BRIEF COMMUNICATION
https://doi.org/10.1038/s41477-018-0259-x

nature plants

Rapid improvement of domestication traits in an orphan crop by genome editing

Zachary H. Lemmon[1], Nathan T. Reem[2,5], Justin Dalrymple[1,5], Sebastian Soyk[1,5], Kerry E. Swartwood[2], Daniel Rodriguez-Leal[1], Joyce Van Eck[2,3*] and Zachary B. Lippman[1,4*]

Genome editing holds great promise for increasing crop productivity, and there is particular interest in advancing breeding in orphan crops, which are often burdened by undesirable characteristics resembling wild relatives. We developed genomic resources and efficient transformation in the orphan Solanaceae crop 'groundcherry' (Physalis pruinosa) and used clustered regularly interspaced short palindromic repeats (CRISPR)–CRISPR-associated protein-9 nuclease (Cas9) (CRISPR–Cas9) to mutate orthologues of tomato domestication and improvement genes that control plant architecture, flower production and fruit size, thereby improving these major productivity traits. Thus, translating knowledge from model crops enables rapid creation of targeted allelic diversity and novel breeding germplasm in distantly related orphan crops.

There has been extensive discussion on leveraging genome-editing technologies to improve staple crops[1], yet their application to regionally important plants grown for subsistence purposes is equally exciting, especially in developing countries. Such 'orphan crops' are relatively unknown and typically have not experienced intensive selection for domestication and improvement. Thus, orphan crops are less productive, untenable at larger agricultural scales, and benefit less from basic research[2]. Genome-editing technologies, such as the broadly successful clustered regularly interspaced short palindromic repeats (CRISPR)–CRISPR-associated protein-9 nuclease (Cas9) (CRISPR–Cas9) provide opportunities to address these deficiencies, with primary goals to increase quality and yield, improve adaptation and expand geographical ranges of cultivation. The *Solanaceae* family contains many orphan crops alongside several well-characterized model crops, such as the tomato (*Solanum lycopersicum*), potato (*Solanum tuberosum*) and pepper (*Capsicum annuum*). This strong foundation of genetic, developmental and genomic knowledge makes the *Solanaceae* an excellent platform for translating genome editing to orphan crops.

We focused on the orphan crop *Physalis pruinosa* (groundcherry), a wild *Solanaceae* that is more distantly related to the tomato than the pepper, and which is grown in Central and South America for its subtly sweet berries[3,4]. Barriers to higher productivity and wider cultivation include a wild sprawling growth habit and small ~1 g fruits that drop to the ground due to strong stem abscission (Fig. 1a–g). These undesirable characteristics parallel the wild ancestor of the tomato, *Solanum pimpinellifolium*, for which selection allowed major improvements in shoot architecture, flower production and fruit size[5,6] (Fig. 1h–m). Although groundcherry and related *Physalis* species have the same chromosome number as most *Solanaceae* ($n = 12$)[7], several challenges remain for gene editing, including the absence of reference genomes, limited information on gene content and function, and several architectural and fruit development traits that differ from the tomato (Fig. 1). However, considering its phylogenetic relationship with the tomato, its diploid genome and the fact that key developmental and productivity genes have similar functions across model *Solanaceae*[8], we hypothesized that *Physalis* orthologues of select tomato domestication and improvement genes could be modified through editing for immediate improvements.

A major obstacle for CRISPR–Cas9 plant genome editing is lack of efficient tissue culture and transformation methodologies[9]. For editing of *P. pruinosa* to be realized, we developed *Agrobacterium tumefaciens*-mediated transformation modelled after our tomato methodology (Supplementary Methods), and evaluated editing by targeting the orthologue of the tomato leaf development gene *Sl-AGO7* (where '*Sl-*' relates to *S. lycopersicum*). This was previously used to test CRISPR–Cas9 in the tomato, because mutations result in conspicuous narrowing of the leaves and floral organs[10]. First-generation (T_0) plants were chimeric for *Ppr-AGO7* mutations (where '*Ppr-*' relates to *P. pruinosa*) and, like tomato *Sl-ago7*^{CR} chimeric T_0 plants (where CR indicates CRISPR–Cas9-induced), the leaves and petals were narrower than in the wild type (Supplementary Fig. 1a,b), indicating efficient editing in the groundcherry.

To expand *Physalis* genomic resources beyond a leaf transcriptome from the related *Physalis peruviana*, which lacked orthologues of several tomato domestication and improvement genes[6,11–13], we generated Illumina whole-genome and RNA sequencing de novo assemblies from vegetative and reproductive tissues (Supplementary Methods). Analysis[14] revealed 93 and 82% complete benchmarking universal single-copy orthologues for the transcriptome and genome assembly, respectively, and 12,993 orthologues of tomato genes were reconstructed with at least 90% coding sequence coverage.

With these tools, our first efforts focused on modifying groundcherry shoot architecture to contain its weedy growth habit. In the tomato, selection for mutations in florigen flowering pathway genes allowed major improvements in plant architecture and yield[15]. In particular, a classical missense mutation in the antflorigen *SELF-PRUNING* (*SP*) gene (*sp*^{classic}) provided compact 'determinate' growth that translated to a burst of flowers and fruits, thereby enabling large-scale field production[16]. *SP* encodes a flowering repressor that modulates sympodial growth—a hallmark *Solanaceae* and many other plant species—and mutations in the pepper *SP* orthologue cause a similar acceleration of sympodial cycling and shoot termination[8].

We searched our groundcherry assemblies for *SP* homologues and related florigen family members, and phylogenetic analysis

[1]Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, USA. [2]The Boyce Thompson Institute, Ithaca, NY, USA. [3]Plant Breeding and Genetics Section, School of Integrative Plant Science, Cornell University, Ithaca, NY, USA. [4]Howard Hughes Medical Institute, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, USA. [5]These authors contributed equally: Nathan T. Reem, Justin Dalrymple, Sebastian Soyk. *e-mail: jv27@cornell.edu; lippman@cshl.edu

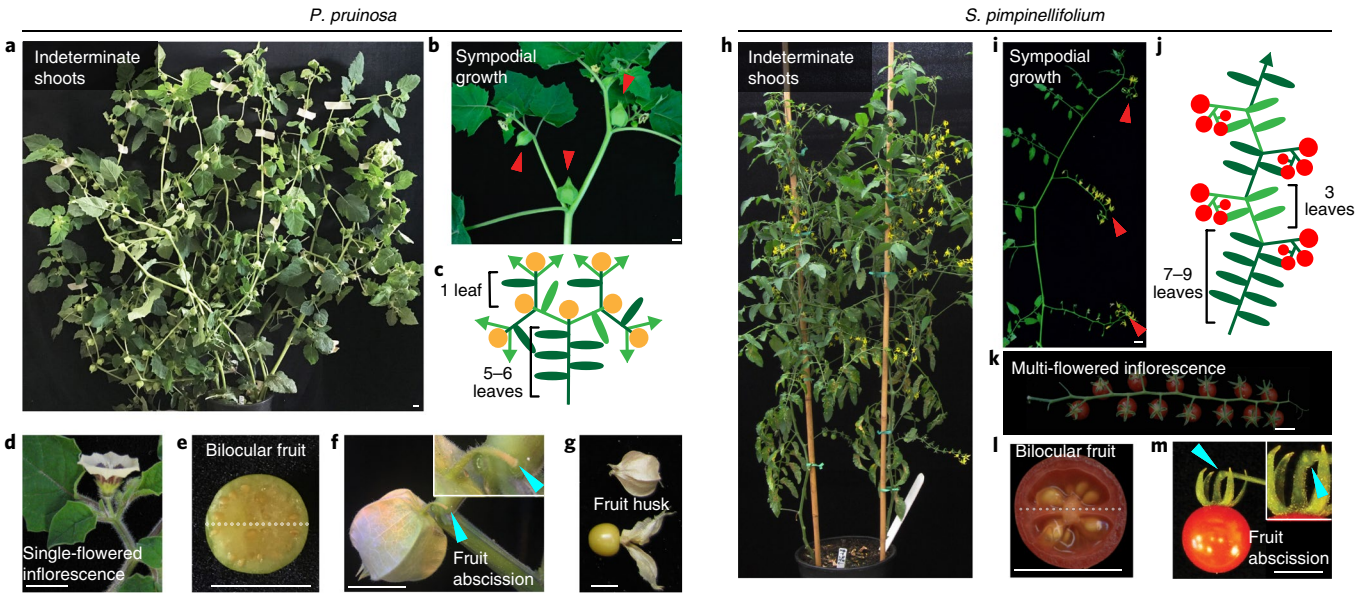

Fig. 1 | The orphan *Solanaceae* crop *P. pruinosa* (groundcherry) exhibits similar traits to the wild tomato species *S. pimpinellifolium*. **a**,**h**, Both *P. pruinosa* (**a**) and *S. pimpinellifolium* (**h**) develop indeterminate shoot systems due to sympodial growth. **b**–**d**, *P. pruinosa* primary shoots terminate after the production of 5–6 leaves (photo in **b**; schematic in **c**) in a single-flowered inflorescence (**d**), and continue vegetative growth with sympodial shoot units of one leaf. **i**–**k**, Similarly, *S. pimpinellifolium* primary shoots terminate after the production of 7–9 leaves (photo in **i**; schematic in **j**) in a multi-flowered inflorescence (**k**), and vegetative growth continues with sympodial shoot units comprising three leaves and a new inflorescence. **e**,**f**,**l**,**m**, Both *P. pruinosa* (**e** and **f**) and *S. pimpinellifolium* (**l** and **m**) develop small ~1 g bilocular fruits (**e** and **l**) with pronounced abscission zones on the fruit pedicels (**f** and **m**; insets show magnified images). **g**, *P. pruinosa* fruits are encapsulated by an inflated calyx (husk). In **c** and **j**, light and dark green bars represent successive sympodial shoots, ovals are leaves, and coloured circles are maturing fruits. Scale bars = 1 cm. In **b** and **i**, red arrowheads mark inflorescences. In **f** and **m**, cyan arrowheads mark fruit abscission zones. All images are representatives of wild-type *P. pruinosa* and *S. pimpinellifolium* from ten independently repeated experiments with similar results.

revealed a probable orthologue of *SP* detected in both genome and transcriptome assemblies (Fig. 2a). We targeted *Ppr-SP* using two guide RNAs and found several out-of-frame insertion and deletion alleles in multiple chimeric T_0 plants. Notably, these plants were more compact than the wild type, and we validated this phenotype in homozygous and biallelic null *Ppr-sp*^{CR} T_1 progeny plants (Fig. 2b,c). Like the tomato *sp*^{classic} mutant, primary shoot flowering time was unaffected in *Ppr-sp*^{CR} mutants (Fig. 2d). However, unlike the progressive sympodial termination of tomato *sp*^{classic} plants, sympodial growth in *Ppr-sp*^{CR} plants ended immediately after flowering on both primary and axillary shoots, resulting in clusters of three or four fruits per shoot (Fig. 2e–g). Although potentially valuable for specific agronomic conditions, the severity of the *Ppr-sp*^{CR} phenotype suggests that null alleles could limit fruit production, similar to null tomato *sp* alleles that also terminate faster and cause more compact plants compared with the weaker *sp*^{classic} allele[16,17] (not shown).

To identify other targets for plant architecture modification without negative effects on productivity, we used our phylogeny to identify the orthologue of *SELF-PRUNING 5G* (*SP5G*) (*Ppr-SP5G*), another florigen repressor (Fig. 2a)[6]. In the tomato, *SP5G* controls primary and canonical axillary shoot flowering time and is the major contributor to day-length sensitivity in wild tomato species. The groundcherry flowers quickly in long-day conditions, after around five leaves, suggesting little or no day-length sensitivity, which we confirmed by showing that flowering occurred one leaf faster in short-day conditions (Supplementary Fig. 1c). This weak response mirrors the cultivated tomato, where reduced *SP5G* activity was selected during domestication to mitigate day-length sensitivity. CRISPR–Cas9-induced null *sp5g* mutations in the tomato eliminate residual day-length sensitivity, causing even faster primary and axillary shoot flowering[6].

To determine whether similar effects could be achieved in the groundcherry, we targeted *Ppr-SP5G*. Multiple chimeric T_0 individuals showed more compact growth, but not as severe as in *Ppr-sp*^{CR}, and these plants were also more floral (Fig. 2h,i). Surprisingly, homozygous null *Ppr-sp5g*^{CR} T_1 progeny carrying identical independently derived deletion alleles were unaffected for primary shoot flowering (Fig. 2j). In contrast, sympodial shoots terminated more rapidly, but to a lesser degree than *Ppr-sp*^{CR} plants (Fig. 2c,i,k), explaining the difference in compactness. Importantly, this moderate shoot termination resulted in up to 50% higher concentrations of fruits along each shoot, emulating the burst of fruit production in tomato *sp*^{classic} mutants, and this effect became more pronounced with age (Fig. 2k,l). In addition to *Ppr-sp5g*^{CR} plants showing greater agronomic potential than *Ppr-sp*^{CR} plants, phenotypic differences from mutations in orthologous *Solanaceae* florigen family members reflect how these conserved flowering regulators influence species-specific sympodial growth patterns (Fig. 1).

An additional target trait for improving groundcherry productivity is fruit size. In the tomato, mutations in the classical *CLAVATA* (*CLV*)–*WUSCHEL* meristem size pathway were major contributors to boosting fruit size by increasing the locule number from the bilocular fruits of *S. pimpinellifolium* (Fig. 1l)[5,12]. In particular, the *fasciated* (*fas*) locus—an inversion that partially disrupts the promoter of the *SlCLV3* signalling peptide gene—was a major fruit size domestication quantitative trait locus (QTL). In contrast, engineered null alleles of *SlCLV3* are undesirable due to excessive and disorganized production of flower and fruit organs[17]. Rather than creating a range of weak *Ppr-CLV3* promoter alleles, as we recently demonstrated for *SlCLV3*[17], we hypothesized that a subtle increase in the locule number could be achieved by mutating the orthologue of the tomato *CLV1* gene, encoding one of several redundant leucine-rich receptors of the CLV3 peptide,

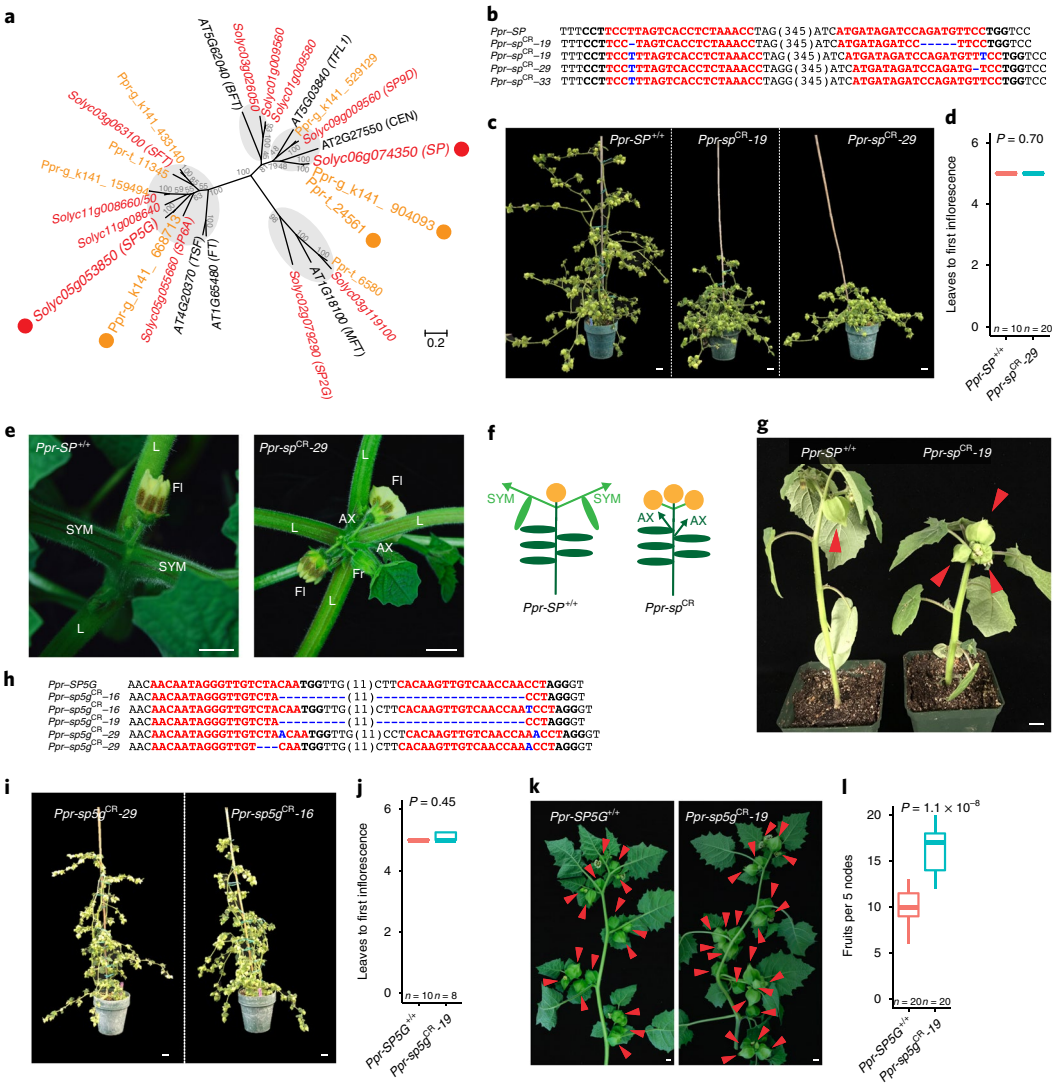

Fig. 2 | CRISPR–Cas9 targeting of domestication and improvement-related genes in *P. pruinosa*. **a**, Phylogenetic tree of *CENTRORADIALIS/TERMINAL FLOWER 1/SELF-PRUNING (CETS)* family genes in the tomato (*S. lycopersicum*), *Arabidopsis* and groundcherry (*P. pruinosa*). Contigs from de novo assemblies of both the *Physalis* genome (*Ppr-g*) and transcriptome (*Ppr-t*) were included in the phylogeny. Tomato and *P. pruinosa SP* and *SP5G* homologues are highlighted by red and orange coloured circles, respectively. Bootstrap values from 100 replicates are indicated on each node. **b**, CRISPR–Cas9 targeting of *SP* in *P. pruinosa*. *Ppr-sp*^{CR} mutant alleles identified by PCR and sequencing in T_0 transgenic plants are shown. **c**, *Ppr-sp*^{CR} stable homozygous null T_1 plants (middle and right) showing compact growth compared with the *Ppr-SP*$^{+/+}$ control (left). **d**, Quantification of the primary shoot flowering time in *Ppr-SP*$^{+/+}$ control and *Ppr-sp*^{CR} T_1 plants. Statistical significance was determined by a two-tailed, two-sample *t*-test. *n*, number of biologically independent plants. **e**,**f**, Photos (**e**) and schematics (**f**) of shoot apices of *Ppr-SP*$^{+/+}$ control (left) and *Ppr-sp*^{CR} T_1 plants (right), showing immediate termination of sympodial shoots into flowers in *Ppr-sp*^{CR}. Vegetative growth continues from axillary shoots. AX, axillary shoot; Fl, flower; Fr, fruit; L, leaf; SYM, sympodial shoot. **g**, Formation of fruit clusters on *Ppr-sp*^{CR} T_1 plants (right) compared with the *Ppr-SP*$^{+/+}$ control (left). Red arrowheads indicate husked fruits. In **c**, **e** and **g**, representative images of *Ppr-sp*^{CR} homozygous T_1 plants from six repeated independent experiments with similar results are shown. **h**, CRISPR–Cas9 targeting of *SP5G* in *P. pruinosa*. *Ppr-sp5g*^{CR} mutant alleles identified by PCR and sequencing in T_0 transgenic plants are shown. **i**, Homozygous null *Ppr-sp5g*^{CR} T_1 plants. Note the more compact growth and fruit clusters in each sympodial unit. **j**, Quantification of primary shoot flowering time in *Ppr-SP5G*$^{+/+}$ control and *Ppr-sp5g*^{CR} T_1 plants. Statistical significance was determined by a two-tailed, two-sample *t*-test. *n*, number of biologically independent plants. **k**, *Ppr-sp5g*^{CR} T_1 plants (right) exhibit shoot units with faster flower and fruit initiation and development compared with *Ppr-SP5G*$^{+/+}$ controls (left). Red arrowheads mark fruits. In **i** and **k**, representative images of *Ppr-sp5g*^{CR} T_1 plants from two repeated independent experiments with similar results are shown. **l**, Quantification of fruits on five consecutive sympodial shoot nodes in *Ppr-SP5G*$^{+/+}$ control and *Ppr-sp5g*^{CR} T_1 plants. Significance was determined by a two-tailed, two-sample *t*-test. *n*, number of biologically independent individual shoots. Backgrounds in **c** and **g** were darkened to highlight whole-plant architectures. In **b** and **h**, guide RNA targets and protospacer-adjacent motif sequences are highlighted in red and bold font, respectively. Deletions and insertions are indicated by blue dashes and blue font, respectively. The sequence gap length is shown in parentheses. Scale bars, 1 cm. For all box plots (**d**, **j** and **l**), the bottom and top of boxes represent the first and third quartile, respectively, the middle line is the median, and the whiskers represent the maximum and minimum values. Hyphenated numbers following mutant allele names in **b** and **h** indicate the individual plant from which the allele was derived.

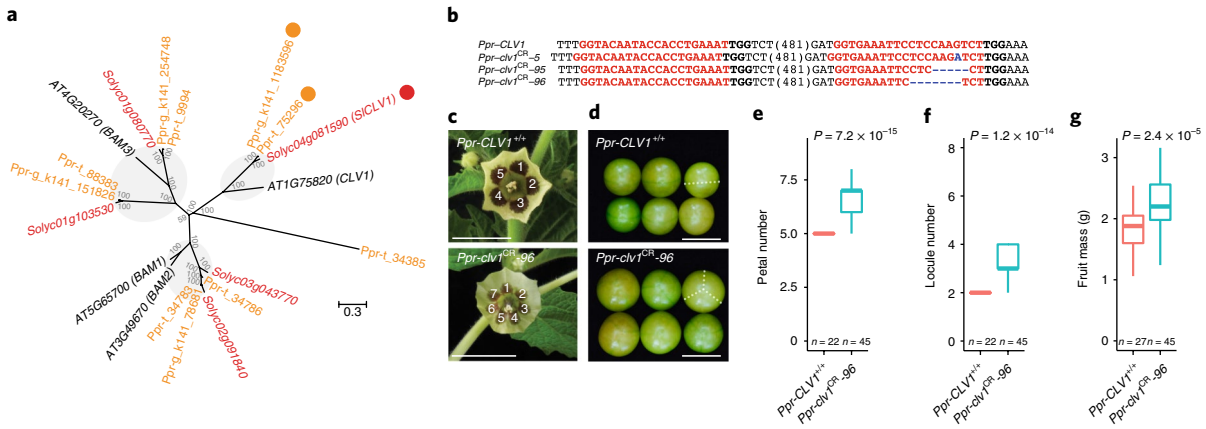

Fig. 3 | CRISPR–Cas9 targeting of domestication and improvement-related gene *CLV1* in *P. pruinosa*. **a**, Phylogenetic tree of *CLV1* homologues in the tomato, *Arabidopsis* and *P. pruinosa*. Tomato and *P. pruinosa CLV1* homologues are highlighted by red and orange coloured circles, respectively. Bootstrap values from 100 replicates are indicated on each node **b**, CRISPR–Cas9 targeting of *CLV1* in *P. pruinosa*. *Ppr-clv1*^{CR} alleles identified by sequencing are shown. **c**,**d**, Flowers (**c**) and fruits (**d**) of *Ppr-clv1*^{CR} T_0 plants (bottom) showing increased petal and locule numbers compared with *Ppr-CLV1*$^{+/+}$ controls (top). Petals are numbered and demarcation of locules is indicated with dotted lines on one representative fruit for each genotype. In **c** and **d**, representative images of *Ppr-clv1*^{CR} T_1 plants from three repeated independent experiments with similar results are shown. **e**–**g**, Quantification of petal (**e**) and locule (**f**) numbers (n = number of biologically independent flowers), and fruit mass (**g**) (n = number of biologically independent fruits) in *Ppr-CLV1*$^{+/+}$ control and *Ppr-clv1*^{CR} T_1 plants. Statistical significance was determined by two-tailed, two-sample t-test. In **b**, guide RNA targets and protospacer-adjacent motif sequences are highlighted in red and bold font, respectively. Deletions and insertions are indicated by blue dashes and blue font, respectively. The sequence gap length is shown in parentheses. Scale bars, 1 cm. For all box plots (**e**–**g**), the bottom and top of boxes represent the first and third quartile, respectively, the middle line is the median, and the whiskers represent the maximum and minimum values. Hyphenated numbers following mutant allele names in **b** indicate the individual plant from which the allele was derived.

and for which CRISPR–Cas9 null mutations moderately increase the locule number compared with *fas*[12]. Phylogenetic analysis revealed a clear orthologue of *Ppr-CLV1* (Fig. 3a), which we targeted with two guide RNAs to produce chimeric plants that produced flowers with more organs compared with the wild type (Fig. 3b). Homozygous null *Ppr-clv1*^{CR} T_1 plants for a 7-base pair deletion validated this effect (Fig. 3b–e and Supplementary Fig. 1d–g), and nearly all fruits developed three locules, translating to a 24% increase in mass (Fig. 3f,g). A similar weak effect on the locule number caused by the tomato *locule number* (*lc*) regulatory mutation of the stem cell regulator *WUSCHEL* was the foundation for the selection of additional QTLs to create a range of larger tomato fruits[5,17]. Such an approach is now possible in the groundcherry by engineering the coding sequence and regulatory mutations in orthologues of other tomato fruit-size QTLs[5,17].

Our study represents the first steps towards improving the groundcherry orphan crop, and we have already extended this work by targeting additional genes that we anticipate will benefit a range of agronomic and consumer quality traits, such as the homologue of the fruit abscission gene *JOINTLESS-2* (ref. [13]) (Supplementary Fig. 1h,i). Beyond these first targets, we are applying multiplex CRISPR–Cas9 of coding and regulatory sequences to simultaneously create qualitative and quantitative variation for multiple traits to rapidly enhance breeding germplasm. While an immediate goal is to elevate the uniquely flavoured fruits of several *Physalis* species alongside the strawberry, blueberry, blackberry and raspberry as the elusive 'fifth berry' crop in large-scale agriculture, we also aim to demonstrate the speed at which other important orphan crops, such as the grass teff, pseudocereal amaranth and legume cowpea, could be improved for regional production and potentially catapulted into mainstream agriculture by exploiting the large knowledge base of related model crops. Success will not be limited by gene-editing technologies per se, which continue to rapidly improve and expand[18], but by the availability of efficient transformation methodologies. Alleviating bottlenecks will require open access to genome-editing technologies for both public and private entities, as well as highly efficient, genotype-independent gene-delivery approaches[19,20].

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

Data availability

Raw data from this study have been submitted to the National Center for Biotechnology Information Sequence Read Archive (http://www.ncbi.nlm.nih.gov/sra) under accession number SRP142654. Transcriptome and genome assemblies have been deposited at the Sol Genomics Network (ftp://ftp.solgenomics.net).

Received: 28 June 2018; Accepted: 21 August 2018;
Published online: 1 October 2018

References

1. Scheben, A., Wolter, F., Batley, J., Puchta, H. & Edwards, D. *New Phytol.* **216**, 682–698 (2017).
2. Bohra, A., Jha, U. C., Kishor, P. B. K., Pandey, S. & Singh, N. P. *Biotechnol. Adv.* **32**, 1410–1428 (2014).
3. Wolff, X. Y. *HortScience* **26**, 1558–1559 (1991).
4. Särkinen, T., Bohs, L., Olmstead, R. G. & Knapp, S. *BMC Evol. Biol.* **13**, 214 (2013).
5. Van der Knaap, E. et al. *Front. Plant Sci.* **5**, 227 (2014).
6. Soyk, S. et al. *Nat. Genet.* **49**, 162–168 (2017).
7. Menzel, M. Y. *Proc. Am. Phil. Soc.* **95**, 132–183 (1951).
8. Paran, I. & van der Knaap, E. *J. Exp. Bot.* **58**, 3841–3852 (2007).
9. Van Eck, J. *Curr. Opin. Biotechnol.* **49**, 35–41 (2018).
10. Brooks, C., Nekrasov, V., Lippman, Z. B. & Van Eck, J. *Plant Physiol.* **166**, 1292–1297 (2014).
11. Garzón-Martínez, Ga, Zhu, Z. I., Landsman, D., Barrero, L. S. & Mariño-Ramírez, L. *BMC Genomics* **13**, 151 (2012).
12. Xu, C. et al. *Nat. Genet.* **47**, 784–792 (2015).
13. Soyk, S. et al. *Cell* **169**, 1142–1155 (2017).
14. Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V. & Zdobnov, E. M. *Bioinformatics* **31**, 3210–3212 (2015).
15. Park, S. J., Eshed, Y. & Lippman, Z. B. *Curr. Opin. Plant Biol.* **17**, 70–77 (2014).
16. Pnueli, L. et al. *Development* **125**, 1979–1989 (1998).
17. Rodríguez-Leal, D., Lemmon, Z. H., Man, J., Bartlett, M. E. & Lippman, Z. B. *Cell* **171**, 470–480 (2017).
18. Cermak, T. et al. *Plant Cell* **29**, 1196–1217 (2017).
19. Altpeter, F. et al. *Plant Cell* **28**, 1510–1520 (2016).
20. Lowe, K. et al. *Plant Cell* **28**, 1998–2015 (2016).

Acknowledgements
We thank members of the Lippman and Van Eck laboratories for valuable discussions. We thank T. Mulligan, S. Vermylen, A. Krainer and S. Qiao for assistance with plant care. We thank C. Pei for assistance with CRISPR analysis. This research was supported by a National Science Foundation Postdoctoral Research Fellowship in Biology grant (IOS-1523423) to Z.H.L., and the National Science Foundation Plant Genome Research Program (IOS-1732253) to J.V.E. and Z.B.L.

Author contributions
Z.H.L., N.T.R., J.D., J.V.E. and Z.B.L. designed and planned the experiments. All authors performed the experiments and collected the data. Z.H.L performed all bioinformatics analyses. All authors analysed the data. Z.H.L., J.V.E. and Z.B.L. designed the research. Z.H.L., S.S., J.V.E. and Z.B.L. wrote the paper with input from all authors.

Competing interests
The authors declare no competing interests.

Additional information
Supplementary information is available for this paper at https://doi.org/10.1038/s41477-018-0259-x.

Reprints and permissions information is available at www.nature.com/reprints.

Correspondence and requests for materials should be addressed to J.V. or Z.B.L.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

nature research

Corresponding author(s): J. van Eck and Z.B. Lippman

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	no software was used
Data analysis	R v3.4.3, RStudio v1.1.383, Excel v14.4.1, Benchling, clustalw2 v2.1, pal2nal v14, PhyML v20120412, FastQC v0.11.2, BUSCO v3.0.2, Trinity v2.2.0, megahit v1.1.1-2-g02102e1, Trinotate v2.0.2

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

Data Availability:

Raw data from this study have been submitted to the NCBI Sequence Read Archive (SRA; http://www.ncbi.nlm.nih.gov/sra) under accession number SRP142654. Transcriptome and genome assemblies have been deposited at the Sol Genomics Network (ftp://ftp.solgenomics.net).

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 ☐ Ecological, evolutionary & environmental 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 sizes were defined based on previous work that found to be sufficient for determining statistically significant results. At least 6 independent replicates were analyzed in each experiment for data presented in Figure 2. n values and statistical tests are presented
Data exclusions	No data excluded
Replication	Number of biologically independent replicates is presented in the Figure legends and Online Methods, and replicated experiments yielded similar results. Individual replicates (e.g. plants, shoots, flowers and fruits) are indicated and at least 6 independent replicates analyzed for each experiment.
Randomization	n/a
Blinding	n/a

Reporting for specific materials, systems and methods

Materials & experimental systems

n/a	Involved in the study
☒	☐ Unique biological materials
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology
☐	☒ Animals and other organisms
☒	☐ Human research participants

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Animals and other organisms

Policy information about studies involving animals; ARRIVE guidelines recommended for reporting animal research

Laboratory animals	n/a
Wild animals	n/a
Field-collected samples	n/a