De novo domestication of wild tomato using genome editing

Agustin Zsögön1,7, Tomáš Čermák2,6,7, Emmanuel Rezende Naves1, Marcela Morato Notini3, Kai H Edel4, Stefan Weinl4, Luciano Freschi5, Daniel F Voytas2, Jörg Kudla48 & Lázaro Eustáquio Pereira Peres3

Breeding of crops over millennia for yield and productivity1 has led to reduced genetic diversity. As a result, beneficial traits of wild species, such as disease resistance and stress tolerance, have been lost2. We devised a CRISPR–Cas9 genome engineering strategy to combine agronomically desirable traits with useful traits present in wild lines. We report that editing of six loci that are important for yield and productivity in present-day tomato crop lines enabled de novo domestication of wild *Solanum pinnellifolium*. Engineered *S. pinnellifolium* morphology was altered, together with the size, number and nutritional value of the fruits. Compared with the wild parent, our engineered lines have a threefold increase in fruit size and a tenfold increase in fruit number. Notably, fruit lycopene accumulation is improved by 500% compared with the widely cultivated *S. lycopersicum*. Our results pave the way for molecular breeding programs to exploit the genetic diversity present in wild plants.

Tomato (*S. lycopersicum*) is the most important vegetable fruit worldwide, with annual production of 100 million tons3. The domestication process from the putative ancestral progenitor, *S. pinnellifolium*, which produces pea-sized fruits, to modern tomato varieties is well described4. However, despite the increases in yield conferred by domestication, the breeding focus on yield has been accompanied by a loss of genetic diversity and reduced nutritional value and taste5.

Many domestication traits have Mendelian inheritance patterns and involve loss-of-function or gain-of-function mutations6 (Table 1 and Supplementary Table 1). This means that it should be possible to recreate these traits in a suitable genetic background with CRISPR–Cas9 genome editing technology7. Although the first CRISPR–Cas9 applications created deletions, modern variants of CRISPR-based genome editing technologies can produce targeted insertions, exchange amino acids and modulate gene expression. Therefore, genome editing could be used to domesticate wild tomato plants and reinitiate lost but desirable traits, including nutritional features or stress tolerance, with yield potential and other agronomically valuable characteristics8.

We previously identified a suite of loci that have shaped the morphology and agronomic potential of current cultivars of tomato, maize, rice and other crops and proposed a reverse genetic approach for the de novo domestication of novel crops9 (Table 1 and Supplementary Table 1). In tomato, at least six loci important for key domestication traits have been identified: general plant growth habit (*SELF-PRUNING*)10, fruit shape (*OVATE*)11 and size (*FASCIAED* and *FRUIT WEIGHT 2.2*)12,13, fruit number (*MULTIFLORA*)14, and nutritional quality (*LYCOPENE BETA CYCLASE*)15. We set out to create a novel crop derived from *S. pinnellifolium* by targeting this set of genes using a multiplex CRISPR–Cas9 approach to generate loss-of-function alleles. We constructed a single CRISPR–Cas9 plant transformation vector, pTc321 (Supplementary Note 1), which harbored six single guide RNAs (gRNAs) targeting specific sequences in the coding regions of all six genes (Supplementary Fig. 1). Using this vector, we generated ten primary T0 lines, of which three were grown to maturity. T1 seeds were harvested from plant 3, which showed an oval fruit phenotype, indicative of successful editing of the *ovate* locus, and determinate growth habit, indicative of loss of function of the *self-pruning* gene. Sequencing of all six targeted loci in 50 T1 lines revealed that four of the six targeted loci were successfully edited in all 50 lines and harbored indel mutations (Supplementary Tables 2 and 3). The four edited genes were *SELF-PRUNING* (*SP*), *OVATE* (*O*), *FRUIT WEIGHT 2.2* (*FW2.2*) and *LYCOPENE BETA CYCLASE* (*CycB*). For all four edited genes, we recovered only edited alleles and did not detect any wild-type (*WT*) alleles in the T1 generation (SupplementaryDatasets 1 and 2). However, we did not recover any mutations in either *FASCIAED* (*FAS*) or *MULTIFLORA* (*MULT*). In the case of *FAS* we identified a G-to-A substitution in the *S. pinnellifolium* genome at the gRNA target site (designed based on the *S. lycopersicum* genome) and the targeted *S. pinnellifolium* sequence, which may have prevented editing (Supplementary Tables 2 and 3).

To address the specificity of our multiplex editing approach, we sequenced the two most closely related off-target loci (as determined by in silico analysis using Geneious R11 program) for each gRNA in two pTc321 plant 3 lines (Supplementary Fig. 2). We did not...

1 Departamento de Biologia Vegetal, Universidade Federal de Viçosa, Viçosa, Brazil. 2 Department of Genetics, Cell Biology and Development, Center for Genome Engineering, University of Minnesota, Minneapolis, Minnesota, USA. 3 Departamento de Ciências Biológicas, Escola Superior de Agricultura “Luiz de Queiroz,” Universidade de São Paulo, Piracicaba, Brazil. 4 Institut für Biologie und Biotechnologie der Pflanzen, Universität Münster, Münster, Germany. 5 Instituto de Biociências, Universidade de São Paulo, São Paulo, Brazil. 6 Present address: Inari Agriculture, Cambridge, Massachusetts, USA. 7 These authors contributed equally to this work. Correspondence should be addressed to J.K. (j.kudla@uni-muenster.de) or L.E.P.P. (lazaro.peres@usp.br).

Received 19 April; accepted 12 September; published online 1 October 2018; doi:10.1038/nbt.4272
observe editing of either off-target locus in any of the sequenced lines (Supplementary Dataset 1).

Next we examined the phenotypes of two T₂ plant lines (designated as line 3-5 and line 3-11). The parental T₁ plants of both 3-5 and 3-11 contained homozygous deletions in the second exon of the gene encoding SP (Fig. 1a). Breeding an sp allele with a single nucleotide polymorphism (SNP) in the coding region into tomato cultivars was instrumental in enabling the mechanical harvest of fruits16. Similarly, loss of SP function in S. pimpinellifolium resulted in compact plants with reduced height, reduced number of sympodial units and determinate growth when compared with WT S. pimpinellifolium (Fig. 1b–g). In the fruit-shape-determining OVATE (O) locus, genome editing had induced distinct homozygous deletions in the first exon of the gene in both T₁ parents (Fig. 1h). As predicted, loss of O gene function in both alleles caused an oval fruit shape (Fig. 1i–m). The elongated form of fruits occurs through reduced internal fruit pressure and is associated with less rain-induced fruit cracking, a detrimental trait in tomato production17. In the FRUIT WEIGHT 2.2 (FW2.2) locus, we identified two independent homozygous deletions that disrupted the coding frame of exon 2 (Supplementary Fig. 3). Surprisingly, despite these mutations, we did not detect any discernible change in fruit sizes in these T₂ mutant lines compared with WT S. pimpinellifolium (Supplementary Table 4).

During the course of our experimental program, we learned that the gene locus responsible for the multinocular fruit phenotype of FAS is not encoded by a YABBY transcription factor, as was previously proposed18, but instead results from a mutation in CLAVATA3 (CLV3)12. Therefore, we carried out a second round of genome engineering to generate multiplex-edited plants in which the selection of target genes focused on modulating fruit size (FW2.2, FAS) and number (MULT), as well as nutritional value (LYCOPENE BETACE CYCLASE, CycB). We designed a CRISPR–Cas9 plant transformation vector named pTC603 (Supplementary Note 2), which harbored eight gRNAs targeting two sites in the coding region of each of these four genes (Supplementary Fig. 1). Three successfully transformed T₀ lines were identified (Supplementary Fig. 4). Sequencing of T₀ plants 5 and 8 identified loss-of-function mutations in all four targeted loci in both lines (Supplementary Fig. 5 and Supplementary Dataset 2). Target locus analysis in 28 individual T₁ representatives derived from both T₀ plants 5 and 8 revealed that in every T₁ plant all four target loci were successfully edited (Supplementary Dataset 1). In CycB and FW2.2 we identified two different mutation events and in CLV3 three different mutation events homozygous mutations or heterozygous (biallelic) combinations of mutant alleles. Either type of mutation resulted in a loss of function of the edited gene. For MULT, we recovered 5 T₁ plants possessing two WT alleles and 23 heterozygous T₁ plants (Supplementary Table 5). We examined the two most closely related genomic sequences for each gRNA (as determined by in silico analysis using the Geneious R11 program) and did not observe off-target editing in ten analyzed T₁ plants (Supplementary Fig. 2). To summarize, in two separate experiments we engineered two sets of four target domestication genes simultaneously and generated a diverse range of combinations of 15 independent loss-of-function alleles (Supplementary Tables 3 and 4).

We phenotyped the second set of edited plant lines. Introduction of a loss-of-function MULT allele into cultivated tomato resulted in a higher number of fruits per truss, conferring enhanced yield19. In our edited plant lines, we identified a single heterozygous mutation (2-bp deletion) in exon 2 (Fig. 2a and Supplementary Fig. 5) of MULT. Edited plants had the predicted branched inflorescence phenotype (T₁, Fig. 2b–d; T₀, Supplementary Fig. 6). Both T₁ lines also harbored homozygous indel mutations in both the first and the second exon of the FW2.2 gene (Supplementary Fig. 5). Moreover, T₁ lines 5 and 8 also carried biallelic deletions in exon 1 and a heterozygous deletion in exon 2 of the CLV3 gene (Fig. 2e), which corresponded with phenotypes that have been attributed to the FAS locus (Fig. 2f,g). Characterization of fruit morphology revealed a fourfold increase in fruit locule number and a fruit weight increased up to 200% compared with S. pimpinellifolium (Supplementary Fig. 2i,j). Considering the absence of a discernible fruit phenotype in the lines in which FW2.2 but not FAS was mutated (Supplementary Fig. 3 and Supplementary Table 4), we conclude that increased fruit size (Fig. 2j,k) is due to the loss of function of CLV3. Since the classical fw2.2 mutation, which affects fruit size, causes a heterochronic alteration in the FW2.2 expression pattern, we suggest that changes in FW2.2 expression, but not loss of function, affect tomato fruit size13.

Nutritional value and taste have largely been neglected in conventional breeding14. The content of lycopene and β-carotene largely determines the nutritional value of tomato20. Anti-inflammatory properties and the reduction of cardiovascular and cancer risks have also been correlated with dietary intake of lycopene21. In addition, while β-carotene occurs in sufficient amounts in many vegetables, lycopene is generally not present in sufficient concentration22. In cultivated tomato, the activity of lycopene β-cyclase converts a substantial fraction of lycopene into β-carotene, thereby reducing the content of beneficial lycopene15. While cultivated cherry tomato fruits accumulate around 60–120 mg kg⁻¹ lycopene23,24, this antioxidant accumulates to levels of up to 270 mg kg⁻¹ in S. pimpinellifolium25,26. The genetic bases determining lycopene accumulation in tomato are only poorly understood and most likely polygenic27. Accordingly, previous attempts to enhance lycopene

<table>
<thead>
<tr>
<th>Crop species</th>
<th>Gene target</th>
<th>Function</th>
<th>Mutation type</th>
<th>Genetic effect</th>
<th>Phenotypic outcome</th>
<th>Refs.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Maize</td>
<td>Tb1</td>
<td>TCP-family transcription factor</td>
<td>Retrotransposon insertion in regulatory region</td>
<td>Gain of function</td>
<td>Inhibition of side branching, altering source–sink relations and increasing yield</td>
<td>31,32</td>
</tr>
<tr>
<td></td>
<td>lg1</td>
<td>Squamosa-promoter binding protein</td>
<td>Retrotransposon insertion</td>
<td>Loss of function</td>
<td>Leaf is upright due to absent ligules and auricles</td>
<td>28,33</td>
</tr>
<tr>
<td></td>
<td>tga1</td>
<td>SBP-box transcription factor</td>
<td>SNP altering single amino acid</td>
<td>Gain of function</td>
<td>Changes encased to naked kernels</td>
<td>34</td>
</tr>
<tr>
<td></td>
<td>ZmCCT</td>
<td>CCT domain-containing protein genes</td>
<td>Retrotransposon insertion in regulatory region</td>
<td>Loss of function</td>
<td>Reduction of photoperiod sensitivity</td>
<td>35,36</td>
</tr>
<tr>
<td>Soybean</td>
<td>DT1</td>
<td>CETS family of regulatory genes</td>
<td>SNPs altering amino acids</td>
<td>Loss of function</td>
<td>Changes growth from indeterminate to determinate, producing a shorter, more compact plant</td>
<td>37,38</td>
</tr>
<tr>
<td></td>
<td>GA20ox</td>
<td>Gibberellin biosynthesis enzyme</td>
<td>Variation in promoter region</td>
<td>Loss of function</td>
<td>Seed weight</td>
<td>39</td>
</tr>
<tr>
<td></td>
<td>SHAT1-5</td>
<td>NAC-family transcription factor</td>
<td>20-bp deletion disrupting a repressive element</td>
<td>Gain of function</td>
<td>Increased secondary wall biosynthesis promoting thickening of fiber cap cells, leading to reduced shattering</td>
<td>40</td>
</tr>
</tbody>
</table>

*Actually targets the close ortholog GmFT2.
content of commercial tomato yielded only moderate success or resulted in enhanced lycopene accumulation at the cost of reduced β-carotene abundance\textsuperscript{15,28}. We targeted the CycB gene in both experiments to evaluate whether we could exploit the lycopene accumulation phenotype of \textit{S. pimpinellifolium}. In all four edited \textit{S. pimpinellifolium} lines, we detected biallelic

**Figure 1** Plant morphology and fruit shape in \textit{de novo} domesticated \textit{S. pimpinellifolium} plants. (a) Genomic sequence showing the gRNA target site (underlined) and the resulting missense mutation in the SP gene. The PAM sequence is depicted in bold. (b,c) Representative WT (b) and \textit{T}_2 sp mutant (c) plants with determinate growth habit; bars, 30 cm. (d,e) Detailed view of the vegetative branch with successive sympodial units (inflorescence plus three leaves; white arrows) in WT (d) and \textit{T}_2 sp mutant (e); bars, 5 cm. (f,g) Height (f) and sympodial index (g) alterations in WT and sp mutant plants. (h) Genomic sequence showing the site of gRNA targeting and the resulting missense mutation in the \textit{O} gene. (i–l) Developmental series of early fruits of WT (i) and \textit{T}_2 ovate (o) mutant (j) plants (bars, 1 cm) and their respective unripe fruits in the plant (k,l); bars, 5 cm. (m) Fruit length/width ratio in WT and line \textit{3-5} and \textit{3-11} mutant plants. Two-tailed t-test (WT vs. lines): **\(P < 0.01\) and ***\(P < 0.001\); \(n = 6\) plants for height and sympodial index and \(n = 90\) fruits for fruit length/width ratio. Data are depicted in box plots: box, interquartile range (IQR); whiskers, 1.5 × IQR; center line, mean.
or homozygous loss-of-function mutations in the first exon of CycB (Fig. 3a and Supplementary Figs. 5 and 7). These genetic alterations resulted in flowers with orange antheridial cones (due to lycopene accumulation) rather than the yellow anthers of S. pimpinellifolium lines 5 and 8 (Fig. 3b–d) and resulted in deep red fruits (Fig. 3e,f).

Quantitative determination of the carotenoid content in fruits of cherry

Figure 2 Flower number and fruit size in de novo domesticated S. pimpinellifolium plants. (a) MULT genomic sequence showing the site of gRNA targeting and the resulting missense mutations. (b,c) Representative WT (b) and T1 multiflora (mult) (c) mutant inflorescences. Note the increased number of petals and their intense yellow color due to the mutated fas and cycB alleles targeted in the same vector. Bars, 2 cm. (d) Number of flowers per inflorescence in WT and two different T1 mutants. (e) clv3 genomic sequence (of the fas allele) showing the site of gRNA targeting and the resulting missense mutations. (f,g) Macroscopic flower morphology in WT (f) and T1 multi-petal fas (g) mutant plants; bars, 2 cm. (h,i) Fruit locules (h) and weight (i) in WT and two different mutant plant lines (derived from events 5 and 8). Increased locule number led to increased fruit size in both fas line 5 T1 (j) and fas line 8 T1 (k) mutant plant lines. Scale bars, 1 cm. Two-tailed t-test (WT vs. lines): ***P < 0.001; n = 60 inflorescences for flower number per inflorescence and n = 90 fruits for fruit locule number and weight. Data are depicted in box plots: box, IQR; whiskers, 1.5 × IQR; center line, mean.
SUPPLEMENTARY FIG. 9. Important traits such as Brix (Cambridge Univ. Press, 1996).

**Methods**, including statements of data availability and any associated accession codes and references, are available in the online version of the paper.

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

**Acknowledgments**
We are grateful to S. Schülke for technical assistance. This work was supported by funding from the Agency for the Support and Evaluation of Graduate Education (CAPES, Brazil), the National Council for Scientific and Technological Development (CNPq, Brazil) and Foundation for Research Assistance of the São Paulo State (FAPESP, Brazil), and the German Federal Ministry of Education and Research (BMBF, Germany). We thank CAPES for studentships granted to E.R.N. and FAPESP for the studentship granted to M.M.N. (2013/12209-1). L.F. was supported by FAPESP grant 2013/18056-2. FAPESP and BMBF provided a grant for L.E.P. (2015/50220-2) and J.K. (031B0334). L.E.P. acknowledges a grant from CNPq (grant 307040/2014-3).

**Author Contributions**
A.Z., T.C., D.F.V., J.K. and L.E.P. designed the study. A.Z., T.C., M.M.N., E.R.N., K.H.E., S.W. and L.F. performed experiments. A.Z., T.C. and K.H.E. analyzed data. A.Z., K.H.E., J.K. and L.E.P. prepared the manuscript. All authors have revised and approved the final version of the manuscript.

**Competing Interests**
After completion of this work in the laboratory of D.F.V., T.C. became an employee of Inari Agriculture, a company that uses novel technologies for crop breeding. D.F.V. is a founder and Chief Science Officer of Calyxt, a company applying genome editing to plants.

Reprints and permissions information is available online at http://www.nature.com/reprints/index.html. Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.


12. Xu, C.


17. Römer, S.

18. Li, C.


23. Xu, C.

24. Xu, C.


28. Li, C.

29. Lu, X.

30. Li, C.

31. Lu, X.

32. Xu, C.

33. Xu, C.

34. Xu, C.

35. Lu, X.

36. Xu, C.

37. Tian, Z.

38. Tian, Z.

39. Lu, X.

40. Dong, Y.
ONLINE METHODS

Molecular cloning and plant transformation. CRISPR–Cas9 plant transformation vectors were constructed using protocols 3A and 5 of Cermák et al.7 The vector pTC321 (35S::Cas9-P2A-AtCas9, 35S::gRNA-array) expressing a Cas9 array of six gRNAs targeting SELF PRUNING (Soly06g074350), OVATE (Soly02g085500), FRUIT WEIGHT 2.2 (Soly02g0997030), Lycopene Beta-CYCLASE (Soly04g04190), MULTIFLORA (Soly02g077390) and FASCILATED/YABB (Soly11g071810) was constructed by direct assembly of gRNAs into the T-DNA vector pDIRECT_22C (Addgene plasmid 93105) (see Supplementary Table 10 for gRNA sequences). Vector pTC603 (35S::Cas9-P2A-AtCas9, CmYLCV-gRNA-array) expressing a Cas9 array of eight gRNAs targeting FASCIATA/CLAVATA3 (Soly11g071810), FRUIT WEIGHT 2.2 (Soly02g0997030), MULTIFLORA (Soly02g077390) and Lycopene Beta-CYCLASE (Soly04g04190) was constructed in two steps. The first two gRNAs were cloned into pMOD_B2103 (Addgene plasmid 91061) and the last two gRNAs into pMOD_C2200 (Addgene plasmid 91082). The gRNA arrays in the resulting plasmids were assembled along with pMOD_A0501 (35S::Cas9-P2A-AtCas9, Addgene plasmid 91011) into the T-DNA vector pTRANS_220 (Addgene plasmid 91115). The final array contains the gRNAs in the following order: gRNAs 41 and 42 targeting FASCIATA/CLAV, gRNAs 15 and 16 targeting FW2.2, gRNAs 19 and 20 targeting MULTIFLORA and gRNAs 17 and 18 targeting CycB (Supplementary Fig. 1 and Supplementary Table 10).

The pTC321 (Supplementary Note 1) and pTC603 (Supplementary Note 2) vectors were introduced into the Agrobacterium tumefaciens strain LB4404 by electroporation. Agrobacteria were cultured in 3 mL of liquid LB medium supplemented with 50 mg L⁻¹ kanamycin and 50 mg L⁻¹ rifampicin and incubated at 28 °C for 24 h at 120 r.p.m. From this culture, 500 µL were added to 50 mL of fresh LB medium with the same antibiotics as above. The culture was incubated overnight under the same conditions and then centrifuged at 2000g for 15 min at room temperature. The pellet was resuspended in liquid Murashige and Skoog (MS) medium41 to an OD₅₀₀ of 0.25–0.3; acetylseringone was added to a final concentration of 100 µM before plant inoculation.

S. pimpinellifolium (LA1589) was transformed as previously published42 with the following modifications: seeds were surface-sterilized by shaking in 20 mL 30% (v/v) commercial bleach (2.7% sodium hypochlorite) supplemented with two drops of commercial detergent for 15 min, followed by three rinses with sterile water. Seeds were germinated on semi-solid MS medium (supplemented with 0.6 g/L agar) and incubated at 25 ± 1 °C in the dark for 4 d. After this period, seeds were transferred and maintained at 25 ± 1 °C under long-day conditions (16 h light/8 h dark) with 45 µmol photons m⁻² s⁻¹ PAR irradiance.

Leaf explants (1–2 cm²) were isolated from 2-week-old in vitro plants and placed with the abaxial side down onto MS medium containing 1 mg L⁻¹ trans-zeatin, 0.1 mg L⁻¹ indole-3-acetic acid (IAA) and 100 µM acetylseringone. Two drops of Agrobacterium suspension were applied per explant and plates were then incubated at room temperature for 10 min. Thereafter, excess bacterial suspension was removed, and explants were blotted dry on sterile filter paper. Plates were maintained in the dark at 25 °C for 2 d for co-cultivation. Explants were transferred to plates with MS medium containing 1 mg L⁻¹ trans-zeatin, 0.1 mg L⁻¹ IAA, 300 mg L⁻¹ timentin, 100 mg L⁻¹ kanamycin. After 5 weeks, shoot primordia emerging from callus tissue were isolated and transferred to shoot elongation medium (MS medium supplemented with 0.5 mg L⁻¹ trans-zeatin, 300 mg L⁻¹ timentin, 100 mg L⁻¹ kanamycin) to recover full plants. After acclimation, these plants were grown in a greenhouse at 30 °C/26 °C day/night temperature and 60–75% ambient relative humidity, 11.5–13 h (winter/summer) photoperiod, sunlight 250–350 µmol photons m⁻² s⁻¹ PAR irradiance, attained by a reflecting mesh (Aluminet, Polysack Indústrias Ltda, Leme, Piracicaba, SP, Brazil), and automatic irrigation four times a day for fruit set.

Seeds from transformed plants were germinated in 350-mL pots with a 1:1 mixture of commercial potting mix Basaplant (Base Agro, Artur Nogueira, SP, Brazil) and expanded vermiculite supplemented with 1 g L⁻¹ 10:10:10 NPK and 4 g L⁻¹ dolomite limestone (MgCO₃ plus CaCO₃). Upon appearance of the first true leaf, seedlings of each genotype were transplanted to 10-L pots containing the soil mix described above, except for NPK supplementation, which was increased to 8 g L⁻¹. After transplanting, plants were sprayed twice at 14-d intervals with 1 g L⁻¹ Peters 20–20–20 leaf fertilizer. Acclimatized plants were pollinated by hand, producing T₁ seeds and the following generations.

Seeds and vectors are available upon request from the corresponding authors (see https://www.uni-muenster.de/Biologie.IBBP/aksgudla/Plasmids.html).

Genotyping of T₀ and T₁ plants. The gRNA target sites were amplified by PCR directly from leaf tissue with the Phire Plant Direct PCR Master Mix following the Dilution & Storage protocol (Thermo Fisher Scientific) or from extracted genomic DNA with Phusion polymerase (New England BioLabs), using primers listed in Supplementary Table 11. Primers and dNTPs were removed using ExoSAP (0.5 U exonuclease I + 0.25 U FastAP thermosensitive alkaline phosphatase in 35 µL of PCR, incubated 30 min at 37 °C and heat-inactivated for 5 min at 95 °C) and the purified PCR products were directly sequenced. Heterozygous and biallelic mutations were identified as overlapping sequence traces. These samples were cloned and several clones sequenced to determine the sequences of individual alleles.

Off-target identification and analysis. The “Find CRISPR sites” function in Genious R11 was used to search for off-target binding sites in S. pimpinellifolium genome43 (https://solgenomics.net/organism/Solanum_pimpinellifo-lium/genome). A maximum of eight mismatches and zero indels were allowed between the on- and off-target sequences. The two top scoring off-target sites were PCR amplified and sequenced. Primers can be found in Supplementary Table 12.

Plant material and growth conditions. Plants were cultivated in semicon-rolled conditions in a greenhouse in Viçosa (642 m asl, 20° 45’ S, 43° 1’ W), Minas Gerais, Brazil. Wild-type S. pimpinellifolium plants were grown alongside T₀ of pTC321-1, T₀ of pTC321-1 and T₁ of pTC603-transformed plants during the months of October 2016 to January 2017, August 2017 to November 2017, August 2017 to November 2017 and August 2017 to November 2017, respectively. For plants of genera T₁ of pTC321 and T₁ of pTC603, the main stem was trained on bamboo sticks and side branches were pruned, leaving a total of three inflorescences per plant. The greenhouse temperature ranged between 24 and 26 °C, with a 13-h/11-h (day/night) photoperiod and an average irradiance of 800 µmol photons m⁻² s⁻¹. Seeds were germinated in plastic trays with commercial substrate Trostrato and supplemented with 1 g L⁻¹ 10:10:10 NPK and 4 g L⁻¹ dolomite limestone (MgCO₃ plus CaCO₃). Weekly foliar fertilization was carried out using 2 g L⁻¹ Biofert leaf fertilizer. Upon appearance of the first true leaves, seedlings of each genotype were transplanted to 5-L pots. The new pots were filled with substrate as described above, except for the NPK supplementation, which was increased to 8 g L⁻¹. Irrigation was performed twice a day in a controlled manner, so that each vessel received the same volume of water.

Carotenoid quantification and sampling. Fruits were harvested at the ‘red ripe’ stage sequentially over a period of 2 weeks. They were cut in half and the seeds were removed, after which the fruits were immediately frozen in liquid nitrogen and stored in a freezer at −80 °C until analysis. Fruit carotenoid extraction was carried out following an adaptation of a previously published protocol44. Frozen fruits were ground in a ball mill (Retsch, model MM400) and 200-mg aliquots subsequently collected in 2-mL microtubes. During the extraction, samples were kept cooled on ice in an environment with reduced luminosity (30–50 µmol photons m⁻² s⁻¹ irradiance). 100 µL of saturated NaCl solution (370 g/L) were added to each sample and mixed by vortexing for 30 s. 200 µL of dichloromethane were added next and mixed by vortexing. Finally, 500 µL of hexane/ethyl ether (1:1) were added and the sample homogenized by vortexing. Samples were then centrifuged (Eppendorf 5415R micro-centrifuge) (4 °C, 13,000g for 5 min). The supernatant was then collected and transferred to a 2-mL amber microtube. Addition of hexane/ethyl ether (1:1), vortex agitation, centrifugation and collection of the upper organic phase (supernatant) were repeated three times, until whitening of the samples was observed in the microtube, showing the appropriate extraction of most carotenoids. Samples were completely dried in a vacuum concentrator (CentriVap, Labconco), re-suspended in 300 µL of ethyl acetate and mixed for 30 s. All supernatant fractions were combined, completely dried by vacuum, filtered through a 0.45 µm membrane filter and suspended with 100 µL of ethyl acetate. Chromatography was
carried out on Agilent Technologies series 1100 HPLC system on a normal-phase Phenomenex column (Luna C18; 250 × 4.6 mm; 5 µm particle diameter) with a flow rate of 1 mL min⁻¹ and temperature 25 °C. The mobile phase was a gradient of ethyl acetate (A) and acetonitrile:water 9:1 (v/v) (B): 0–4 min: 20% A; 4–30 min: 20–65% A; 30–35 min: 65% A; 35–40 min: 65–20% A. Eluted compounds were detected between 340 and 700 nm and quantified at 450 nm. The endogenous metabolite concentration was obtained by comparing the peak areas of the chromatograms with commercial standards (Sigma-Aldrich).

The extraction of corolla and anther samples were performed in a similar way to that of fruits, but with the following modifications: after harvesting, the samples were lyophilized and stored in a refrigerator until the extraction process. The period between freeze-drying and carotenoid extraction was less than 10 d. Lyophilized samples were ground in reaction tubes with the help of a plastic pestle. Anther samples contained on average 14 mg of dry mass, corolla samples 10 mg. Samples were rehydrated in 300 µL ultrapure water and subjected to the extraction process and HPLC quantification.

Brix determination. The total content of soluble solids in fruits (measured as Brix) was determined for the ‘red ripe’ fruits stage by means of a digital bench refractometer (Instrutherm Model RTD-45). First the apparatus was calibrated with distilled water; after drying the prism, the liquid of each fruit was placed in the prism of the refractometer and thus quantified. The Brix of each replicate (each plant) was composed of analyzes of many fruits.

Phenotypic parameters of the fruits. The phenotypic parameters of the fruits (mass, length and equatorial width) were determined using precision balances (Micronal B160) and a digital caliper (MTX) using at least 10 fruits per plant.

Statistics. The quantitative parameters studied here show continuous variation and were therefore analyzed using parametric tests: Student’s t-test was done on GraphPad; ANOVA and Tukey HSD tests were performed using VassarStats (http://vassarstats.net). Percentage data were converted to the inverse function (1/X) before analysis. All information on replication, statistical test and presentation are given in the respective figure legends.

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

Data availability. The data supporting the findings of this study are available in the paper and the accompanying Supplementary Information files. Specifically, sequences supporting on- and off-target analyses are available in Figures 1–3, Supplementary Figures 2, 3, 5 and 7, and Supplementary Tables 3 and 4. Raw sequences and alignments for the above mentioned figures and tables are available in Supplementary Datasets 1 and 2. Complete and annotated sequences of pTC321.gb and pTC603.gb are in Supplementary Notes 1 and 2, respectively.

41. Pino, L.E. et al. The Rg1 allele as a valuable tool for genetic transformation of the tomato ‘Micro-Tom’ model system. Plant Methods 6, 23 (2010).
Reporting Summary

Nature Research 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 Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistical parameters

When statistical analyses are reported, confirm that the following items are present in the relevant location (e.g. 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
☐ An indication of 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 statistics 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
☐ Clearly defined error bars
☐ State explicitly what error bars represent (e.g. SD, SE, CI)

Our web collection on statistics for biologists may be useful.

Software and code

Policy information about availability of computer code

Data collection
The data collected were basically growth measurements and carotenoid determination, which did need software to be collected.

Data analysis
Statistical analyses were performed using Excel from Windows Office v. 2010 Package. DNA analyses were performed using Geneious v. 9 or DNAStar Lasergene Version 15.

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/reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research 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 list of figures that have associated raw data
- A description of any restrictions on data availability

Supplementary Information files. Specifically, sequences supporting on- and off-target analyses are available in Figure 2–4, Supplementary Figure 2, 3, 5 and 7 as Nature Biotechnology: doi:10.1038/nbt.4272
Field-specific reporting

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

☒ Life sciences    ☐ Behavioural & social sciences

For a reference copy of the document with all sections, see nature.com/authors/policies/ReportingSummary-flat.pdf

Life sciences

Study design

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

| Sample size | Sample size was initially determined based on the number of successful plants harbouring the target mutation. From this point, growth measurements and molecules determinations were based on statistically relevant number of repetitions for ANOVA (please see materials and methods and figure legends). |
| Data exclusions | No data were excluded during growth analysis and analysis of carotenoid content. |
| Replication | As usual in growth measurements of plants, they were grown in controlled environmental conditions and experiments were performed in different and independent group of plants. |
| Randomization | Two sets of plants were used depending on the genes targeted in each vector. The growth parameters and carotenoid determination were performed independently in these two set of plants. |
| Blinding | Data were collected in randomized samples of plants size, organ length and height and tissues for carotenoid quantification |

Materials & experimental systems

Policy information about availability of materials

n/a Involved in the study
☒ ☐ Unique materials
☒ ☐ Antibodies
☒ ☐ Eukaryotic cell lines
☒ ☐ Research animals
☒ ☐ Human research participants

Unique materials

Obtaining unique materials | There are no restrictions on availability of unique materials. Both transformation vectors used and CRISPR/Cas9 plants obtained are available for research use, upon request. |

Method-specific reporting

n/a Involved in the study
☒ ☐ ChIP-seq
☒ ☐ Flow cytometry
☒ ☐ Magnetic resonance imaging

Nature Biotechnology: doi:10.1038/nbt.4272