001) were cloned into a GoldenGate Level 1 acceptor (pICH47732 and pICH47742). A Level 2 binary vector, pICSL002203, containing the Cas9-expression cassette and kanamycin-resistance-expression (nptII) cassette, was used as the destination vector to generate the *Sl7-DR2*-CRISPR–Cas9 construct. sgRNA efficiency was tested by co-transformation of tomato using *Agrobacterium rhizogenes* (strain A7ATCC15834)<sup>24</sup>. The sequences of exon 2 of *Sl7-DR2* were amplified by PCR directly from hairy roots with the Phire Plant Direct PCR Master Mix following the manufacturer’s instructions (Thermo Scientific) using primers flanking the sgRNA target sequences (forward: TGTTTCACTGGGCTGGTTTACG and reverse: GAGAAGTCTTTCACC

ATGTCACGA). Stable transformations were conducted after the sgRNA efficiency check<sup>25</sup>.

**Tomato stable transformation.** The *Sl7-DR2*-CRISPR–Cas9 construct was transformed into *Agrobacterium tumefaciens* (strain AGL1) for stable transformation, which was undertaken using cotyledons as initial explants.

**Screening of *Sl7-DR2*-knockout lines.** DNA was isolated from the finely ground powder of leaf tissues using DNeasy Plant Mini Kits (Qiagen) following the manufacturer’s instructions. Five independent *Sl7-DR2*-knockout lines were obtained by genotyping with primers flanking the sgRNA target sequences (forward: TGTTTCACTGGGCTGGTTTACG and reverse: GAGAAGTCTTTCACCATGTCACGA) and confirmed by sequencing.

**RT-qPCR analysis.** Total RNA was extracted from tomato leaf tissues using the Trizol method (Sigma-Aldrich). DNase I (Roche)-treated RNA was reverse transcribed using SuperScript<sup>™</sup> III (Invitrogen). SYBR Green JumpStart Taq ReadyMix (Sigma-Aldrich) was used to perform all the RT-qPCR reactions using the X96 Touch Real-Time PCR Detection System (Biorad). Data were analysed using CFX Maestro Software. *SlActin* (*Solyc03g078400*) was selected as the house-keeping reference gene. The relative expression of genes was calculated by the  $\Delta\text{Ct}$  method. Gene-specific primers were designed using National Center for Biotechnology Information (NCBI) primer Basic Local Alignment Search Tool (BLAST) (<https://www.ncbi.nlm.nih.gov/tools/primer-blast/>), listed in Supplementary Table 1. All primers used in this study were synthesized by Sigma-Aldrich.

**MALDI-imaging analysis.** Cryosectioning of fruit was undertaken as described by Dong et al.<sup>26</sup>. Fresh immature green fruit (about 16 days after anthesis) was flash frozen in liquid nitrogen and then embedded with M1 embedding matrix (Thermo Scientific) on a flat metal holder on dry ice. The embedded tissues were transferred to a CryoStar NX70 Microtome (Thermo Scientific) and thermally equilibrated at –18 °C for at least 3 h. The tissues were cut into 35- $\mu\text{m}$ -thick sections and thaw-mounted on Superfrost Plus slides (Thermo Scientific), followed by vacuum drying in a desiccator.

Optical images were taken using a Canon 5D Mark IV camera with a Canon MP-E 65 mm f/2.8 1–5x Macro Photo lens (Canon) at 1:1 ratio. Raw image files were processed with Capture One photo editing software (Capture One).

Sections were covered with 2,5-dihydroxybenzoic acid matrix (DHB) using a SunCollect MALDI Sprayer (SunChrome) with a DHB solution of 10  $\text{mg ml}^{-1}$  in 80% methanol/0.05% trifluoroacetic acid (TFA) to a density of approximately 3  $\mu\text{g mm}^{-2}$ .

MALDI imaging was performed with a Synapt G2-Si mass spectrometer with a MALDI source (Waters) equipped with a 2.5 kHz Nd:YAG (neodymium-doped yttrium aluminum garnet) laser operated at 355 nm. The slides were fixed in the instrument metal holder and were scanned with a flat-bed scanner (Canon). The images were used to generate pattern files and acquisition methods in the HDImaging software version 1.4 (Waters) with the following parameters: area of a complete section approximately 400  $\text{mm}^2$ , laser beam diameter at low setting (60  $\mu\text{m}$ ) with 105  $\mu\text{m}$  step size, resulting in approximately 36,000 pixels per section, MALDI-MS-positive sensitivity mode,  $m/z$  50–1,200, scan time 0.5 s, laser repetition rate 1 kHz, laser energy 200. For ion-mobility measurements, the same parameters were used in MALDI-high-definition mass spectrometry (HDMS) mode with the following additional tune page settings: trap d.c. bias, 45.0; transfer wave velocity ( $\text{m s}^{-1}$ ), 315; IMS (ion mobility spectrometry) wave height (V), 40.0; variable wave velocity enabled with linear ramp, wave velocity start ( $\text{m s}^{-1}$ ): 1,500.0; wave velocity end ( $\text{m s}^{-1}$ ), 200.0. Red phosphorus clusters were used for instrument calibration and lock mass correction. Total scan time for a complete section was 10–12 h, and the lock mass was acquired every 10 min for 2 s.

The MS raw files were processed in HDI1.4 with the following parameters: detection of the 2,000 most-abundant peaks,  $m/z$  window 0.05, MS resolution

10,000, lock mass 526.554 (red phosphorous cluster). The processed data were loaded into HDI1.4 and normalized by total ion content. Images were generated using the HotMetal2 colour scale and exported as png image files. Compounds of interest, 7-DHC, cholesterol and  $\alpha$ -tomatine were identified by comparison with the drift time and mass of authentic standards analysed on the same instrument. The masses detected for 7-DHC, vitamin D<sub>3</sub> and cholesterol during MALDI are listed in Supplementary Table 2. It has been reported that cholesterol is susceptible to laser-induced oxidation during MALDI time of flight MS<sup>27</sup>, and 7-DHC has an even higher tendency for non-enzymatic autoxidation<sup>28,29</sup>. Among the peaks of standards generated during MALDI, taking into account their specificity and relative abundance, 367.33, 365.32 and 363.31 were selected as representative masses for 7-DHC and 369.35 and 1034.55 were selected as representative masses for cholesterol and  $\alpha$ -tomatine, respectively (Supplementary Table 2).

**Sterol analysis.** The methods for sterol extraction and analysis were modified from Jäpelt et al.<sup>14</sup>. For consistency of extraction efficiency and comparability among samples, all extracts were prepared with the same mass of tissue per volume of extraction solvents. Freeze-dried material (about 20 mg) was weighed into a 2 ml Eppendorf tube and mixed with 100  $\mu$ l 60% potassium hydroxide (Sigma-Aldrich), 500  $\mu$ l 96% ethanol (Sigma-Aldrich) and 300  $\mu$ l 15% ascorbic acid (Sigma-Aldrich). To avoid thermal isomerization of vitamin D<sub>3</sub> to previtamin D<sub>3</sub>, we applied overnight cold saponification rather than hot saponification. The sample preparation process was carefully maintained at or below room temperature. The tubes were shaken for approximately 18 h at 22 °C in a thermoshaker (Eppendorf). Twenty percent ethyl acetate in pentane (v/v) (750  $\mu$ l) was added and shaken for 30 min on a flat shaker, followed by centrifugation at 2000  $\times$ g for 5 min at room temperature. The organic layer was transferred into a new 2 ml Eppendorf tube. The extraction steps were repeated twice. Total extracts were washed with 500  $\mu$ l of 0.1 mol l<sup>-1</sup> hydrochloric acid by inverting tubes 30 times to completely remove the alkali. The upper layer was transferred into a 2 ml Eppendorf tube following centrifugation at 1000  $\times$ g for 2 min. Total extracts were evaporated to dryness using a Genevac EZ-2 Elite Evaporator with the programme of 'Very Low BP Mix'. The residue was finally redissolved in 200  $\mu$ l methanol and filtered through 0.22  $\mu$ m nylon Corning® Costar® Spin-X® tube filter (Sigma-Aldrich). The samples were stored at -80 °C until analysis.

Sterol compounds were identified by comparison with the retention time and mass spectrometry spectra of authentic standards (Sigma-Aldrich) analysed on the same instrument, and were quantified based on calibration curves generated with authentic standards running along with the samples using Xcalibur Quan Browser software (Version 4.3). LC analysis was undertaken on a Dionex UltiMate (Thermo Scientific) equipped with a thermostated column compartment. The chromatographic separation was done on a 50  $\times$  2.1 mm 2.6  $\mu$ m Kinetex F5 column (Phenomenex) at a flow rate of 0.6 ml min<sup>-1</sup>. Solvents were 0.2% formic acid and 25% acetonitrile in Milli-Q water (v/v) (A) versus 100% methanol (B). The gradient programme was as follows: 60% B for 0.5 min, a linear gradient to 85% B for 7 min, a linear gradient to 100% B for 0.5 min, isocratic elution for 1 min and 0.5 min linear gradient back to 60% B and re-equilibration for 3.5 min, giving a total run time of 13 min. The column was maintained at 40 °C. Five microlitre samples were injected. MS was performed using a Q Exactive Orbitrap Mass Spectrometer (Thermo Scientific) with an atmospheric pressure chemical ionization source. The MS was set up to collect full scans at 70,000 resolution from *m/z* 180–2,000 and data-dependent MS<sub>2</sub> of the top 4 ions, at an isolation width of *m/z* 4.0, 30% normalized collision energy. These ions were then ignored for 5 s in favour of the next-most-abundant ion; isotope peaks were also ignored. Data-dependent MS<sub>2</sub> analysis was at 17,500 resolution, maximum ion time of 50 ms, automatic gain control target of 1  $\times$  10<sup>5</sup> ions. MS scans had a maximum ion time of 50 ms and automatic gain control target of 3  $\times$  10<sup>6</sup> ions. Spray chamber conditions were 231 °C capillary temperature, 21.25 units sheath gas, 5 units aux gas, no spare gas, 4  $\mu$ A current, 363 °C probe heater temperature and 50V S-lens RF. Samples were run in randomized order to avoid systematic bias. Xcalibur software (version 4.3, Thermo Scientific) was used for instrument control and data acquisition.

**SGA analysis.** SGA extraction was performed as described by Itkin et al.<sup>30</sup>. Briefly, 20 mg of freeze-dried samples (leaf or fruit) were sonicated in 1 ml 100% methanol, incubated on ice for 1 h and centrifuged at 1,700  $\times$ g for 10 min. The supernatants were collected, centrifuged at 1,700  $\times$ g for 3 min and filtered with a 0.22  $\mu$ m filter. The sample extracts were stored at -80 °C until the analysis.  $\alpha$ -tomatine was identified and quantified by comparing the retention time and MS spectrum to the authentic standard (Sigma-Aldrich). Esculeoside A was identified on the basis of its MS spectrum compared with previously published spectra and relative retention time.

Chromatographic separation was performed on a 50  $\times$  2.1 mm 2.6  $\mu$ m Kinetex EVO C18 column (Phenomenex) at a flow rate of 0.6 ml min<sup>-1</sup>. Solvents were 0.1% formic acid in Milli-Q water (v/v) (A) versus 100% acetonitrile (B). The gradient programme was as follows: a linear gradient from 2% B to 40% B for 4 min, a linear gradient to 95% B for 2 min, isocratic elution for 1 min and a 0.1 min linear gradient back to 2% B and re-equilibration for 2.1 min, giving a total run time of 9.2 min. MS was performed using a Q Exactive Orbitrap Mass Spectrometer (Thermo Scientific) with an electrospray ionization (ESI) source. All other settings were the same as those described above for sterol analysis.

**UV treatment.** Fruits were cut into 1 mm slices before UV exposure. Leaf or fruit tissues were exposed to UVB light (3.2 mW cm<sup>-2</sup>) for 1 h at 20 cm below the inverted ultraviolet transilluminator (302 nm; Analytik-Jena). Samples were immediately frozen in liquid nitrogen after treatment for subsequent analysis.

**Statistical analysis.** All the experiments in this paper were repeated at least three times independently, and results from representative datasets are presented. All numerical values are presented as means  $\pm$  s.e.m. GraphPad Prism (version 9.2.0) was used for the statistical analysis. Statistical differences between wild-type and mutants were calculated using two-tailed *t*-tests with or without Welch's correction, for comparing groups with or without unequal variance, respectively. Related information is listed in the Source Data.

**Reporting Summary.** Further information on research design is available in the Nature Research Reporting Summary linked to this article.

## Data availability

Materials generated and analysed in this study are available from the corresponding author upon request. A reporting summary for this paper is available as a Supplementary information file. Source data are provided with this paper.

Received: 22 September 2021; Accepted: 12 April 2022;

Published online: 23 May 2022

## References

- McCullum, E. V., Simmonds, N., Becker, J. E. & Shipley, P. G. Studies on experimental rickets. XXI. An experimental demonstration of the existence of a vitamin which promotes calcium deposition. *J. Biol. Chem.* **53**, 293–312 (1922).
- Bikle, D. & Christakos, S. New aspects of vitamin D metabolism and action—addressing the skin as source and target. *Nat. Rev. Endocrinol.* **16**, 234–252 (2020).
- Garland, C. & Garland, F. Do sunlight and vitamin D reduce the likelihood of colon cancer? *Int. J. Epidemiol.* **9**, 227–231 (1980).
- Zhou, Z., Zhou, R., Zhang, Z. & Li, K. The association between vitamin D status, vitamin D supplementation, sunlight exposure, and Parkinson's disease: a systematic review and meta-analysis. *Med. Sci. Monit.* **25**, 666–674 (2019).
- Anglin, R. E., Samaan, Z., Walter, S. D. & McDonald, S. D. Vitamin D deficiency and depression in adults: systematic review and meta-analysis. *Br. J. Psychiatry* **202**, 100–107 (2013).
- Llewellyn, D. J. et al. Vitamin D and risk of cognitive decline in elderly persons. *Arch. Intern. Med.* **170**, 1135–1141 (2010).
- Sommer, I. et al. Vitamin D deficiency as a risk factor for dementia: a systematic review and meta-analysis. *BMC Geriatrics* **17**, 1–13 (2017).
- Pizzini, A. et al. Impact of vitamin D deficiency on COVID-19—a prospective analysis from the CovILD registry. *Nutrients* **12**, 2775 (2020).
- Holick, M. F. et al. Photosynthesis of previtamin D<sub>3</sub> in human skin and the physiological consequences. *Science* **210**, 203–205 (1980).
- Jäpelt, R. B. & Jakobsen, J. Vitamin D in plants: a review of occurrence, analysis, and biosynthesis. *Front. Plant Sci.* **4**, 136 (2013).
- van Schoor, N. & Lips, P. Chapter 59—worldwide vitamin D status. In *Vitamin D* 4th edn (ed. Feldman, D.) 15–40 (Academic Press, 2018).
- Black, L. J., Lucas, R. M., Sherriff, J. L., Björn, L. O. & Bornman, J. F. In pursuit of vitamin D in plants. *Nutrients* **9**, 136 (2017).
- Jäpelt, R. B., Silvestro, D., Smedsgaard, J., Jensen, P. E. & Jakobsen, J. LC-MS/MS with atmospheric pressure chemical ionisation to study the effect of UV treatment on the formation of vitamin D<sub>3</sub> and sterols in plants. *Food Chem.* **129**, 217–225 (2011).
- Heaney, R. P., Recker, R. R., Grote, J., Horst, R. L. & Armas, L. A. Vitamin D<sub>3</sub> is more potent than vitamin D<sub>2</sub> in humans. *J. Clin. Endocrinol. Metab.* **96**, E447–E452. (2011).
- Boland, R., Skliar, M., Curino, A. & Milanesi, L. Vitamin D compounds in plants. *Plant Sci.* **164**, 357–369 (2003).
- Sonawane, P. D. et al. Plant cholesterol biosynthetic pathway overlaps with phytosterol metabolism. *Nat. Plants* **3**, 1–13 (2016).
- Jäpelt, R. B., Silvestro, D., Smedsgaard, J., Jensen, P. E. & Jakobsen, J. Quantification of vitamin D<sub>3</sub> and its hydroxylated metabolites in wax leaf nightshade (*Solanum glaucophyllum* Desf.), tomato (*Solanum lycopersicum* L.) and bell pepper (*Capsicum annuum* L.). *Food Chem.* **138**, 1206–1211 (2013).
- Choe, S. et al. Lesions in the sterol D<sup>7</sup> reductase gene of *Arabidopsis* cause dwarfism due to a block in brassinosteroid biosynthesis. *Plant J.* **21**, 431–443 (2000).
- Prabhu, A. V., Luu, W., Sharpe, L. J. & Brown, A. J. Cholesterol-mediated degradation of 7-dehydrocholesterol reductase switches the balance from cholesterol to vitamin D synthesis. *J. Biol. Chem.* **291**, 8363–8373 (2016).
- Cárdenas, P. D. et al. Pathways to defense metabolites and evading fruit bitterness in genus *Solanum* evolved through 2-oxoglutarate-dependent dioxygenases. *Nat. Commun.* **10**, 5169 (2019).

21. SACN (Scientific Advisory Committee on Nutrition). *Vitamin D and Health*. <https://www.gov.uk/government/publications/sacn-vitamin-d-and-health-report> (2016).
22. U.S. Department of Agriculture, Agricultural Research Service. *Nutrient Intakes from Food and Beverages: Mean Amounts Consumed per Individual, by Gender and Age, What We Eat in America, NHANES 2017-2018*. <http://www.ars.usda.gov/nea/bhnrc/fsrg> (2020).
23. Ballester, A. R. et al. Biochemical and molecular analysis of pink tomatoes: deregulated expression of the gene encoding transcription factor SLMYB12 leads to pink tomato fruit color. *Plant Physiol.* **152**, 71–84 (2010).
24. Ron, M. et al. Hairy root transformation using *Agrobacterium rhizogenes* as a tool for exploring cell type-specific gene expression and function using tomato as a model. *Plant Physiol.* **166**, 455–469 (2014).
25. Galdon-Armero, J., Arce-Rodriguez, L., Downie, M., Li, J. & Martin, C. A scanning electron micrograph-based resource for identification of loci involved in epidermal development in tomato: elucidation of a new function for the Mixta-like transcription factor in leaves. *Plant Cell* **32**, 1414–1433 (2020).
26. Dong, Y. et al. High mass resolution, spatial metabolite mapping enhances the current plant gene and pathway discovery toolbox. *New Phytol.* **228**, 1986–2002 (2020).
27. McAvey, K. M. et al. Laser-induced oxidation of cholesterol observed during MALDI-TOF mass spectrometry. *J. Am. Soc. Mass. Spectrom.* **22**, 659–669 (2011).
28. Xu, L., Davis, T. A. & Porter, N. A. Rate constants for peroxidation of polyunsaturated fatty acids and sterols in solution and in liposomes. *J. Am. Chem. Soc.* **131**, 13037–13044 (2009).
29. Xu, L., Korade, Z. & Porter, N. A. Oxysterols from free radical chain oxidation of 7-dehydrocholesterol: product and mechanistic studies. *J. Am. Chem. Soc.* **132**, 2222–2232 (2010).
30. Itkin, M. et al. GLYCOALKALOID METABOLISM1 is required for steroidal alkaloid glycosylation and prevention of phytotoxicity in tomato. *Plant Cell* **23**, 4507–4525 (2011).

## Acknowledgements

This study was supported by the joint ERA CAPS Regulome project: BBSRC BB/N005023/1 (A. Scarano, F.D. and C.M.); the Institute Strategic Programmes 'Understanding and Exploiting Plant and Microbial Secondary Metabolism' (BB/J004596/1) and 'Molecules from Nature' (BB/P012523/1), CEPAMS Funding (Project CPM19), and a Proof of Concept Award from the High Value Compounds from Plants Network in Industrial Biotechnology and Bioenergy (POC-OCT17-04) from the UK Biotechnology and Biological Sciences Research Council (J.L., N.M.G., F.D., G.S., L.H., C.d.O.M. and C.M.), the European-funded COST ACTION FA1106, QualityFruit, which supported F.D. with an STSM award, and the European Union's Horizon 2020 research and innovation programme through the TomGEM project under grant

agreement number 679796 (A. Scarano and C.M.). A. Santino and A. Scarano were in part supported by CNR-DiSBA project NutrAge (project number 7022) and the European-funded COST Action CA18111 (PlantEd, Genome editing in plants). F.D. was in part supported by BIOTECH-Cisget project from the Italian Ministry of Agriculture (MiPAAF) and by SMART-BREED project (A0375E0166, POR FESR LAZIO 2014 – 2020). We are grateful to P. Brett for help with metabolite analysis. We also thank S. Lanham-New for introducing us to this important topic and for enthusiastic support.

## Author contributions

C.M., A. Santino and R.M. designed the project, N.M.G. and F.D. built the vectors for genome editing and tested the guide efficiency in hairy roots, Y.Y. undertook tomato transformation, J.L. and A. Scarano genotyped edited plants and undertook all the metabolite analysis with assistance from L.H., G.S. performed the MALDI-imaging with assistance from C.d.O.M. and J.L. and K.N. did the UV experiments. J.L. and C.M. drafted the manuscript and prepared the figures with all authors contributing to and approving the final version.

## Competing interests

The authors declare no competing interests.

## Additional information

**Extended data** is available for this paper at <https://doi.org/10.1038/s41477-022-01154-6>.

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41477-022-01154-6>.

**Correspondence and requests for materials** should be addressed to Cathie Martin.

**Peer review information** *Nature Plants* thanks Pan Liao and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

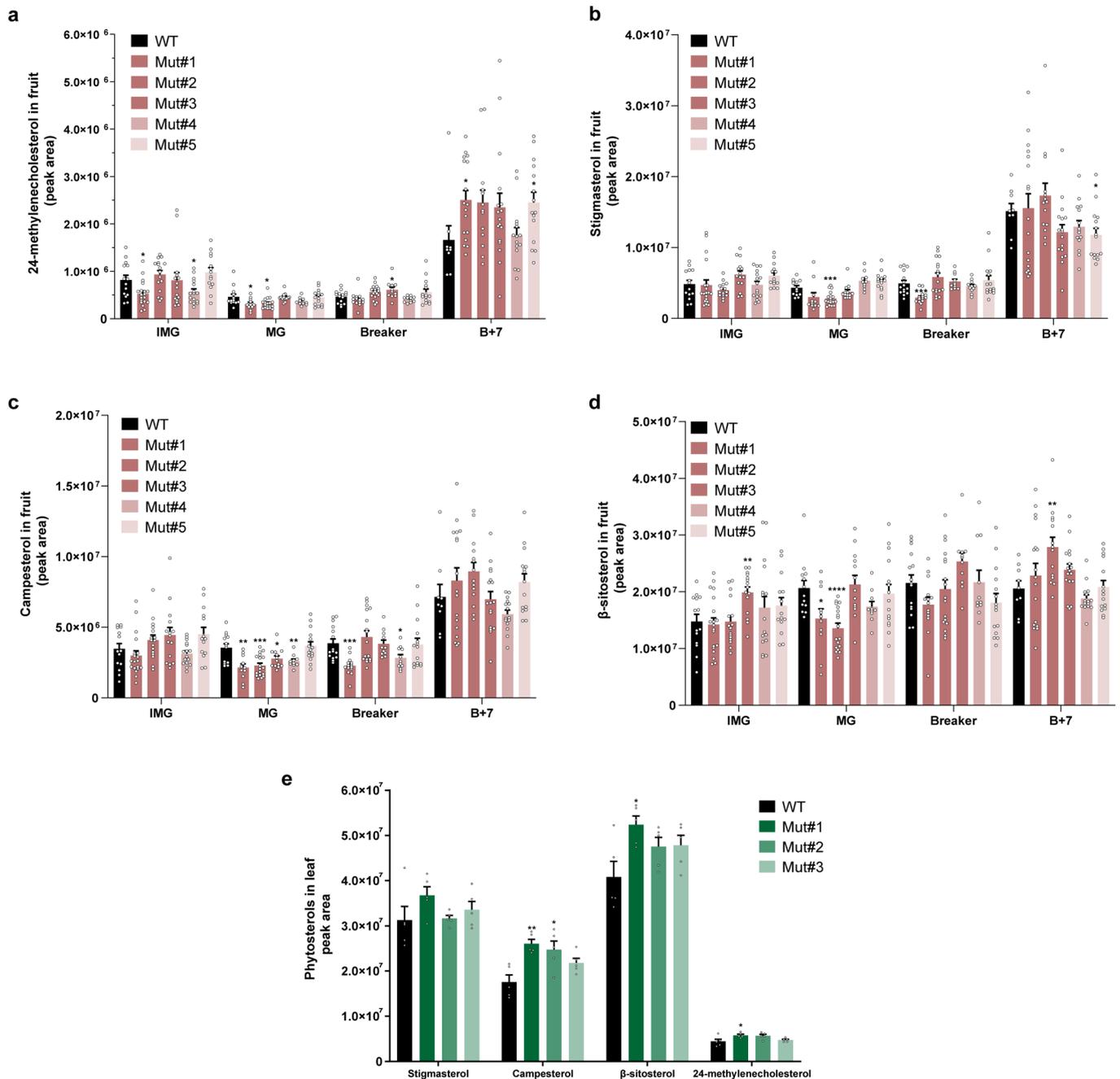
**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**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 license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022



**Extended Data Fig. 1 | Relative contents of four representative phytosterols (stigmasterol, campesterol,  $\beta$ -sitosterol and 24-methylenecholesterol) of wild-type (WT) and *S17-DR2* knock-out fruit during ripening (a-d) and leaves (e). (mean  $\pm$  s.e.m). a-d, from left to right, n = 14, 19, 16, 15, 16, 14, 13, 11, 18, 13, 10, 15, 15, 14, 17, 11, 11, 15, 9, 17, 14, 17, 15, and 15 biologically independent fruit samples; e, n = 5 biologically independent leaf samples for each line. Statistical significance between WT and mutants at each fruit ripening stage (a-d) in leaves (e) was assessed using two-tailed *t*-tests ( $^*P \leq 0.05$ ;  $^{**}P \leq 0.01$ ;  $^{***}P \leq 0.001$ ;  $^{****}P \leq 0.0001$ ). See source data for *P* values where relevant.**

## Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our [Editorial Policies](#) and the [Editorial Policy Checklist](#).

### Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a Confirmed

- The exact sample size ( $n$ ) for each experimental group/condition, given as a discrete number and unit of measurement
- A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- The statistical test(s) used AND whether they are one- or two-sided  
*Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
- A description of all covariates tested
- A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- For null hypothesis testing, the test statistic (e.g.  $F$ ,  $t$ ,  $r$ ) with confidence intervals, effect sizes, degrees of freedom and  $P$  value noted  
*Give  $P$  values as exact values whenever suitable.*
- For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- Estimates of effect sizes (e.g. Cohen's  $d$ , Pearson's  $r$ ), indicating how they were calculated

*Our web collection on [statistics for biologists](#) contains articles on many of the points above.*

### Software and code

Policy information about [availability of computer code](#)

#### Data collection

All presented data have been acquired using existing and routinely used software. Primers were designed with NCBI primer BLAST tool. Quantitative real-time PCR data were collected by X96 Touch™ Real-Time PCR Detection System (Biorad). Liquid chromatographic and mass spectrometry data were collected by the Dionex UltiMate (Thermo Fisher Scientific) and Q Exactive Orbitrap Mass Spectrometer (Thermo Scientific). Optical images were taken using a Canon 5D Mark IV camera with a Canon MP-E 65mm f/2.8 1-5x Macro Photo lens (Canon Inc, Ōta, Tokyo, Japan). MALDI imaging was performed with a Synapt G2-Si mass spectrometer with a MALDI source (Waters, Wilmslow, UK) equipped with a 2.5 kHz Nd:YAG laser operated at 355 nm.

#### Data analysis

Bio-Rad CFX Maestro Software. Capture One photo editing software (Capture One, Frederiksberg, Denmark). High Definition Imaging (HDI)1.4. Xcalibur software (version 4.3, Thermo Scientific). GraphPad (version 9.2.0). Geneious (version 10.2.6).

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio [guidelines for submitting code & software](#) for further information.

## Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our [policy](#)

Statistical Source Data with exact P values underlying Figures 1c-d, Figures 2b-e, Supplementary Figures 1d-e, 1g-h, Supplementary Figures 2a-b and Extended Data Figure 2 are provided as Source Data. Materials generated in this study are available from the corresponding author upon request.

## Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

- Life sciences       Behavioural & social sciences       Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see [nature.com/documents/nr-reporting-summary-flat.pdf](https://www.nature.com/documents/nr-reporting-summary-flat.pdf)

## Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	Sample sizes were chosen based on published studies, previous experience on the similar experiments in our lab and the availability of samples. At least three independent biological replicates were applied in each experiment. Sample sizes were described in figure legends where relevant. No statistical methods were used to predetermine sample size.
Data exclusions	No data were excluded from the analysis.
Replication	All the replication were successfully performed independently. Five independent transgenic lines were generated. At least three independent CRISPR-Cas9 knock-out lines were included in each analysis.
Randomization	Plants were grown in a randomized design. Sample analysis order has been further independently randomized for metabolic profiling.
Blinding	For leaf and fruit sample harvesting, corresponding metabolic profiling and MALDI imaging, samples were labeled with IDs denoting their randomised run order for both wildtype and mutant lines. Blinding was not used in the remaining experiments in this study because these kinds of experiments generally did not require blinding.

## Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

### Materials & experimental systems

n/a	Involved in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> Antibodies
<input checked="" type="checkbox"/>	<input type="checkbox"/> Eukaryotic cell lines
<input checked="" type="checkbox"/>	<input type="checkbox"/> Palaeontology and archaeology
<input checked="" type="checkbox"/>	<input type="checkbox"/> Animals and other organisms
<input checked="" type="checkbox"/>	<input type="checkbox"/> Human research participants
<input checked="" type="checkbox"/>	<input type="checkbox"/> Clinical data
<input checked="" type="checkbox"/>	<input type="checkbox"/> Dual use research of concern

### Methods

n/a	Involved in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> ChIP-seq
<input checked="" type="checkbox"/>	<input type="checkbox"/> Flow cytometry
<input checked="" type="checkbox"/>	<input type="checkbox"/> MRI-based neuroimaging