Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Beyond Pepper: A Four-Tier Framework for Transferring Capsaicinoid Biosynthesis to 17 Crop Species, with Genome-Wide Synteny to Capsicum

Priyanca Ford

scovillesplice@gmail.com

Scoville Splice LLC https://orcid.org/0000-0003-0395-1752

Research Article

Keywords: Capsicum annuum, capsaicinoid biosynthesis, cross-species pathway transfer, synteny, tomato, eggplant, Pun1/AT3, pAMT, CRISPR-Cas9 guide design, transcript-based capacity model

Posted Date: November 26th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11875808/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: The authors declare no competing interests.

SCOVILLESPLICE.FARM

SCOVILLE SPLICE RESEARCH SERIES

Beyond Pepper: A Four-Tier Framework for Transferring Capsaicinoid Biosynthesis to 17 Crop Species, with Genome-Wide Synteny to Capsicum

A computed framework that groups 17 crop entries into four tiers by relatedness to Capsicum, with pathway matches, collinear blocks, construct and guide designs and a transcript-based capacity model.

P.I. Ford

Scoville Splice LLC · ORCID: 0000-0003-0395-1752

DOI: 10.5281/zenodo.24511314 License: CC BY 4.0 **Patents:** US 64/172,827; US 64/188,197; US 64/192,062; US 64/199,105

Figshare: figshare/35015800 Harvard Dataverse: DVN/GO8MRF OSF: osf.io/bep9x Hugging Face: capsaicinoid-data

GitHub: paper-02 GenBank: QB274170–QB274216 Website: cross-species-capsaicinoid

Package record. 278 claims traced to a file or a cited source (223 computed, 39 definition, 12 design assumption, 4 design target). 28 tables, 10 figures, 89 data files. Data fingerprint `631e3e12ff37`. Rerun every result with `bash run_all.sh`.



Tomato cross-section (generated image). Scoville Splice LLC image library.

ABSTRACT

Capsaicinoids are made by pepper (*Capsicum*), and a crop that carries the phenylpropanoid and acyl-branch steps needs two added genes, pAMT and Pun1/AT3. This paper sorts 17 crop entries into 4 tiers by relatedness to *Capsicum* and computes what a transfer would involve. Of 97 pepper pathway proteins, 78 have a close tomato match. The pAMT and Pun1/AT3 reference proteins match tomato at 85.9 to 86.3% and 72.6 to 73.8% identity. Collinear blocks

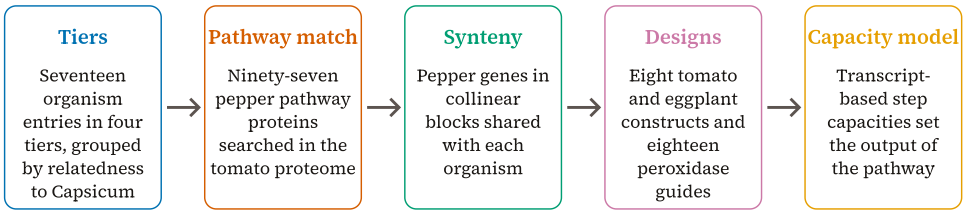
hold 51.3% of pepper genes shared with tomato and 52.6% with eggplant; tier 4 organisms share 0.1 to 10.7%. The design set has 8 tomato and eggplant constructs that carry pAMT and Pun1/AT3 and 18 guides, each with one perfect genome site, on 6 peroxidase genes. A transcript-based capacity model puts the tomato placenta output at 4.7 to 15.9% of pepper placenta. In tomato and strawberry a native phenylpropanoid step sets the output in every scenario, and the added genes leave it unchanged. All results are computed. No experimental data are reported in this paper.

Keywords: Capsicum annuum · capsaicinoid biosynthesis · cross-species pathway transfer · synteny · tomato · eggplant · Pun1/AT3 · pAMT · CRISPR-Cas9 guide design · transcript-based capacity model

KEY POINTS

- Seventeen crop entries sit in four tiers; pepper genes in collinear blocks fall from 52.6% in tier 1 to 0.1% in tier 4.
- Of 97 pepper pathway proteins, 78 have a close tomato match.
- A tomato or eggplant design adds two genes, pAMT and Pun1/AT3, to native phenylpropanoid and acyl-branch steps.
- The design set holds 18 guides on six peroxidase genes, each with 1 perfect genome site.
- In the capacity model a native tomato or strawberry step sets the output in every scenario.

Marks under numbers: computed · design target · design assumption · literature value · definition. The evidence table in Appendix A lists the source file of every one.



P02 · GA · data/figures/GA.csv

Graphical abstract. The framework in five steps: tiers, pathway matches in tomato, synteny with pepper, construct and guide designs, and the transcript-based capacity model. Drawn from the steps file. Data: `data/figures/GA.csv`; script: `code/06_supplementary.py`.

1. Introduction

Capsaicinoids are made by plants of the genus *Capsicum*. The Pun1 locus of pepper encodes an acyltransferase that is needed for pungency (Stewart et al., 2005). The pathway has two branches (pathway model of 10.5281/zenodo.24511300). The phenylpropanoid branch, PAL, C4H, 4CL, HCT, a methylation step and a reductase (VR), forms vanillin, which the aminotransferase pAMT converts to vanillylamine. The acyl branch, BCAT, KAS and a reductase, forms a branched fatty acyl-CoA. Pun1/AT3 joins vanillylamine and the acyl-CoA. A crop that carries both branches needs two added genes, pAMT and Pun1/AT3.

This paper asks how closely related candidate crops are to pepper and what a transfer of the two genes would involve. It makes five computed contributions. First, a four-tier framework for 17 organism entries with 57 designs. Second, a search of 97 pepper pathway proteins in the tomato proteome. Third, collinear blocks shared by pepper and 15 of the 17 entries. Fourth, construct sequences and guide designs for tomato and eggplant. Fifth, a transcript-based model of the step capacities of the pathway in tomato placenta, strawberry fruit and eggplant fruit peel.

The design set follows the companion study of ten pepper cultivars (10.5281/zenodo.24511300), which models the pathway and ranks Pun1/AT3 substitutions, and the guide method of 10.5281/zenodo.24511324 and 10.5281/zenodo.24511350.

2. Methods

2.1 Computational basis

The pepper reference is the RefSeq assembly GCF_002878395.1 (*Capsicum annuum* UCD10Xv1.1; Hulse-Kemp et al., 2018). Tomato is RefSeq GCF_000188115.5 (SL3.1). Eggplant is GCA_057556415.1 (ASM5755641v1), a genome without gene models. The assemblies of the other organisms are listed in Table 5 and Table S6, and every accession is in `inputs/input_sources.csv`. Protein searches use DIAMOND 2.2.6 (Buchfink et al., 2015), protein-to-genome placement uses miniprot 0.18-r281 (Li, 2023) and quantification uses Salmon 2.8.0 (Patro et al., 2017) (Table S16). Analysis scripts use the Python standard library only and run on Python 3.9 or later.

2.2 Tiers and designs

The framework lists 17 organism entries in 4 tiers: tier 1 holds tomato, eggplant and *Nicotiana benthamiana*, tier 2 the other Solanaceae crops, tier 3 fruit and leaf crops of three other eudicot families, and tier 4 carrot, *Brassica oleracea* and the two grasses. Tiers are design groups: they set the order in which a transfer would be tackled, from the closest relatives to the most distant. *Brassica oleracea* fills two entries, leaf and head types and broccoli, so the 17 entries cover 16 species in 7 families. Each entry has 2 to 5 designs, 57 in all. A design target is a heat level in Scoville heat units (SHU) set by the designer, from 100 to 1,000,000 SHU; it is a target and not a model result (`inputs/design_targets.csv`). Ground cherry and tomatillo are grouped in tier 2 by relatedness (Solanaceae).

2.3 Pathway proteins in tomato

The 97 pepper pathway proteins are listed in `inputs/orthologs/pathway_97_pepper.tsv`: PAL, C4H, 4CL, HCT, methylation enzymes (COMT-class and CCoAOMT-type), KAS, BCAT, FabG-type reductases and acyltransferases related to Pun1/AT3. Each was searched with DIAMOND blastp against the RefSeq tomato proteome of GCF_000188115.5, and the best hit by bit score was kept (Table S1). Matches were grouped by the identity of the best hit: close matches (identity 71.1% or more in this set), weaker matches (66.8% or less) and acyltransferase-family matches for the queries related to Pun1/AT3.

The reference searches used four pAMT-type queries (O82521.2, NP_001311635.1, PHU21971.1 and PHT38654.1) and four Pun1/AT3-type queries (AAV66309.1, NP_001311698.1, PHU26684.1 and PHT56199.1) against the same tomato proteome with DIAMOND 2.2.6 in default mode at an E-value cut-off of 1e-5, keeping up to five hits per query (Table 3). The query proteins are in `inputs/orthologs/search/`.

2.4 Collinear blocks

Gene order is the order of protein-coding genes by start position on each chromosome-scale sequence, taken from the annotation as published; sequences shorter than 10 Mb (organelles) are left out. The longest protein of each gene represents it. Proteins were searched in both directions with DIAMOND 2.2.6 (E-value 1e-10, up to 5 targets, sensitive mode). A one-to-one ortholog is a reciprocal best hit by bit score. Blocks follow the logic of MCScanX (Wang et al., 2012) with our own implementation: two consecutive orthologs join a block when at most 10 genes lie between them in both genomes and they keep the same order (direct block) or the reverse order (inverted block); a block needs at least 5 orthologs. Blocks are found one at a time, best first, so each gene belongs to at most one block. The measure is the share of the 27,958 pepper genes on the chromosomes that lie in a block. Sensitivity covers maximum gaps of 5, 10 and 25 genes, an index that counts orthologs only, and 20 shuffles of the positions of the orthologs in one genome (a control that returns no blocks if the method finds none by chance). The tomato side has 25,158 genes.

Tomato and eggplant were counted a second way from the stored ortholog and block tables (`code/02_synteny.py`; Table S2). For the other organisms a second implementation was run in addition

to the main one: it reads the annotation with its own parser, counts a sequence as a chromosome when it holds at least 1000 genes, runs DIAMOND in default mode at an E-value cut-off of 1e-5 and uses its own best-chain code. Eggplant has a second route through the tomato genes. The scripts are in `research_record/`.

Genomes without gene models (eggplant, *N. benthamiana* and petunia) were handled by placement: miniprot aligned every pepper protein to the genome, the best placement of each protein with an identity of 0.5 or more (or 0.8 or more in the strict setting) became a locus, the loci were translated and ordered along each chromosome-scale sequence, and the same ortholog and block code ran on them. This gives 32,356 eggplant loci at identity 0.5 and 21,355 at identity 0.8. For two organisms a close relative stands in: *N. tabacum* for *N. benthamiana* (alongside the placement on *N. benthamiana* itself) and diploid *Fragaria vesca* for octoploid *F.* × *ananassa* (alongside the octoploid). For potato the chromosome-scale Hera haplotype and the scaffold-level SolTub_3.0 assembly were both used; for SolTub_3.0 every scaffold with at least 50 genes was kept.

2.5 Pathway genes and blocks

For each of the 97 pathway proteins on a chromosome the analysis records whether its tomato or eggplant ortholog lies in a block and the share of its 10 flanking genes (five on each side) that lie in a block (Table S4 and Table S5).

2.6 Constructs

The eight construct sequences are stored in `data/results/constructs.fasta`. Both strands were translated in all three frames. The pAMT reading frame starts at the first 12 residues of the pAMT reference O82521.2 (459 residues), and the Pun1/AT3 reading frame at the first 12 residues of AAV66309.1 (440 residues); each runs to the next stop. The Pun1/AT3 allele is read from the differences between the translated reading frame and the reference.

2.7 Guide design and scoring

Three peroxidase genes of tomato (SlPrx1, SlPrx34 and SlPrx72) have RefSeq models on the proteins NP_001233970.2, NP_001234644.2 and NP_001296734.1. The eggplant counterparts (SmPrx1, SmPrx34 and SmPrx72) are placements of the same three proteins on the eggplant genome, taken from the best miniprot locus (Table S7). Every SpCas9 site with an NGG PAM whose 20-base protospacer and PAM lie in one coding exon was listed; candidates with a GC content of 40 to 70%, no TTTT and no run of five equal bases were kept (293 candidates, Table S8 and Table S9). The C scanner (`code/guide_design/scan.c`) listed every genome site of each candidate with an NGG PAM, on both strands, with up to 4 mismatches. The score of a guide is the Hsu et al. (2013) aggregate score: 100 / (100 + the sum of its site scores) × 100; the score of one site is the product of $(1 - w)$ over its mismatch positions, where w is the position weight (Table S18), times $1 / (((19 - d) / 19) \times 4 + 1)$ for sites with more than one mismatch, where d is the mean distance between mismatches, times $1 / n^2$ for n mismatches (`code/04_guides.py`). Per gene, 3 guides were chosen from the candidates that have exactly one perfect genome site and cut in the first 60% of the coding sequence, highest score first, each at least 100 coding bases from the guides already chosen. A second scanner (numpy, no shared code) and an exact text search of each guide with its PAM on both strands gave the same sites; every check of that comparison passed (`research_record/outputs/guides/verify_report.txt`: ALL PASS). On-target activity scores are not computed.

2.8 Capacity model

Each step of the pathway has a capacity in model units: the transcripts per million (TPM) of its gene family times 1000 / 1580.295, so one thousand units equal the Pun1 transcript level of pepper placenta at 24 days post anthesis (DPA). The pathway is one chain: PAL, C4H, 4CL, HCT, a methylation step (CCoAOMT-type or COMT-class), VR (cinnamoyl-CoA reductase), pAMT and Pun1/AT3. The output J is the smallest capacity along the chain. The acyl branch adds one limit: capsaicin and dihydrocapsaicin together make at least 0.85 of the total, so J is at most the KAS capacity divided by that share. This rule reproduces the optimum of the stored linear programme of the 44-reaction model in every one of its 150 stored scenario rows to within 0.006 units (Table S12).

The added genes pAMT and Pun1/AT3 are design assumptions: their capacity is a multiple x of the pepper placenta level, 1.0, with 0.25 and 0.1 also run. Pun1/AT3 capacity is further scaled by the factors 1.0, 1.35, 1.65 and 2.1, which stand for the Pun1/AT3 designs (Table S15).

The step capacities come from public RNA-seq runs. Tomato is PRJNA1348326 (cultivar M82 placenta: 3 replicates each of mature green, breaker and ripe fruit), strawberry is PRJNA1203349 (*F.* × *ananassa* fruit, three unripe and three ripe replicates, reads mapped to the *F. vesca* transcripts of GCF_000184155.1) and eggplant is PRJNA1374051 (fruit peel, six runs, reads mapped to the placed eggplant coding sequences). The first 3,000,000 read pairs of every run were quantified with Salmon 2.8.0 against the transcripts of the species (selective alignment, automatic library type). A transcript belongs to a step when the text of its protein annotation matches the step's patterns (Table S19). A step is represented in two ways: the representative transcript, the one with the highest total TPM over all runs, and the family sum of all matching transcripts; each value is the mean of the replicates. The reference for comparison is pepper placenta at 34 DPA, whose output computed by the same model is 56.42 units (representative) and 67.28 units (family sum). The sensitivity analysis moves one input at a time across its range: step capacities between 0.5 and 1.5 times the baseline, the added-gene level between 0.1 and 1.0, the Pun1 multiplier between 1.0 and 2.1 and the main-capsaicinoid share between 0.7 and 0.95.

2.9 Pepper COMT transcript

The class-I COMT transcript XP_047270260.1 was quantified in the public pepper placenta series PRJNA789050 (*C. chinense* accession SL08, five stages at 4, 24, 34, 44 and 54 DPA, one library per stage).

3. Results

3.1 Tiers and organism entries

Table 1 lists the 17 entries. Tier 1 has 3 entries (tomato, eggplant and *N. benthamiana*), tier 2 has 4 (potato, ground cherry, tomatillo and petunia), tier 3 has 5 (strawberry, cucumber, melon, apple and lettuce) and tier 4 has 5 (carrot, two *Brassica oleracea* entries, corn and rice). The highest design target is 1,000,000 SHU in tier 1 for tomato and eggplant and 10,000 SHU in tier 4. The tiers carry computed synteny with pepper (Section 3.4): the entries with a value span 40.6 to 52.6% in tier 1, 45.5 to 46.3% in tier 2, 11.8 to 18.8% in tier 3 and 0.1 to 10.7% in tier 4. The lowest tier 2 value is above the highest tier 3 value, and the lowest tier 3 value is above the highest tier 4 value. Ground cherry and tomatillo (2 entries) stay in tier 2, grouped by relatedness.

Table 1. Organism entries and tiers. Seventeen organism entries in four tiers with the number of designs, the highest design target in Scoville heat units (SHU) and the share of pepper genes in collinear blocks (maximum gap of 10 genes, at least 5 orthologs per block) shared with the genome compared. Ground cherry and tomatillo are grouped with their tier by relatedness. Data: `data/tables/T01.csv`*; script:* `code/02_synteny.py`*.*

Tier	Organism	Species	Family	Designs	Highest design target (SHU)	Pepper genes in blocks (%)	Genome compared
Tier 1	Tomato	Solanum lycopersicum	Solanaceae	5	1,000,000	51.3	Solanum lycopersicum, RefSeq annotation
Tier 1	Eggplant	Solanum melongena	Solanaceae	5	1,000,000	52.6	Solanum melongena, pepper-protein placement
Tier 1	Nicotiana benthamiana (model host)	Nicotiana benthamiana	Solanaceae	3	500,000	40.6	Nicotiana benthamiana, pepper-protein placement
Tier 2	Potato	Solanum tuberosum	Solanaceae	5	500,000	46.3	Solanum tuberosum (Hera hap1), submitter annotation, 12 chromosomes
Tier 2	Ground cherry	Physalis peruviana	Solanaceae	3	100,000	not computed	grouped with its tier by relatedness (same family as Tier 1 and Tier 2 crops)
Tier 2	Tomatillo	Physalis philadelphica	Solanaceae	3	100,000	not computed	grouped with its tier by relatedness (same family as Tier 1 and Tier 2 crops)

Tier	Organism	Species	Family	Designs	Highest design target (SHU)	Pepper genes in blocks (%)	Genome compared
Tier 2	Petunia	Petunia x hybrida	Solanaceae	2	100,000	45.5	Petunia x hybrida, pepper-protein placement
Tier 3	Strawberry	Fragaria x ananassa	Rosaceae	4	100,000	18.8	Fragaria vesca (diploid stand-in), RefSeq annotation
Tier 3	Cucumber	Cucumis sativus	Cucurbitaceae	3	50,000	14.0	Cucumis sativus, RefSeq annotation
Tier 3	Melon	Cucumis melo	Cucurbitaceae	3	50,000	14.0	Cucumis melo, RefSeq annotation
Tier 3	Apple	Malus domestica	Rosaceae	3	50,000	16.4	Malus domestica, RefSeq annotation
Tier 3	Lettuce	Lactuca sativa	Asteraceae	3	50,000	11.8	Lactuca sativa, RefSeq annotation
Tier 4	Carrot	Daucus carota	Apiaceae	3	10,000	10.7	Daucus carota, RefSeq annotation
Tier 4	Brassica (leaf and head types)	Brassica oleracea	Brassicaceae	3	10,000	1.5	Brassica oleracea, RefSeq annotation
Tier 4	Broccoli	Brassica oleracea var. italica	Brassicaceae	3	10,000	1.5	Brassica oleracea, RefSeq annotation
Tier 4	Corn	Zea mays	Poaceae	3	10,000	0.1	Zea mays, RefSeq annotation
Tier 4	Rice	Oryza sativa	Poaceae	3	10,000	0.3	Oryza sativa, RefSeq annotation

3.2 Pathway proteins in tomato

Of the 97 pepper pathway proteins, 78 have a close tomato match (identity 71.1 to 98.3%), 7 a weaker match (54.7 to 66.8%) and 12 are acyltransferase-family queries related to Pun1/AT3 (41.9 to 88.9%) (Table 2, Figure 1, Table S1). All 6 PAL queries match tomato at 76.1 to 95.2%, all 11 4CL queries at 83.1 to 95.8% and all 9 FabG-type reductase queries at 77.1 to 90.1%. The C4H query (1 protein) has a weaker match of 66.8% identity over 94.9% of its length. The HCT, methylation, KAS and BCAT queries have close matches in 8 of 9, 20 of 22, 18 of 19 and 6 of 8 queries.

Table 2. Tomato matches of the pepper pathway proteins. Best tomato protein for each of 97 pepper pathway proteins, summarised by gene. A close match has an identity of 71.1% or more, a weaker match 66.8% or less; acyltransferase-family matches are the 12 queries related to Pun1/AT3. Data: `data/tables/T02.csv`*; script:* `code/01_orthologs.py`*.*

Gene	Enzyme	Pepper queries	Close matches	Weaker matches	Acyltransferase-family matches	Identity range (%)
PAL	phenylalanine ammonia-lyase	6	6	0	0	76.1 to 95.2
C4H	cinnamate 4-hydroxylase	1	0	1	0	66.8 to 66.8
4CL	4-coumarate-CoA ligase	11	11	0	0	83.1 to 95.8
HCT	hydroxycinnamoyl transferase	9	8	1	0	60.9 to 95.2
COMT	O-methyltransferase (COMT and CCoAOMT classes)	22	20	2	0	56.8 to 98.3
KAS	3-oxoacyl-ACP synthase	19	18	1	0	61.1 to 93.7
BCAT	branched-chain aminotransferase	8	6	2	0	54.7 to 94.0
FAR/FabG	3-oxoacyl-ACP reductase (FabG-type)	9	9	0	0	77.1 to 90.1
Pun1/CS-related	acyltransferase queries related to Pun1/AT3	12	0	0	12	41.9 to 88.9

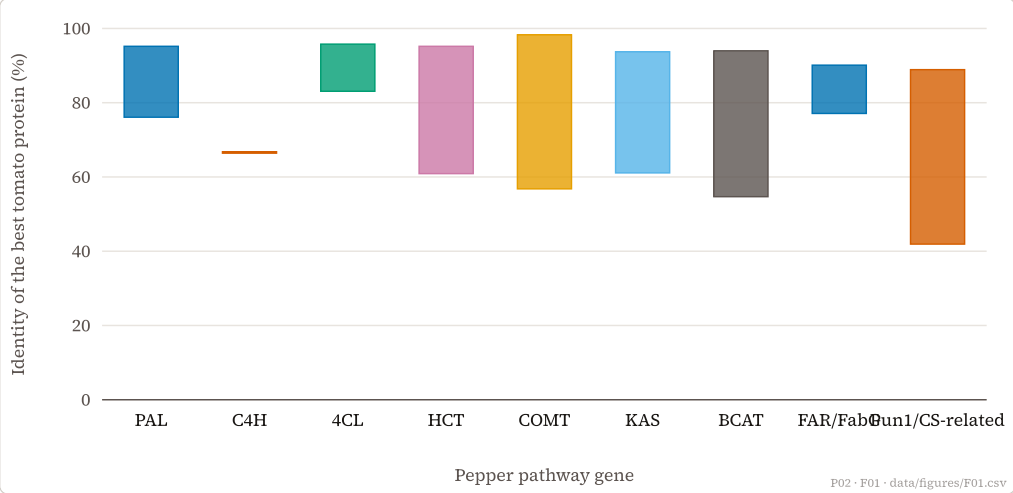


Figure 1. Tomato matches of the pepper pathway proteins. Each bar spans the lowest to the highest identity of the best tomato protein over the pepper queries of one gene (97 queries in all). DIAMOND blastp, best hit by bit score for each query. Data: `data/figures/F01.csv`; script: `code/01_orthologs.py`.

The reference searches for the two genes a tomato design adds are in Table 3. The four pAMT-type queries have the same best tomato protein, XP_004244777.1, annotated as a gamma-aminobutyrate transaminase 2-like protein, at 85.9 to 86.3% identity over 98.9% of the query. Together the pAMT-type queries hit 5 tomato proteins at 80.3 to 86.3%, all annotated as gamma-aminobutyrate transaminases. The four Pun1/AT3-type queries have acylsugar acyltransferase 3-like proteins as best hits at 72.6 to 73.8% identity over the full length of the query (100% coverage). They hit 5 tomato proteins at 43.8 to 74.4%. The design therefore adds pAMT and Pun1/AT3 as genes, next to the native phenylpropanoid and acyl-branch steps found in the first search.

Table 3. pAMT and Pun1/AT3 reference searches against the tomato proteome. Best tomato protein for four pAMT-type and four Pun1/AT3-type pepper reference proteins (DIAMOND, E-value 1e-5, default mode). Best means the highest bit score. Data: `data/tables/T03.csv`; script: `code/01_orthologs.py`.

Query family	Pepper query	Best tomato protein	Identity (%)	Query coverage (%)	E-value	Tomato annotation
pAMT (vanillin aminotransferase)	pAMT_ref_O82521.2	XP_004244777.1	86.3	98.9	4.17e-299	gamma aminobutyrate transaminase 2-like [Solanum lycopersicum]
pAMT (vanillin aminotransferase)	NP_001311635.1	XP_004244777.1	86.3	98.9	5.08e-300	gamma aminobutyrate transaminase 2-like [Solanum lycopersicum]
pAMT (vanillin aminotransferase)	PHU21971.1	XP_004244777.1	86.3	98.9	4.17e-299	gamma aminobutyrate transaminase 2-like [Solanum lycopersicum]
pAMT (vanillin aminotransferase)	PHT38654.1	XP_004244777.1	85.9	98.9	6.90e-298	gamma aminobutyrate transaminase 2-like [Solanum lycopersicum]
Pun1/AT3 (capsaicin synthase)	AT3_Pun1_ref_AAV66309.1	XP_004232634.1	72.9	100	1.48e-227	acylsugar acyltransferase 3-like [Solanum lycopersicum]
Pun1/AT3 (capsaicin synthase)	NP_001311698.1	XP_004232634.1	72.6	100	2.43e-226	acylsugar acyltransferase 3-like [Solanum lycopersicum]
Pun1/AT3 (capsaicin synthase)	PHU26684.1	XP_004232635.1	73.5	100	4.63e-228	acylsugar acyltransferase 3-like [Solanum lycopersicum]
Pun1/AT3 (capsaicin synthase)	PHT56199.1	XP_004232635.1	73.8	100	3.98e-229	acylsugar acyltransferase 3-like [Solanum lycopersicum]

3.3 Collinear blocks shared with tomato and eggplant

Pepper has 27,958 protein-coding genes on its chromosomes. Of these, 18,330 (65.56%) have a one-to-one tomato ortholog. A total of 14,350 genes (51.33%) lie in 520 blocks, 265 direct and 255 inverted. The median block holds 12 orthologs and the largest 520 (Figure 3). Of the orthologs, 78.3% lie in a block (Table 4). Recounting the tomato and eggplant block statistics from the stored ortholog and block tables gives the stored value for all recounted quantities. The share is 49.97% at a maximum gap of 5 genes and 53.34% at 25 genes (Figure 2, Table S2); with the gene index counting orthologs only it is 80.21%. Shuffling the positions of the orthologs leaves 0 genes in blocks.

The eggplant comparison uses 32,356 placed loci. In all, 20,104 pepper genes (71.91%) have an eggplant ortholog, and 14,712 genes (52.62%) lie in 523 blocks (252 direct, 271 inverted). At the stricter identity filter of 0.8 the comparison has 15,354 orthologs and 11,880 genes (42.49%) in 469 blocks. The share is 50.97% at a maximum gap of 5 and 54.75% at 25. The second route through the tomato genes gives 50.1% and, at identity 0.8, 44.8%, within 2.52 and 2.31 points of the main method.

Table 4. Pepper genes in collinear blocks shared with tomato and eggplant. Blocks need at least 5 one-to-one orthologs with at most 10 genes between neighbours in both genomes. Percentages are of 27,958 pepper protein-coding genes on the 12 chromosomes; the eggplant gene set is a placement of pepper proteins on the genome. Data: `data/tables/T04.csv`*; script:* `code/02_synteny.py`*.*

Comparison	One-to-one orthologs	Pepper genes with an ortholog (%)	Pepper genes in blocks	Pepper genes in blocks (%)	Blocks (direct / inverted)	Median / largest block (anchors)	Orthologs in blocks (%)
Tomato	18,330	65.56	14,350	51.33	520 (265 / 255)	12 / 520	78.3
Eggplant (identity 0.5 or more)	20,104	71.91	14,712	52.62	523 (252 / 271)	12 / 443	73.2
Eggplant (identity 0.8 or more)	15,354	54.92	11,880	42.49	469 (226 / 243)	12 / 459	77.4

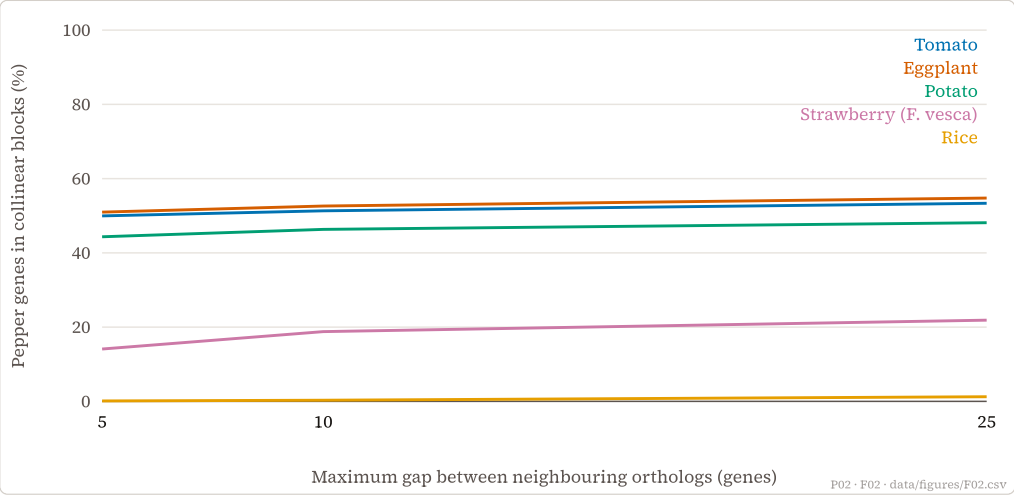


Figure 2. Synteny result at three gap settings. Pepper genes in blocks at maximum gaps of 5, 10 and 25 genes for five comparisons. Same block search at each gap setting. Data: `data/figures/F02.csv`; script: `code/02_synteny.py`.

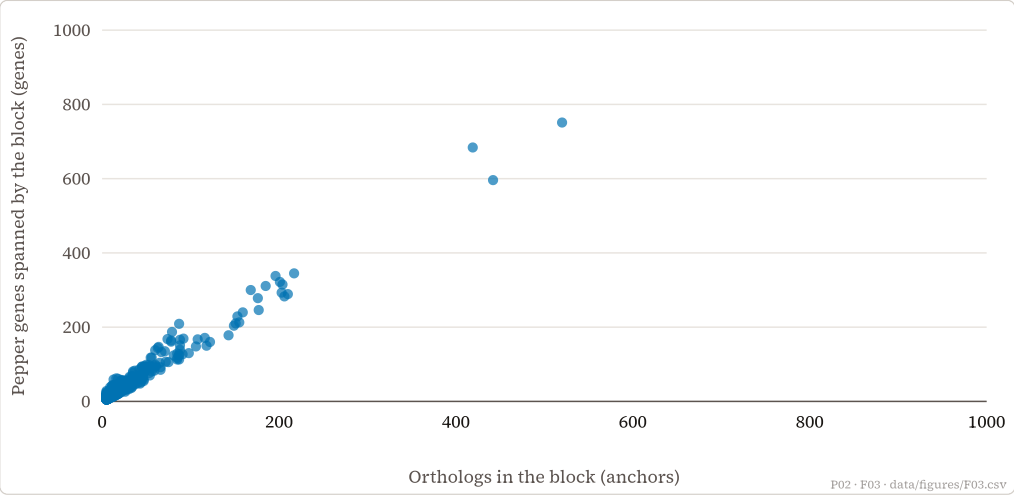


Figure 3. The 520 pepper-tomato blocks. Each point is one block: its number of orthologs and the number of pepper genes it spans. Block table of the pepper-tomato comparison. Data: `data/figures/F03.csv`; script: `code/02_synteny.py`.

Counting by homology alone gives higher numbers. By homology alone, 94.5% of pepper genes have any tomato protein hit and 96.4% any eggplant hit; 65.6% and 71.9% have a one-to-one ortholog (Table S3). The

blocks measure collinearity, which is the quantity used for the tiers.

Of the 97 pathway proteins, 81 lie on a chromosome (68 distinct genes). Of these, 59 have a tomato ortholog, 46 of them in a block; 59 have an eggplant ortholog and 49 of those lie in a block. On average 51.5% of the flanking genes of a pathway gene lie in a tomato block and 52.5% in an eggplant block, close to the genome-wide shares (Table S4, Table S5).

3.4 Collinear blocks across the organisms

Table 5 and Figure 4 give the share of pepper genes in blocks for the other comparisons. In tier 2, potato (Hera chromosomes) shares 46.3% (40.88% for the scaffold-level SolTub_3.0, a lower bound) and petunia 45.5%. In tier 1, *N. benthamiana* shares 40.6% by placement and the annotated *N. tabacum* 39.81%. In tier 3, strawberry shares 18.8% (diploid *F. vesca*; 4.64% for the octoploid), apple 16.4%, cucumber 14.0%, melon 14.0% and lettuce 11.8%. In tier 4, carrot shares 10.7%, *Brassica oleracea* 1.5%, corn 0.1% and rice 0.3%. Every shuffled-position control returned 0 blocks. Against the main method the second method differs by at most 0.28 points (octoploid strawberry) and by 0.10 points or less in the other annotated comparisons.

Table 5. Pepper genes in collinear blocks for every comparison. Percent of pepper genes in blocks at maximum gaps of 5, 10 and 25 genes, with the assembly, the annotation basis and, where run, a second method at a gap of 10. Data: `data/tables/T05.csv`*; script:* `code/02_synteny.py`.

Organism entry	Species compared	Assembly	Basis	Gap 5 (%)	Gap 10 (%)	Gap 25 (%)	Blocks (gap 10)	Second method (gap 10, %)
Tomato	Solanum lycopersicum	GCF_000188115.5	RefSeq annotation	49.97	51.33	53.34	520	51.37
Eggplant	Solanum melongena (identity 0.5 or more)	GCA_057556415.1	pepper-protein placement	50.97	52.62	54.75	523	50.1
Eggplant	Solanum melongena (identity 0.8 or more)	GCA_057556415.1	pepper-protein placement		42.49		469	44.8
Potato	Solanum tuberosum (Hera hap1)	GCA_020169585.1	submitter annotation, 12 chromosomes	44.32	46.34	48.14	612	46.34
Potato	Solanum tuberosum (SolTub_3.0), lower bound	GCF_000226075.1	RefSeq annotation, scaffolds with at least 50 genes	39.9	40.88	41.75	438	40.9
N. benthamiana (model host)	Nicotiana tabacum (stand-in), stand-in	GCF_000715075.1	RefSeq annotation	30.73	39.81	44.27	826	39.91
N. benthamiana (model host)	Nicotiana benthamiana (identity 0.5 or more)	GCA_034376525.1	pepper-protein placement	32.85	40.56	44.92	857	
N. benthamiana (model host)	Nicotiana benthamiana (identity 0.8 or more)	GCA_034376525.1	pepper-protein placement	17.29	25.36	30.63	649	
Petunia	Petunia x hybrida (identity 0.5 or more)	GCA_046563445.1	pepper-protein placement	43.0	45.46	47.56	539	
Petunia	Petunia x hybrida (identity 0.8 or more)	GCA_046563445.1	pepper-protein placement	25.21	28.37	30.31	408	
Strawberry	Fragaria vesca (diploid stand-in), stand-in	GCF_000184155.1	RefSeq annotation	14.11	18.77	21.86	468	18.75
Strawberry	Fragaria x ananassa (octoploid)	GCA_050043575.1	submitter annotation	1.14	4.64	12.77	201	4.36
Apple	Malus domestica	GCF_042453785.1	RefSeq annotation	9.58	16.35	22.53	516	16.38
Cucumber	Cucumis sativus	GCF_000004075.3	RefSeq annotation	8.6	13.99	19.02	437	14.05
Melon	Cucumis melo	GCF_025177605.1	RefSeq annotation	8.47	14.02	19.11	441	14.05
Lettuce	Lactuca sativa	GCF_002870075.5	RefSeq annotation	5.46	11.84	18.94	425	11.82
Carrot	Daucus carota	GCF_001625215.2	RefSeq annotation	5.45	10.66	18.35	374	10.69
Rice	Oryza sativa	GCF_034140825.1	RefSeq annotation	0.11	0.31	1.24	16	0.33
Corn	Zea mays	GCF_902167145.1	RefSeq annotation	0.0	0.11	0.35	5	0.09
Brassica and Broccoli	Brassica oleracea	GCF_000695525.1	RefSeq annotation	0.47	1.48	4.95	72	

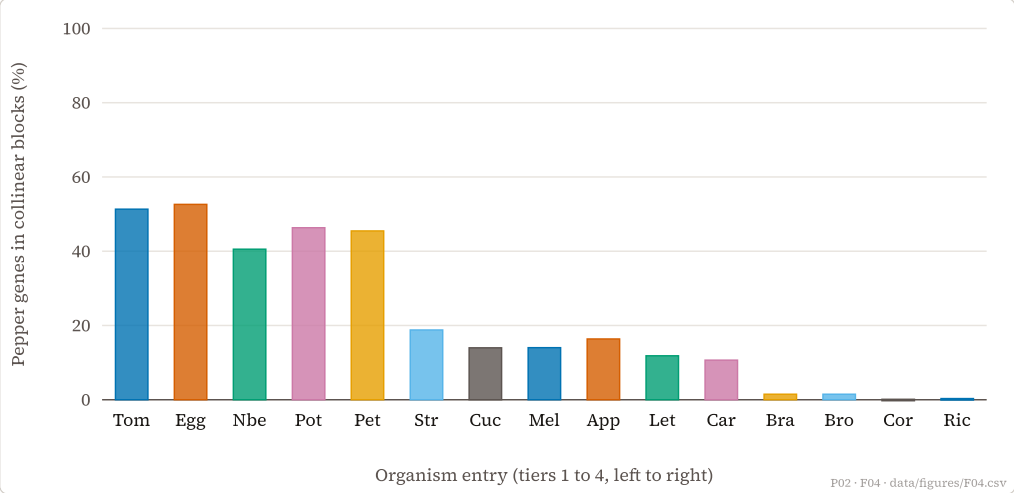


Figure 4. Pepper genes in collinear blocks shared with each organism. Percent of the 27,958 pepper genes on chromosomes that lie in a block shared with the genome compared (maximum gap of 10 genes, at least 5 orthologs per block). Organisms are in tier order: Tom, tomato; Egg, eggplant; Nbe, N. benthamiana; Pot, potato; Pet, petunia; Str, strawberry; Cuc, cucumber; Mel, melon; App, apple; Let, lettuce; Car, carrot; Bra, Brassica oleracea (leaf and head types); Bro, broccoli (same genome); Cor, corn; Ric, rice. Ground cherry and tomatillo are not plotted. Best-chain block search on one-to-one orthologs. Data: `data/figures/F04.csv`; script: `code/02_synteny.py`.

3.5 Construct sequences

Table 6 lists the 8 constructs: 4 for tomato (SS-TOM-01 to SS-TOM-04) and 4 for eggplant (SS-EGG-01 to SS-EGG-04). Each carries the pAMT reading frame and the Pun1/AT3 reading frame in the forward direction (Figure 5). The pAMT frame has 466 residues, the 459 residues of the reference and 7 further residues before the stop, and the Pun1/AT3 frame has 447 residues, the 440 residues of the reference and 7 further residues. The Pun1/AT3 allele is the wild type in 4 constructs (designs 01 and 02 of each host), S39L in 2 (design 03) and S39L+L345G in 2 (design 04). The designs 01 and 02 of each host have the same sequence, so the 8 constructs hold 6 distinct sequences; they differ in design target. The tomato constructs are 4,076 bp (designs 01 and 02) and 5,692 bp (03 and 04); the eggplant constructs are 4,065 and 5,670 bp. The longer constructs carry 2 further reading frames of 163 and 137 residues. The substituted positions of Pun1/AT3 rank 1 (S39L) and 2 (L345G) in the ESM2 pocket scan of the companion design (Lin et al., 2023; 10.5281/zenodo.24511300), at 7.34 and 7.31 Å from the catalytic His169 and Asp173, which places both in the second shell.

Table 6. Construct sequences. Length, open reading frames of pAMT and Pun1/AT3 and the Pun1/AT3 allele of the eight construct sequences. Coordinates are 1-based on the stored sequence. Data: `data/tables/T06.csv`; script: `code/03_constructs.py`.

Construct	Host	Length (bp)	Pun1/AT3 allele	pAMT ORF	Pun1/AT3 ORF	Further ORFs (100 residues or more)	Note
SS-TOM-01	tomato	4,076	wild type	699-2099	2452-3795	0	
SS-TOM-02	tomato	4,076	wild type	699-2099	2452-3795	0	same sequence as SS-TOM-01
SS-TOM-03	tomato	5,692	S39L	699-2099	2452-3795	2	
SS-TOM-04	tomato	5,692	S39L+L345G	699-2099	2452-3795	2	
SS-EGG-01	eggplant	4,065	wild type	689-2089	2441-3784	0	
SS-EGG-02	eggplant	4,065	wild type	689-2089	2441-3784	0	same sequence as SS-EGG-01
SS-EGG-03	eggplant	5,670	S39L	689-2089	2441-3784	2	
SS-EGG-04	eggplant	5,670	S39L+L345G	689-2089	2441-3784	2	

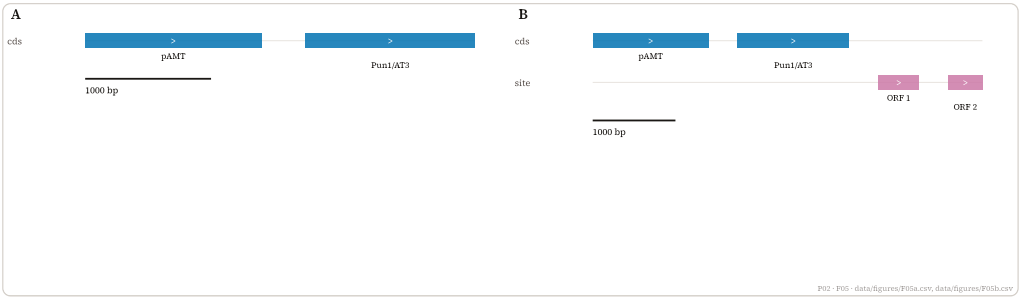


Figure 5. Reading frames in two tomato constructs. A, SS-TOM-01 (and SS-TOM-02, the same sequence). B, SS-TOM-04. The pAMT and Pun1/AT3 reading frames and, in B, two further reading frames; the scale bar is in base pairs. Six-frame translation of the stored sequence against the reference proteins. Data: `data/figures/F05a.csv`, `data/figures/F05b.csv`; script: `code/03_constructs.py`.

3.6 Peroxidase guides

The guides target three peroxidase genes of tomato and three of eggplant, each with an ATG start, a stop at the end and no internal stop in its model (Table S7). The tomato models are the RefSeq coding sequences of 999, 996 and 996 bases. The eggplant models are placements of the same proteins with identities of 91.6%, 93.3% and 89.8%. The three eggplant models have coding lengths of 996, 996 and 999 bases. For SmPrx34 the best tomato hit of the eggplant model is XP_004243525.1 at 94.9% identity, just ahead of the SlPrx34 protein at 93.4% (Table S17), as expected for a cluster of close paralogues.

The scan scored 293 candidates (Table S8, Table S9, Figure 7). In SlPrx34 17 of 47 candidates have exactly one perfect genome site, as against all 62 candidates in SlPrx1. From these, 3 guides per gene were chosen, 18 in all (Table 7, Figure 6). Each guide has 1 perfect genome site, at the intended position in the gene. The MIT scores run from 99.7 to 100.0. No guide has a site with one mismatch (0 sites). Sites with two or three mismatches number 22 and belong to 7 guides (Table S10); 160 sites have four mismatches. The SlPrx72 and SmPrx72 guides g2 have the same 20-base protospacer, `GTGATAATAGCATCAACGGC`, which lies on the minus strand in tomato and on the plus strand in eggplant.

Table 7. Eighteen peroxidase guides. Each guide has exactly one perfect genome site. The cut position is the coding base before the cut. The MIT score is the Hsu et al. (2013) aggregate specificity score (0 to 100). Data: `data/tables/T07.csv`; script: `code/04_guides.py`.

Guide	Gene	Genome position (strand)	Protospacer 5' to 3'	PAM	Cut after coding base	GC (%)	Perfect sites	Sites with 1 / 2 / 3 / 4 mismatches	MIT score
SlPrx1-g1	SlPrx1	NC_015438.3:93,248,788-93,248,807 (-)	TTCGGCCCTTGGGCACGTAC	TGG	123 of 999	65	1	0 / 0 / 0 / 0	100.0
SlPrx1-g2	SlPrx1	NC_015438.3:93,249,398-93,249,417 (-)	TCGTCCATCTCTGCGTCCCG	TGG	447 of 999	65	1	0 / 0 / 0 / 3	100.0
SlPrx1-g3	SlPrx1	NC_015438.3:93,248,910-93,248,929 (+)	GACGGTTCCATCCTCATTTC	TGG	259 of 999	50	1	0 / 0 / 0 / 7	100.0
SlPrx34-g1	SlPrx34	NC_015444.3:59,079,764-59,079,783 (+)	GCAGTAGAAGCTGAGTGCCC	TGG	349 of 996	60	1	0 / 2 / 1 / 7	99.9
SlPrx34-g2	SlPrx34	NC_015444.3:59,079,404-59,079,423 (+)	GGAATAGCGGGTTCTAGCTA	TGG	67 of 996	50	1	0 / 1 / 2 / 9	99.9
SlPrx34-g3	SlPrx34	NC_015444.3:59,079,518-59,079,537 (-)	AGCAATGCAGCTGCAAGAGA	TGG	167 of 996	50	1	0 / 1 / 2 / 30	99.9
SlPrx72-g1	SlPrx72	NC_015440.3:8,112,565-8,112,584 (+)	CACTGCATACCCCGGTCCAC	CGG	430 of 996	65	1	0 / 0 / 0 / 1	100.0
SlPrx72-g2	SlPrx72	NC_015440.3:8,112,918-8,112,937 (-)	GTGATAATAGCATCAACGGC	AGG	254 of 996	45	1	0 / 0 / 0 / 6	100.0
SlPrx72-g3	SlPrx72	NC_015440.3:8,112,447-8,112,466 (-)	TGCCTCTCATGGTTTAAATC	AGG	562 of 996	40	1	0 / 2 / 1 / 18	99.9
SmPrx1-g1	SmPrx1	CM172108.1:5,542,608-5,542,627 (+)	TTCGGCTCTTGGACACGTAC	TGG	121 of 996	55	1	0 / 0 / 0 / 1	100.0
SmPrx1-g2	SmPrx1	CM172108.1:5,542,070-5,542,089 (-)	CTGACCTGGTCTGTGCCCAC	GGG	440 of 996	65	1	0 / 0 / 0 / 1	100.0
SmPrx1-g3	SmPrx1	CM172108.1:5,542,476-5,542,495 (+)	GCCTGCACCAGCAATGAGGA	TGG	253 of 996	60	1	0 / 0 / 0 / 6	100.0

Guide	Gene	Genome position (strand)	Protospacer 5' to 3'	PAM	Cut after coding base	GC (%)	Perfect sites	Sites with 1 / 2 / 3 / 4 mismatches	MIT score
SmPrx34-g1	SmPrx34	CM172110.1:34,753,250-34,753,269 (+)	GACTGGAAAAGTTACTCGTC	GGG	503 of 996	45	1	0 / 0 / 0 / 8	100.0
SmPrx34-g2	SmPrx34	CM172110.1:34,753,922-34,753,941 (+)	AAATGCATCCTGAGCAAGGC	AGG	180 of 996	50	1	0 / 0 / 0 / 9	100.0
SmPrx34-g3	SmPrx34	CM172110.1:34,753,545-34,753,564 (+)	CTGCGCATGACACAACTCCA	GGG	356 of 996	55	1	0 / 0 / 2 / 13	99.9
SmPrx72-g1	SmPrx72	CM172110.1:15,348,260-15,348,279 (-)	CAGTTGAATGACATCTCTGG	TGG	408 of 999	45	1	0 / 1 / 1 / 5	100.0
SmPrx72-g2	SmPrx72	CM172110.1:15,348,094-15,348,113 (+)	GTGATAATAGCATCAACGGC	AGG	256 of 999	45	1	0 / 0 / 0 / 7	100.0
SmPrx72-g3	SmPrx72	CM172110.1:15,347,292-15,347,311 (-)	CGCACAATGGATTCAACATT	AGG	122 of 999	40	1	0 / 1 / 5 / 29	99.7

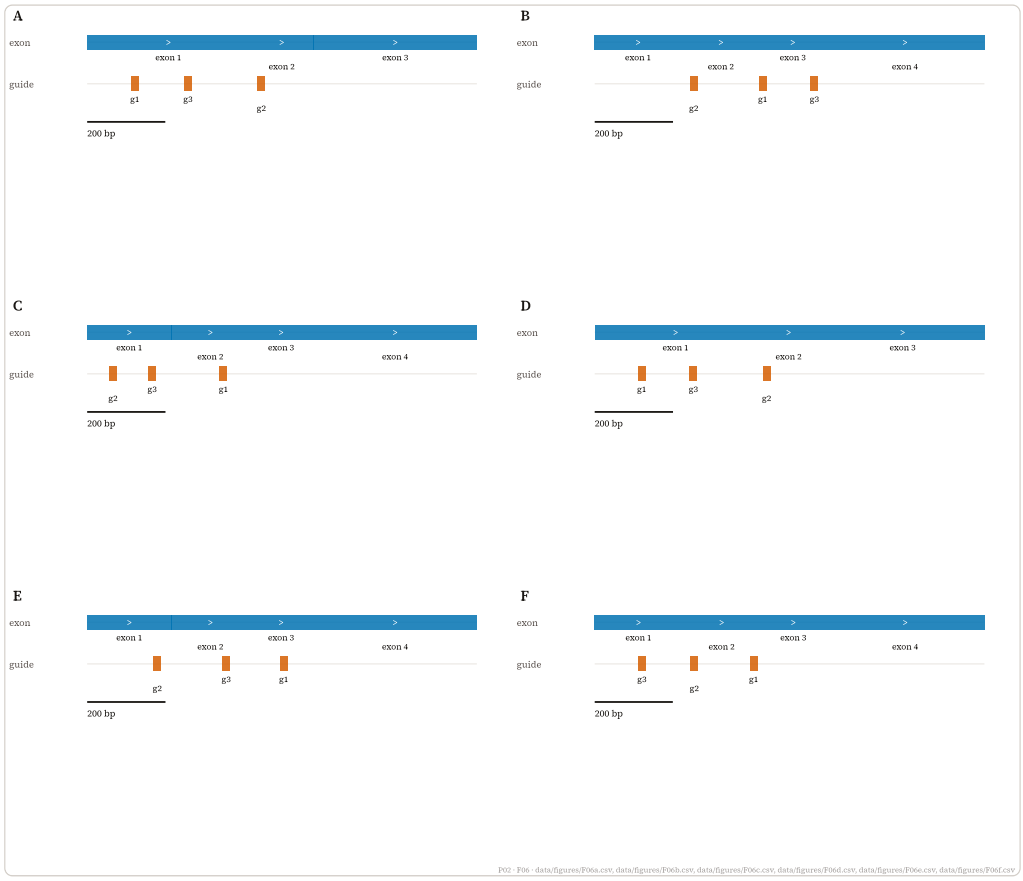


Figure 6. Guide cut sites along six peroxidase coding sequences. A, SlPrx1. B, SlPrx34. C, SlPrx72. D, SmPrx1. E, SmPrx34. F, SmPrx72. Exons and the three guides of each gene along the coding sequence; the scale bar is in base pairs. Positions are coding-sequence coordinates computed from the exon coordinates and the cut positions. Data: `data/figures/F06a.csv`, `data/figures/F06b.csv`, `data/figures/F06c.csv`, `data/figures/F06d.csv`, `data/figures/F06e.csv`, `data/figures/F06f.csv`; script: `code/04_guides.py`.

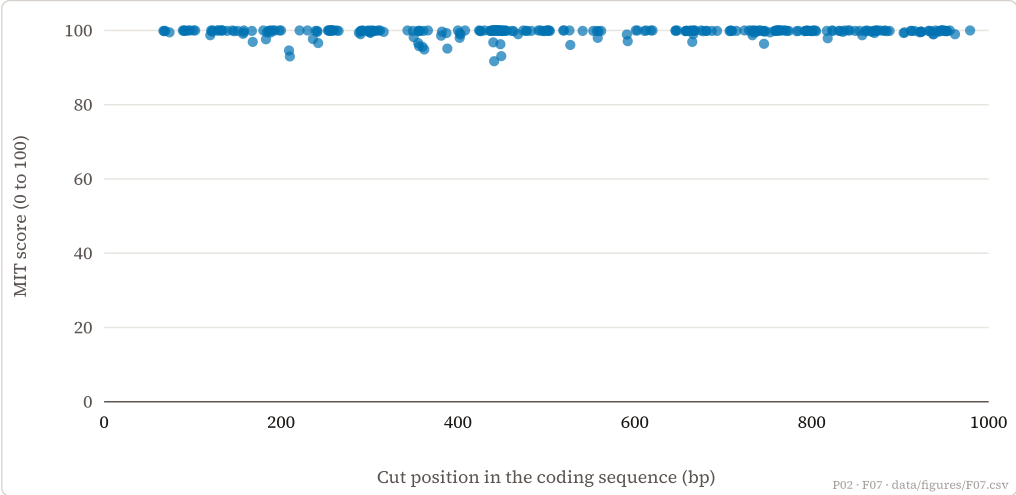


Figure 7. Specificity of the 293 candidate guides. Each point is one candidate guide: its cut position along the coding sequence and its MIT score; the six genes are plotted together. Scores from the whole-genome scan. Data: `data/figures/F07.csv`; script: `code/04_guides.py`.

3.7 Capacity model

The pathway steps peak at different stages in tomato placenta (Table S11). With representative transcripts, PAL has 105.94 units at mature green and 15.31 at breaker, whereas HCT has 2.66 and 140.53 units. The output of the pathway (Table 8, Figure 8) is limited by a native step in every dataset, stage and statistic. In tomato placenta it is 2.7 units (limited by HCT) and 5.1 units (C4H) at mature green, 7.2 (methylation) and 10.7 (C4H) at breaker, and 4.1 (methylation) and 6.3 (C4H) at ripe, which is 4.7 to 15.9% of pepper placenta at 34 DPA (56.42 and 67.28 units). The replicates of tomato placenta at breaker give 5.04 to 7.45 units (Table S13).

In strawberry fruit HCT limits the output in all four combinations: 5.3 and 8.1 units in unripe fruit and 9.0 and 11.6 units in ripe fruit, 9.5 to 17.2% of pepper placenta. The strawberry replicates differ widely (0.94 to 14.12 units in unripe fruit). In eggplant fruit peel the output is 21.8 units with representative transcripts, limited by the acyl branch through the KAS capacity, and 34.6 units with family sums, limited by HCT: 38.6% and 51.4% of pepper placenta. The six eggplant replicates range from 5.78 to 31.26 units. Read mapping rates are 71.6 to 93.7% for tomato, 65.5 to 91.6% for strawberry and 60.0 to 72.0% for eggplant (Table S14).

Table 8. Modelled output of the capsaicinoid pathway. Output J in model units for representative and family-sum transcript capacities, with the limiting step. The percentage compares J with pepper placenta at 34 days post anthesis (56.42 and 67.28 units). Added pAMT and Pun1/AT3 are at the pepper level (design assumption). Data: `data/tables/T08.csv`; script: `code/05_design_model.py`.

Dataset	Stage	J, representative (units)	Limiting step	J, family sum (units)	Limiting step	% of pepper, representative	% of pepper, family sum
Tomato placenta	mature green	2.7	HCT	5.1	C4H	4.7	7.5
Tomato placenta	breaker	7.2	methylation	10.7	C4H	12.7	15.9
Tomato placenta	ripe	4.1	methylation	6.3	C4H	7.3	9.4
Strawberry fruit	unripe	5.3	HCT	8.1	HCT	9.5	12.0
Strawberry fruit	ripe	9.0	HCT	11.6	HCT	15.9	17.2
Eggplant fruit peel	fruit peel	21.8	KAS (acyl branch)	34.6	HCT	38.6	51.4

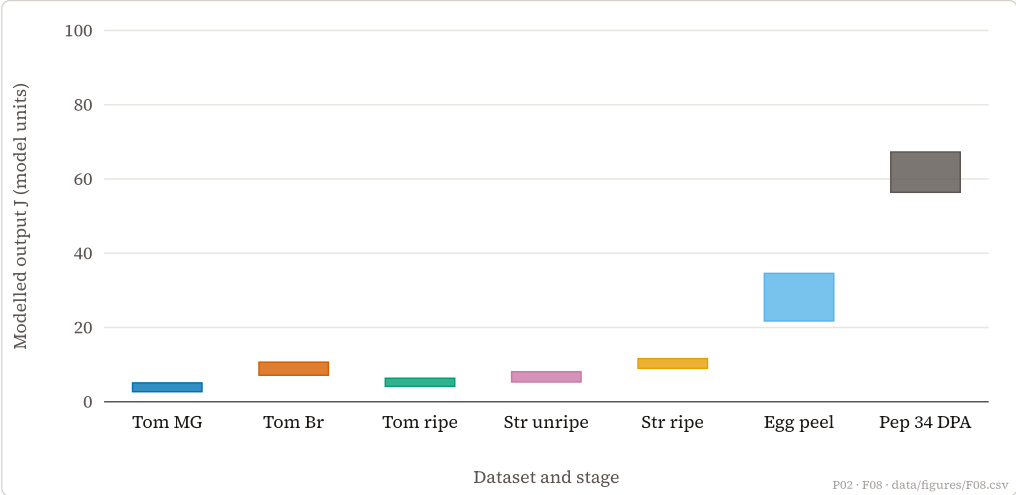


Figure 8. Modelled output of the pathway. Each bar spans the output J for representative and family-sum transcript capacities, with added pAMT and Pun1/AT3 at the pepper level (design assumption). Tom, tomato placenta; Str, strawberry fruit; Egg, eggplant fruit peel; MG, mature green; Br, breaker; Pep 34 DPA, pepper placenta at 34 days post anthesis. Smallest step capacity along the chain, with the acyl-branch limit. Data: `data/figures/F08.csv`; script: `code/05_design_model.py`.

The added genes leave the output unchanged. Across 72 scenario rows, pAMT or Pun1/AT3 is the limiting step in 0 rows, and the four Pun1 multipliers give the same output in 72 rows. In 24 of 24 groups the output is the same at the added-gene levels 1.0, 0.25 and 0.1. The sensitivity analysis shows which inputs matter (Figure 9). In tomato placenta at breaker only the methylation capacity (output 3.58 to 8.59 units) and the C4H capacity (4.29 to 7.17 units) move the output from its baseline of 7.17 units. In ripe strawberry only the HCT capacity does (4.50 to 13.49 units around 8.99). In eggplant peel the KAS capacity (10.88 to 28.01 units), the HCT capacity, the VR capacity and the assumed main-capsaicinoid share (26.43 to 19.48 units) move it from 21.77 units.

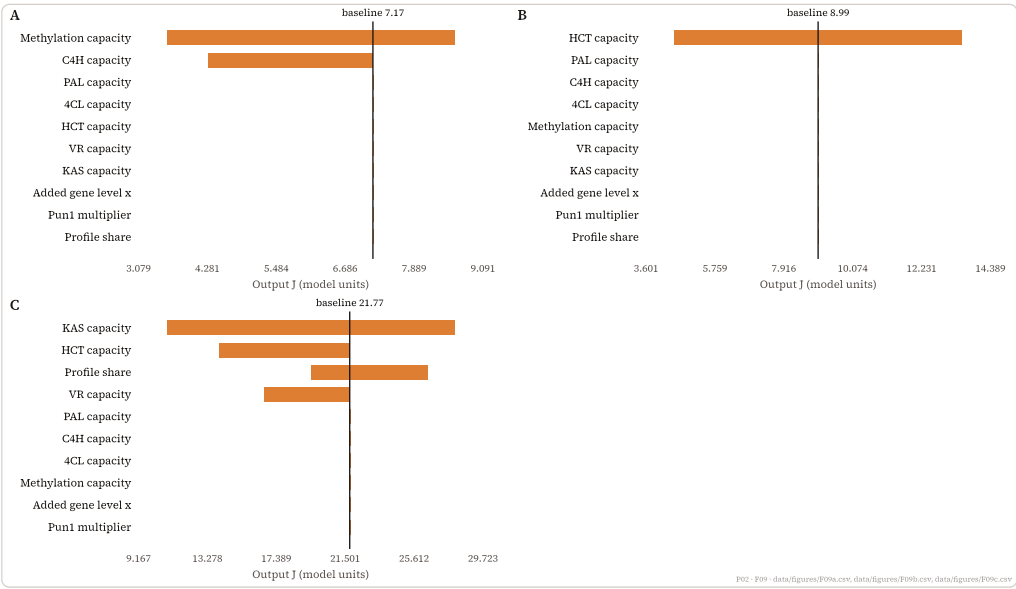


Figure 9. Sensitivity of the modelled output to each input. A, tomato placenta at breaker. B, strawberry fruit, ripe. C, eggplant fruit peel. Each bar spans the output when one input moves across its range (Table S15) and the others stay at baseline; representative transcripts, CCoAOMT-type methylation. One-at-a-time analysis of the capacity model. Data: `data/figures/F09a.csv`, `data/figures/F09b.csv`, `data/figures/F09c.csv`; script: `code/05_design_model.py`.

3.8 COMT transcript in pepper placenta

The class-I COMT transcript XP_047270260.1 is at 535.2 TPM at 24 DPA and 166.3 TPM at 34 DPA, a fall of 68.9% (Table 9, Figure 10). The fall continues to 29.8 TPM at 54 DPA. The series has one library per stage.

Table 9. Pepper COMT transcript in placenta. Transcripts per million of the class-I COMT transcript XP_047270260.1 at five stages (one library per stage). Data: `data/tables/T09.csv`*; script:* `code/06_supplementary.py`.

Days post anthesis	COMT transcript (TPM)
4	0.3
24	535.2
34	166.3
44	76.1
54	29.8

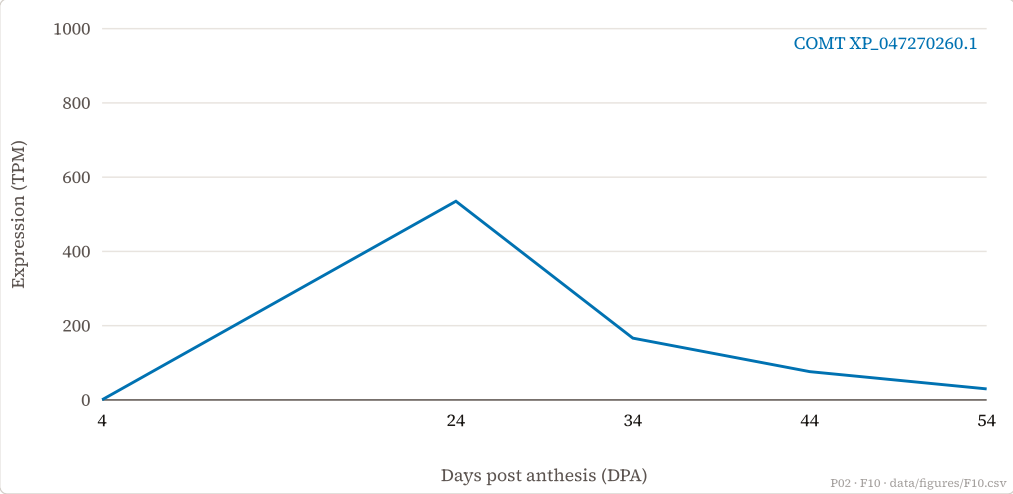


Figure 10. Pepper COMT transcript in placenta. Transcripts per million of the class-I COMT transcript at five stages, one library per stage. Salmon quantification of the public placenta series. Data: `data/figures/F10.csv`; script: `code/06_supplementary.py`.

4. Limitations

The eggplant, *N. benthamiana* and petunia gene sets are placements of pepper proteins on genomes, not annotations. Block thresholds are choices; sensitivity covers the gap, the identity filter and a second method. The capacity model reads capacity from transcript level and returns the smallest step, without regulation or kinetics; its added-gene levels are design assumptions, and its tissues and stages differ between crops. Guide specificity is the MIT score; on-target activity is not scored. The added segment of constructs 03 and 04 is described by length and reading frames.

5. Data and code availability

Everything behind this paper is in the data package: the claims table, the data for every table and figure, the scripts that regenerate them, the inputs, the research record and a backup. `bash run_all.sh` reruns every step and compares each regenerated file with the stored file; `bash verify.sh` checks the checksums. Full mode (`bash run_all.sh --full`) repeats the whole-genome guide scan; the genome downloads, the block searches and the RNA-seq runs are described in `code/heavy/NEEDS.txt` and their scripts are in `research_record/`. The paper and data are licensed under CC BY 4.0 and the code under the MIT licence. The public input datasets are listed in `inputs/input_sources.csv`.

Public records of this paper: Zenodo 10.5281/zenodo.24511314; Figshare: figshare/35015800; Harvard Dataverse: DVN/GO8MRF; OSF: osf.io/bep9x; Hugging Face: capsaicinoid-data; GitHub: paper-02; GenBank: QB274170–QB274216; Website: cross-species-capsaicinoid.

Intellectual Property. The engineered constructs, cultivar designs, and methodologies described in this paper are the subject of pending United States patent applications (USPTO Provisional Application Nos. 64/172,827; 64/188,197; 64/192,062; and 64/199,105).

Experimental status. All results reported in this paper are computational. Experimental work on these designs is under way at the authors' laboratory facilities in Montecito, California, and Geneva, Switzerland. Those data are not part of this paper, are shared only with commercial partners under agreement, and are planned for publication next year.

6. References

1. Buchfink, B., Xie, C., & Huson, D. H. (2015). Fast and sensitive protein alignment using DIAMOND. *Nature Methods, 12*(1), 59–60. doi:10.1038/nmeth.3176
2. Ford, P.I. (2026). In Silico Design of Ten Capsicum annuum Cultivars Spanning 0.87 to 2.96 Million Scoville Heat Units. *Scoville Splice Research Series*. doi:10.5281/zenodo.24511300
3. Ford, P.I. (2026). Umami-Enriched Capsicum: In Silico Design of Glutamate Accumulation by Glutamate Decarboxylase Knockout and Glutamate Dehydrogenase Expression. *Scoville Splice Research Series*. doi:10.5281/zenodo.24511324
4. Ford, P.I. (2026). Trait Expansion in Capsicum: CRISPR Guide Design for Nine Trait Groups with Gene Assignment and Genome-Wide Off-Target Profiling. *Scoville Splice Research Series*. doi:10.5281/zenodo.24511350
5. Hsu, P. D., Scott, D. A., Weinstein, J. A., Ran, F. A., Konermann, S., Agarwala, V., ... & Zhang, F. (2013). DNA targeting specificity of RNA-guided Cas9 nucleases. *Nature Biotechnology, 31*(9), 827–832. doi:10.1038/nbt.2647
6. Hulse-Kemp, A. M., Maheshwari, S., Stoffel, K., et al. (2018). Reference quality assembly of the 3.5-Gb genome of *Capsicum annuum* from a single linked-read library. *Horticulture Research, 5*, 4. doi:10.1038/s41438-017-0011-0
7. Li, H. (2023). Protein-to-genome alignment with miniprot. *Bioinformatics, 39*(1), btad014. doi:10.1093/bioinformatics/btad014
8. Lin, Z., Akin, H., Rao, R., et al. (2023). Evolutionary-scale prediction of atomic-level protein structure with a language model. *Science, 379*(6637), 1123–1130. doi:10.1126/science.ade2574
9. Patro, R., Duggal, G., Love, M. I., Irizarry, R. A., & Kingsford, C. (2017). Salmon provides fast and bias-aware quantification of transcript expression. *Nature Methods, 14*(4), 417–419. doi:10.1038/nmeth.4197
10. Stewart, C., Kang, B., Liu, K., Mazourek, M., Moore, S. L., Yoo, E. Y., Kim, B., Paran, I., & Jahn, M. M. (2005). The *Pun1* gene for pungency in pepper encodes a putative acyltransferase. *The Plant Journal, 42*(5), 675–688. doi:10.1111/j.1365-313X.2005.02410.x
11. Wang, Y., Tang, H., DeBarry, J. D., Tan, X., Li, J., Wang, X., ... & Paterson, A. H. (2012). MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic Acids Research, 40*(7), e49. doi:10.1093/nar/gkr1293

Supplementary data

The supplementary tables hold the full data behind the paper. Table S1 lists the 97 pathway proteins with their tomato matches. Table S2 to Table S6 hold the synteny results: sensitivity and the shuffled control, four ways of counting, the pathway proteins and blocks, and every comparison with its assembly and second method. Table S7 to Table S10 and Table S18 hold the guide design: the target gene models, the candidates per gene, all 293 candidate guides, the genome sites with one to three mismatches and the MIT position weights. Table S11 to Table S15 and Table S19 hold the capacity model: step capacities, agreement with the stored solver output, replicates, RNA-seq runs, model inputs and the assignment of transcripts to steps. Table S16 lists software and versions and Table S17 the check of the eggplant gene models against the tomato proteome. Every table is stored as a CSV file in `data/tables/`, and the scripts that write them are in `code/`. The file map below lists every data, input, code and research-record file.

Table S1. All 97 pepper pathway proteins and their tomato matches. Best tomato protein by bit score for each query: identity, coverage, E-value, bit score and match class. Data: `data/tables/S01.csv`*; script:* `code/01_orthologs.py`*.*

Pepper protein	Gene	Tomato protein	Identity (%)	Coverage (%)	E-value	Bit score	Match class
NP_001311532.1	PAL	NP_001307538.1	94.6	100.6	0.0	1328	close match
XP_016554784.1	PAL	XP_004249558.1	95.2	99.7	0.0	1340	close match
XP_016565628.1	PAL	NP_001307530.1	76.1	101.9	0.0	1001	close match
XP_016573925.1	PAL	NP_001307530.1	92.5	100.0	0.0	1298	close match
XP_047251840.1	PAL	NP_001307538.1	94.9	100.3	0.0	1338	close match
XP_047251841.1	PAL	NP_001307538.1	94.9	100.3	0.0	1338	close match
NP_001311984.1	C4H	XP_004240122.1	66.8	94.9	6e-237	660	weaker match
NP_001311759.1	4CL	NP_001333770.1	90.3	100.6	0.0	992	close match
XP_016545240.2	4CL	NP_001333776.1	84.9	99.8	0.0	915	close match
XP_016561322.1	4CL	XP_004232175.1	95.8	99.6	0.0	1044	close match
XP_016561323.1	4CL	XP_004232175.1	91.3	104.9	0.0	979	close match
XP_016570682.1	4CL	XP_004234005.1	83.8	100.0	0.0	881	close match
XP_016571691.1	4CL	XP_004245937.1	83.1	100.7	0.0	950	close match
XP_016578426.2	4CL	NP_001333774.1	95.0	99.6	0.0	1027	close match
XP_016579621.2	4CL	XP_004242790.1	88.6	101.6	0.0	968	close match
XP_047265503.1	4CL	NP_001333794.1	90.8	99.7	0.0	1007	close match
XP_047265504.1	4CL	NP_001333794.1	91.7	99.6	0.0	1014	close match
XP_047265606.1	4CL	XP_004236342.1	83.9	100.4	0.0	914	close match
NP_001311756.1	HCT	XP_004235891.1	95.2	100.0	4.5e-314	848	close match
XP_016540288.2	HCT	NP_001234850.2	60.9	100.7	1.5199999999999998e-188	531	weaker match
XP_016546636.1	HCT	XP_004250998.1	89.5	99.8	3.06e-298	809	close match
XP_016548258.1	HCT	XP_004251197.1	90.3	100.9	1.92e-302	820	close match
XP_016551171.2	HCT	XP_004244558.1	82.7	101.1	1.12e-219	604	close match
XP_016562978.1	HCT	XP_004235891.1	95.2	100.0	4.5e-314	848	close match
XP_016562979.1	HCT	XP_004235891.1	95.2	100.0	4.5e-314	848	close match
XP_016579751.1	HCT	NP_001234850.2	88.3	99.8	1.3599999999999999e-285	776	close match
XP_047260600.1	HCT	XP_004250074.1	87.2	100.0	2.53e-91	273	close match
NP_001311511.1	COMT	XP_004230766.1	95.2	100.4	6.1e-172	473	close match
NP_001311774.1	COMT	XP_004235028.1	88.3	100.0	1.07e-241	659	close match
XP_016545476.1	COMT	NP_001306101.1	91.2	100.6	1.87e-249	679	close match
XP_016553826.1	COMT	XP_004237700.1	83.2	98.9	1.2400000000000002e-163	455	close match
XP_016553827.1	COMT	XP_004237700.1	83.6	98.7	9.14e-131	370	close match
XP_016561061.1	COMT	XP_004231856.1	97.9	100.0	2.15e-176	484	close match
XP_016561062.2	COMT	XP_004231856.1	98.3	99.6	1.25e-175	482	close match
XP_016561063.1	COMT	XP_004231856.1	97.9	100.0	3.57e-175	481	close match
XP_016570346.1	COMT	XP_004234518.1	75.2	100.0	1.03e-122	348	close match
XP_016571727.2	COMT	XP_004234518.1	78.0	62.1	6.16e-118	341	close match
XP_016576832.2	COMT	XP_004242403.1	79.8	99.0	6.55e-123	352	close match
XP_047258539.1	COMT	NP_001306101.1	72.6	100.8	2.8e-183	511	close match
XP_047258760.1	COMT	XP_004235028.1	88.0	90.8	3.03e-63	196	close match
XP_047260147.1	COMT	XP_004235028.1	66.7	107.3	1.1399999999999998e-52	169	weaker match
XP_047260702.1	COMT	XP_004235028.1	84.7	100.0	8.229999999999999e-230	629	close match
XP_047261854.1	COMT	XP_004234518.1	82.8	99.1	1.8e-137	385	close match
XP_047264644.1	COMT	XP_004235028.1	56.8	86.6	1.8799999999999997e-54	176	weaker match
XP_047266718.1	COMT	XP_004234518.1	78.1	94.1	4.1199999999999995e-48	153	close match
XP_047270260.1	COMT	XP_004242403.1	83.4	99.7	5.65e-223	612	close match
XP_047270582.1	COMT	XP_004242403.1	74.6	102.2	9.200000000000001e-99	290	close match
XP_047270907.1	COMT	NP_001306101.1	77.3	95.3	4.0099999999999996e-92	273	close match
XP_047270933.1	COMT	NP_001306101.1	71.1	100.0	1.8699999999999998e-52	168	close match

Pepper protein	Gene	Tomato protein	Identity (%)	Coverage (%)	E-value	Bit score	Match class
NP_001311587.1	KAS	XP_010324549.1	88.4	100.0	3.37e-249	680	close match
NP_001311771.1	KAS	XP_004235286.1	87.7	100.0	1.4300000000000001e-276	753	close match
XP_016554211.1	KAS	XP_004243689.1	85.8	91.7	4.92e-310	848	close match
XP_016554212.1	KAS	XP_004243689.1	86.6	102.4	0.0	919	close match
XP_016559523.1	KAS	XP_004233105.1	93.2	100.0	0.0	871	close match
XP_016562487.2	KAS	XP_010318480.1	91.9	100.2	0.0	1015	close match
XP_016565864.2	KAS	XP_004245499.1	85.9	100.0	4.58e-308	838	close match
XP_016580597.2	KAS	XP_004243502.1	91.6	98.9	1.71e-314	852	close match
XP_016580598.2	KAS	XP_004243502.1	86.7	104.3	2.82e-293	798	close match
XP_016580599.2	KAS	XP_010323849.1	93.7	100.0	1.69e-293	796	close match
XP_016580601.2	KAS	XP_010323849.1	93.7	100.0	1.69e-293	796	close match
XP_047258241.1	KAS	XP_004244796.1	82.5	94.2	2.75e-110	323	close match
XP_047259333.1	KAS	XP_004251704.1	61.1	109.7	4.8e-70	221	weaker match
XP_047259825.1	KAS	XP_004251704.1	79.1	103.4	5.66e-312	851	close match
XP_047260439.1	KAS	XP_004251704.1	79.1	103.4	5.66e-312	851	close match
XP_047269425.1	KAS	XP_004243502.1	92.3	100.0	5.11e-317	858	close match
XP_047269877.1	KAS	XP_004245499.1	89.6	100.0	1.39e-109	321	close match
XP_047269878.1	KAS	XP_004245499.1	87.3	100.0	9.1e-315	855	close match
XP_047269897.1	KAS	XP_004245499.1	86.5	100.0	4.12e-311	845	close match
NP_001311869.1	BCAT	XP_010323427.1	77.7	97.7	2.2700000000000002e-210	582	close match
XP_016560903.1	BCAT	XP_004231952.1	94.0	100.0	7.33e-288	781	close match
XP_016568949.1	BCAT	XP_010323427.1	77.7	97.9	2.0800000000000003e-208	577	close match
XP_047250607.1	BCAT	XP_004230170.1	93.1	99.8	0.0	956	close match
XP_047256925.1	BCAT	XP_010314598.1	84.7	100.3	4.24e-238	652	close match
XP_047257659.1	BCAT	XP_010323427.1	60.8	102.4	2.61e-127	369	weaker match
XP_047260758.1	BCAT	XP_010314598.1	83.3	100.0	2.22e-58	184	close match
XP_047267041.1	BCAT	XP_010323427.1	54.7	121.1	3.3599999999999997e-69	214	weaker match
NP_001312005.1	FAR/FabG	XP_004241501.1	90.1	100.3	8.229999999999999e-191	527	close match
XP_016540579.2	FAR/FabG	XP_004249129.1	83.0	93.8	1.93e-171	476	close match
XP_016543643.2	FAR/FabG	XP_004249319.1	84.4	101.4	7.8299999999999995e-171	474	close match
XP_016543644.1	FAR/FabG	XP_004249318.1	84.1	102.5	1.02e-171	476	close match
XP_016543953.1	FAR/FabG	XP_004249129.1	89.3	97.6	9.930000000000001e-185	509	close match
XP_016549820.1	FAR/FabG	XP_019067123.1	82.3	100.0	1.79e-162	451	close match
XP_016549987.1	FAR/FabG	XP_004252844.1	83.0	100.4	8.8e-160	444	close match
XP_016552215.1	FAR/FabG	XP_004252379.1	83.0	100.0	4.6e-165	458	close match
XP_047257356.1	FAR/FabG	XP_004252379.1	77.1	100.4	3.36e-153	428	close match
XP_016546825.1	Pun1/CS-related	XP_004251378.4	62.1	100.7	2.16e-183	517	acyltransferase-family match
XP_016546866.1	Pun1/CS-related	XP_004251382.3	64.5	98.9	1.3499999999999998e-197	554	acyltransferase-family match
XP_016550707.1	Pun1/CS-related	XP_004252959.1	63.3	109.3	6.89e-132	384	acyltransferase-family match
XP_016556301.2	Pun1/CS-related	XP_004246214.1	41.9	95.0	2.2e-94	291	acyltransferase-family match
XP_016562171.1	Pun1/CS-related	XP_004251382.3	48.7	98.9	3e-125	371	acyltransferase-family match
XP_016574525.1	Pun1/CS-related	NP_001316321.1	65.3	102.1	2.14e-201	564	acyltransferase-family match
XP_016576957.1	Pun1/CS-related	XP_004251378.4	62.3	101.2	1.43e-193	543	acyltransferase-family match
XP_016579552.2	Pun1/CS-related	XP_025887953.1	81.6	99.7	7.89e-192	532	acyltransferase-family match

Pepper protein	Gene	Tomato protein	Identity (%)	Coverage (%)	E-value	Bit score	Match class
XP_047253533.1	Pun1/CS-related	XP_004249102.1	88.9	97.7	1.57e-248	680	acyltransferase-family match
XP_047257634.1	Pun1/CS-related	XP_004251378.4	47.1	116.2	6.599999999999999e-82	252	acyltransferase-family match
XP_047259156.1	Pun1/CS-related	XP_004251383.1	56.1	86.9	8.55e-105	309	acyltransferase-family match
XP_047260008.1	Pun1/CS-related	XP_004251378.4	66.3	101.6	2.0899999999999997e-89	268	acyltransferase-family match

Table S2. Sensitivity of the synteny result and the shuffled-position control. Pepper genes in blocks at maximum gaps of 5, 10 and 25 genes, with the gene index counting orthologs only, and the mean of 20 shuffled-position controls. Data: `data/tables/S02.csv`*; script:* `code/02_synteny.py`*.*

Comparison	Setting	Genes in blocks	Blocks	Mean genes in blocks (shuffled)	Pepper genes in blocks (%)
Tomato	max gap 5	13971	656		49.97
Tomato	max gap 5, positions shuffled (20 shuffles)			0	0.0
Tomato	max gap 10	14350	520		51.33
Tomato	max gap 10, positions shuffled (20 shuffles)			0	0.0
Tomato	max gap 25	14914	440		53.34
Tomato	max gap 25, positions shuffled (20 shuffles)			0	0.0
Tomato	max gap 10, orthologs only in the gene index	14702	462		80.21
Eggplant (identity 0.5 or more)	max gap 5	14249	700		50.97
Eggplant (identity 0.5 or more)	max gap 5, positions shuffled (20 shuffles)			0	0.0
Eggplant (identity 0.5 or more)	max gap 10	14712	523		52.62
Eggplant (identity 0.5 or more)	max gap 10, positions shuffled (20 shuffles)			0	0.0
Eggplant (identity 0.5 or more)	max gap 25	15308	458		54.75
Eggplant (identity 0.5 or more)	max gap 25, positions shuffled (20 shuffles)			0	0.0
Eggplant (identity 0.5 or more)	max gap 10, orthologs only in the gene index	15089	471		75.05

Table S3. Four ways of counting. A, pepper genes with any protein hit (homology); B, with a one-to-one ortholog; C, in a collinear block; D, orthologs in a block. Denominators are given in the table. Data: `data/tables/S03.csv`*; script:* `code/02_synteny.py`*.*

Organism	Way of counting	Genes	Denominator	Percent
tomato	A. pepper genes with any protein hit (E<=1e-10): homology only, not synteny	26432	27958	94.5
tomato	B. pepper genes with a one-to-one (RBH) ortholog	18330	27958	65.6
tomato	C. pepper genes inside a collinear block (main definition)	14350	27958	51.3
tomato	D. RBH orthologs that lie inside a collinear block	14350	18330	78.3
tomato	C at max gap 5 genes (sensitivity)	13971	27958	49.97
tomato	C at max gap 10 genes (sensitivity)	14350	27958	51.33
tomato	C at max gap 25 genes (sensitivity)	14914	27958	53.34
tomato	null: positions shuffled, max gap 5 (mean of 20)	0	27958	0.0
tomato	null: positions shuffled, max gap 10 (mean of 20)	0	27958	0.0
tomato	null: positions shuffled, max gap 25 (mean of 20)	0	27958	0.0
eggplant	A. pepper genes with any protein hit (E<=1e-10): homology only, not synteny	26957	27958	96.4
eggplant	B. pepper genes with a one-to-one (RBH) ortholog	20104	27958	71.9
eggplant	C. pepper genes inside a collinear block (main definition)	14712	27958	52.6
eggplant	D. RBH orthologs that lie inside a collinear block	14712	20104	73.2
eggplant	C at max gap 5 genes (sensitivity)	14249	27958	50.97
eggplant	C at max gap 10 genes (sensitivity)	14712	27958	52.62
eggplant	C at max gap 25 genes (sensitivity)	15308	27958	54.75
eggplant	null: positions shuffled, max gap 5 (mean of 20)	0	27958	0.0
eggplant	null: positions shuffled, max gap 10 (mean of 20)	0	27958	0.0
eggplant	null: positions shuffled, max gap 25 (mean of 20)	0	27958	0.0

Table S4. Pathway proteins in the synteny analysis. Of the 97 pathway proteins, how many lie on chromosomes, have an ortholog and lie in a block; mean share of the ten flanking genes inside a block. Data: `data/tables/S04.csv`*; script:* `code/02_synteny.py`*.*

Measure	Value
pathway proteins in the list	97
of which on unplaced scaffolds (excluded from chromosome analysis)	16
of which on the 12 chromosomes	81
unique pepper genes on chromosomes	68
tomato RBH ortholog found	59
tomato RBH ortholog inside a collinear block	46
eggplant placed-locus RBH found	59
eggplant RBH inside a collinear block	49

Measure	Value
mean % of the 10 flanking genes (+/-5) inside a tomato block	51.5
mean % of the 10 flanking genes (+/-5) inside an eggplant block	52.5

Table S5. Pathway proteins and collinear blocks. For each pathway protein on a chromosome: whether its tomato and eggplant ortholog lies in a block and the share of its ten flanking genes inside a block. Data: `data/tables/S05.csv`*; script:* `code/02_synteny.py`.

Pepper protein	Gene	Pepper chromosome	Tomato ortholog in a block	Eggplant ortholog in a block	Flanking genes in tomato blocks (%)	Flanking genes in eggplant blocks (%)
NP_001311511.1	COMT	NC_061118.1	yes	yes	50.0	50.0
NP_001311532.1	PAL	NC_061119.1	yes	yes	30.0	30.0
NP_001311587.1	KAS	NC_061111.1	no	no	20.0	20.0
NP_001311756.1	HCT	NC_061113.1	yes	yes	60.0	80.0
NP_001311759.1	4CL	NC_061113.1	yes	yes	60.0	60.0
NP_001311771.1	KAS	NC_061113.1	yes	yes	40.0	50.0
NP_001311774.1	COMT	NC_061113.1	yes	yes	20.0	30.0
NP_001311869.1	BCAT	NC_061114.1	no	yes	70.0	70.0
NP_001311984.1	C4H	NC_061116.1	no	yes	80.0	80.0
NP_001312005.1	FAR/FabG	NC_061116.1	yes	yes	60.0	70.0
XP_016540288.2	HCT	NC_061119.1	no ortholog	no ortholog	90.0	70.0
XP_016540579.2	FAR/FabG	NC_061111.1	no ortholog	no ortholog	60.0	60.0
XP_016543643.2	FAR/FabG	NC_061120.1	yes	yes	60.0	60.0
XP_016543644.1	FAR/FabG	NC_061120.1	yes	yes	70.0	70.0
XP_016543953.1	FAR/FabG	NC_061120.1	yes	yes	90.0	90.0
XP_016545240.2	4CL	NC_061120.1	no	no	30.0	20.0
XP_016545476.1	COMT	NC_061120.1	yes	yes	60.0	60.0
XP_016546636.1	HCT	NC_061121.1	yes	yes	60.0	70.0
XP_016548258.1	HCT	NC_061121.1	yes	yes	30.0	30.0
XP_016549820.1	FAR/FabG	NC_061122.1	yes	yes	40.0	50.0
XP_016549987.1	FAR/FabG	NC_061122.1	yes	yes	30.0	40.0
XP_016550707.1	Pun1/CS-related	NC_061122.1	yes	yes	60.0	60.0
XP_016551171.2	HCT	NC_061111.1	yes	yes	50.0	50.0
XP_016552215.1	FAR/FabG	NC_061122.1	no	no	0.0	0.0
XP_016553826.1	COMT	NC_061114.1	yes	yes	70.0	70.0
XP_016553827.1	COMT	NC_061114.1	yes	yes	70.0	70.0
XP_016554211.1	KAS	NC_061117.1	no	no	0.0	0.0
XP_016554212.1	KAS	NC_061117.1	no	no	0.0	0.0
XP_016554784.1	PAL	NC_061120.1	yes	yes	70.0	70.0
XP_016556301.2	Pun1/CS-related	NC_061116.1	no	no ortholog	70.0	60.0
XP_016559523.1	KAS	NC_061112.1	yes	yes	60.0	60.0
XP_016560903.1	BCAT	NC_061112.1	yes	yes	70.0	70.0
XP_016561061.1	COMT	NC_061112.1	yes	yes	50.0	50.0
XP_016561062.2	COMT	NC_061112.1	yes	yes	50.0	50.0
XP_016561063.1	COMT	NC_061112.1	no ortholog	no ortholog	60.0	60.0
XP_016561322.1	4CL	NC_061112.1	yes	yes	30.0	50.0
XP_016561323.1	4CL	NC_061112.1	yes	yes	30.0	50.0
XP_016562171.1	Pun1/CS-related	NC_061112.1	no	no	90.0	60.0
XP_016562487.2	KAS	NC_061113.1	yes	yes	60.0	60.0
XP_016562978.1	HCT	NC_061113.1	yes	yes	60.0	80.0
XP_016562979.1	HCT	NC_061113.1	yes	yes	60.0	80.0

Pepper protein	Gene	Pepper chromosome	Tomato ortholog in a block	Eggplant ortholog in a block	Flanking genes in tomato blocks (%)	Flanking genes in eggplant blocks (%)
XP_016565628.1	PAL	NC_061113.1	no ortholog	no ortholog	50.0	40.0
XP_016565864.2	KAS	NC_061111.1	no ortholog	no ortholog	60.0	60.0
XP_016568949.1	BCAT	NC_061114.1	no	yes	70.0	70.0
XP_016570346.1	COMT	NC_061114.1	yes	yes	50.0	50.0
XP_016570682.1	4CL	NC_061114.1	yes	yes	70.0	80.0
XP_016571691.1	4CL	NC_061111.1	yes	yes	60.0	60.0
XP_016571727.2	COMT	NC_061114.1	no ortholog	no ortholog	70.0	60.0
XP_016573925.1	PAL	NC_061115.1	yes	yes	90.0	90.0
XP_016574525.1	Pun1/CS-related	NC_061115.1	yes	no ortholog	50.0	10.0
XP_016576832.2	COMT	NC_061116.1	no ortholog	no ortholog	60.0	60.0
XP_016578426.2	4CL	NC_061116.1	yes	yes	90.0	90.0
XP_016579552.2	Pun1/CS-related	NC_061117.1	yes	yes	70.0	70.0
XP_016579621.2	4CL	NC_061117.1	no	no	80.0	80.0
XP_016579751.1	HCT	NC_061117.1	yes	yes	30.0	40.0
XP_016580597.2	KAS	NC_061117.1	no ortholog	no ortholog	30.0	30.0
XP_016580598.2	KAS	NC_061117.1	no ortholog	no ortholog	30.0	30.0
XP_016580599.2	KAS	NC_061117.1	no ortholog	no ortholog	30.0	30.0
XP_016580601.2	KAS	NC_061117.1	no ortholog	no ortholog	30.0	30.0
XP_047250607.1	BCAT	NC_061118.1	yes	yes	80.0	80.0
XP_047251840.1	PAL	NC_061119.1	yes	yes	30.0	30.0
XP_047251841.1	PAL	NC_061119.1	yes	yes	30.0	30.0
XP_047253533.1	Pun1/CS-related	NC_061120.1	yes	yes	70.0	60.0
XP_047256925.1	BCAT	NC_061122.1	yes	yes	40.0	40.0
XP_047257356.1	FAR/FabG	NC_061122.1	no ortholog	no ortholog	20.0	30.0
XP_047257634.1	Pun1/CS-related	NC_061122.1	no ortholog	no ortholog	30.0	40.0
XP_047257659.1	BCAT	NC_061122.1	no ortholog	no ortholog	50.0	60.0
XP_047264644.1	COMT	NC_061113.1	no ortholog	no ortholog	40.0	40.0
XP_047265503.1	4CL	NC_061113.1	yes	yes	80.0	80.0
XP_047265504.1	4CL	NC_061113.1	yes	yes	80.0	80.0
XP_047265606.1	4CL	NC_061113.1	yes	yes	70.0	70.0
XP_047266718.1	COMT	NC_061114.1	no ortholog	yes	20.0	10.0
XP_047267041.1	BCAT	NC_061114.1	no ortholog	no ortholog	60.0	80.0
XP_047269425.1	KAS	NC_061116.1	no	no	0.0	0.0
XP_047269877.1	KAS	NC_061111.1	no ortholog	no ortholog	40.0	40.0
XP_047269878.1	KAS	NC_061111.1	no	no	40.0	40.0
XP_047269897.1	KAS	NC_061111.1	no ortholog	no ortholog	50.0	50.0
XP_047270260.1	COMT	NC_061116.1	yes	no ortholog	50.0	60.0
XP_047270582.1	COMT	NC_061116.1	no ortholog	yes	60.0	50.0
XP_047270907.1	COMT	NC_061116.1	no ortholog	no ortholog	60.0	60.0
XP_047270933.1	COMT	NC_061116.1	no ortholog	no	30.0	30.0

Table S6. All synteny comparisons. Every comparison with its assembly, annotation basis, identity filter, one-to-one orthologs, blocks, any-hit share and the second method. Data: `data/tables/S06.csv`*; script:* `code/02_synteny.py`*.*

Comparison	Entry	Species	Assembly	Basis	Identity filter	Role	Genes with 1:1 ortholog (%)	Gap 5 (%)	Gap 10 (%)	Gap 25 (%)	Blocks (gap 10)	Orthologs in blocks (%)	Genes with any hit (%)	Secor meth gap 1 (%)
tomato	Tomato	Solanum lycopersicum	GCF_000188115.5	RefSeq annotation		main	65.56	49.97	51.33	53.34	520	78.3	94.54	51.37
eggplant_id0.5	Eggplant	Solanum melongena	GCA_057556415.1	pepper-protein placement	0.5	main	71.91	50.97	52.62	54.75	523	73.18	96.43	50.1
eggplant_id0.8	Eggplant	Solanum melongena	GCA_057556415.1	pepper-protein placement	0.8	strict filter	54.92		42.49		469	77.37		44.8
potato_hera	Potato	Solanum tuberosum (Hera hap1)	GCA_020169585.1	submitter annotation, 12 chromosomes		main	61.89	44.32	46.34	48.14	612	74.9	95.98	46.34
potato_soltub3	Potato	Solanum tuberosum (SolTub_3.0)	GCF_000226075.1	RefSeq annotation, scaffolds with at least 50 genes		lower bound	49.12	39.9	40.88	41.75	438	83.2	90.42	40.9
tabacum	N. benthamiana (model host)	Nicotiana tabacum (stand-in)	GCF_000715075.1	RefSeq annotation		stand-in	64.08	30.73	39.81	44.27	826	62.1	95.41	39.91
benthamiana_id0.5	N. benthamiana (model host)	Nicotiana benthamiana	GCA_034376525.1	pepper-protein placement	0.5	main	68.11	32.85	40.56	44.92	857	59.6	95.48	
benthamiana_id0.8	N. benthamiana (model host)	Nicotiana benthamiana	GCA_034376525.1	pepper-protein placement	0.8	strict filter	45.61	17.29	25.36	30.63	649	55.6	80.15	
petunia_id0.5	Petunia	Petunia x hybrida	GCA_046563445.1	pepper-protein placement	0.5	main	67.91	43.0	45.46	47.56	539	67.0	95.3	
petunia_id0.8	Petunia	Petunia x hybrida	GCA_046563445.1	pepper-protein placement	0.8	strict filter	41.12	25.21	28.37	30.31	408	69.0	77.33	
strawberry_vesca	Strawberry	Fragaria vesca (diploid stand-in)	GCF_000184155.1	RefSeq annotation		stand-in	43.17	14.11	18.77	21.86	468	43.5	88.54	18.75
strawberry_ananassa	Strawberry	Fragaria x ananassa (octoploid)	GCA_050043575.1	submitter annotation		main	42.83	1.14	4.64	12.77	201	10.8	87.66	4.36
apple	Apple	Malus domestica	GCF_042453785.1	RefSeq annotation		main	46.91	9.58	16.35	22.53	516	34.9	91.5	16.38
cucumber	Cucumber	Cucumis sativus	GCF_000004075.3	RefSeq annotation		main	44.21	8.6	13.99	19.02	437	31.6	87.03	14.05
melon	Melon	Cucumis melo	GCF_025177605.1	RefSeq annotation		main	44.22	8.47	14.02	19.11	441	31.7	89.21	14.05
lettuce	Lettuce	Lactuca sativa	GCF_002870075.5	RefSeq annotation		main	47.37	5.46	11.84	18.94	425	25.0	90.84	11.82
carrot	Carrot	Daucus carota	GCF_001625215.2	RefSeq annotation		main	47.71	5.45	10.66	18.35	374	22.3	91.96	10.69
rice	Rice	Oryza sativa	GCF_034140825.1	RefSeq annotation		main	39.22	0.11	0.31	1.24	16	0.8	85.69	0.33
maize	Corn	Zea mays	GCF_902167145.1	RefSeq annotation		main	38.58	0.0	0.11	0.35	5	0.3	84.03	0.09
brassica	Brassica and Broccoli	Brassica oleracea	GCF_000695525.1	RefSeq annotation		main	43.52	0.47	1.48	4.95	72	3.4	89.1	

Table S7. Target peroxidase genes. Gene models of the three tomato and three eggplant peroxidases used for guide design, with locus, coding length and model checks. Data: `data/tables/S07.csv`; script: `code/04_guides.py`.

Organism	Gene	Query protein	Contig (strand)	Exons	Locus	Coding bases	Residues	Model source	Identity to query (%)	ATG start	Stop at end	Internal stops
tomato	SlPrx1	NP_001233970.2	NC_015438.3 (+)	3	93,248,667-93,250,066	999	332	RefSeq GFF	100.0 (RefSeq protein)	yes	yes	0
tomato	SlPrx72	NP_001296734.1	NC_015440.3 (-)	4	8,111,393-8,113,910	996	331	RefSeq GFF	100.0 (RefSeq protein)	yes	yes	0
tomato	SlPrx34	NP_001234644.2	NC_015444.3 (+)	4	59,079,353-59,080,603	996	331	RefSeq GFF	100.0 (RefSeq protein)	yes	yes	0
eggplant	SmPrx1	NP_001233970.2	CM172108.1 (-)	3	5,541,420-5,542,745	996	331	miniprot rank-1	91.6 (placement identity)	yes	yes	0
eggplant	SmPrx34	NP_001234644.2	CM172110.1 (-)	4	34,752,686-34,754,118	996	331	miniprot rank-1	93.3 (placement identity)	yes	yes	0
eggplant	SmPrx72	NP_001296734.1	CM172110.1 (+)	4	15,347,172-15,351,122	999	332	miniprot rank-1	89.8 (placement identity)	yes	yes	0

Table S8. Candidate guides per gene. Candidates, candidates with exactly one perfect genome site, candidates cutting in the first 60% of the coding sequence and guides selected. Data: `data/tables/S08.csv`; script: `code/04_guides.py`.

Gene	Candidates	With exactly one perfect site	Cut in first 60% of coding sequence	Guides selected
SlPrx1	62	62	32	3
SlPrx34	47	17	12	3
SlPrx72	38	36	19	3
SmPrx1	56	56	32	3
SmPrx34	48	37	24	3
SmPrx72	42	42	14	3

Table S9. All 293 candidate guides. Every candidate with cut position, GC content, genome sites at 0 to 4 mismatches and MIT score; the selected guides are named. Data: `data/tables/S09.csv`; script: `code/04_guides.py`.

Gene	Candidate 20-mer	PAM	Cut after coding base	GC (%)	Perfect sites	1 mm	2 mm	3 mm	4 mm	MIT score	Selected as
SlPrx1	GTGACTATGGTATTCGGCCA	AGG	91	50	1	0	0	0	4	100.0	
SlPrx1	TGACTATGGTATTCGGCCAA	GGG	92	45	1	0	0	0	22	99.9	
SlPrx1	AAAATCCAACACGAGTCCCT	TGG	97	45	1	0	0	0	13	100.0	
SlPrx1	TTCGGCCAAGGGACTCGTGT	TGG	103	60	1	0	1	1	8	100.0	
SlPrx1	TTCGGCCCTTGGGCACGTAC	TGG	123	65	1	0	0	0	0	100.0	SlPrx1-g1
SlPrx1	TTTATTCCAGTACGTGCCCA	AGG	128	45	1	0	0	0	3	100.0	
SlPrx1	TTATTCCAGTACGTGCCCAA	GGG	129	45	1	0	0	0	7	100.0	
SlPrx1	GAACTATGGATTCGGCCCTT	GGG	133	50	1	0	0	0	6	100.0	
SlPrx1	TGAACTATGGATTCGGCCCT	TGG	134	50	1	0	0	0	6	100.0	
SlPrx1	CCTCACCGTTGATTGAACTA	TGG	147	45	1	0	0	1	28	99.8	
SlPrx1	CCATAGTTCAATCAACGGTG	AGG	158	45	1	0	0	2	12	100.0	
SlPrx1	AACAATCCTGGTGCCACTGT	TGG	188	50	1	0	0	1	16	100.0	
SlPrx1	TCTGATCCAACAGTGGCACC	AGG	193	55	1	0	0	0	6	100.0	
SlPrx1	AAGTGCATTCTCAACAATCC	TGG	200	40	1	0	0	0	14	100.0	
SlPrx1	TGTTTCGTACAAGGTTGTGA	CGG	241	40	1	0	0	2	37	99.9	
SlPrx1	GCCAGTACCAGAAATGAGGA	TGG	255	50	1	0	0	1	11	99.9	
SlPrx1	GACGGTTCCATCCTCATTTC	TGG	259	50	1	0	0	0	7	100.0	SlPrx1-g3

Gene	Candidate 20-mer	PAM	Cut after coding base	GC (%)	Perfect sites	1 mm	2 mm	3 mm	4 mm	MIT score	Selected as
SlPrx1	CAGTGCCAGTACCAGAAATG	AGG	259	50	1	0	0	0	7	100.0	
SlPrx1	TCCATCCTCATTTCTGGTAC	TGG	265	45	1	0	0	0	41	99.9	
SlPrx1	CCTCTCAAATTGGAATTCGG	AGG	293	45	1	0	0	3	16	100.0	
SlPrx1	AACCTCGAATCCTCTCAAAT	TGG	303	40	1	0	0	0	16	100.0	
SlPrx1	CCTCCGAATTCCAATTTGAG	AGG	304	45	1	0	0	1	13	99.9	
SlPrx1	TTCCAATTTGAGAGGATTCG	AGG	312	40	1	0	0	1	26	100.0	
SlPrx1	CAAATTGAAGCTGTTTGTCC	TGG	355	40	1	0	0	4	30	99.9	
SlPrx1	TCAGCACATGAAACAACTCC	AGG	362	45	1	8	17	15	22	94.9	
SlPrx1	ATTGACCTGGTCTGTCCCCA	CGG	441	55	1	0	0	0	10	100.0	
SlPrx1	TTGACCTGGTCTGTCCCCAC	GGG	442	60	1	0	0	0	6	100.0	
SlPrx1	GTCCATCTCTGCGTCCCGTG	GGG	445	65	1	0	0	0	4	100.0	
SlPrx1	CGTCCATCTCTGCGTCCCGT	GGG	446	65	1	0	0	0	4	100.0	
SlPrx1	TCGTCCATCTCTGCGTCCCG	TGG	447	65	1	0	0	0	3	100.0	SlPrx1-g2
SlPrx1	GTCCCCACGGGACGCAGAGA	TGG	454	70	1	0	0	0	6	100.0	
SlPrx1	TCAGACACATCTAATCTGCC	AGG	493	45	1	0	0	1	43	100.0	
SlPrx1	GCTGAAGAATTGGCATGCTG	AGG	603	50	1	0	0	0	15	100.0	
SlPrx1	ATAGCCTGTAGCTGAAGAAT	TGG	613	40	1	0	0	1	8	100.0	
SlPrx1	CATGCCAATTCTTCAGCTAC	AGG	620	45	1	0	0	0	8	100.0	
SlPrx1	TACAATTTCAACTCCACTGG	TGG	646	40	1	0	0	3	17	99.9	
SlPrx1	TGAAGGGTCAGGGCCACCAG	TGG	648	65	1	0	0	0	4	100.0	
SlPrx1	TTGCATCTATTGAAGGGTCA	GGG	658	40	1	0	0	0	56	99.9	
SlPrx1	GTTGCATCTATTGAAGGGTC	AGG	659	45	1	0	0	3	20	99.9	
SlPrx1	TGCTTGAAGCTGAGAAAGAA	AGG	681	40	1	0	0	3	47	99.9	
SlPrx1	TTATGTCCACAAAACGGAGA	TGG	715	40	1	0	0	3	23	99.9	
SlPrx1	CGGAGATGGCTCGAAGCGTG	TGG	729	65	1	0	0	0	6	100.0	
SlPrx1	TGGCTCGAAGCGTGTGGCAC	TGG	735	65	1	0	0	1	3	100.0	
SlPrx1	AAGCGTGTGGCACTGGACAC	TGG	742	60	1	0	0	0	5	100.0	
SlPrx1	CCTCAAGTTAGAGAAATACG	AGG	774	40	1	0	0	0	30	99.9	
SlPrx1	CCTCGTATTTCTCTAACTTG	AGG	785	40	1	0	0	2	31	99.9	
SlPrx1	TCTCTAACTTGAGGAATGGT	CGG	794	40	1	0	0	1	27	99.9	
SlPrx1	CTCTAACTTGAGGAATGGTC	GGG	795	45	1	0	0	0	11	100.0	
SlPrx1	TCTAACTTGAGGAATGGTCG	GGG	796	45	1	0	0	7	23	99.9	
SlPrx1	GAGGAATGGTCGGGGAATTC	TGG	804	55	1	0	0	0	29	100.0	
SlPrx1	AAGCATCGGTCCACAATATC	TGG	823	45	1	0	0	0	8	100.0	
SlPrx1	TGGAATCAGACCAGATATTG	TGG	824	40	1	0	0	4	18	99.9	
SlPrx1	AAACACCTTGGTCGAAGCAT	CGG	837	45	1	0	0	1	20	100.0	
SlPrx1	GTGGACCGATGCTTCGACCA	AGG	843	60	1	0	0	0	9	100.0	
SlPrx1	ATACCTTTGGACAAACACCT	TGG	849	40	1	0	0	2	32	99.9	
SlPrx1	CGACCAAGGTGTTTGTCCAA	AGG	857	50	1	0	1	19	143	98.7	
SlPrx1	CCCTGAGGCCTAAATACCTT	TGG	862	50	1	0	0	2	4	100.0	
SlPrx1	TCCAAAGGTATTTAGGCCTC	AGG	872	45	1	0	0	2	7	100.0	
SlPrx1	CCAAAGGTATTTAGGCCTCA	GGG	873	45	1	0	0	3	56	99.9	
SlPrx1	CAAAGGTATTTAGGCCTCAG	GGG	874	45	1	0	0	1	15	100.0	
SlPrx1	TCAATCCAAGAAATCCCCTG	AGG	877	45	1	0	0	0	16	100.0	
SlPrx1	TTAGGCCTCAGGGGATTTCT	TGG	883	50	1	0	0	3	11	100.0	
SlPrx34	GGAATAGCGGGTTCTAGCTA	TGG	67	50	1	0	1	2	9	99.9	SlPrx34-g2
SlPrx34	CTGCAAGAGATGGAGCGTTT	GGG	157	50	1	2	1	2	33	99.1	
SlPrx34	GCTGCAAGAGATGGAGCGTT	TGG	158	55	1	2	3	4	12	99.5	
SlPrx34	AGCAATGCAGCTGCAAGAGA	TGG	167	50	1	0	1	2	30	99.9	SlPrx34-g3
SlPrx34	ATTTCCACGATTGCTTTGTC	AGG	209	40	3	6	2	9	32	94.6	
SlPrx34	TAGTGCTCGAAGTGAAGTTC	AGG	241	45	1	1	0	1	11	99.6	
SlPrx34	AAGCCTCTCAGCGTTTGATT	AGG	290	45	2	1	0	2	9	99.0	
SlPrx34	GTTCCTAATCAAACGCTGAG	AGG	298	45	3	0	0	3	14	99.9	

Gene	Candidate 20-mer	PAM	Cut after coding base	GC (%)	Perfect sites	1 mm	2 mm	3 mm	4 mm	MIT score	Selected as
SlPrx34	AGAGGCTTCTCATTCATTGA	TGG	316	40	1	1	5	0	44	99.6	
SlPrx34	GCAGTAGAAGCTGAGTGCCC	TGG	349	60	1	0	2	1	7	99.9	SlPrx34-g1
SlPrx34	CCGCACAAGAGACAACTCCA	GGG	355	55	2	3	9	6	12	96.6	
SlPrx34	TCCGCACAAGAGACAACTCC	AGG	356	55	2	4	10	19	22	95.8	
SlPrx34	CCCTGGAGTTGTCTCTTGTG	CGG	366	55	2	0	1	0	7	100.0	
SlPrx34	TTGTGCGGATATTGTTGCCT	TGG	381	45	1	2	2	7	10	98.6	
SlPrx34	AACAGAGTCTCTAGCAACCA	AGG	387	45	2	2	1	3	23	99.3	
SlPrx34	GGTTGCTAGAGACTCTGTTG	TGG	402	50	3	3	1	2	22	98.0	
SlPrx34	CAGTTGGAACCTTCCAGTAA	GGG	424	45	2	0	0	3	13	100.0	
SlPrx34	CCAGTTGGAACCTTCCAGTA	AGG	425	50	2	0	0	4	20	99.9	
SlPrx34	CCTTACTGGAAGGTTCCAAC	TGG	436	50	2	0	0	10	32	99.9	
SlPrx34	CTCCCATCTCTTCTACCAGT	TGG	440	50	1	3	4	3	26	96.8	
SlPrx34	GTTCCAACTGGTAGAAGAGA	TGG	448	45	3	4	2	6	24	96.3	
SlPrx34	TTCCAACTGGTAGAAGAGAT	GGG	449	40	1	0	0	2	31	99.9	
SlPrx34	TGGGAGAATATCAAACGCCT	CGG	468	45	1	1	1	0	11	99.0	
SlPrx34	GATGTTTGCCAAGGCTTCCG	AGG	474	55	3	0	0	1	2	100.0	
SlPrx34	ATCAAACGCCTCGGAAGCCT	TGG	477	55	3	0	0	1	27	99.9	
SlPrx34	CGGAGGAGGGATGTTTGCCA	AGG	483	60	2	0	1	1	13	99.9	
SlPrx34	AAAAGTTACTTGTCGGAGGA	GGG	496	40	2	0	1	4	30	99.9	
SlPrx34	GAAAAGTTACTTGTCGGAGG	AGG	497	45	2	0	4	3	27	99.8	
SlPrx34	CTGGAAAAGTTACTTGTCGG	AGG	500	45	2	0	0	4	7	100.0	
SlPrx34	GGCAAAAGACGTCTGGAGAC	TGG	519	55	3	0	0	0	5	100.0	
SlPrx34	CCTTGCTGGCAAAAGACGTC	TGG	526	55	3	0	0	4	8	100.0	
SlPrx34	GGGTCTTGACCTAAAAGACT	TGG	558	45	4	5	0	2	19	98.0	
SlPrx34	CGGGCAATGAGAGACTCCAA	TGG	591	55	1	1	3	0	7	98.9	
SlPrx34	GCATATTCGCTGTCTAGAGA	TGG	665	45	3	5	1	3	14	96.9	
SlPrx34	AGGATCCATTTCGACAATTG	TGG	732	40	3	0	0	4	18	100.0	
SlPrx34	CAACACCACAATTGTCGAAA	TGG	738	40	3	0	1	4	54	99.8	
SlPrx34	CAGCTTGTGCTCAGAAGAAA	AGG	799	45	2	0	0	2	17	99.9	
SlPrx34	AGCTTGTGCTCAGAAGAAAA	GGG	800	40	2	0	0	6	23	99.9	
SlPrx34	TTGTCAAGGCTGCATCAGAT	TGG	817	45	3	0	2	2	12	99.9	
SlPrx34	TGTTGTCGCACTTGTTGTCA	AGG	831	45	3	0	2	4	23	99.9	
SlPrx34	TCACTGATCCTTGTACTAGC	TGG	868	45	1	0	0	10	362	99.6	
SlPrx34	TTCATCAACCAGCTAGTACA	AGG	871	40	1	2	0	3	20	99.3	
SlPrx34	TGCCGAATTTGGTTTAGCAA	TGG	912	40	1	0	1	3	29	99.9	
SlPrx34	GGAAAGATTGAAGTGAAGAC	CGG	943	40	1	0	2	7	45	99.8	
SlPrx34	CCTAATCTCACCAGCAGAGC	CGG	951	55	2	1	0	0	9	99.8	
SlPrx34	GAAGTGAAGACCGGCTCTGC	TGG	952	60	1	0	4	0	8	99.9	
SlPrx34	CCGGCTCTGCTGGTGAGATT	AGG	962	60	2	1	0	3	26	99.0	
SlPrx72	AGTTGAGCATCTACCAAGTT	TGG	69	40	1	0	0	2	29	99.9	
SlPrx72	CGCACAATGGATTCAACATT	AGG	120	40	1	1	2	2	38	98.7	
SlPrx72	AAGAAAAGACGAAGAACAGC	CGG	186	40	1	0	1	2	58	99.8	
SlPrx72	GTGATAATAGCATCAACGGC	AGG	254	45	1	0	0	0	6	100.0	SlPrx72-g2
SlPrx72	TGATAATAGCATCAACGGCA	GGG	255	40	1	0	0	0	7	100.0	
SlPrx72	CCAGCCAATGAAAGATTATC	TGG	288	40	1	1	0	0	19	99.4	
SlPrx72	CCAGATAATCTTTCATTGGC	TGG	299	40	1	1	1	0	17	99.7	
SlPrx72	AATCTTTCATTGGCTGGAGA	TGG	305	40	1	0	2	3	23	99.9	
SlPrx72	ACTTGGGATCGCATCAACTG	CGG	343	50	1	0	0	0	9	100.0	
SlPrx72	AATAACATCTCGAGTGGCTA	AGG	400	40	1	0	0	1	19	100.0	
SlPrx72	CACTGCATACCCCGGTCCAC	CGG	430	65	1	0	0	0	1	100.0	SlPrx72-g1
SlPrx72	CCCAATTCCACTGCATACCC	CGG	438	55	1	0	0	0	6	100.0	
SlPrx72	CGGTGGACCGGGGTATGCAG	TGG	442	70	1	0	0	0	1	100.0	
SlPrx72	ACCGGGGTATGCAGTGGAAT	TGG	448	55	1	0	0	0	6	100.0	

Gene	Candidate 20-mer	PAM	Cut after coding base	GC (%)	Perfect sites	1 mm	2 mm	3 mm	4 mm	MIT score	Selected as
SlPrx72	CCGGGGTATGCAGTGGAATT	GGG	449	55	1	0	0	0	31	100.0	
SlPrx72	CGGGGTATGCAGTGGAATTG	GGG	450	55	1	0	1	1	34	99.9	
SlPrx72	GTGGAATTGGGGAGATTAGA	TGG	461	45	1	0	1	3	51	99.9	
SlPrx72	AGCCTGATTTAAACCATGAG	AGG	553	40	1	0	1	2	22	99.9	
SlPrx72	TGCCTCTCATGGTTTAAATC	AGG	562	40	1	0	2	1	18	99.9	SlPrx72-g3
SlPrx72	TGTTGAGACTTGGATCCACT	GGG	665	45	1	0	1	5	72	99.8	
SlPrx72	TTGTTGAGACTTGGATCCAC	TGG	666	45	1	0	0	1	16	100.0	
SlPrx72	GCATACGTCTTGTTGAGACT	TGG	675	45	1	0	0	0	21	100.0	
SlPrx72	CCATGTTGATGGCTATTCTT	GGG	734	40	1	0	1	3	386	99.4	
SlPrx72	TCCATGTTGATGGCTATTCT	TGG	735	40	1	0	1	4	40	99.9	
SlPrx72	TGTTATTGGGTCCATGTTGA	TGG	745	40	1	0	0	6	60	99.8	
SlPrx72	CCCAAGAATAGCCATCAACA	TGG	745	45	1	0	1	1	18	99.9	
SlPrx72	CAAAAGCCCTAGGTGTTATT	GGG	758	40	1	0	0	2	31	99.9	
SlPrx72	TCAAAAGCCCTAGGTGTTAT	TGG	759	40	1	0	0	1	27	99.9	
SlPrx72	ACATGGACCCAATAACACCT	AGG	762	45	1	0	0	1	18	100.0	
SlPrx72	CATGGACCCAATAACACCTA	GGG	763	45	1	0	0	2	11	100.0	
SlPrx72	TACACATTGTCAAAAGCCCT	AGG	768	40	1	0	0	2	31	100.0	
SlPrx72	CCAAAACTTGCAAAACGGAA	TGG	805	40	1	0	1	30	136	99.8	
SlPrx72	TTCACAGACCAAAGGTCTAA	AGG	854	40	1	0	1	2	26	99.9	
SlPrx72	ATGACTAAATTGGGCCGAGT	TGG	935	45	1	1	2	2	38	99.8	
SlPrx72	TACCCGTCTTCACACCAACT	CGG	938	50	1	1	0	0	5	99.0	
SlPrx72	GGGCCGAGTTGGTGTGAAGA	CGG	946	60	2	0	0	0	5	100.0	
SlPrx72	GGCCGAGTTGGTGTGAAGAC	GGG	947	60	2	0	0	2	87	100.0	
SlPrx72	GGTGTGAAGACGGGTAAGAA	TGG	956	50	1	0	1	1	15	100.0	
SmPrx1	GTGACTATGGTATTAGGCCA	AGG	89	45	1	0	0	2	16	100.0	
SmPrx1	TGACTATGGTATTAGGCCAA	GGG	90	40	1	0	0	1	23	99.9	
SmPrx1	AAAATCCAACACGAGTCCCT	TGG	95	45	1	0	0	1	37	100.0	
SmPrx1	TTAGGCCAAGGGACTCGTGT	TGG	101	55	1	0	0	3	64	100.0	
SmPrx1	TTCGGCTCTTGGACACGTAC	TGG	121	55	1	0	0	0	1	100.0	SmPrx1-g1
SmPrx1	TGAACGATGGATTCGGCTCT	TGG	132	50	1	0	0	0	10	100.0	
SmPrx1	CGTTGATTGAACGATGGATT	CGG	139	40	1	0	0	4	25	99.9	
SmPrx1	TCTCACCGTTGATTGAACGA	TGG	145	45	1	0	0	0	4	100.0	
SmPrx1	CGAATCCATCGTTCAATCAA	CGG	151	40	1	0	0	4	29	99.9	
SmPrx1	AGCAATCCTGGTGCCACTGT	TGG	186	55	1	0	0	2	16	99.9	
SmPrx1	TCTGATCCAACAGTGGCACC	AGG	191	55	1	0	0	0	9	100.0	
SmPrx1	AAGTGCATTCTGAGCAATCC	TGG	198	45	1	0	0	0	13	100.0	
SmPrx1	AACCTCGGACGAAACAATCA	TGG	221	45	1	0	0	3	13	100.0	
SmPrx1	TTCCATGATTGTTTCGTCCG	AGG	230	45	1	0	0	0	13	100.0	
SmPrx1	GGATGGAACCATCACAACCT	CGG	236	50	1	0	0	6	1146	97.7	
SmPrx1	TGTTTCGTCCGAGGTTGTGA	TGG	239	50	1	0	0	2	18	99.9	
SmPrx1	GCCTGCACCAGCAATGAGGA	TGG	253	60	1	0	0	0	6	100.0	SmPrx1-g3
SmPrx1	CAGTGCCTGCACCAGCAATG	AGG	257	60	1	0	0	0	13	100.0	
SmPrx1	GATGGTTCCATCCTCATTGC	TGG	257	50	1	0	0	1	12	100.0	
SmPrx1	TCCATCCTCATTGCTGGTGC	AGG	263	55	1	0	0	1	9	100.0	
SmPrx1	CCTCTCAACAAGGAATTCGG	AGG	291	50	1	0	0	1	67	99.9	
SmPrx1	AACCTCGTATCCTCTCAACA	AGG	301	45	1	0	0	4	70	99.4	
SmPrx1	CCTCCGAATTCCTTGTTGAG	AGG	302	50	1	0	0	3	38	99.8	
SmPrx1	TTCCTTGTTGAGAGGATACG	AGG	310	45	1	0	0	2	14	100.0	
SmPrx1	TCAGCACATGAAACAACTCC	GGG	360	45	1	7	14	21	31	95.6	
SmPrx1	GTCAGCACATGAAACAACTC	CGG	361	45	1	0	0	3	19	99.9	
SmPrx1	ACTGACCTGGTCTGTGCCCA	CGG	439	60	1	0	0	2	19	99.9	
SmPrx1	CTGACCTGGTCTGTGCCCAC	GGG	440	65	1	0	0	0	1	100.0	SmPrx1-g2
SmPrx1	AGTCCATCTCTGCGTCCCGT	GGG	444	60	1	0	0	0	3	100.0	

Gene	Candidate 20-mer	PAM	Cut after coding base	GC (%)	Perfect sites	1 mm	2 mm	3 mm	4 mm	MIT score	Selected as
SmPrx1	TAGTCCATCTCTGCGTCCCG	TGG	445	60	1	0	0	1	14	100.0	
SmPrx1	GTGCCCACGGGACGCAGAGA	TGG	452	70	1	0	0	0	4	100.0	
SmPrx1	GCAGACACATCTAATCTGCC	AGG	491	50	1	0	0	0	7	100.0	
SmPrx1	GCTGAAGAATTGGCATGCTG	AGG	601	50	1	0	0	1	19	100.0	
SmPrx1	ATAGCCTGTAGCTGAAGAAT	TGG	611	40	1	0	0	0	35	99.9	
SmPrx1	CATGCCAATTCTTCAGCTAC	AGG	618	45	1	0	0	1	15	100.0	
SmPrx1	TGAAGGATCAGGGCCTCCAG	TGG	646	60	1	0	0	1	13	100.0	
SmPrx1	GATGCATCTATTGAAGGATC	AGG	657	40	1	0	0	4	56	99.8	
SmPrx1	GGACATAATGCTTGAAGCTG	AGG	687	45	1	0	0	1	27	99.9	
SmPrx1	CAAGCATTATGTCCCCAAAA	CGG	707	40	1	0	0	2	43	99.8	
SmPrx1	AAGCATTATGTCCCCAAAAC	GGG	708	40	1	0	0	1	16	100.0	
SmPrx1	AGCATTATGTCCCCAAAACG	GGG	709	45	1	0	0	1	13	99.9	
SmPrx1	GCTTCGAGGCATCCCCGTTT	TGG	710	60	1	0	0	1	26	99.9	
SmPrx1	GTCTAGTGCCACGCGCTTCG	AGG	724	65	1	0	0	0	2	100.0	
SmPrx1	AAGCGCGTGGCACTAGACAC	CGG	740	60	1	0	0	0	1	100.0	
SmPrx1	CCTCAAGTTAGAGAAATACG	AGG	772	40	1	0	0	6	33	99.9	
SmPrx1	CCTCGTATTTCTCTAACTTG	AGG	783	40	1	0	0	1	94	99.8	
SmPrx1	TCTCTAACTTGAGGAATGGT	CGG	792	40	1	0	0	5	75	99.8	
SmPrx1	CTCTAACTTGAGGAATGGTC	GGG	793	45	1	0	0	0	24	99.9	
SmPrx1	TCTAACTTGAGGAATGGTCG	GGG	794	45	1	0	0	3	54	99.9	
SmPrx1	GAGGAATGGTCGGGGAATTC	TGG	802	55	1	0	0	0	9	100.0	
SmPrx1	AAACACCTTGGTTGAAGCAT	CGG	835	40	1	0	0	10	145	99.6	
SmPrx1	GTGGACCGATGCTTCAACCA	AGG	841	55	1	0	0	0	9	100.0	
SmPrx1	CCCTGAGACCTAAATATCTT	TGG	860	40	1	0	0	5	55	99.8	
SmPrx1	TTAGGTCTCAGGGGATTTCT	TGG	881	45	1	0	0	1	21	99.9	
SmPrx1	TTGAGATTCGGAGTCGAATT	TGG	905	40	1	0	4	21	252	99.5	
SmPrx1	AGTCGAATTTGGAAGATCCA	TGG	916	40	1	0	0	6	30	99.9	
SmPrx34	GGAATAGTGGGTTCTAGCCA	TGG	68	50	1	0	5	5	15	99.8	
SmPrx34	AGTTGAGCTGCAGCTGACCA	TGG	74	55	1	1	4	0	7	99.5	
SmPrx34	AGCAAGGCAGGTGCAAGAGA	TGG	168	55	1	5	0	21	54	96.9	
SmPrx34	AAATGCATCCTGAGCAAGGC	AGG	180	50	1	0	0	0	9	100.0	SmPrx34-g2
SmPrx34	CTCTTGCACCTGCCTTGCTC	AGG	183	60	1	4	0	0	15	97.6	
SmPrx34	GTGGAAATGCATCCTGAGCA	AGG	184	50	4	2	0	4	10	99.5	
SmPrx34	ATTTCCACGATTGCTTTGTC	AGG	210	40	5	8	3	4	25	93.0	
SmPrx34	TCGTGCTCGAAGTGAAATTC	AGG	242	45	1	4	0	1	12	96.6	
SmPrx34	AAGCCTCTCAGCGTTTGATT	AGG	291	45	5	0	0	7	25	99.9	
SmPrx34	GTTCCTAATCAAACGCTGAG	AGG	299	45	5	0	0	5	12	100.0	
SmPrx34	GTAGTGGAAGCTGAATGCCC	TGG	350	55	1	4	7	3	13	98.2	
SmPrx34	CTGCGCATGACACAACTCCA	GGG	356	55	1	0	0	2	13	99.9	SmPrx34-g3
SmPrx34	TCTGCGCATGACACAACTCC	AGG	357	55	1	0	0	7	30	99.9	
SmPrx34	ATGCGCAGATATTGTTGCCT	TGG	382	45	1	0	3	1	21	99.7	
SmPrx34	AACTGAGTCTCTAGCAACCA	AGG	388	45	1	5	4	1	19	95.1	
SmPrx34	GGTTGCTAGAGACTCAGTTG	TGG	403	50	1	6	1	1	26	98.9	
SmPrx34	CCAGTTGGAACTTTCCAATA	AGG	426	40	1	0	8	4	29	99.8	
SmPrx34	CCTTATTGGAAAGTTCCAAC	TGG	437	40	1	0	3	6	27	99.9	
SmPrx34	CTTCCATCTCTTCTACCAGT	TGG	441	45	2	10	4	3	37	91.7	
SmPrx34	GTTCCAACTGGTAGAAGAGA	TGG	449	45	4	8	0	5	46	93.1	
SmPrx34	ATCGAACGCTTCAGAAGCTT	TGG	478	45	1	0	0	4	10	99.9	
SmPrx34	CTGGAAAAGTTACTCGTCGG	GGG	501	50	1	0	4	1	15	99.9	
SmPrx34	ACTGGAAAAGTTACTCGTCG	GGG	502	45	1	0	0	3	15	100.0	
SmPrx34	GACTGGAAAAGTTACTCGTC	GGG	503	45	1	0	0	0	8	100.0	SmPrx34-g1
SmPrx34	AGACTGGAAAAGTTACTCGT	CGG	504	40	1	0	4	1	31	99.9	
SmPrx34	GGCAAAAGATGTCTGGAGAC	TGG	520	50	1	0	0	1	23	100.0	

Gene	Candidate 20-mer	PAM	Cut after coding base	GC (%)	Perfect sites	1 mm	2 mm	3 mm	4 mm	MIT score	Selected as
SmPrx34	CTTTGCTGGCAAAAGATGTC	TGG	527	45	1	4	0	4	19	96.1	
SmPrx34	TTTCAGGTCAAGACCTTTGC	TGG	541	45	1	0	0	8	22	99.9	
SmPrx34	ACAACAGGACCAAGTCTTTC	AGG	557	45	1	0	0	5	19	99.9	
SmPrx34	AGGTCTTGACCTGAAAGACT	TGG	559	45	1	0	2	6	30	99.8	
SmPrx34	CGGGCAGTGAGAGACTCCAA	TGG	592	60	2	3	0	0	11	97.1	
SmPrx34	GCATATTCTCTGTCTAGAGA	TGG	666	40	4	1	8	4	46	99.0	
SmPrx34	AGGATCCATTTCGACAATAG	TGG	733	40	1	4	1	1	16	98.7	
SmPrx34	ACTATTGTCGAAATGGATCC	TGG	746	40	1	4	2	2	106	96.4	
SmPrx34	TCAAATTTACTGGAACTGCC	AGG	753	40	4	1	0	1	32	99.5	
SmPrx34	CAGCTTGTGCTTAAGAAAAG	AGG	800	40	1	0	0	8	30	99.8	
SmPrx34	TTGTCAAGGCTGCATCAGAT	TGG	818	45	5	2	0	1	56	97.9	
SmPrx34	TGTTGTCGCACTTGTTGTCA	AGG	832	45	5	0	0	2	49	99.9	
SmPrx34	TGAGTGATCCTTGTGCTAGC	TGG	869	50	1	0	4	1	19	99.9	
SmPrx34	TTCATCAACCAGCTAGCACA	AGG	872	45	1	0	1	3	19	99.9	
SmPrx34	CTCCATAGCTAGACCAAATT	CGG	904	40	1	0	1	13	269	99.3	
SmPrx34	TGCCGAATTTGGTCTAGCTA	TGG	913	45	1	0	1	7	64	99.9	
SmPrx34	TGGTCTAGCTATGGAGAAAA	TGG	922	40	1	0	1	6	87	99.5	
SmPrx34	GGTCTAGCTATGGAGAAAAT	GGG	923	40	1	0	1	4	65	99.7	
SmPrx34	GTCTAGCTATGGAGAAAATG	GGG	924	40	1	0	0	19	93	99.7	
SmPrx34	GGGAAGATTGAAGTCAAGAC	CGG	944	45	1	0	5	8	65	99.7	
SmPrx34	TCTGATCTCACCAGCAGAGC	CGG	952	55	1	0	1	4	5	99.9	
SmPrx34	GAAGTCAAGACCGGCTCTGC	TGG	953	60	1	0	9	0	5	99.8	
SmPrx72	CGCACAATGGATTCAACATT	AGG	122	40	1	0	1	5	29	99.7	SmPrx72-g3
SmPrx72	AAGAAGAGACGGAGAACAGC	AGG	188	50	1	1	0	1	41	99.4	
SmPrx72	GTGATAATAGCATCAACGGC	AGG	256	45	1	0	0	0	7	100.0	SmPrx72-g2
SmPrx72	CCAGCCAATGAAAGATTATC	TGG	290	40	1	0	1	5	35	99.9	
SmPrx72	CCAGATAATCTTTCATTGGC	TGG	301	40	1	1	1	1	39	99.4	
SmPrx72	AATCTTTCATTGGCTGGTGA	TGG	307	40	1	0	1	4	44	99.9	
SmPrx72	AATGACATCTCTGGTGGCTA	TGG	402	45	1	1	0	1	16	99.3	
SmPrx72	CAGTTGAATGACATCTCTGG	TGG	408	45	1	0	1	1	5	100.0	SmPrx72-g1
SmPrx72	CCCAATTCCACTGCATACCT	TGG	440	50	1	0	0	0	13	100.0	
SmPrx72	TGGTGGACCAAGGTATGCAG	TGG	444	55	1	0	0	0	11	100.0	
SmPrx72	ACCAAGGTATGCAGTGGAAT	TGG	450	45	1	0	0	0	21	99.9	
SmPrx72	CCAAGGTATGCAGTGGAATT	GGG	451	45	1	0	0	2	28	99.9	
SmPrx72	GGTATGCAGTGGAATTGGGA	AGG	455	50	1	0	0	4	23	99.9	
SmPrx72	GTGGAATTGGGAAGGTTAGA	TGG	463	45	1	0	1	4	89	99.8	
SmPrx72	TTGGGTCCACTGGGTTTGTA	GGG	658	50	1	0	0	1	32	99.9	
SmPrx72	CTTGGGTCCACTGGGTTTGT	AGG	659	55	1	0	1	2	99	99.9	
SmPrx72	CTTTAGCCCTACAAACCCAG	TGG	663	50	1	0	0	1	17	100.0	
SmPrx72	TGTTTAAACTTGGGTCCACT	GGG	667	40	1	0	0	0	30	99.9	
SmPrx72	TTGTTTAAACTTGGGTCCAC	TGG	668	40	1	0	0	1	17	100.0	
SmPrx72	GCATAGGCCTTGTTTAAACT	TGG	677	40	1	0	2	9	99	99.8	
SmPrx72	AGTGGACCCAAGTTTAAACA	AGG	681	40	1	0	2	4	36	99.9	
SmPrx72	TTGTTGTAATTGGGCTGCAT	AGG	693	40	1	0	0	2	49	99.9	
SmPrx72	CCATGTTGATGGCTATTCTT	GGG	736	40	1	0	1	4	29	99.9	
SmPrx72	TCCATGTTGATGGCTATTCT	TGG	737	40	1	0	2	3	223	99.6	
SmPrx72	CCCAAGAATAGCCATCAACA	TGG	747	45	1	0	1	5	22	99.9	
SmPrx72	TGTCATTGGGTCCATGTTGA	TGG	747	45	1	0	0	10	67	99.7	
SmPrx72	CAAAGGCCCGTGGTGTCATT	GGG	760	55	1	0	0	0	5	100.0	
SmPrx72	TCAAAGGCCCGTGGTGTCAT	TGG	761	55	1	0	0	1	7	100.0	
SmPrx72	ACATGGACCCAATGACACCA	CGG	764	50	1	0	0	1	20	100.0	
SmPrx72	CATGGACCCAATGACACCAC	GGG	765	55	1	0	0	2	15	100.0	
SmPrx72	TAGACATTGTCAAAGGCCCG	TGG	770	50	1	0	0	0	22	100.0	

Gene	Candidate 20-mer	PAM	Cut after coding base	GC (%)	Perfect sites	1 mm	2 mm	3 mm	4 mm	MIT score	Selected as
SmPrx72	CCAAAGGTCCAAAAGTACTG	TGG	864	45	1	0	0	5	126	99.9	
SmPrx72	TTCAAAGGCCTTAGAATTGC	TGG	885	40	1	0	0	3	35	99.8	
SmPrx72	TTTATGGGCCAGCAATTCTA	AGG	888	40	1	0	0	1	30	99.9	
SmPrx72	GACCCGGCCCAATTTAGTCA	TGG	924	55	1	1	0	0	9	99.6	
SmPrx72	ATGCCATGACTAAATTGGGC	CGG	932	45	1	0	0	2	12	99.9	
SmPrx72	TGCCATGACTAAATTGGGCC	GGG	933	50	1	0	1	0	13	99.9	
SmPrx72	ATGACTAAATTGGGCCGGGT	CGG	937	50	1	1	1	1	15	99.0	
SmPrx72	TGCCCATCTTGACACCGACC	CGG	940	60	1	0	0	0	4	100.0	
SmPrx72	GGGCCGGGTCGGTGTCAAGA	TGG	948	70	1	0	0	1	2	100.0	
SmPrx72	GGCCGGGTCGGTGTCAAGAT	GGG	949	65	1	0	0	1	6	100.0	
SmPrx72	GGAAATATTCGTAGAGACTG	TGG	979	40	1	0	0	0	19	100.0	

Table S10. Genome sites with 1 to 3 mismatches to the 18 guides. Every site of the whole-genome scan with one to three mismatches to a selected guide. Data: `data/tables/S10.csv`*; script:* `code/04_guides.py`*.*

Guide	Organism	Contig	Position of the 23-mer	Strand	Mismatches	Overlapping tomato gene
SlPrx72-g3	tomato	NC_015438.3	63,673,788	+	2	LOC101248706
SlPrx34-g1	tomato	NC_015439.3	55,426,138	+	2	LOC101267754
SlPrx72-g3	tomato	NC_015439.3	54,265,868	-	2	LOC101257228
SlPrx72-g3	tomato	NC_015440.3	9,919,641	-	3	
SlPrx34-g1	tomato	NC_015441.3	54,138,925	+	2	
SlPrx34-g2	tomato	NC_015441.3	41,791,749	+	3	
SlPrx34-g1	tomato	NC_015444.3	59,648,638	+	3	LOC101246070
SlPrx34-g2	tomato	NC_015444.3	59,648,277	+	2	LOC101246070
SlPrx34-g3	tomato	NC_015444.3	2,471,020	+	3	LOC101265042
SlPrx34-g3	tomato	NC_015444.3	59,648,388	-	2	LOC101246070
SlPrx34-g2	tomato	NC_015445.3	21,449,385	+	3	
SlPrx34-g3	tomato	NC_015448.3	3,191,297	+	3	LOC101259600
SmPrx72-g3	eggplant	CM172108.1	8,390,218	+	3	
SmPrx72-g3	eggplant	CM172108.1	18,305,557	-	3	
SmPrx72-g1	eggplant	CM172109.1	59,964,605	-	3	
SmPrx72-g1	eggplant	CM172109.1	71,340,375	-	2	
SmPrx72-g3	eggplant	CM172109.1	71,339,971	-	2	
SmPrx72-g3	eggplant	CM172112.1	11,702,563	+	3	
SmPrx34-g3	eggplant	CM172114.1	101,189,746	-	3	
SmPrx34-g3	eggplant	CM172117.1	18,655,076	-	3	
SmPrx72-g3	eggplant	CM172117.1	63,803,807	-	3	
SmPrx72-g3	eggplant	CM172118.1	102,265,438	-	3	

Table S11. Step capacities. TPM and capacity of every pathway step for the three datasets, stages and both transcript statistics. Data: `data/tables/S11.csv`*; script:* `code/05_design_model.py`*.*

Dataset	Stage	Statistic	Step	TPM	Capacity (units)
Tomato placenta	mature green	representative	PAL	167.419	105.94
Tomato placenta	mature green	representative	C4H	4.487	2.84
Tomato placenta	mature green	representative	4CL	7.892	4.99
Tomato placenta	mature green	representative	HCT	4.204	2.66
Tomato placenta	mature green	representative	CCoAOMT	68.171	43.14
Tomato placenta	mature green	representative	COMT	7.615	4.82

Dataset	Stage	Statistic	Step	TPM	Capacity (units)
Tomato placenta	mature green	representative	VR	72.645	45.97
Tomato placenta	mature green	representative	BCAT	6.474	4.10
Tomato placenta	mature green	representative	KAS	18.615	11.78
Tomato placenta	mature green	representative	FAS	39.412	24.94
Tomato placenta	mature green	representative	FAD2	1661.536	1051.41
Tomato placenta	mature green	representative	POX	365.870	231.52
Tomato placenta	mature green	family sum	PAL	233.505	147.76
Tomato placenta	mature green	family sum	C4H	7.993	5.06
Tomato placenta	mature green	family sum	4CL	10.355	6.55
Tomato placenta	mature green	family sum	HCT	11.850	7.50
Tomato placenta	mature green	family sum	CCoAOMT	134.065	84.84
Tomato placenta	mature green	family sum	COMT	41.257	26.11
Tomato placenta	mature green	family sum	VR	239.908	151.81
Tomato placenta	mature green	family sum	BCAT	56.902	36.01
Tomato placenta	mature green	family sum	KAS	95.368	60.35
Tomato placenta	mature green	family sum	FAS	40.252	25.47
Tomato placenta	mature green	family sum	FAD2	2185.761	1383.13
Tomato placenta	mature green	family sum	POX	772.892	489.08
Tomato placenta	breaker	representative	PAL	24.188	15.31
Tomato placenta	breaker	representative	C4H	13.570	8.59
Tomato placenta	breaker	representative	4CL	113.281	71.68
Tomato placenta	breaker	representative	HCT	222.073	140.53
Tomato placenta	breaker	representative	CCoAOMT	11.330	7.17
Tomato placenta	breaker	representative	COMT	53.985	34.16
Tomato placenta	breaker	representative	VR	589.053	372.75
Tomato placenta	breaker	representative	BCAT	564.589	357.27
Tomato placenta	breaker	representative	KAS	226.018	143.02
Tomato placenta	breaker	representative	FAS	42.703	27.02
Tomato placenta	breaker	representative	FAD2	244.638	154.81
Tomato placenta	breaker	representative	POX	39.068	24.72
Tomato placenta	breaker	family sum	PAL	178.466	112.93
Tomato placenta	breaker	family sum	C4H	16.885	10.68
Tomato placenta	breaker	family sum	4CL	114.038	72.16
Tomato placenta	breaker	family sum	HCT	224.805	142.26

Dataset	Stage	Statistic	Step	TPM	Capacity (units)
Tomato placenta	breaker	family sum	CCoAOMT	36.334	22.99
Tomato placenta	breaker	family sum	COMT	57.643	36.48
Tomato placenta	breaker	family sum	VR	751.223	475.37
Tomato placenta	breaker	family sum	BCAT	671.258	424.77
Tomato placenta	breaker	family sum	KAS	358.765	227.02
Tomato placenta	breaker	family sum	FAS	45.631	28.87
Tomato placenta	breaker	family sum	FAD2	416.427	263.51
Tomato placenta	breaker	family sum	POX	200.044	126.59
Tomato placenta	ripe	representative	PAL	15.731	9.95
Tomato placenta	ripe	representative	C4H	7.711	4.88
Tomato placenta	ripe	representative	4CL	126.867	80.28
Tomato placenta	ripe	representative	HCT	290.280	183.69
Tomato placenta	ripe	representative	CCoAOMT	6.507	4.12
Tomato placenta	ripe	representative	COMT	252.084	159.52
Tomato placenta	ripe	representative	VR	140.707	89.04
Tomato placenta	ripe	representative	BCAT	1653.257	1046.17
Tomato placenta	ripe	representative	KAS	303.648	192.15
Tomato placenta	ripe	representative	FAS	35.729	22.61
Tomato placenta	ripe	representative	FAD2	446.955	282.83
Tomato placenta	ripe	representative	POX	17.321	10.96
Tomato placenta	ripe	family sum	PAL	43.088	27.27
Tomato placenta	ripe	family sum	C4H	9.979	6.31
Tomato placenta	ripe	family sum	4CL	127.395	80.61
Tomato placenta	ripe	family sum	HCT	291.340	184.36
Tomato placenta	ripe	family sum	CCoAOMT	13.609	8.61
Tomato placenta	ripe	family sum	COMT	252.535	159.80
Tomato placenta	ripe	family sum	VR	315.518	199.66
Tomato placenta	ripe	family sum	BCAT	1745.283	1104.40
Tomato placenta	ripe	family sum	KAS	465.365	294.48
Tomato placenta	ripe	family sum	FAS	39.596	25.06
Tomato placenta	ripe	family sum	FAD2	483.529	305.97
Tomato placenta	ripe	family sum	POX	142.074	89.90
Strawberry fruit	unripe	representative	PAL	312.766	197.92
Strawberry fruit	unripe	representative	C4H	129.517	81.96

Dataset	Stage	Statistic	Step	TPM	Capacity (units)
Strawberry fruit	unripe	representative	4CL	102.113	64.62
Strawberry fruit	unripe	representative	HCT	8.432	5.34
Strawberry fruit	unripe	representative	CCoAOMT	126.895	80.30
Strawberry fruit	unripe	representative	COMT	1723.967	1090.91
Strawberry fruit	unripe	representative	VR	247.665	156.72
Strawberry fruit	unripe	representative	BCAT	123.237	77.98
Strawberry fruit	unripe	representative	KAS	149.569	94.65
Strawberry fruit	unripe	representative	FAS	107.515	68.03
Strawberry fruit	unripe	representative	FAD2	0.000	0.00
Strawberry fruit	unripe	representative	POX	4.451	2.82
Strawberry fruit	unripe	family sum	PAL	408.602	258.56
Strawberry fruit	unripe	family sum	C4H	129.517	81.96
Strawberry fruit	unripe	family sum	4CL	119.174	75.41
Strawberry fruit	unripe	family sum	HCT	12.798	8.10
Strawberry fruit	unripe	family sum	CCoAOMT	126.895	80.30
Strawberry fruit	unripe	family sum	COMT	1760.591	1114.09
Strawberry fruit	unripe	family sum	VR	627.047	396.79
Strawberry fruit	unripe	family sum	BCAT	250.815	158.71
Strawberry fruit	unripe	family sum	KAS	298.415	188.83
Strawberry fruit	unripe	family sum	FAS	208.799	132.13
Strawberry fruit	unripe	family sum	FAD2	0.000	0.00
Strawberry fruit	unripe	family sum	POX	194.972	123.38
Strawberry fruit	ripe	representative	PAL	170.348	107.80
Strawberry fruit	ripe	representative	C4H	163.140	103.23
Strawberry fruit	ripe	representative	4CL	68.353	43.25
Strawberry fruit	ripe	representative	HCT	14.210	8.99
Strawberry fruit	ripe	representative	CCoAOMT	149.970	94.90
Strawberry fruit	ripe	representative	COMT	721.144	456.34
Strawberry fruit	ripe	representative	VR	150.226	95.06
Strawberry fruit	ripe	representative	BCAT	93.139	58.94
Strawberry fruit	ripe	representative	KAS	64.763	40.98
Strawberry fruit	ripe	representative	FAS	79.143	50.08
Strawberry fruit	ripe	representative	FAD2	0.000	0.00
Strawberry fruit	ripe	representative	POX	164.237	103.93

Dataset	Stage	Statistic	Step	TPM	Capacity (units)
Strawberry fruit	ripe	family sum	PAL	284.929	180.30
Strawberry fruit	ripe	family sum	C4H	163.140	103.23
Strawberry fruit	ripe	family sum	4CL	146.221	92.53
Strawberry fruit	ripe	family sum	HCT	18.328	11.60
Strawberry fruit	ripe	family sum	CCoAOMT	149.970	94.90
Strawberry fruit	ripe	family sum	COMT	784.739	496.58
Strawberry fruit	ripe	family sum	VR	453.110	286.72
Strawberry fruit	ripe	family sum	BCAT	171.047	108.24
Strawberry fruit	ripe	family sum	KAS	140.599	88.97
Strawberry fruit	ripe	family sum	FAS	140.873	89.14
Strawberry fruit	ripe	family sum	FAD2	0.000	0.00
Strawberry fruit	ripe	family sum	POX	703.176	444.97
Eggplant fruit peel	fruit peel	representative	PAL	410.964	260.06
Eggplant fruit peel	fruit peel	representative	C4H	110.496	69.92
Eggplant fruit peel	fruit peel	representative	4CL	168.541	106.65
Eggplant fruit peel	fruit peel	representative	HCT	44.258	28.01
Eggplant fruit peel	fruit peel	representative	CCoAOMT	78.692	49.80
Eggplant fruit peel	fruit peel	representative	COMT	21.870	13.84
Eggplant fruit peel	fruit peel	representative	VR	52.731	33.37
Eggplant fruit peel	fruit peel	representative	BCAT	42.782	27.07
Eggplant fruit peel	fruit peel	representative	KAS	29.240	18.50
Eggplant fruit peel	fruit peel	representative	FAS	65.293	41.32
Eggplant fruit peel	fruit peel	representative	FAD2	102.599	64.92
Eggplant fruit peel	fruit peel	representative	POX	699.286	442.50
Eggplant fruit peel	fruit peel	family sum	PAL	625.644	395.90
Eggplant fruit peel	fruit peel	family sum	C4H	189.470	119.90
Eggplant fruit peel	fruit peel	family sum	4CL	174.021	110.12
Eggplant fruit peel	fruit peel	family sum	HCT	54.649	34.58
Eggplant fruit peel	fruit peel	family sum	CCoAOMT	117.960	74.64
Eggplant fruit peel	fruit peel	family sum	COMT	22.169	14.03
Eggplant fruit peel	fruit peel	family sum	VR	265.730	168.15
Eggplant fruit peel	fruit peel	family sum	BCAT	63.008	39.87
Eggplant fruit peel	fruit peel	family sum	KAS	101.043	63.94
Eggplant fruit peel	fruit peel	family sum	FAS	65.293	41.32

Dataset	Stage	Statistic	Step	TPM	Capacity (units)
Eggplant fruit peel	fruit peel	family sum	FAD2	102.634	64.95
Eggplant fruit peel	fruit peel	family sum	POX	1225.693	775.61

Table S12. Agreement of the recomputed output with the stored solver output. Recomputed J (smallest step capacity, with the acyl-branch limit) against the stored linear-programme optimum for 150 scenario rows. Data: `data/tables/S12.csv`*; script:* `code/05_design_model.py`*.*

Dataset	Stage	Statistic	Methylation enzyme	Added level x	Pun1 multiplier	Stored solver optimum	Recomputed J	Difference	Agree
Eggplant fruit peel	peel	family_sum	CCoAOMT	0.1	1.0	34.58	34.58	0.002	yes
Eggplant fruit peel	peel	family_sum	CCoAOMT	0.1	1.35	34.58	34.58	0.002	yes
Eggplant fruit peel	peel	family_sum	CCoAOMT	0.1	1.65	34.58	34.58	0.002	yes
Eggplant fruit peel	peel	family_sum	CCoAOMT	0.1	2.1	34.58	34.58	0.002	yes
Eggplant fruit peel	peel	family_sum	CCoAOMT	0.25	1.0	34.58	34.58	0.002	yes
Eggplant fruit peel	peel	family_sum	CCoAOMT	0.25	1.35	34.58	34.58	0.002	yes
Eggplant fruit peel	peel	family_sum	CCoAOMT	0.25	1.65	34.58	34.58	0.002	yes
Eggplant fruit peel	peel	family_sum	CCoAOMT	0.25	2.1	34.58	34.58	0.002	yes
Eggplant fruit peel	peel	family_sum	CCoAOMT	1.0	1.0	34.58	34.58	0.002	yes
Eggplant fruit peel	peel	family_sum	CCoAOMT	1.0	1.35	34.58	34.58	0.002	yes
Eggplant fruit peel	peel	family_sum	CCoAOMT	1.0	1.65	34.58	34.58	0.002	yes
Eggplant fruit peel	peel	family_sum	CCoAOMT	1.0	2.1	34.58	34.58	0.002	yes
Eggplant fruit peel	peel	family_sum	COMT	0.1	1.0	14.03	14.03	0.002	yes
Eggplant fruit peel	peel	family_sum	COMT	0.1	1.35	14.03	14.03	0.002	yes
Eggplant fruit peel	peel	family_sum	COMT	0.1	1.65	14.03	14.03	0.002	yes
Eggplant fruit peel	peel	family_sum	COMT	0.1	2.1	14.03	14.03	0.002	yes
Eggplant fruit peel	peel	family_sum	COMT	0.25	1.0	14.03	14.03	0.002	yes
Eggplant fruit peel	peel	family_sum	COMT	0.25	1.35	14.03	14.03	0.002	yes
Eggplant fruit peel	peel	family_sum	COMT	0.25	1.65	14.03	14.03	0.002	yes
Eggplant fruit peel	peel	family_sum	COMT	0.25	2.1	14.03	14.03	0.002	yes
Eggplant fruit peel	peel	family_sum	COMT	1.0	1.0	14.03	14.03	0.002	yes
Eggplant fruit peel	peel	family_sum	COMT	1.0	1.35	14.03	14.03	0.002	yes
Eggplant fruit peel	peel	family_sum	COMT	1.0	1.65	14.03	14.03	0.002	yes
Eggplant fruit peel	peel	family_sum	COMT	1.0	2.1	14.03	14.03	0.002	yes
Eggplant fruit peel	peel	representative	CCoAOMT	0.1	1.0	21.77	21.77	0.002	yes
Eggplant fruit peel	peel	representative	CCoAOMT	0.1	1.35	21.77	21.77	0.002	yes

Dataset	Stage	Statistic	Methylation enzyme	Added level x	Pun1 multiplier	Stored solver optimum	Recomputed J	Difference	Agree
Eggplant fruit peel	peel	representative	CCoAOMT	0.1	1.65	21.77	21.77	0.002	yes
Eggplant fruit peel	peel	representative	CCoAOMT	0.1	2.1	21.77	21.77	0.002	yes
Eggplant fruit peel	peel	representative	CCoAOMT	0.25	1.0	21.77	21.77	0.002	yes
Eggplant fruit peel	peel	representative	CCoAOMT	0.25	1.35	21.77	21.77	0.002	yes
Eggplant fruit peel	peel	representative	CCoAOMT	0.25	1.65	21.77	21.77	0.002	yes
Eggplant fruit peel	peel	representative	CCoAOMT	0.25	2.1	21.77	21.77	0.002	yes
Eggplant fruit peel	peel	representative	CCoAOMT	1.0	1.0	21.77	21.77	0.002	yes
Eggplant fruit peel	peel	representative	CCoAOMT	1.0	1.35	21.77	21.77	0.002	yes
Eggplant fruit peel	peel	representative	CCoAOMT	1.0	1.65	21.77	21.77	0.002	yes
Eggplant fruit peel	peel	representative	CCoAOMT	1.0	2.1	21.77	21.77	0.002	yes
Eggplant fruit peel	peel	representative	COMT	0.1	1.0	13.84	13.84	0.001	yes
Eggplant fruit peel	peel	representative	COMT	0.1	1.35	13.84	13.84	0.001	yes
Eggplant fruit peel	peel	representative	COMT	0.1	1.65	13.84	13.84	0.001	yes
Eggplant fruit peel	peel	representative	COMT	0.1	2.1	13.84	13.84	0.001	yes
Eggplant fruit peel	peel	representative	COMT	0.25	1.0	13.84	13.84	0.001	yes
Eggplant fruit peel	peel	representative	COMT	0.25	1.35	13.84	13.84	0.001	yes
Eggplant fruit peel	peel	representative	COMT	0.25	1.65	13.84	13.84	0.001	yes
Eggplant fruit peel	peel	representative	COMT	0.25	2.1	13.84	13.84	0.001	yes
Eggplant fruit peel	peel	representative	COMT	1.0	1.0	13.84	13.84	0.001	yes
Eggplant fruit peel	peel	representative	COMT	1.0	1.35	13.84	13.84	0.001	yes
Eggplant fruit peel	peel	representative	COMT	1.0	1.65	13.84	13.84	0.001	yes
Eggplant fruit peel	peel	representative	COMT	1.0	2.1	13.84	13.84	0.001	yes
Strawberry fruit	ripe	family_sum	CCoAOMT	1.0	1.0	11.60	11.60	0.002	yes
Strawberry fruit	ripe	family_sum	CCoAOMT	1.0	1.35	11.60	11.60	0.002	yes
Strawberry fruit	ripe	family_sum	CCoAOMT	1.0	2.1	11.60	11.60	0.002	yes
Strawberry fruit	ripe	family_sum	COMT	1.0	1.0	11.60	11.60	0.002	yes
Strawberry fruit	ripe	family_sum	COMT	1.0	1.35	11.60	11.60	0.002	yes
Strawberry fruit	ripe	family_sum	COMT	1.0	2.1	11.60	11.60	0.002	yes
Strawberry fruit	ripe	representative	CCoAOMT	1.0	1.0	8.99	8.99	0.002	yes
Strawberry fruit	ripe	representative	CCoAOMT	1.0	1.35	8.99	8.99	0.002	yes
Strawberry fruit	ripe	representative	CCoAOMT	1.0	2.1	8.99	8.99	0.002	yes
Strawberry fruit	ripe	representative	COMT	1.0	1.0	8.99	8.99	0.002	yes
Strawberry fruit	ripe	representative	COMT	1.0	1.35	8.99	8.99	0.002	yes
Strawberry fruit	ripe	representative	COMT	1.0	2.1	8.99	8.99	0.002	yes
Strawberry fruit	unripe	family_sum	CCoAOMT	1.0	1.0	8.10	8.10	0.002	yes
Strawberry fruit	unripe	family_sum	CCoAOMT	1.0	1.35	8.10	8.10	0.002	yes
Strawberry fruit	unripe	family_sum	CCoAOMT	1.0	2.1	8.10	8.10	0.002	yes
Strawberry fruit	unripe	family_sum	COMT	1.0	1.0	8.10	8.10	0.002	yes
Strawberry fruit	unripe	family_sum	COMT	1.0	1.35	8.10	8.10	0.002	yes

Dataset	Stage	Statistic	Methylation enzyme	Added level x	Pun1 multiplier	Stored solver optimum	Recomputed J	Difference	Agree
Strawberry fruit	unripe	family_sum	COMT	1.0	2.1	8.10	8.10	0.002	yes
Strawberry fruit	unripe	representative	CCoAOMT	1.0	1.0	5.34	5.34	0.004	yes
Strawberry fruit	unripe	representative	CCoAOMT	1.0	1.35	5.34	5.34	0.004	yes
Strawberry fruit	unripe	representative	CCoAOMT	1.0	2.1	5.34	5.34	0.004	yes
Strawberry fruit	unripe	representative	COMT	1.0	1.0	5.34	5.34	0.004	yes
Strawberry fruit	unripe	representative	COMT	1.0	1.35	5.34	5.34	0.004	yes
Strawberry fruit	unripe	representative	COMT	1.0	2.1	5.34	5.34	0.004	yes
Tomato placenta	br	family_sum	CCoAOMT	0.1	1.0	10.68	10.68	0.000	yes
Tomato placenta	br	family_sum	CCoAOMT	0.1	1.35	10.68	10.68	0.000	yes
Tomato placenta	br	family_sum	CCoAOMT	0.1	1.65	10.68	10.68	0.000	yes
Tomato placenta	br	family_sum	CCoAOMT	0.1	2.1	10.68	10.68	0.000	yes
Tomato placenta	br	family_sum	CCoAOMT	0.25	1.0	10.68	10.68	0.000	yes
Tomato placenta	br	family_sum	CCoAOMT	0.25	1.35	10.68	10.68	0.000	yes
Tomato placenta	br	family_sum	CCoAOMT	0.25	1.65	10.68	10.68	0.000	yes
Tomato placenta	br	family_sum	CCoAOMT	0.25	2.1	10.68	10.68	0.000	yes
Tomato placenta	br	family_sum	CCoAOMT	1.0	1.0	10.68	10.68	0.005	yes
Tomato placenta	br	family_sum	CCoAOMT	1.0	1.35	10.68	10.68	0.000	yes
Tomato placenta	br	family_sum	CCoAOMT	1.0	1.65	10.68	10.68	0.000	yes
Tomato placenta	br	family_sum	CCoAOMT	1.0	2.1	10.68	10.68	0.000	yes
Tomato placenta	br	family_sum	COMT	1.0	1.0	10.68	10.68	0.005	yes
Tomato placenta	br	representative	CCoAOMT	0.1	1.0	7.17	7.17	0.000	yes
Tomato placenta	br	representative	CCoAOMT	0.1	1.35	7.17	7.17	0.000	yes
Tomato placenta	br	representative	CCoAOMT	0.1	1.65	7.17	7.17	0.000	yes
Tomato placenta	br	representative	CCoAOMT	0.1	2.1	7.17	7.17	0.000	yes
Tomato placenta	br	representative	CCoAOMT	0.25	1.0	7.17	7.17	0.000	yes
Tomato placenta	br	representative	CCoAOMT	0.25	1.35	7.17	7.17	0.000	yes
Tomato placenta	br	representative	CCoAOMT	0.25	1.65	7.17	7.17	0.000	yes
Tomato placenta	br	representative	CCoAOMT	0.25	2.1	7.17	7.17	0.000	yes
Tomato placenta	br	representative	CCoAOMT	1.0	1.0	7.17	7.17	0.000	yes
Tomato placenta	br	representative	CCoAOMT	1.0	1.35	7.17	7.17	0.000	yes
Tomato placenta	br	representative	CCoAOMT	1.0	1.65	7.17	7.17	0.000	yes
Tomato placenta	br	representative	CCoAOMT	1.0	2.1	7.17	7.17	0.000	yes
Tomato placenta	br	representative	COMT	1.0	1.0	8.59	8.59	0.003	yes
Tomato placenta	mg	family_sum	CCoAOMT	0.1	1.0	5.06	5.06	0.000	yes
Tomato placenta	mg	family_sum	CCoAOMT	0.1	1.35	5.06	5.06	0.000	yes

Dataset	Stage	Statistic	Methylation enzyme	Added level x	Pun1 multiplier	Stored solver optimum	Recomputed J	Difference	Agree
Tomato placenta	mg	family_sum	CCoAOMT	0.1	1.65	5.06	5.06	0.000	yes
Tomato placenta	mg	family_sum	CCoAOMT	0.1	2.1	5.06	5.06	0.000	yes
Tomato placenta	mg	family_sum	CCoAOMT	0.25	1.0	5.06	5.06	0.000	yes
Tomato placenta	mg	family_sum	CCoAOMT	0.25	1.35	5.06	5.06	0.000	yes
Tomato placenta	mg	family_sum	CCoAOMT	0.25	1.65	5.06	5.06	0.000	yes
Tomato placenta	mg	family_sum	CCoAOMT	0.25	2.1	5.06	5.06	0.000	yes
Tomato placenta	mg	family_sum	CCoAOMT	1.0	1.0	5.06	5.06	0.002	yes
Tomato placenta	mg	family_sum	CCoAOMT	1.0	1.35	5.06	5.06	0.000	yes
Tomato placenta	mg	family_sum	CCoAOMT	1.0	1.65	5.06	5.06	0.000	yes
Tomato placenta	mg	family_sum	CCoAOMT	1.0	2.1	5.06	5.06	0.000	yes
Tomato placenta	mg	family_sum	COMT	1.0	1.0	5.06	5.06	0.002	yes
Tomato placenta	mg	representative	CCoAOMT	0.1	1.0	2.66	2.66	0.000	yes
Tomato placenta	mg	representative	CCoAOMT	0.1	1.35	2.66	2.66	0.000	yes
Tomato placenta	mg	representative	CCoAOMT	0.1	1.65	2.66	2.66	0.000	yes
Tomato placenta	mg	representative	CCoAOMT	0.1	2.1	2.66	2.66	0.000	yes
Tomato placenta	mg	representative	CCoAOMT	0.25	1.0	2.66	2.66	0.000	yes
Tomato placenta	mg	representative	CCoAOMT	0.25	1.35	2.66	2.66	0.000	yes
Tomato placenta	mg	representative	CCoAOMT	0.25	1.65	2.66	2.66	0.000	yes
Tomato placenta	mg	representative	CCoAOMT	0.25	2.1	2.66	2.66	0.000	yes
Tomato placenta	mg	representative	CCoAOMT	1.0	1.0	2.66	2.66	0.000	yes
Tomato placenta	mg	representative	CCoAOMT	1.0	1.35	2.66	2.66	0.000	yes
Tomato placenta	mg	representative	CCoAOMT	1.0	1.65	2.66	2.66	0.000	yes
Tomato placenta	mg	representative	CCoAOMT	1.0	2.1	2.66	2.66	0.000	yes
Tomato placenta	mg	representative	COMT	1.0	1.0	2.66	2.66	0.000	yes
Tomato placenta	ripe	family_sum	CCoAOMT	0.1	1.0	6.31	6.31	0.000	yes
Tomato placenta	ripe	family_sum	CCoAOMT	0.1	1.35	6.31	6.31	0.000	yes
Tomato placenta	ripe	family_sum	CCoAOMT	0.1	1.65	6.31	6.31	0.000	yes
Tomato placenta	ripe	family_sum	CCoAOMT	0.1	2.1	6.31	6.31	0.000	yes
Tomato placenta	ripe	family_sum	CCoAOMT	0.25	1.0	6.31	6.31	0.000	yes
Tomato placenta	ripe	family_sum	CCoAOMT	0.25	1.35	6.31	6.31	0.000	yes
Tomato placenta	ripe	family_sum	CCoAOMT	0.25	1.65	6.31	6.31	0.000	yes
Tomato placenta	ripe	family_sum	CCoAOMT	0.25	2.1	6.31	6.31	0.000	yes

Dataset	Stage	Statistic	Methylation enzyme	Added level x	Pun1 multiplier	Stored solver optimum	Recomputed J	Difference	Agree
Tomato placenta	ripe	family_sum	CCoAOMT	1.0	1.0	6.31	6.31	0.005	yes
Tomato placenta	ripe	family_sum	CCoAOMT	1.0	1.35	6.31	6.31	0.000	yes
Tomato placenta	ripe	family_sum	CCoAOMT	1.0	1.65	6.31	6.31	0.000	yes
Tomato placenta	ripe	family_sum	CCoAOMT	1.0	2.1	6.31	6.31	0.000	yes
Tomato placenta	ripe	family_sum	COMT	1.0	1.0	6.31	6.31	0.005	yes
Tomato placenta	ripe	representative	CCoAOMT	0.1	1.0	4.12	4.12	0.000	yes
Tomato placenta	ripe	representative	CCoAOMT	0.1	1.35	4.12	4.12	0.000	yes
Tomato placenta	ripe	representative	CCoAOMT	0.1	1.65	4.12	4.12	0.000	yes
Tomato placenta	ripe	representative	CCoAOMT	0.1	2.1	4.12	4.12	0.000	yes
Tomato placenta	ripe	representative	CCoAOMT	0.25	1.0	4.12	4.12	0.000	yes
Tomato placenta	ripe	representative	CCoAOMT	0.25	1.35	4.12	4.12	0.000	yes
Tomato placenta	ripe	representative	CCoAOMT	0.25	1.65	4.12	4.12	0.000	yes
Tomato placenta	ripe	representative	CCoAOMT	0.25	2.1	4.12	4.12	0.000	yes
Tomato placenta	ripe	representative	CCoAOMT	1.0	1.0	4.12	4.12	0.002	yes
Tomato placenta	ripe	representative	CCoAOMT	1.0	1.35	4.12	4.12	0.000	yes
Tomato placenta	ripe	representative	CCoAOMT	1.0	1.65	4.12	4.12	0.000	yes
Tomato placenta	ripe	representative	CCoAOMT	1.0	2.1	4.12	4.12	0.000	yes
Tomato placenta	ripe	representative	COMT	1.0	1.0	4.88	4.88	0.001	yes

Table S13. Output per replicate. Smallest step capacity of each replicate library (representative transcripts, CCoAOMT-type methylation). Data: `data/tables/S13.csv`*; script:* `code/05_design_model.py`*.*

Dataset	Stage	Replicate	Chain minimum (units)	Limiting step
Tomato placenta	mature green	1	1.86	C4H
Tomato placenta	mature green	2	3.51	HCT
Tomato placenta	mature green	3	1.89	HCT
Tomato placenta	breaker	1	5.04	methylation
Tomato placenta	breaker	2	6.82	C4H
Tomato placenta	breaker	3	7.45	methylation
Tomato placenta	ripe	1	4.58	C4H
Tomato placenta	ripe	2	2.89	methylation
Tomato placenta	ripe	3	4.67	methylation
Strawberry fruit	unripe	1	0.94	HCT
Strawberry fruit	unripe	2	0.95	HCT

Dataset	Stage	Replicate	Chain minimum (units)	Limiting step
Strawberry fruit	unripe	3	14.12	HCT
Strawberry fruit	ripe	1	21.68	HCT
Strawberry fruit	ripe	2	5.30	HCT
Strawberry fruit	ripe	3	0.00	HCT
Eggplant fruit peel	fruit peel	1	22.04	C4H
Eggplant fruit peel	fruit peel	2	5.78	methylation
Eggplant fruit peel	fruit peel	3	16.96	C4H
Eggplant fruit peel	fruit peel	4	31.26	VR
Eggplant fruit peel	fruit peel	5	28.36	HCT
Eggplant fruit peel	fruit peel	6	21.90	HCT

Table S14. RNA-seq runs and mapping rates. Read pairs processed and percent mapped for every run used by the model. Data: `data/tables/S14.csv`*; script:* `code/05_design_model.py`*.*

Dataset	Run	Read pairs processed	Mapped (%)	Salmon version
tomato placenta	SRR35858318	3000000	87.1	2.8.0
tomato placenta	SRR35858319	3000000	92.3	2.8.0
tomato placenta	SRR35858320	3000000	90.8	2.8.0
tomato placenta	SRR35858321	3000000	92.2	2.8.0
tomato placenta	SRR35858322	3000000	90.8	2.8.0
tomato placenta	SRR35858323	3000000	80.6	2.8.0
tomato placenta	SRR35858325	3000000	71.6	2.8.0
tomato placenta	SRR35858326	3000000	93.7	2.8.0
tomato placenta	SRR35858327	3000000	83.1	2.8.0
strawberry fruit	SRR31895405	3000000	79.3	2.8.0
strawberry fruit	SRR31895406	3000000	65.5	2.8.0
strawberry fruit	SRR31895407	3000000	87.3	2.8.0
strawberry fruit	SRR31895408	3000000	91.6	2.8.0
strawberry fruit	SRR31895409	3000000	91.5	2.8.0
strawberry fruit	SRR31895410	3000000	90.5	2.8.0
eggplant fruit peel	SRR36307035	3000000	70.2	2.8.0
eggplant fruit peel	SRR36307036	3000000	68.5	2.8.0
eggplant fruit peel	SRR36307037	3000000	66.0	2.8.0
eggplant fruit peel	SRR36307038	3000000	72.0	2.8.0

Dataset	Run	Read pairs processed	Mapped (%)	Salmon version
eggplant fruit peel	SRR36307039	3000000	60.0	2.8.0
eggplant fruit peel	SRR36307040	3000000	61.4	2.8.0

Table S15. Model inputs. Every input of the capacity model with its kind. The added-gene levels, the Pun1 multipliers, the product-profile share and the sensitivity ranges are design assumptions. Data: `data/tables/S15.csv`*; script:* `code/06_supplementary.py`*.*

Parameter	Value	Unit	Kind	Description
pun1_reference_units	1000	model units	definition	One thousand model units equal the Pun1 transcript level of pepper placenta at 24 days post anthesis (Pun1 TPM 1580.295)
pepper_pun1_tpm_24dpa	1580.295	TPM	computed	Pun1 transcripts per million in pepper placenta at 24 DPA, public series PRJNA789050 (one library per stage), from the pepper expression table of the companion design
pepper_pamt_tpm_24dpa	2238.72	TPM	computed	pAMT transcripts per million in pepper placenta at 24 DPA, same series
added_gene_level_x_high	1.0	fraction of pepper level	design assumption	Added pAMT and Pun1/AT3 reach the pepper placenta transcript level (x = 1.0)
added_gene_level_x_mid	0.25	fraction of pepper level	design assumption	Added genes at one quarter of the pepper level
added_gene_level_x_low	0.1	fraction of pepper level	design assumption	Added genes at one tenth of the pepper level
pun1_multiplier_1	1.0	factor	design assumption	Pun1/AT3 capacity factor of the wild-type design
pun1_multiplier_2	1.35	factor	design assumption	Pun1/AT3 capacity factor of the S39L design
pun1_multiplier_3	1.65	factor	design assumption	Pun1/AT3 capacity factor of the S39L+L345G design with peroxidase knockout
pun1_multiplier_4	2.1	factor	design assumption	Largest Pun1/AT3 capacity factor of the design set
main_capsaicinoid_share_min	0.85	fraction	design assumption	Capsaicin plus dihydrocapsaicin make at least this share of the total capsaicinoids; limits J to the KAS capacity divided by this share
capsaicin_to_dihydrocapsaicin_min	1.8	ratio	design assumption	Lower bound of the capsaicin to dihydrocapsaicin ratio in the 44-reaction model
capsaicin_to_dihydrocapsaicin_max	3.0	ratio	design assumption	Upper bound of the capsaicin to dihydrocapsaicin ratio in the 44-reaction model

Parameter	Value	Unit	Kind	Description
capacity_range_low	0.5	factor	design assumption	Lower end of the range of each step capacity in the sensitivity analysis
capacity_range_high	1.5	factor	design assumption	Upper end of the range of each step capacity in the sensitivity analysis
profile_share_range_low	0.7	fraction	design assumption	Lower end of the main-capsaicinoid share in the sensitivity analysis
profile_share_range_high	0.95	fraction	design assumption	Upper end of the main-capsaicinoid share in the sensitivity analysis

Table S16. Software and versions. Programs, versions and what each was used for. Data: `data/tables/S16.csv`; script: `code/06_supplementary.py`.

Tool	Version	Use
DIAMOND	2.2.6	blastp searches, both directions (E-value 1e-10, up to 5 targets, sensitive mode); reference searches at E-value 1e-5, default mode
miniprot	0.18-r281	placement of pepper proteins on the eggplant, N. benthamiana and petunia genomes; placement of tomato proteins on the eggplant genome
Salmon	2.8.0	quantification of RNA-seq reads (selective alignment, automatic library type)
C scanner (code/guide_design/scan.c)	own program	whole-genome scan for guide sites with up to four mismatches, strict NGG, both strands
Linear programme of the 44-reaction model	stored solver output	optimum of the pathway model for every stored scenario row
Python	3.9 or newer	standard library only for steps 1 to 6 and the build tools

Table S17. Eggplant gene models against the tomato proteome. DIAMOND hits (up to four per query) of the three eggplant peroxidase models against the tomato proteome. Data: `data/tables/S17.csv`; script: `code/06_supplementary.py`.

Eggplant gene	Tomato protein	Identity (%)	Alignment length	Query coverage (%)	E-value	Bit score
SmPrx1	NP_001233970.2	91.9	332	100	2.55e-222	608
SmPrx1	XP_004233442.2	61.8	319	95.8	2.30e-135	387
SmPrx1	XP_004230619.1	51.1	319	95.8	5.29e-103	305
SmPrx1	XP_019067378.1	50.8	319	95.8	2.62e-100	298
SmPrx34	XP_004243525.1	94.9	331	100	3.40e-229	625
SmPrx34	NP_001234644.2	93.4	331	100	7.67e-226	617
SmPrx34	XP_004243459.1	83.0	330	99.7	2.45e-194	537
SmPrx34	XP_004243458.1	82.4	330	99.7	2.85e-193	534
SmPrx72	NP_001296734.1	89.8	332	100	4.52e-216	592
SmPrx72	XP_004231908.1	82.2	321	96.4	3.86e-192	531
SmPrx72	XP_004242911.2	69.7	300	90.4	7.46e-149	422
SmPrx72	XP_010315910.1	67.6	306	92.2	1.94e-141	403

Table S18. MIT position weights. The twenty position weights of the Hsu et al. (2013) score, PAM-distal end first, from the scoring code. Data: `data/tables/S18.csv`*; script:* `code/04_guides.py`*.*

Position (1 = PAM-distal)	Weight
1	0
2	0
3	0.014
4	0
5	0
6	0.395
7	0.317
8	0
9	0.389
10	0.079
11	0.445
12	0.508
13	0.613
14	0.851
15	0.732
16	0.828
17	0.615
18	0.804
19	0.685
20	0.583

Table S19. Step assignment of the capacity model. Annotation text patterns that assign transcripts to each pathway step, the number of transcripts matched and the representative protein, for each dataset. Data: `data/tables/S19.csv`*; script:* `code/05_design_model.py`*.*

Dataset	Step	Annotation text patterns (regular expressions)	Transcripts matched	Representative protein	Description
Tomato placenta	PAL	phenylalanine ammonia-lyase	12	NP_001307538.1	phenylalanine ammonia-l 3
Tomato placenta	C4H	trans-cinnamate 4-monooxygenase | cinnamate 4-hydroxylase	3	XP_004242477.1	trans-cinnamate 4-monooxygenase isoform X
Tomato placenta	4CL	4-coumarate--CoA ligase	4	NP_001333770.1	4-coumarate--CoA ligase
Tomato placenta	HCT	shikimate O-hydroxycinnamoyltransferase | \bhydroxycinnamoyltransferase\b	3	XP_004232772.1	shikimate O-hydroxycinnamoyltransfe

Dataset	Step	Annotation text patterns (regular expressions)	Transcripts matched	Representative protein	Description
Tomato placenta	CCoAOMT	caffeoyl-CoA O-methyltransferase	5	XP_004248757.1	caffeoyl-CoA O-methyltransferase 5
Tomato placenta	COMT	caffeic acid 3-O-methyltransferase	4	XP_010312333.1	caffeic acid 3-O-methyltransferase 1
Tomato placenta	VR	cinnamoyl-CoA reductase	17	NP_001234612.1	cinnamoyl-CoA reductase
Tomato placenta	BCAT	branched-chain.*aminotransferase	7	XP_010314598.1	branched-chain-amino-ac aminotransferase 2, chloroplastic
Tomato placenta	KAS	3-oxoacyl-\[acyl-carrier-protein\] synthase	10	XP_004235286.1	3-oxoacyl-[acyl-carrier-protein] synthase 3 A, chloroplastic
Tomato placenta	FAS	3-oxoacyl-\[acyl-carrier-protein\] reductase	3	XP_004241501.1	3-oxoacyl-[acyl-carrier-protein] reductase 4
Tomato placenta	FAD2	delta\(12\)-fatty-acid desaturase FAD2 | delta\(12\)-acyl-lipid-desaturase	6	XP_004228665.1	delta(12)-fatty-acid desaturase FAD2
Tomato placenta	POX	\bperoxidase\b	110	XP_004232712.1	peroxidase 42
Strawberry fruit	PAL	phenylalanine ammonia-lyase	2	XP_004307376.1	PREDICTED: phenylalanin ammonia-lyase 1
Strawberry fruit	C4H	trans-cinnamate 4-monooxygenase | cinnamate 4-hydroxylase	1	XP_004294725.1	PREDICTED: trans-cinnam 4-monooxygenase
Strawberry fruit	4CL	4-coumarate--CoA ligase	3	XP_004309949.1	PREDICTED: 4-coumarate- CoA ligase 2-like
Strawberry fruit	HCT	shikimate O-hydroxycinnamoyltransferase | \bhydroxycinnamoyltransferase\b	10	XP_004303231.1	PREDICTED: shikimate O-hydroxycinnamoyltransfe like
Strawberry fruit	CCoAOMT	caffeoyl-CoA O-methyltransferase	1	XP_004290595.1	PREDICTED: caffeoyl-CoA methyltransferase
Strawberry fruit	COMT	caffeic acid 3-O-methyltransferase	4	XP_004307899.1	PREDICTED: caffeic acid 3 methyltransferase
Strawberry fruit	VR	cinnamoyl-CoA reductase	10	XP_004291772.1	PREDICTED: cinnamoyl-C reductase 1
Strawberry fruit	BCAT	branched-chain.*aminotransferase	7	XP_004291004.1	PREDICTED: branched-cha amino-acid aminotransfe 2, chloroplastic-like isofor X2
Strawberry fruit	KAS	3-oxoacyl-\[acyl-carrier-protein\] synthase	5	XP_004287663.1	PREDICTED: 3-oxoacyl-[ac carrier-protein] synthase I chloroplastic-like
Strawberry fruit	FAS	3-oxoacyl-\[acyl-carrier-protein\] reductase	4	XP_004294537.1	PREDICTED: 3-oxoacyl-[ac carrier-protein] reductase chloroplastic-like

Dataset	Step	Annotation text patterns (regular expressions)	Transcripts matched	Representative protein	Description
Strawberry fruit	FAD2	delta\(12\)-fatty-acid desaturase FAD2 | delta\(12\)-acyl-lipid-desaturase	0		
Strawberry fruit	POX	\bperoxidase\b	63	XP_004290489.1	PREDICTED: peroxidase 1
Eggplant fruit peel	PAL	phenylalanine ammonia-lyase	3	EGG|CM172116.1:13650325-13652815(-)|gene-PAL3	phenylalanine ammonia-l 3
Eggplant fruit peel	C4H	trans-cinnamate 4-monooxygenase | cinnamate 4-hydroxylase	2	EGG|CM172113.1:94850170-94851733(-)|gene-LOC101244196	trans-cinnamate 4-monooxygenase isoform
Eggplant fruit peel	4CL	4-coumarate--CoA ligase	4	EGG|CM172110.1:86451045-86454626(+)|gene-4CL	4-coumarate--CoA ligase
Eggplant fruit peel	HCT	shikimate O-hydroxycinnamoyltransferase | \bhydroxycinnamoyltransferase\b	2	EGG|CM172110.1:86659198-86663600(-)|gene-LOC101256271	shikimate O-hydroxycinnamoyltransfe
Eggplant fruit peel	CCoAOMT	caffeoyl-CoA O-methyltransferase	2	EGG|CM172109.1:70811268-70812401(+)|gene-LOC101265690	caffeoyl-CoA O-methyltransferase
Eggplant fruit peel	COMT	caffeic acid 3-O-methyltransferase	4	EGG|CM172110.1:95623903-95626732(+)|gene-LOC101251452	caffeic acid 3-O-methyltransferase
Eggplant fruit peel	VR	cinnamoyl-CoA reductase	10	EGG|CM172111.1:84043073-84049901(-)|gene-LOC101268681	cinnamoyl-CoA reductase
Eggplant fruit peel	BCAT	branched-chain.*aminotransferase	5	EGG|CM172109.1:71853843-71859319(+)|gene-LOC101245812	branched-chain amino ac aminotransferase 2, chloroplastic-like
Eggplant fruit peel	KAS	3-oxoacyl-\[acyl-carrier-protein\] synthase	7	EGG|CM172109.1:63057322-63061822(+)|gene-LOC101267959	3-oxoacyl-[acyl-carrier-protein] synthase I, chloroplastic
Eggplant fruit peel	FAS	3-oxoacyl-\[acyl-carrier-protein\] reductase	1	EGG|CM172113.1:90635217-90641793(-)|gene-LOC101255205	3-oxoacyl-[acyl-carrier-protein] reductase 4
Eggplant fruit peel	FAD2	delta\(12\)-fatty-acid desaturase FAD2 | delta\(12\)-acyl-lipid-desaturase	2	EGG|CM172108.1:106945975-106947126(-)|gene-LOC101259966	delta(12)-fatty-acid desaturase FAD2
Eggplant fruit peel	POX	\bperoxidase\b	73	EGG|CM172109.1:77410775-77412338(+)|gene-LOC101253648	peroxidase 42

Supplementary data: file map

Path	Contents
`data/tables/T01.csv`	Table 1: Organism entries and tiers
`data/tables/T02.csv`	Table 2: Tomato matches of the pepper pathway proteins
`data/tables/T03.csv`	Table 3: pAMT and Pun1/AT3 reference searches against the tomato proteome
`data/tables/T04.csv`	Table 4: Pepper genes in collinear blocks shared with tomato and eggplant

Path	Contents
`data/tables/T05.csv`	Table 5: Pepper genes in collinear blocks for every comparison
`data/tables/T06.csv`	Table 6: Construct sequences
`data/tables/T07.csv`	Table 7: Eighteen peroxidase guides
`data/tables/T08.csv`	Table 8: Modelled output of the capsaicinoid pathway
`data/tables/T09.csv`	Table 9: Pepper COMT transcript in placenta
`data/tables/S01.csv`	Table S1: All 97 pepper pathway proteins and their tomato matches
`data/tables/S02.csv`	Table S2: Sensitivity of the synteny result and the shuffled-position control
`data/tables/S03.csv`	Table S3: Four ways of counting
`data/tables/S04.csv`	Table S4: Pathway proteins in the synteny analysis
`data/tables/S05.csv`	Table S5: Pathway proteins and collinear blocks
`data/tables/S06.csv`	Table S6: All synteny comparisons
`data/tables/S07.csv`	Table S7: Target peroxidase genes
`data/tables/S08.csv`	Table S8: Candidate guides per gene
`data/tables/S09.csv`	Table S9: All 293 candidate guides
`data/tables/S10.csv`	Table S10: Genome sites with 1 to 3 mismatches to the 18 guides
`data/tables/S11.csv`	Table S11: Step capacities
`data/tables/S12.csv`	Table S12: Agreement of the recomputed output with the stored solver output
`data/tables/S13.csv`	Table S13: Output per replicate
`data/tables/S14.csv`	Table S14: RNA-seq runs and mapping rates
`data/tables/S15.csv`	Table S15: Model inputs
`data/tables/S16.csv`	Table S16: Software and versions
`data/tables/S17.csv`	Table S17: Eggplant gene models against the tomato proteome
`data/tables/S18.csv`	Table S18: MIT position weights
`data/tables/S19.csv`	Table S19: Step assignment of the capacity model
`data/figures/GA.csv`	Data behind the graphical abstract
`data/figures/F01.csv`	Data behind Figure 1
`data/figures/F02.csv`	Data behind Figure 2
`data/figures/F03.csv`	Data behind Figure 3
`data/figures/F04.csv`	Data behind Figure 4
`data/figures/F05a.csv`	Data behind Figure 5
`data/figures/F05b.csv`	Data behind Figure 5
`data/figures/F06a.csv`	Data behind Figure 6
`data/figures/F06b.csv`	Data behind Figure 6
`data/figures/F06c.csv`	Data behind Figure 6
`data/figures/F06d.csv`	Data behind Figure 6
`data/figures/F06e.csv`	Data behind Figure 6
`data/figures/F06f.csv`	Data behind Figure 6
`data/figures/F07.csv`	Data behind Figure 7
`data/figures/F08.csv`	Data behind Figure 8
`data/figures/F09a.csv`	Data behind Figure 9
`data/figures/F09b.csv`	Data behind Figure 9
`data/figures/F09c.csv`	Data behind Figure 9
`data/figures/F10.csv`	Data behind Figure 10
`data/results/ortholog_status_counts.csv`	Number of pepper pathway proteins by match class
`data/results/ortholog_by_gene.csv`	Tomato matches summarised by pathway gene: queries, match classes, identity range and the top bit-score hit
`data/results/reference_search_best_hits.csv`	Best tomato hit of each pAMT-type and Pun1/AT3-type reference query
`data/results/reference_search_summary.csv`	Identity ranges of the reference searches and the number of distinct tomato proteins hit
`data/results/synteny_recount_check.csv`	Recount of the tomato and eggplant block statistics from the stored ortholog and block tables, beside the stored summary
`data/results/synteny_main_metrics.csv`	Main synteny metrics for tomato and eggplant (identity 0.5 and 0.8)
`data/results/synteny_sensitivity.csv`	Pepper genes in blocks at maximum gaps 5, 10 and 25, ortholog-only index and shuffled-position control

Path	Contents
data/results/synteny_organisms.csv	Synteny results for every comparison (20 rows)
data/results/synteny_second_method_check.csv	Main method against the second method at a gap of 10 genes, with the absolute difference
data/results/synteny_tier_summary.csv	Range and median of pepper genes in blocks per tier
data/results/pathway_synteny_summary.csv	Summary of the 97 pathway proteins in the synteny analysis
data/results/tier_table.csv	Organism entries with tier, designs, highest design target and gap-10 synteny
data/results/design_counts.csv	Counts of designs, entries, species, families and design targets
data/results/construct_counts.csv	Counts of constructs by host, allele and reading frames
data/results/construct_orfs.csv	Open reading frames of pAMT and Pun1/AT3 and further reading frames in the eight constructs
data/results/construct_summary.csv	Length, Pun1/AT3 allele and reading-frame coordinates of the eight constructs
data/results/pun1_substitutions.csv	Codon and residue at Pun1/AT3 positions 39 and 345 in each construct, with the companion ESM2 pocket-scan ranks
data/results/constructs.fasta	The eight construct sequences with host, Pun1/AT3 allele and length in the header
data/results/guide_table.csv	The 18 selected guides: position, strand, PAM, cut position, site counts and MIT score
data/results/guides.fasta	The 18 guide sequences with genome position, PAM and MIT score
data/results/guide_recount_check.csv	Recount of site counts, scores and selection from the stored scan, beside the stored candidate table
data/results/guide_summary.csv	Counts and ranges over the 18 selected guides
data/results/candidate_counts.csv	Candidates per gene and how many pass each selection step
data/results/target_genes.csv	Gene models of the six target peroxidases with model checks
data/results/offtarget_sites.csv	Every genome site with one to three mismatches to a selected guide
data/results/eggplant_models_vs_tomato_proteome.csv	DIAMOND hits of the eggplant peroxidase models against the tomato proteome
data/results/mit_position_weights.csv	Position weights of the MIT score
data/results/model_step_assignment.csv	Annotation patterns, transcripts matched and representative protein of each pathway step
data/results/model_capacities.csv	TPM and capacity of every pathway step by dataset, stage and statistic
data/results/model_invariance.csv	Counts of scenario rows in which the added genes or the Pun1 multipliers change the output
data/results/model_output.csv	Modelled output J for every dataset, stage, statistic, methylation enzyme and added-gene level
data/results/model_main_output.csv	Modelled output J at added-gene level 1.0 with CCoAOMT-type methylation and the percentage of pepper
data/results/model_solver_check.csv	Recomputed J beside the stored solver optimum for 150 scenario rows
data/results/model_replicates.csv	Chain minimum for each replicate library
data/results/model_sensitivity.csv	One-at-a-time sensitivity of J for three models
data/results/pepper_reference_output.csv	Output J of pepper placenta at 34 days post anthesis, the reference of the percentages
data/results/mapping_summary.csv	Range of mapping rates per dataset
data/results/pepper_comt.csv	Pepper COMT transcript TPM by stage
data/results/pepper_comt_fall.csv	Fall of the COMT transcript between stages
data/results/software_versions.csv	Programs, versions and use
inputs/input_sources.csv	Public genome assemblies, annotations, RNA-seq series and reference proteins used by the research scripts
inputs/design_targets.csv	The 57 designs: organism, tier, design target in SHU, Pun1/AT3 variant and the assumed Pun1 capacity multiplier
inputs/model_inputs.csv	Every input of the capacity model with its kind
inputs/orthologs/*	Best tomato hit of 97 pepper pathway proteins, all 472 hits, the pathway list and the pAMT and Pun1/AT3 reference searches with their query proteins
inputs/synteny/tables/*	Ortholog pairs, block tables, anchor membership, summaries and sensitivity runs of the pepper-tomato and pepper-eggplant comparisons
inputs/synteny/organisms/*	Results of the block search for every other organism
inputs/constructs/*	Construct sequences SS-TOM-01 to 04 and SS-EGG-01 to 04
inputs/pepper_comt/*	Pepper COMT transcript TPM by stage and the Pun1/AT3 pocket-scan ranks of the companion design
inputs/guide_design/*	Gene models, the candidate lists, the whole-genome scan hits, the scored candidates and the selected guides
inputs/design_model/*	TPM of every pathway step for the three datasets, run lists, mapping rates and the stored solver output
code/01_orthologs.py	Tomato matches of the pathway proteins and the reference searches
code/02_synteny.py	Synteny recount for tomato and eggplant; all comparisons; tier table
code/03_constructs.py	Construct reading frames and Pun1/AT3 alleles
code/04_guides.py	Guide recount from the stored scan; scoring and selection

Path	Contents
`code/05_design_model.py`	Capacity model of the pathway and its sensitivity
`code/06_supplementary.py`	Pepper COMT series; graphical abstract; supplementary tables; table index; data dictionary
`code/07_full_guide_scan.py`	Full-mode guide scan repeated from the genomes
`code/guide_design/scan.c`	C program: scans a genome for sites with up to four mismatches to each guide (strict NGG PAM, both strands)
`code/heavy/NEEDS.txt`	What full mode needs and the order of the full-scale research scripts
`research_record/scripts/synteny/*`	Synteny pipeline for pepper against tomato and eggplant: gene models, miniprot placement, DIAMOND, orthologs, blocks, validation, definitions
`research_record/scripts/organisms/*`	Genome downloads and the block search for the other organisms (two implementations), placement runs and the table of results
`research_record/scripts/guides/*`	Gene models, candidate guides, C scanner, scoring and selection, second scanner and the check script
`research_record/scripts/design_model/*`	RNA-seq download, Salmon runs, TPM of the pathway steps, eggplant gene set and the model runs
`research_record/scripts/orthologs/*`	Tomato proteome identity check, reference-search script, pepper COMT recount and construct translation check
`research_record/outputs/*`	Raw outputs: block validation, per-organism results, placement results and the second-scanner report
`research_record/second_script_rerun/*`	Independent second implementations: block search and eggplant placement route, tomato, potato, cucumber and rice synteny, and the model solver, with their outputs

Cite this paper

Ford, P.I. (2026). Beyond Pepper: A Four-Tier Framework for Transferring Capsaicinoid Biosynthesis to 17 Crop Species, with Genome-Wide Synteny to Capsicum. Scoville Splice Research Series. https://doi.org/10.5281/zenodo.24511314

```
@misc{Ford2026_P02,
  author = {Ford, P.I.},
  title = {Beyond Pepper: A Four-Tier Framework for Transferring Capsaicinoid Biosynthesis to
17 Crop Species, with Genome-Wide Synteny to Capsicum},
  year = {2026},
  publisher = {Scoville Splice LLC},
  series = {Scoville Splice Research Series},
  doi = {10.5281/zenodo.24511314},
  url = {https://doi.org/10.5281/zenodo.24511314}
}
```

Also in `CITATION.cff`, `citation.bib` and `citation.ris` in the data package.

Related papers in this series

ID	Title	DOI
P01	In Silico Design of Ten Capsicum annuum Cultivars Spanning 0.87 to 2.96 Million Scoville Heat Units	10.5281/zenodo.24511300
P05	Umami-Enriched Capsicum: In Silico Design of Glutamate Accumulation by Glutamate Decarboxylase Knockout and Glutamate Dehydrogenase Expression	10.5281/zenodo.24511324
P12	Trait Expansion in Capsicum: CRISPR Guide Design for Nine Trait Groups with Gene Assignment and Genome-Wide Off-Target Profiling	10.5281/zenodo.24511350

Appendix A. Evidence table

Claim ID	Claim	Value	Kind	Source file	Location
P02-C001	Organism entries in the framework	17	computed	data/results/design_counts.csv	row organism entries
P02-C002	Number of tiers	4	definition	data/results/tier_table.csv	column tier, values 1 to 4
P02-C003	Pepper pathway proteins searched in the tomato proteome	97	computed	data/results/ortholog_status_counts.csv	row ALL
P02-C004	Queries with a close tomato match	78	computed	data/results/ortholog_status_counts.csv	row CLOSE MATCH
P02-C005	Identity range of the best tomato hit of four pAMT-type queries	85.9 to 86.3%	computed	data/results/reference_search_summary.csv	row pAMT, columns best_hit_identity_min and max
P02-C006	Identity range of the best tomato hit of four Pun1/AT3-type queries	72.6 to 73.8%	computed	data/results/reference_search_summary.csv	row Pun1/AT3, columns best_hit_identity_min and max
P02-C007	Pepper genes in blocks shared with tomato	51.3%	computed	data/results/synteny_main_metrics.csv	row pepper_vs_tomato, column pepper_genes_in_blocks_pct
P02-C008	Pepper genes in blocks shared with eggplant (identity 0.5 or more)	52.6%	computed	data/results/synteny_main_metrics.csv	row pepper_vs_eggplant, column pepper_genes_in_blocks_pct
P02-C009	Range of pepper genes in blocks for tier 4 entries	0.1 to 10.7%	computed	data/tables/T01.csv	tier 4 rows, column Pepper genes in blocks (%)
P02-C010	Construct sequences	8	computed	data/results/construct_counts.csv	row construct sequences
P02-C011	Selected guides	18	computed	data/results/guide_summary.csv	row selected guides
P02-C012	Target peroxidase genes	6	computed	data/results/guide_summary.csv	row genes
P02-C013	Tomato placenta modelled output as a share of pepper placenta	4.7 to 15.9%	computed	data/results/model_main_output.csv	tomato rows, columns pct_of_pepper_representative and pct_of_pepper_family_sum
P02-C014	Highest tier 1 share of pepper genes in blocks	52.6%	computed	data/tables/T01.csv	tier 1 rows, column Pepper genes in blocks (%)
P02-C015	Lowest tier 4 share of pepper genes in blocks	0.1%	computed	data/tables/T01.csv	tier 4 rows, column Pepper genes in blocks (%)
P02-C016	Pepper pathway proteins searched	97	computed	data/results/ortholog_status_counts.csv	row ALL
P02-C017	Perfect genome sites per guide	1	computed	data/results/guide_summary.csv	row perfect genome sites per guide, lowest
P02-C018	Designs in the framework	57	computed	data/results/design_counts.csv	row designs
P02-C019	Organism entries with a computed synteny value	15	computed	data/results/design_counts.csv	row entries with a computed synteny value
P02-C020	Pepper protein-coding genes on the 12 chromosomes (denominator of every percentage)	27,958	computed	inputs/synteny/tables/T1_main_synteny_metrics.csv	row pepper_vs_tomato, column pepper_universe
P02-C021	Tomato protein-coding genes on the chromosomes	25,158	computed	inputs/synteny/tables/T1_main_synteny_metrics.csv	row pepper_vs_tomato, column other_universe
P02-C022	Eggplant loci placed from pepper proteins (identity 0.5 or more)	32,356	computed	inputs/synteny/tables/T1_main_synteny_metrics.csv	row pepper_vs_eggplant, column other_universe

Claim ID	Claim	Value	Kind	Source file	Location
P02-C023	Eggplant loci placed from pepper proteins (identity 0.8 or more)	21,355	computed	`inputs/synteny/tables/T1_eggplant_identity_0.8.csv`	row pepper_vs_eggplant, column other_universe
P02-C024	DIAMOND version	2.2.6	definition	`data/results/software_versions.csv`	row DIAMOND
P02-C025	miniprot version	0.18-r281	definition	`data/results/software_versions.csv`	row miniprot
P02-C026	Salmon version	2.8.0	definition	`data/results/software_versions.csv`	row Salmon
P02-C027	Pepper reference assembly accession	GCF_002878395.1	definition	`inputs/input_sources.csv`	row Capsicum annuum reference genome UCD10Xv1.1 (NCBI RefSeq)
P02-C028	Tomato assembly accession	GCF_000188115.5	definition	`inputs/input_sources.csv`	row Solanum lycopersicum genome SL3.1 (NCBI RefSeq)
P02-C029	Eggplant assembly accession	GCA_057556415.1	definition	`inputs/input_sources.csv`	row Solanum melongena genome ASM5755641v1 (genome only)
P02-C030	Python version of the analysis scripts	3.9	definition	`data/results/software_versions.csv`	row Python
P02-C031	Distinct species among the organism entries	16	computed	`data/results/design_counts.csv`	row distinct species
P02-C032	Plant families among the organism entries	7	computed	`data/results/design_counts.csv`	row families
P02-C033	Smallest number of designs per entry	2	computed	`data/results/tier_table.csv`	column designs, minimum
P02-C034	Largest number of designs per entry	5	computed	`data/results/tier_table.csv`	column designs, maximum
P02-C035	Lowest design target (SHU)	100	design target	`data/results/design_counts.csv`	row lowest design target (SHU)
P02-C036	Highest design target (SHU)	1,000,000	design target	`data/results/design_counts.csv`	row highest design target (SHU)
P02-C037	Lowest identity of a close match (%)	71.1	computed	`data/tables/S01.csv`	close-match rows, column Identity (%)
P02-C038	Highest identity of a weaker match (%)	66.8	computed	`data/tables/S01.csv`	weaker-match rows, column Identity (%)
P02-C039	E-value cut-off of the reference searches	1e-5	definition	`data/results/software_versions.csv`	row DIAMOND
P02-C040	Smallest sequence counted as a chromosome	10 Mb	definition	`research_record/scripts/synteny/01_gene_models.py`	line MIN_CHROM_BP
P02-C041	E-value cut-off of the ortholog search	1e-10	definition	`research_record/scripts/synteny/04_diamond.sh`	line --evalue
P02-C042	Targets kept per query in the ortholog search	5	definition	`research_record/scripts/synteny/04_diamond.sh`	line --max-target-seqs
P02-C043	Maximum gap between neighbours in a block (genes)	10	definition	`inputs/synteny/tables/T1_main_synteny_metrics.csv`	row pepper_vs_tomato, column max_gap
P02-C044	Minimum orthologs per block	5	definition	`inputs/synteny/tables/T1_main_synteny_metrics.csv`	row pepper_vs_tomato, column min_anchors
P02-C045	Position shuffles of the control	20	definition	`inputs/synteny/tables/T4_sensitivity_and_null.csv`	column null_shuffles

Claim ID	Claim	Value	Kind	Source file	Location
P02-C046	Minimum genes per sequence in the second method	1000	definition	research_record/scripts/organisms/syn_B.py	line MING
P02-C047	E-value cut-off of the second method	1e-5	definition	research_record/scripts/organisms/syn_B.py	line --evalue
P02-C048	Identity filter of the main placement	0.5	definition	inputs/synteny/organisms/organism_results.csv	column identity_filter
P02-C049	Identity filter of the strict placement	0.8	definition	inputs/synteny/organisms/organism_results.csv	column identity_filter
P02-C050	Minimum genes per SolTub_3.0 scaffold	50	definition	research_record/scripts/organisms/run_A.sh	line run potato, option --mingenes
P02-C051	Flanking genes checked per pathway gene	10	definition	inputs/synteny/tables/T3_pathway_97_synteny.csv	column neighbours_checked
P02-C052	Residues of the reference used to find a reading frame	12	definition	code/03_constructs.py	constant 12 in find_orf
P02-C053	Residues of the pAMT reference	459	computed	data/results/construct_orfs.csv	column reference_residues, pAMT rows
P02-C054	Residues of the Pun1/AT3 reference	440	computed	data/results/construct_orfs.csv	column reference_residues, Pun1/AT3 rows
P02-C055	GC range of candidate guides	40 to 70%	definition	code/04_guides.py	text of the candidate filters (GC 40 to 70%)
P02-C056	Candidate guides	293	computed	data/results/guide_recount_check.csv	row candidates whose site counts equal the stored counts, column stored
P02-C057	Mismatches allowed in the genome scan	4	definition	code/guide_design/scan.c	usage line, argument maxmm
P02-C058	Guides per gene	3	definition	data/results/candidate_counts.csv	column guides_selected
P02-C059	Share of the coding sequence where the cut must lie	60%	definition	code/04_guides.py	text of the selection rule
P02-C060	Minimum distance between the cuts of two guides of one gene (coding bases)	100	definition	code/04_guides.py	text of the selection rule
P02-C061	Result of the second-method check of the guides	ALL PASS	computed	research_record/outputs/guides/verify_report.txt	last line
P02-C062	Units of the pepper Pun1 transcript level	1000	definition	inputs/model_inputs.csv	row pun1_reference_units
P02-C063	Pun1 TPM of pepper placenta at 24 DPA	1580.295	computed	inputs/model_inputs.csv	row pepper_pun1_tpm_24dpa
P02-C064	Days post anthesis of the pepper reference stage	24	definition	inputs/model_inputs.csv	description of row pepper_pun1_tpm_24dpa
P02-C065	Minimum share of the two main capsaicinoids	0.85	design assumption	inputs/model_inputs.csv	row main_capsaicinoid_share_min
P02-C066	Reactions of the stored pathway model	44	definition	code/05_design_model.py	text of the model description (44-reaction model)
P02-C067	Stored solver rows compared with the recomputed output	150	computed	data/tables/tables_index.csv	caption of S12
P02-C068	Tolerance of the solver comparison (model units)	0.006	definition	data/results/model_solver_check.csv	header of column agree_within_0.006
P02-C069	Added-gene level of the main scenario	1.0	design assumption	inputs/model_inputs.csv	row added_gene_level_x_high
P02-C070	Added-gene level, middle scenario	0.25	design assumption	inputs/model_inputs.csv	row added_gene_level_x_mid
P02-C071	Added-gene level, lowest scenario	0.1	design assumption	inputs/model_inputs.csv	row added_gene_level_x_low

Claim ID	Claim	Value	Kind	Source file	Location
P02-C072	Pun1 multiplier 1	1.0	design assumption	`inputs/model_inputs.csv`	row pun1_multiplier_1
P02-C073	Pun1 multiplier 2	1.35	design assumption	`inputs/model_inputs.csv`	row pun1_multiplier_2
P02-C074	Pun1 multiplier 3	1.65	design assumption	`inputs/model_inputs.csv`	row pun1_multiplier_3
P02-C075	Pun1 multiplier 4	2.1	design assumption	`inputs/model_inputs.csv`	row pun1_multiplier_4
P02-C076	Tomato RNA-seq BioProject	PRJNA1348326	definition	`inputs/input_sources.csv`	row Public M82 tomato placenta RNA-seq series (mature green, breaker, ripe; 3 replicates each)
P02-C077	Replicates per tomato stage	3	computed	`data/results/model_replicates.csv`	rows of one tomato stage
P02-C078	Strawberry RNA-seq BioProject	PRJNA1203349	definition	`inputs/input_sources.csv`	row Public Fragaria x ananassa fruit RNA-seq series (unripe and ripe; 3 replicates each)
P02-C079	Fragaria vesca assembly accession	GCF_000184155.1	definition	`inputs/input_sources.csv`	row Fragaria vesca FraVesHawaii1 (NCBI RefSeq)
P02-C080	Eggplant RNA-seq BioProject	PRJNA1374051	definition	`inputs/input_sources.csv`	row Public eggplant fruit-peel RNA-seq series (6 runs)
P02-C081	Read pairs used from each run	3,000,000	definition	`data/tables/S14.csv`	column Read pairs processed
P02-C082	Days post anthesis of the pepper reference output	34	definition	`data/results/pepper_reference_output.csv`	column J_pepper_placenta_34_dpa header
P02-C083	Pepper placenta output J, representative transcripts	56.42	computed	`data/results/pepper_reference_output.csv`	row representative
P02-C084	Pepper placenta output J, family-sum transcripts	67.28	computed	`data/results/pepper_reference_output.csv`	row family_sum
P02-C085	Lower end of the capacity range	0.5	design assumption	`inputs/model_inputs.csv`	row capacity_range_low
P02-C086	Upper end of the capacity range	1.5	design assumption	`inputs/model_inputs.csv`	row capacity_range_high
P02-C087	Lower end of the profile share range	0.7	design assumption	`inputs/model_inputs.csv`	row profile_share_range_low
P02-C088	Upper end of the profile share range	0.95	design assumption	`inputs/model_inputs.csv`	row profile_share_range_high
P02-C089	Pepper placenta RNA-seq BioProject	PRJNA789050	definition	`inputs/input_sources.csv`	row Public pepper placenta RNA-seq series (Capsicum chinense SL08; five stages)
P02-C090	Days post anthesis of the pepper placenta stages	4, 24, 34, 44 and 54	definition	`data/results/pepper_comt.csv`	column stage_dpa
P02-C091	Entries in tier 1	3	computed	`data/results/design_counts.csv`	row tier 1 entries
P02-C092	Entries in tier 2	4	computed	`data/results/design_counts.csv`	row tier 2 entries
P02-C093	Entries in tier 3	5	computed	`data/results/design_counts.csv`	row tier 3 entries
P02-C094	Entries in tier 4	5	computed	`data/results/design_counts.csv`	row tier 4 entries
P02-C095	Highest design target in tier 1 (SHU)	1,000,000	design target	`data/tables/T01.csv`	tier 1 rows, column Highest design target (SHU)
P02-C096	Highest design target in tier 4 (SHU)	10,000	design target	`data/tables/T01.csv`	tier 4 rows, column Highest design target (SHU)
P02-C097	Range of pepper genes in blocks, tier 1	40.6 to 52.6%	computed	`data/tables/T01.csv`	tier 1 rows, column Pepper genes in blocks (%)
P02-C098	Range of pepper genes in blocks, tier 2	45.5 to 46.3%	computed	`data/tables/T01.csv`	tier 2 rows, column Pepper genes in blocks (%)
P02-C099	Range of pepper genes in blocks, tier 3	11.8 to 18.8%	computed	`data/tables/T01.csv`	tier 3 rows, column Pepper genes in blocks (%)
P02-C100	Range of pepper genes in blocks, tier 4	0.1 to 10.7%	computed	`data/tables/T01.csv`	tier 4 rows, column Pepper genes in blocks (%)

Claim ID	Claim	Value	Kind	Source file	Location
P02-C101	Entries grouped with their tier by relatedness only	2	computed	data/results/design_counts.csv	row entries grouped with their tier by relatedness only
P02-C102	Identity range of close matches (%)	71.1 to 98.3	computed	data/tables/S01.csv	close-match rows, column Identity (%)
P02-C103	Queries with a weaker tomato match	7	computed	data/results/ortholog_status_counts.csv	row WEAKER MATCH
P02-C104	Identity range of weaker matches (%)	54.7 to 66.8	computed	data/tables/S01.csv	weaker-match rows, column Identity (%)
P02-C105	Acyltransferase-family queries	12	computed	data/results/ortholog_status_counts.csv	row ACYLTRANSFERASE MATCH
P02-C106	Identity range of acyltransferase-family matches (%)	41.9 to 88.9	computed	data/tables/S01.csv	acyltransferase-family rows, column Identity (%)
P02-C107	PAL queries	6	computed	data/tables/T02.csv	row PAL, column Pepper queries
P02-C108	Identity range of the PAL queries	76.1 to 95.2%	computed	data/tables/T02.csv	row PAL, column Identity range (%)
P02-C109	4CL queries	11	computed	data/tables/T02.csv	row 4CL, column Pepper queries
P02-C110	Identity range of the 4CL queries	83.1 to 95.8%	computed	data/tables/T02.csv	row 4CL, column Identity range (%)
P02-C111	FabG-type queries	9	computed	data/tables/T02.csv	row FAR/FabG, column Pepper queries
P02-C112	Identity range of the FabG-type queries	77.1 to 90.1%	computed	data/tables/T02.csv	row FAR/FabG, column Identity range (%)
P02-C113	C4H queries	1	computed	data/tables/T02.csv	row C4H, column Pepper queries
P02-C114	Identity of the C4H match (%)	66.8	computed	data/tables/S01.csv	row C4H, column Identity (%)
P02-C115	Coverage of the C4H match (%)	94.9	computed	data/tables/S01.csv	row C4H, column Coverage (%)
P02-C116	HCT queries with a close match	8	computed	data/tables/T02.csv	row HCT, column Close matches
P02-C117	HCT queries	9	computed	data/tables/T02.csv	row HCT, column Pepper queries
P02-C118	Methylation queries with a close match	20	computed	data/tables/T02.csv	row COMT, column Close matches
P02-C119	Methylation queries	22	computed	data/tables/T02.csv	row COMT, column Pepper queries
P02-C120	KAS queries with a close match	18	computed	data/tables/T02.csv	row KAS, column Close matches
P02-C121	KAS queries	19	computed	data/tables/T02.csv	row KAS, column Pepper queries
P02-C122	BCAT queries with a close match	6	computed	data/tables/T02.csv	row BCAT, column Close matches
P02-C123	BCAT queries	8	computed	data/tables/T02.csv	row BCAT, column Pepper queries
P02-C124	Query coverage of the best pAMT-type hit (%)	98.9	computed	data/results/reference_search_best_hits.csv	pAMT rows, column query_coverage_pct
P02-C125	Distinct tomato proteins hit by the pAMT-type queries	5	computed	data/results/reference_search_summary.csv	row pAMT, column distinct_tomato_proteins_hit
P02-C126	Identity range of all pAMT-type hits	80.3 to 86.3%	computed	data/results/reference_search_summary.csv	row pAMT, columns any_hit_identity_min and max
P02-C127	Query coverage of the best Pun1/AT3-type hits (%)	100	computed	data/results/reference_search_best_hits.csv	Pun1/AT3 rows, column query_coverage_pct
P02-C128	Distinct tomato proteins hit by the Pun1/AT3-type queries	5	computed	data/results/reference_search_summary.csv	row Pun1/AT3, column distinct_tomato_proteins_hit

Claim ID	Claim	Value	Kind	Source file	Location
P02-C129	Identity range of all Pun1/AT3-type hits	43.8 to 74.4%	computed	data/results/reference_search_summary.csv	row Pun1/AT3, columns any_hit_identity_min and max
P02-C130	One-to-one pepper-tomato orthologs	18,330	computed	data/results/synteny_main_metrics.csv	row pepper_vs_tomato, column one_to_one_orthologs
P02-C131	Pepper genes with a tomato ortholog (%)	65.56	computed	data/results/synteny_main_metrics.csv	row pepper_vs_tomato, column pepper_with_ortholog_pct
P02-C132	Pepper genes in blocks shared with tomato	14,350	computed	data/results/synteny_main_metrics.csv	row pepper_vs_tomato, column pepper_genes_in_blocks
P02-C133	Pepper genes in blocks shared with tomato (%)	51.33	computed	data/results/synteny_main_metrics.csv	row pepper_vs_tomato, column pepper_genes_in_blocks_pct
P02-C134	Blocks shared by pepper and tomato	520	computed	data/results/synteny_main_metrics.csv	row pepper_vs_tomato, column blocks
P02-C135	Direct pepper-tomato blocks	265	computed	data/results/synteny_main_metrics.csv	row pepper_vs_tomato, column blocks_direct
P02-C136	Inverted pepper-tomato blocks	255	computed	data/results/synteny_main_metrics.csv	row pepper_vs_tomato, column blocks_inverted
P02-C137	Median orthologs per pepper-tomato block	12	computed	data/results/synteny_main_metrics.csv	row pepper_vs_tomato, column median_anchors_per_block
P02-C138	Largest pepper-tomato block (orthologs)	520	computed	data/results/synteny_main_metrics.csv	row pepper_vs_tomato, column largest_block_anchors
P02-C139	Pepper-tomato orthologs that lie in a block (%)	78.3	computed	data/tables/T04.csv	row Tomato, column Orthologs in blocks (%)
P02-C140	Recounted quantities equal the stored summary	all	computed	data/results/synteny_recount_check.csv	column agree, all rows yes
P02-C141	Pepper genes in blocks shared with tomato at a gap of 5 (%)	49.97	computed	data/results/synteny_sensitivity.csv	row Tomato, max gap 5
P02-C142	Pepper genes in blocks shared with tomato at a gap of 25 (%)	53.34	computed	data/results/synteny_sensitivity.csv	row Tomato, max gap 25
P02-C143	Share of orthologs in blocks with an ortholog-only index, tomato (%)	80.21	computed	data/results/synteny_sensitivity.csv	row Tomato, orthologs only
P02-C144	Genes in blocks after shuffling positions	0	computed	data/results/synteny_sensitivity.csv	rows with shuffled positions, column genes_in_blocks_shuffled_mean
P02-C145	One-to-one pepper-eggplant orthologs	20,104	computed	data/results/synteny_main_metrics.csv	row pepper_vs_eggplant, column one_to_one_orthologs
P02-C146	Pepper genes with an eggplant ortholog (%)	71.91	computed	data/results/synteny_main_metrics.csv	row pepper_vs_eggplant, column pepper_with_ortholog_pct
P02-C147	Pepper genes in blocks shared with eggplant	14,712	computed	data/results/synteny_main_metrics.csv	row pepper_vs_eggplant, column pepper_genes_in_blocks
P02-C148	Pepper genes in blocks shared with eggplant (%)	52.62	computed	data/results/synteny_main_metrics.csv	row pepper_vs_eggplant, column pepper_genes_in_blocks_pct
P02-C149	Blocks shared by pepper and eggplant	523	computed	data/results/synteny_main_metrics.csv	row pepper_vs_eggplant, column blocks
P02-C150	Direct pepper-eggplant blocks	252	computed	data/results/synteny_main_metrics.csv	row pepper_vs_eggplant, column blocks_direct
P02-C151	Inverted pepper-eggplant blocks	271	computed	data/results/synteny_main_metrics.csv	row pepper_vs_eggplant, column blocks_inverted

Claim ID	Claim	Value	Kind	Source file	Location
P02-C152	One-to-one pepper-eggplant orthologs at identity 0.8	15,354	computed	data/results/synteny_main_metrics.csv	row pepper_vs_eggplant_id0.8, column one_to_one_orthologs
P02-C153	Pepper genes in blocks shared with eggplant at identity 0.8	11,880	computed	data/results/synteny_main_metrics.csv	row pepper_vs_eggplant_id0.8, column pepper_genes_in_blocks
P02-C154	Pepper genes in blocks shared with eggplant at identity 0.8 (%)	42.49	computed	data/results/synteny_main_metrics.csv	row pepper_vs_eggplant_id0.8, column pepper_genes_in_blocks_pct
P02-C155	Blocks shared by pepper and eggplant at identity 0.8	469	computed	data/results/synteny_main_metrics.csv	row pepper_vs_eggplant_id0.8, column blocks
P02-C156	Pepper genes in blocks shared with eggplant at a gap of 5 (%)	50.97	computed	data/results/synteny_sensitivity.csv	row Eggplant, max gap 5
P02-C157	Pepper genes in blocks shared with eggplant at a gap of 25 (%)	54.75	computed	data/results/synteny_sensitivity.csv	row Eggplant, max gap 25
P02-C158	Pepper genes in blocks shared with eggplant, second route (%)	50.1%	computed	data/results/synteny_organisms.csv	row eggplant_id0.5, column second_method_gap10_pct
P02-C159	Pepper genes in blocks shared with eggplant at identity 0.8, second route (%)	44.8%	computed	data/results/synteny_organisms.csv	row eggplant_id0.8, column second_method_gap10_pct
P02-C160	Difference of the two eggplant routes at identity 0.5 (points)	2.52	computed	data/results/synteny_second_method_check.csv	row eggplant_id0.5, column absolute_difference
P02-C161	Difference of the two eggplant routes at identity 0.8 (points)	2.31	computed	data/results/synteny_second_method_check.csv	row eggplant_id0.8, column absolute_difference
P02-C162	Pepper genes with any tomato protein hit (%)	94.5	computed	data/tables/S03.csv	row tomato, definition A
P02-C163	Pepper genes with any eggplant protein hit (%)	96.4	computed	data/tables/S03.csv	row eggplant, definition A
P02-C164	Pepper genes with a one-to-one tomato ortholog, rounded (%)	65.6	computed	data/tables/S03.csv	row tomato, definition B
P02-C165	Pepper genes with a one-to-one eggplant ortholog, rounded (%)	71.9	computed	data/tables/S03.csv	row eggplant, definition B
P02-C166	Pathway proteins on the 12 chromosomes	81	computed	data/results/pathway_synteny_summary.csv	row of which on the 12 chromosomes
P02-C167	Unique pepper genes among the chromosomal pathway proteins	68	computed	data/results/pathway_synteny_summary.csv	row unique pepper genes on chromosomes
P02-C168	Pathway proteins with a tomato ortholog	59	computed	data/results/pathway_synteny_summary.csv	row tomato RBH ortholog found
P02-C169	Pathway proteins whose tomato ortholog lies in a block	46	computed	data/results/pathway_synteny_summary.csv	row tomato RBH ortholog inside a collinear block
P02-C170	Pathway proteins with an eggplant ortholog	59	computed	data/results/pathway_synteny_summary.csv	row eggplant placed-locus RBH found
P02-C171	Pathway proteins whose eggplant ortholog lies in a block	49	computed	data/results/pathway_synteny_summary.csv	row eggplant RBH inside a collinear block
P02-C172	Mean share of flanking genes inside a tomato block (%)	51.5	computed	data/results/pathway_synteny_summary.csv	row mean % of the 10 flanking genes inside a tomato block
P02-C173	Mean share of flanking genes inside an eggplant block (%)	52.5	computed	data/results/pathway_synteny_summary.csv	row mean % of the 10 flanking genes inside an eggplant block
P02-C174	Pepper genes in blocks shared with potato (%)	46.3	computed	data/tables/T01.csv	row Potato, column Pepper genes in blocks (%)

Claim ID	Claim	Value	Kind	Source file	Location
P02-C175	Pepper genes in blocks shared with potato, SolTub_3.0 (%)	40.88	computed	data/tables/T05.csv	row Solanum tuberosum (SolTub_3.0), column Gap 10 (%)
P02-C176	Pepper genes in blocks shared with petunia (%)	45.5	computed	data/tables/T01.csv	row Petunia, column Pepper genes in blocks (%)
P02-C177	Pepper genes in blocks shared with N. benthamiana (%)	40.6	computed	data/tables/T01.csv	row Nicotiana benthamiana, column Pepper genes in blocks (%)
P02-C178	Pepper genes in blocks shared with N. tabacum (%)	39.81	computed	data/tables/T05.csv	row Nicotiana tabacum, column Gap 10 (%)
P02-C179	Pepper genes in blocks shared with F. vesca (%)	18.8	computed	data/tables/T01.csv	row Strawberry, column Pepper genes in blocks (%)
P02-C180	Pepper genes in blocks shared with octoploid strawberry (%)	4.64	computed	data/tables/T05.csv	row Fragaria x ananassa, column Gap 10 (%)
P02-C181	Pepper genes in blocks shared with apple (%)	16.4	computed	data/tables/T01.csv	row Apple, column Pepper genes in blocks (%)
P02-C182	Pepper genes in blocks shared with cucumber (%)	14.0	computed	data/tables/T01.csv	row Cucumber, column Pepper genes in blocks (%)
P02-C183	Pepper genes in blocks shared with melon (%)	14.0	computed	data/tables/T01.csv	row Melon, column Pepper genes in blocks (%)
P02-C184	Pepper genes in blocks shared with lettuce (%)	11.8	computed	data/tables/T01.csv	row Lettuce, column Pepper genes in blocks (%)
P02-C185	Pepper genes in blocks shared with carrot (%)	10.7	computed	data/tables/T01.csv	row Carrot, column Pepper genes in blocks (%)
P02-C186	Pepper genes in blocks shared with Brassica oleracea (%)	1.5	computed	data/tables/T01.csv	row Brassica, column Pepper genes in blocks (%)
P02-C187	Pepper genes in blocks shared with corn (%)	0.1	computed	data/tables/T01.csv	row Corn, column Pepper genes in blocks (%)
P02-C188	Pepper genes in blocks shared with rice (%)	0.3	computed	data/tables/T01.csv	row Rice, column Pepper genes in blocks (%)
P02-C189	Blocks found after shuffling positions, other organisms	0	computed	data/results/synteny_organisms.csv	column null_blocks_gap10, all rows
P02-C190	Largest difference between the two block methods for annotated genomes (points)	0.28	computed	data/results/synteny_second_method_check.csv	row strawberry_ananassa, column absolute_difference
P02-C191	Largest difference of the two block methods in the other annotated comparisons (points)	0.10	computed	data/results/synteny_second_method_check.csv	row tabacum, column absolute_difference
P02-C192	Tomato constructs	4	computed	data/results/construct_counts.csv	row tomato constructs
P02-C193	Eggplant constructs	4	computed	data/results/construct_counts.csv	row eggplant constructs
P02-C194	Residues of the pAMT reading frame	466	computed	data/results/construct_orfs.csv	pAMT rows, column orf_residues
P02-C195	Residues beyond the reference in the pAMT frame	7	computed	data/results/construct_counts.csv	row residues beyond the reference in the pAMT frame
P02-C196	Residues of the Pun1/AT3 reading frame	447	computed	data/results/construct_orfs.csv	Pun1/AT3 rows, column orf_residues
P02-C197	Residues beyond the reference in the Pun1/AT3 frame	7	computed	data/results/construct_counts.csv	row residues beyond the reference in the Pun1/AT3 frame
P02-C198	Constructs with wild-type Pun1/AT3	4	computed	data/results/construct_counts.csv	row constructs with wild-type Pun1/AT3
P02-C199	Constructs with S39L	2	computed	data/results/construct_counts.csv	row constructs with S39L

Claim ID	Claim	Value	Kind	Source file	Location
P02-C200	Constructs with S39L+L345G	2	computed	data/results/construct_counts.csv	row constructs with S39L+L345G
P02-C201	Distinct construct sequences	6	computed	data/results/construct_counts.csv	row distinct sequences
P02-C202	Length of the tomato constructs 01 and 02 (bp)	4,076	computed	data/tables/T06.csv	rows SS-TOM-01 and SS-TOM-02, column Length (bp)
P02-C203	Length of the tomato constructs 03 and 04 (bp)	5,692	computed	data/tables/T06.csv	rows SS-TOM-03 and SS-TOM-04, column Length (bp)
P02-C204	Length of the eggplant constructs 01 and 02 (bp)	4,065	computed	data/tables/T06.csv	rows SS-EGG-01 and SS-EGG-02, column Length (bp)
P02-C205	Length of the eggplant constructs 03 and 04 (bp)	5,670	computed	data/tables/T06.csv	rows SS-EGG-03 and SS-EGG-04, column Length (bp)
P02-C206	Further reading frames of 100 residues or more in constructs 03 and 04	2	computed	data/tables/T06.csv	rows with design 03 or 04, column Further ORFs (100 residues or more)
P02-C207	Residues of the first further reading frame	163	computed	data/results/construct_orfs.csv	row SS-TOM-03, further ORF 1
P02-C208	Residues of the second further reading frame	137	computed	data/results/construct_orfs.csv	row SS-TOM-03, further ORF 2
P02-C209	ESM2 pocket-scan rank of S39L in the companion design	1	computed	data/results/pun1_substitutions.csv	column esm2_pocket_scan_rank, S39L rows
P02-C210	ESM2 pocket-scan rank of L345G in the companion design	2	computed	data/results/pun1_substitutions.csv	column esm2_pocket_scan_rank, L345G rows
P02-C211	Distance of S39 to His169/Asp173 (angstrom)	7.34	computed	data/results/pun1_substitutions.csv	column distance_to_His169_Asp173_angstrom, S39L rows
P02-C212	Distance of L345 to His169/Asp173 (angstrom)	7.31	computed	data/results/pun1_substitutions.csv	column distance_to_His169_Asp173_angstrom, L345G rows
P02-C213	Coding length of SlPrx1 (bases)	999	computed	data/tables/S07.csv	row SlPrx1, column Coding bases
P02-C214	Coding length of SlPrx34 (bases)	996	computed	data/tables/S07.csv	row SlPrx34, column Coding bases
P02-C215	Coding length of SlPrx72 (bases)	996	computed	data/tables/S07.csv	row SlPrx72, column Coding bases
P02-C216	Placement identity of SmPrx1 (%)	91.6	computed	data/tables/S07.csv	row SmPrx1, column Identity to query (%)
P02-C217	Placement identity of SmPrx34 (%)	93.3	computed	data/tables/S07.csv	row SmPrx34, column Identity to query (%)
P02-C218	Placement identity of SmPrx72 (%)	89.8	computed	data/tables/S07.csv	row SmPrx72, column Identity to query (%)
P02-C219	Coding length of SmPrx1 (bases)	996	computed	data/tables/S07.csv	row SmPrx1, column Coding bases
P02-C220	Coding length of SmPrx34 (bases)	996	computed	data/tables/S07.csv	row SmPrx34, column Coding bases
P02-C221	Coding length of SmPrx72 (bases)	999	computed	data/tables/S07.csv	row SmPrx72, column Coding bases
P02-C222	Identity of SmPrx34 to XP_004243525.1 (%)	94.9	computed	data/tables/S17.csv	row SmPrx34, XP_004243525.1, column Identity (%)
P02-C223	Identity of SmPrx34 to the SlPrx34 protein (%)	93.4	computed	data/tables/S17.csv	row SmPrx34, NP_001234644.2, column Identity (%)
P02-C224	Candidate guides scored	293	computed	data/results/guide_summary.csv	row candidates scored
P02-C225	SlPrx34 candidates with exactly one perfect site	17	computed	data/tables/S08.csv	row SlPrx34, column With exactly one perfect site
P02-C226	SlPrx34 candidates	47	computed	data/tables/S08.csv	row SlPrx34, column Candidates
P02-C227	SlPrx1 candidates	62	computed	data/tables/S08.csv	row SlPrx1, column Candidates

Claim ID	Claim	Value	Kind	Source file	Location
P02-C228	Guides per gene	3	computed	data/results/guide_summary.csv	row guides per gene
P02-C229	Lowest MIT score of the guides	99.7	computed	data/results/guide_summary.csv	row MIT score, lowest
P02-C230	Highest MIT score of the guides	100.0	computed	data/results/guide_summary.csv	row MIT score, highest
P02-C231	Genome sites with one mismatch to the 18 guides	0	computed	data/results/guide_summary.csv	row sites with 1 mismatch, all guides
P02-C232	Genome sites with two or three mismatches to the guides	22	computed	data/results/guide_summary.csv	row sites with 2 or 3 mismatches, all guides
P02-C233	Guides with a site at two or three mismatches	7	computed	data/results/guide_summary.csv	row guides with a site at 2 or 3 mismatches
P02-C234	Genome sites with four mismatches to the guides	160	computed	data/results/guide_summary.csv	row sites with 4 mismatches, all guides
P02-C235	Protospacer shared by SlPrx72-g2 and SmPrx72-g2	GTGATAATAGCATCAACGGC	computed	data/results/guides.fasta	records SlPrx72-g2 and SmPrx72-g2
P02-C236	PAL capacity of tomato placenta at mature green (units)	105.94	computed	data/results/model_capacities.csv	row Tomato placenta, mature green, representative, PAL
P02-C237	PAL capacity of tomato placenta at breaker (units)	15.31	computed	data/results/model_capacities.csv	row Tomato placenta, breaker, representative, PAL
P02-C238	HCT capacity of tomato placenta at mature green (units)	2.66	computed	data/results/model_capacities.csv	row Tomato placenta, mature green, representative, HCT
P02-C239	HCT capacity of tomato placenta at breaker (units)	140.53	computed	data/results/model_capacities.csv	row Tomato placenta, breaker, representative, HCT
P02-C240	Tomato placenta output at mature green, representative (units)	2.7	computed	data/tables/T08.csv	row Tomato placenta, mature green, column J, representative
P02-C241	Tomato placenta output at mature green, family sum (units)	5.1	computed	data/tables/T08.csv	row Tomato placenta, mature green, column J, family sum
P02-C242	Tomato placenta output at breaker, representative (units)	7.2	computed	data/tables/T08.csv	row Tomato placenta, breaker, column J, representative
P02-C243	Tomato placenta output at breaker, family sum (units)	10.7	computed	data/tables/T08.csv	row Tomato placenta, breaker, column J, family sum
P02-C244	Tomato placenta output at ripe, representative (units)	4.1	computed	data/tables/T08.csv	row Tomato placenta, ripe, column J, representative
P02-C245	Tomato placenta output at ripe, family sum (units)	6.3	computed	data/tables/T08.csv	row Tomato placenta, ripe, column J, family sum
P02-C246	Tomato placenta output as a share of pepper placenta	4.7 to 15.9%	computed	data/tables/T08.csv	tomato rows, columns % of pepper
P02-C247	Range of the chain minimum over the three tomato breaker replicates (units)	5.04 to 7.45	computed	data/results/model_replicates.csv	tomato breaker rows, column chain_minimum_J
P02-C248	Strawberry unripe output, representative (units)	5.3	computed	data/tables/T08.csv	row Strawberry fruit, unripe, column J, representative
P02-C249	Strawberry unripe output, family sum (units)	8.1	computed	data/tables/T08.csv	row Strawberry fruit, unripe, column J, family sum
P02-C250	Strawberry ripe output, representative (units)	9.0	computed	data/tables/T08.csv	row Strawberry fruit, ripe, column J, representative
P02-C251	Strawberry ripe output, family sum (units)	11.6	computed	data/tables/T08.csv	row Strawberry fruit, ripe, column J, family sum
P02-C252	Strawberry output as a share of pepper placenta	9.5 to 17.2%	computed	data/tables/T08.csv	strawberry rows, columns % of pepper
P02-C253	Range of the chain minimum over the three unripe strawberry replicates (units)	0.94 to 14.12	computed	data/results/model_replicates.csv	strawberry unripe rows, column chain_minimum_J

Claim ID	Claim	Value	Kind	Source file	Location
P02-C254	Eggplant peel output, representative (units)	21.8	computed	data/tables/T08.csv	row Eggplant fruit peel, column J, representative
P02-C255	Eggplant peel output, family sum (units)	34.6	computed	data/tables/T08.csv	row Eggplant fruit peel, column J, family sum
P02-C256	Eggplant output as a share of pepper placenta	38.6% and 51.4%	computed	data/tables/T08.csv	eggplant row, columns % of pepper
P02-C257	Lowest chain minimum over the six eggplant replicates (units)	5.78	computed	data/results/model_replicates.csv	eggplant rows, column chain_minimum_J
P02-C258	Highest chain minimum over the six eggplant replicates (units)	31.26	computed	data/results/model_replicates.csv	eggplant rows, column chain_minimum_J
P02-C259	Mapping rates of the tomato runs	71.6 to 93.7%	computed	data/results/mapping_summary.csv	row tomato placenta
P02-C260	Mapping rates of the strawberry runs	65.5 to 91.6%	computed	data/results/mapping_summary.csv	row strawberry fruit
P02-C261	Mapping rates of the eggplant runs	60.0 to 72.0%	computed	data/results/mapping_summary.csv	row eggplant fruit peel
P02-C262	Scenario rows of the capacity model	72	computed	data/results/model_invariance.csv	row scenario rows
P02-C263	Scenario rows in which pAMT or Pun1/AT3 limits	0	computed	data/results/model_invariance.csv	row rows where pAMT or Pun1/AT3 is the limiting step
P02-C264	Scenario rows with the same output for the four Pun1 multipliers	72	computed	data/results/model_invariance.csv	row rows where the four Pun1 multipliers give the same J
P02-C265	Groups with the same output at the three added-gene levels	24	computed	data/results/model_invariance.csv	row groups where the three added-gene levels give the same J
P02-C266	Groups of three added-gene levels	24	computed	data/results/model_invariance.csv	row groups of three added-gene levels
P02-C267	Tomato breaker output across the methylation capacity range (units)	3.58 to 8.59	computed	data/results/model_sensitivity.csv	row Tomato placenta, Methylation capacity
P02-C268	Tomato breaker output across the C4H capacity range (units)	4.29 to 7.17	computed	data/results/model_sensitivity.csv	row Tomato placenta, C4H capacity
P02-C269	Tomato breaker output at baseline (units)	7.17	computed	data/results/model_sensitivity.csv	row Tomato placenta, Methylation capacity, column J_baseline
P02-C270	Strawberry ripe output across the HCT capacity range (units)	4.50 to 13.49	computed	data/results/model_sensitivity.csv	row Strawberry fruit, HCT capacity
P02-C271	Strawberry ripe output at baseline (units)	8.99	computed	data/results/model_sensitivity.csv	row Strawberry fruit, HCT capacity, column J_baseline
P02-C272	Eggplant output across the KAS capacity range (units)	10.88 to 28.01	computed	data/results/model_sensitivity.csv	row Eggplant fruit peel, KAS capacity
P02-C273	Eggplant output across the profile share range (units)	26.43 to 19.48	computed	data/results/model_sensitivity.csv	row Eggplant fruit peel, Profile share
P02-C274	Eggplant output at baseline (units)	21.77	computed	data/results/model_sensitivity.csv	row Eggplant fruit peel, KAS capacity, column J_baseline
P02-C275	COMT transcript at 24 DPA (TPM)	535.2	computed	data/tables/T09.csv	row 24, column COMT transcript (TPM)
P02-C276	COMT transcript at 34 DPA (TPM)	166.3	computed	data/tables/T09.csv	row 34, column COMT transcript (TPM)
P02-C277	Fall of the COMT transcript between 24 and 34 DPA (%)	68.9	computed	data/results/pepper_comt_fall.csv	row to_dpa 34, column fall_pct
P02-C278	COMT transcript at 54 DPA (TPM)	29.8	computed	data/tables/T09.csv	row 54, column COMT transcript (TPM)

Appendix B. Glossary, abbreviations and units

Term	Meaning	Type
Tier	A group of organism entries ordered by relatedness to Capsicum: tier 1 holds the closest relatives and tier 4 the most distant	term
Organism entry	One row of the framework: a crop with its species, family, tier and designs	term
SHU	Scoville heat units: the scale of pungency	unit
Design target	A heat level in SHU set by the designer; it is a target and not a model result	term
Capsaicinoid	A pungent compound of Capsicum such as capsaicin and dihydrocapsaicin	term
PAL	Phenylalanine ammonia-lyase: first step of the phenylpropanoid branch	abbreviation
C4H	Cinnamate 4-hydroxylase	abbreviation
4CL	4-coumarate-CoA ligase	abbreviation
HCT	Hydroxycinnamoyl transferase	abbreviation
COMT	Caffeic acid O-methyltransferase: a methylation enzyme of the phenylpropanoid branch	abbreviation
CCoAOMT	Caffeoyl-CoA O-methyltransferase: the CoA-ester methylation enzyme	abbreviation
VR	Vanillin reduction step, here cinnamoyl-CoA reductase	abbreviation
pAMT	Putative aminotransferase that converts vanillin to vanillylamine	abbreviation
Pun1/AT3	Acyltransferase of pepper that joins vanillylamine and an acyl-CoA to make a capsaicinoid	term
BCAT	Branched-chain amino acid aminotransferase: first step of the acyl branch	abbreviation
KAS	3-oxoacyl-ACP synthase: step of the acyl branch	abbreviation
FabG	Type of 3-oxoacyl-ACP reductase	term
FAS	Fatty acid synthase step used in the model	abbreviation
FAD2	Delta-12 fatty acid desaturase used in the model	abbreviation
POX	Peroxidase	abbreviation
GABA	Gamma-aminobutyric acid	abbreviation
BAHD	Family of plant acyltransferases	abbreviation
DIAMOND	Fast protein alignment program (blastp mode)	term
blastp	Protein-against-protein sequence search	term
E-value	Expected number of chance hits at least as good as the observed hit	term
Identity	Share of aligned residues that are identical, in percent	term
Bit score	Alignment score used to rank hits	term
Reciprocal best hit	Two proteins that are each other's best hit; used as a one-to-one ortholog	term
Ortholog	Gene in another species that descends from the same ancestral gene; here a reciprocal best hit	term
Synteny	Conserved gene order between two genomes	term

Term	Meaning	Type
Collinear block	A run of at least 5 orthologs in the same or reversed order in two genomes with small gaps	term
Anchor	An ortholog pair that anchors a block	term
Gap	Number of genes between two neighbouring orthologs of a block	term
miniprot	Program that aligns proteins to a genome	term
Placement	A gene set made by aligning proteins of another species to a genome	term
Stand-in	A related species used in place of the entry species	term
Homoeolog	One of the gene copies from the subgenomes of a polyploid	term
Salmon	Program that quantifies transcripts from RNA-seq reads	term
RNA-seq	Sequencing of the RNA of a tissue	abbreviation
TPM	Transcripts per million: a normalised measure of transcript abundance	unit
DPA	Days post anthesis: days after the flower opens	unit
Representative transcript	The matching transcript with the highest total TPM over all runs	term
Family sum	The sum of the TPM of all transcripts that match a step	term
Capacity	Model quantity of a step: TPM times a fixed scale; 1000 units equal the Pun1 transcript level of pepper placenta at 24 DPA	term
Model unit	Unit of capacity and output of the model	unit
J	Modelled output of the pathway: the smallest capacity along the chain, limited by the acyl branch	term
Added-gene level x	Capacity of pAMT and Pun1/AT3 as a multiple of the pepper placenta level (design assumption)	term
Tornado figure	A sensitivity figure: each bar shows how the output moves when one input moves across its range	term
ESM2	Protein language model used for the Pun1/AT3 substitution ranks of the companion design	abbreviation
S39L	Pun1/AT3 substitution of serine 39 by leucine	term
L345G	Pun1/AT3 substitution of leucine 345 by glycine	term
ORF	Open reading frame: a stretch of codons from a start to the next stop	abbreviation
Read-through residues	Residues that a reading frame carries beyond the end of the reference protein before the next stop	term
Protospacer	The 20-nucleotide target sequence of a guide RNA	term
PAM	Protospacer-adjacent motif: the NGG sequence next to the target that SpCas9 needs	abbreviation
SpCas9	Streptococcus pyogenes Cas9 nuclease	abbreviation
MIT score	Hsu et al. (2013) aggregate specificity score of a guide: higher means fewer likely off-target sites	term
GC	Share of guanine and cytosine bases	abbreviation

Term	Meaning	Type
bp	Base pairs	unit
Mb	Megabases: millions of base pairs	unit
RefSeq	NCBI reference sequence database	abbreviation
Paralogue	A gene that arose by duplication within a genome	term
Linear programme	Optimisation of a model with linear constraints; the stored solver output of the 44-reaction model	term
mm	Mismatches between a guide and a genome site	unit

Appendix C. Computational environment

Item	Value
Python	3.15.2
Platform	Darwin x86_64
Libraries	standard library only
Steps and run time	1 code/01_orthologs.py: 0.1 s; 2 code/02_synteny.py: 0.3 s; 3 code/03_constructs.py: 0.1 s; 4 code/04_guides.py: 0.1 s; 5 code/05_design_model.py: 0.1 s; 6 code/06_supplementary.py: 0.1 s
Input files	52 files, SHA-256 in `code/environment.json`
Data fingerprint	`631e3e12ff37245cf8efad68727e3e23780b4633fdde414e5b1cc17e59fa932d`

LEGAL NOTICES

Research use only. All results in this paper are computed designs, predictions and targets. They are not food, a dietary supplement, a drug or a consumer product, and they are not medical, dietary, agronomic, regulatory or financial advice. Concentrated capsaicinoids can cause severe pain and injury. Do not expose any person or animal to material derived from these designs.

No warranty; limitation of liability. This paper, its data and its code are provided "as is" without warranty of any kind. To the fullest extent permitted by law, Scoville Splice LLC and the authors are not liable for any claim, loss or damage arising from their use. Users are responsible for complying with all biosafety, biosecurity, plant-health, food-safety, import and export, and other laws that apply to them, and for obtaining every approval needed before any laboratory, greenhouse or field work.

Licence. The paper and the data are licensed under Creative Commons Attribution 4.0 International (CC BY 4.0); the code is licensed under the MIT licence. These licences cover copyright in the files only and grant no right under any patent, patent application, trade secret or trademark. See `legal/DISCLAIMER.txt`, `legal/TERMS_OF_USE.txt` and `legal/NOTICE.txt` in the data package.

DOI 10.5281/zenodo.24511314 · US provisional applications 64/172,827, 64/188,197, 64/192,062 and 64/199,105 · © 2026 Scoville Splice LLC and P.I. Ford · Published under CC BY 4.0 · scovillesplice.farm · ORCID: 0000-0003-0395-1752