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.

Ember: In Silico Design of a Peroxidase-Knockout Capsicum annuum Cultivar Targeting 866,611 Scoville Heat Units

Priyanca Ford

scovillesplice@gmail.com

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

Research Article

Keywords: Capsicum annuum, Scoville Heat Units, capsaicinoids, class III peroxidase, CRISPR-Cas9 knockout cassette, guide RNA design, flux model, TRPV1 docking

Posted Date: November 26th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11876215/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

Ember: In Silico Design of a Peroxidase-Knockout Capsicum annuum Cultivar Targeting 866,611 Scoville Heat Units

A computed design with a blend target, a flux model, guides for one peroxidase gene, a complete knockout cassette and docking of the five capsaicinoids.

P.I. Ford

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

DOI: 10.5281/zenodo.24511362 License: CC BY 4.0 **Patents:** US 64/172,827; US 64/188,197; US 64/192,062; US 64/199,105

Figshare: figshare/35015830 OSF: osf.io/4xwcd Hugging Face: capsaicinoid-data GitHub: cultivar-ss-0087 Website: ember

Package record. 150 claims traced to a file or a cited source (120 computed, 24 definition, 2 design assumption, 4 design target). 23 tables, 10 figures, 69 data files. Data fingerprint `d3a0165533e3`. Rerun every result with `bash run_all.sh`.



Pepper cross-section. Scoville Splice LLC image library.

ABSTRACT

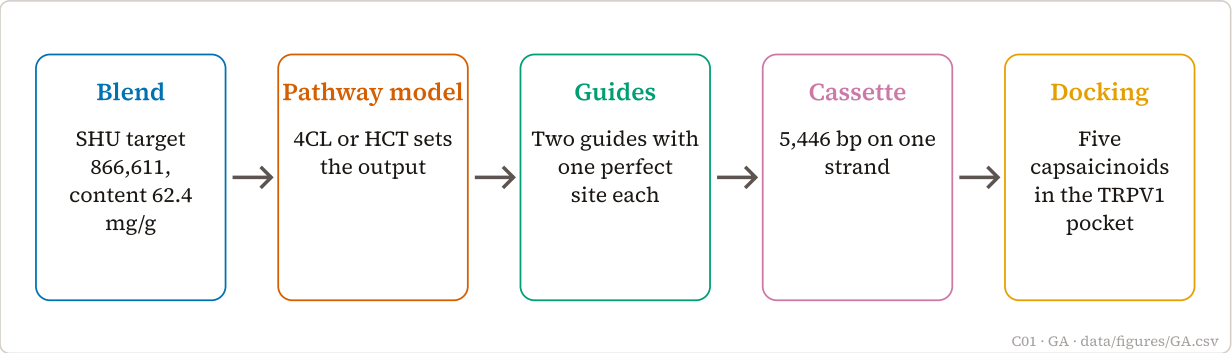
Ember (SS-0087) is a computed design for a *Capsicum annuum* cultivar with a pungency design target of 866,611 Scoville Heat Units (SHU) and one intervention, a knockout of the class III peroxidase gene LOC107856092. For the stated five-compound blend and SHU coefficients, the target equals a capsaicinoid content of 62.4 mg/g of dry weight (6.24%), which is 0.59 of a comparison value of 106 mg/g. A flux model of the capsaicinoid pathway, built from public placenta transcript counts, puts the smallest step capacity at 4CL or HCT; the Pun1/AT3 pool limits the output only below 7.76% to 9.24% of its capacity. Two guide RNAs target LOC107856092, each with one perfect genomic site and MIT scores of 99.3 and 94.7. The complete knockout cassette is 5,446 bp: an AtU6-26 guide unit, the CaMV 35S block in forward orientation, plant-optimised SpCas9 with nuclear localisation signals and the NOS terminator. Docking the five capsaicinoids into the TRPV1 vanilloid pocket places the vanillyl ring within 1.1 Å of the vanillyl ring of the crystal ligand in 13 of 15 runs. All results are computed. No experimental data are reported in this paper.

Keywords: Capsicum annuum · Scoville Heat Units · capsaicinoids · class III peroxidase · CRISPR-Cas9 knockout cassette · guide RNA design · flux model · TRPV1 docking

KEY POINTS

- The design target of 866,611 SHU equals a capsaicinoid content of 62.4 mg/g of dry weight for the stated five-compound blend.
- In the flux model, 4CL or HCT sets the output; the chance that Pun1/AT3 limits it is 0.0 in all four perturbation settings.
- Two guides, each with exactly one perfect genomic site, target LOC107856092; the cassette carries the first, at codon 42.
- The complete 5,446-bp knockout cassette places the forward 35S block, plant-optimised SpCas9 and the NOS terminator on one strand.
- All results are computed from stated design inputs and public data.

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



Graphical abstract. The design in five parts: the blend sets the SHU target, the pathway model ranks the step capacities, two guides target the peroxidase gene, a complete cassette carries the knockout, and docking places the five capsaicinoids in TRPV1. Drawn from the steps file. Data: `data/figures/GA.csv`; script: `code/07_text.py`.

1. Introduction

The pungency of a pepper is rated in Scoville Heat Units (SHU), a scale that began as a dilution rating (Scoville, 1912). SHU can be computed from the amount of each capsaicinoid and its threshold pungency (Todd et al., 1977). Capsaicinoids are made in the placenta of the fruit, and the *Pun1* gene, which encodes a putative acyltransferase, is needed for pungency (Stewart et al., 2005).

Ember (SS-0087) holds rank 1 of 10 by SHU target in the ten-cultivar series (10.5281/zenodo.24511300). It is a single-intervention peroxidase knockout design, called tier A. The design idea of this tier is that the loss of one class III peroxidase leaves more phenylpropanoid precursor for capsaicinoid synthesis. The effect of that change on content is a design assumption of this paper. The paper computes the parts around it.

It reports the capsaicinoid blend and its SHU arithmetic, the placenta expression of the pathway transcripts and of the peroxidase family, a flux model that ranks the capacities of the pathway steps, two guides for LOC107856092 with whole-genome site counts, the complete knockout cassette, and docking of the five capsaicinoids into the vanilloid pocket of TRPV1. Each part has its data, scripts and claims in the package, so that a reader can rerun it.

2. Methods

2.1 Computational basis

Gene models and genome positions are the RefSeq annotation of the GCF_002878395.1 assembly (*Capsicum annuum* UCD10Xv1.1; Hulse-Kemp et al., 2018). The expression series is PRJNA789050. The analysis steps run on Python 3.9 or later with the standard library only. The research record also uses numpy 2.0.2 and scipy 1.13.1. The guide scanner is a C program compiled with `cc` in full mode (`code/heavy/NEEDS.txt`).

2.2 Blend and SHU arithmetic

The cultivar specification (`inputs/cultivar_spec.json`) holds the SHU target of 866,611, five blend fractions that sum to 100% and one SHU coefficient per compound (Table 1). The blend-weighted coefficient is the sum of fraction times coefficient divided by the sum of the fractions. The content in mg/g of dry weight is the SHU target divided by the blend-weighted coefficient, times 1,000. The SHU contribution of a compound is the target times that compound's share of the weighted sum. The target, the fractions and the coefficients are design inputs with no experimental source. The comparison value of 106 mg/g (10.6% of dry weight) is a design assumption that serves as a scale. Values are rounded half up. Table S2 and Figure 2 give the content when one fraction moves by five points (the other fractions are rescaled to keep the sum) or one coefficient moves by ten percent.

2.3 Flux model

The pathway model has 37 metabolites and 44 reactions (`inputs/flux/metabolic_model_37_metabolites_44_reactions.json`). Capsaicinoid synthesis is one chain of 8 steps: PAL, C4H, 4CL, HCT, a methylation step, VR (reduction), pAMT and one reaction of the shared Pun1/AT3 pool. The optimum therefore equals the smallest of the 8 capacities. Step capacity is proportional to the TPM of the step transcript at 34 DPA, scaled so that the Pun1/AT3 paralog XP_016560262.2 at 1580.3 TPM has capacity 1,000; flux units are arbitrary. The *representative* setting uses one transcript per step and the *family sum* setting adds all matching transcripts. The methylation step is assigned to a CCoAOMT-type or to a COMT-class transcript, which gives four settings. The stored model was solved with COBRApy 0.32.1 (GLPK solver; Ebrahim et al., 2013) and again with a second linear-programme solver (SciPy HiGHS); the largest difference between the solver and the chain rule is 5e-09 in all knockout and scaling runs of the research record. The package computes the chain rule in the standard library and compares it with the stored solver values for the Pun1/AT3 sweep (Table S7). The perturbation run draws every capacity from a log-normal distribution with log standard deviation 0.5, 5,000 draws per setting, seed 20261111. The capacity control coefficient is the relative change of the output per relative change of one capacity, from changes of 2% up and down (Table S5). The sensitivity analysis moves one capacity at a time to half and to double its value (Figure 7).

2.4 Expression data

The public placenta series (PRJNA789050) holds reads of *Capsicum chinense* at 4, 24, 34, 44 and 54 days post anthesis (DPA), one library per stage (Table S4). Mapping rates against the *C. annuum* reference are 80.8 to 84.0%. The transcripts per million (TPM) of each transcript are stored in `inputs/expression/`. The peroxidase family table lists 94 records ranked by the mean TPM at 24 and 34 DPA (Table S3). The two Pun1 loci in the reference can share reads, so their TPM are also summed (`data/results/pun1_group.csv`).

2.5 Guide design and scoring

The gene model of LOC107856092 (XP_016556569.1) has 4 exons and a coding sequence of 1,002 bases that encodes 333 residues; its translation from the genome equals the RefSeq protein (identity 1.00). Each guide is a 20-base SpCas9 protospacer with an NGG PAM in the coding exons. The whole genome was scanned with our own C program (strict NGG PAM, both strands, up to four mismatches), and the

MIT specificity score of Hsu et al. (2013) was computed from the hits. The score of one site is the product of (1 − w) over its mismatch positions, where w is the position weight (Table S11), times 1 / (((19 − d) / 19) × 4 + 1) for sites with two or more mismatches, where d is the mean distance between mismatches, times 1 / n^2 for n mismatches, times 100. The score of a guide is 100 / (100 + the sum of its site scores) × 100. The cut lies three bases from the PAM, between guide bases 17 and 18; its coding position is the last coding base before the cut. Sites with up to three mismatches were annotated with the gene they fall in (Table S10). The best strict-NGG match of each guide in three other *Capsicum* genome assemblies was counted with up to two mismatches (Table S12).

2.6 Knockout cassette

The cassette is assembled on one strand from eight parts (Table 7). The promoter of the guide unit is bases 1 to 448 of the *Arabidopsis thaliana* U6-26 gene (GenBank X52528.1). Base 449 of that record is G, the start of the U6 transcript; the guide begins with T, so one G is placed in front of it. The scaffold is the standard 76-base SpCas9 sgRNA scaffold, followed by a run of seven T as the Pol III stop. The 35S block is 435 bases of the cauliflower mosaic virus genome (GenBank V00141.1, bases 7213 to 7647): a promoter part of 221 bases and a leader of 214 bases. SpCas9 with nuclear localisation signals is the coding sequence of GenBank KR029112.1 (protein AKE81020.1), compared with UniProt Q99ZW2. The terminator is 256 bases of the NOS terminator. The assembly script checks the reading frame, the translation, the orientation of the block and the guide position (Table S13); a second script that reads only the FASTA file scans all six frames (research record). Restriction sites are counted on both strands (Table S14).

2.7 Docking

The five capsaicinoids were docked with AutoDock Vina 1.2.5 (Trott and Olson, 2010) into chain A of the TRPV1 structure PDB 5IRX (Gao et al., 2016), residues 335 to 751. Ligands were built from structure strings with RDKit, and receptor and ligands were prepared with Meeko 0.8. The vanilloid-pocket box is a 24 Å cube centred on the centroid of the crystal resiniferatoxin. The pore-facing box is a 30 Å cube centred on the non-water HETATM atoms of the four chains; its centre lies 24.6 Å from the crystal ligand. Each ligand was docked with 3 random seeds per box at exhaustiveness 32. Every top pose was re-scored with the score-only mode of Vina, and the distance from the docked vanillyl ring to the crystal vanillyl ring was computed from the pose files. A Vina score is a scoring-function value for one pose and is not a binding energy.

3. Results

3.1 Blend, SHU arithmetic and series position

Table 1 gives the blend and Figure 1 the SHU contributions. The blend-weighted coefficient is 13,885,000 SHU per gram, which is 0.868 of the capsaicin coefficient. The SHU target of 866,611 equals a capsaicinoid content of 62.4 mg/g of dry weight (6.24%), which is 0.59 of the comparison value of 106 mg/g. The contributions of capsaicin, dihydrocapsaicin, nordihydrocapsaicin, homodihydrocapsaicin and nonivamide are 399,446, 299,585, 85,194, 53,676 and 28,710, and they sum to 866,611 SHU. The content moves most with the capsaicin coefficient: a change of ten percent in either direction gives

59.66 to 65.43 mg/g. Moving the capsaicin fraction by five points gives 61.63 to 63.22 mg/g (Figure 2, Table S2). Table 2 and Figure 3 place the target in the ten-cultivar series, which runs from 866,611 to 2,956,310 SHU; this target holds rank 1 of 10.

Table 1. Blend and SHU contributions. The five-compound blend: fraction, SHU coefficient, SHU contribution and content in mg/g of dry weight. Fractions, coefficients and the SHU target are design inputs. Data: `data/tables/T01.csv`*; script:* `code/01_blend.py`.

Compound	Fraction of the blend (%)	SHU coefficient (SHU per g)	SHU contribution	Content (mg/g dry weight)
capsaicin	40	16,000,000	399,446	24.97
dihydrocapsaicin	30	16,000,000	299,585	18.72
nordihydrocapsaicin	15	9,100,000	85,194	9.36
homodihydrocapsaicin	10	8,600,000	53,676	6.24
nonivamide	5	9,200,000	28,710	3.12
Blend total	100	13,885,000	866,611	62.41

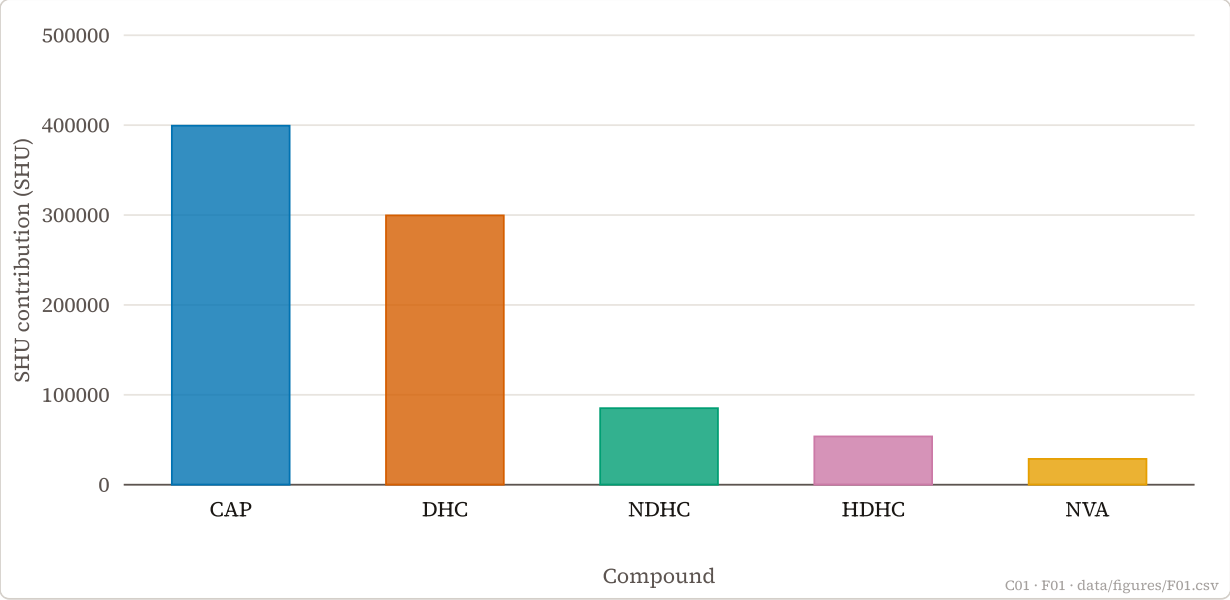


Figure 1. SHU contribution of each capsaicinoid. The SHU design target split over the five compounds of the blend. CAP capsaicin, DHC dihydrocapsaicin, NDHC nordihydrocapsaicin, HDHC homodihydrocapsaicin, NVA nonivamide. Target times the share of fraction times coefficient in the weighted sum. Data: `data/figures/F01.csv`; script: `code/01_blend.py`.

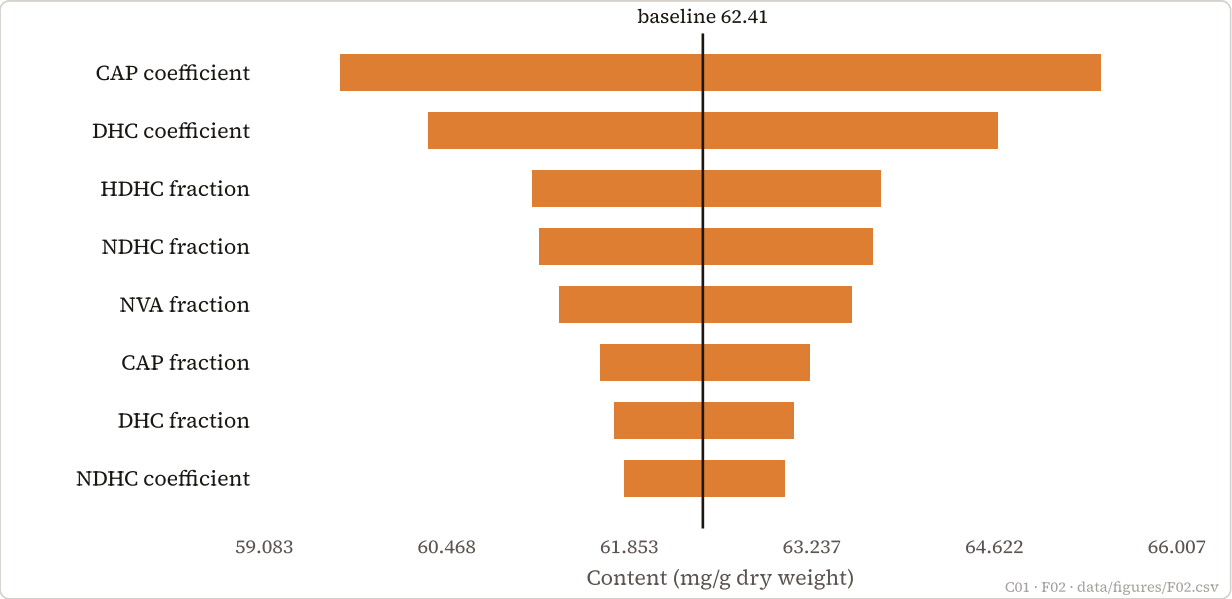


Figure 2. Sensitivity of the content to the blend inputs. Each bar spans the content when one blend fraction moves by five points (others rescaled) or one SHU coefficient moves by ten percent; the eight largest effects are shown. CAP capsaicin, DHC dihydrocapsaicin, NDHC nordihydrocapsaicin, HDHC homodihydrocapsaicin, NVA nonivamide. One-at-a-time analysis of the content formula. Data: `data/figures/F02.csv` ; script: `code/01_blend.py` .

Table 2. Series of ten cultivars. SHU design targets and capsaicinoid content of the ten cultivars of the series, ranked by target. Data: `data/tables/T02.csv` *; script:* `code/01_blend.py` .

Cultivar	Name	SHU design target	Content (mg/g dry weight)	Rank by target
SS-0087	Ember	866,611	62	1
SS-0088	Cinder	1,130,896	81	2
SS-0089	Kindle	1,382,889	98	3
SS-0090	Flare	1,622,590	114	4
SS-0091	Blaze	1,849,999	129	5
SS-0092	Wildfire	2,065,115	143	6
SS-0093	Scorch	2,267,939	156	7
SS-0094	Firestorm	2,458,470	168	8
SS-0095	Solara	2,802,656	190	9
SS-0096	Phoenix	2,956,310	199	10

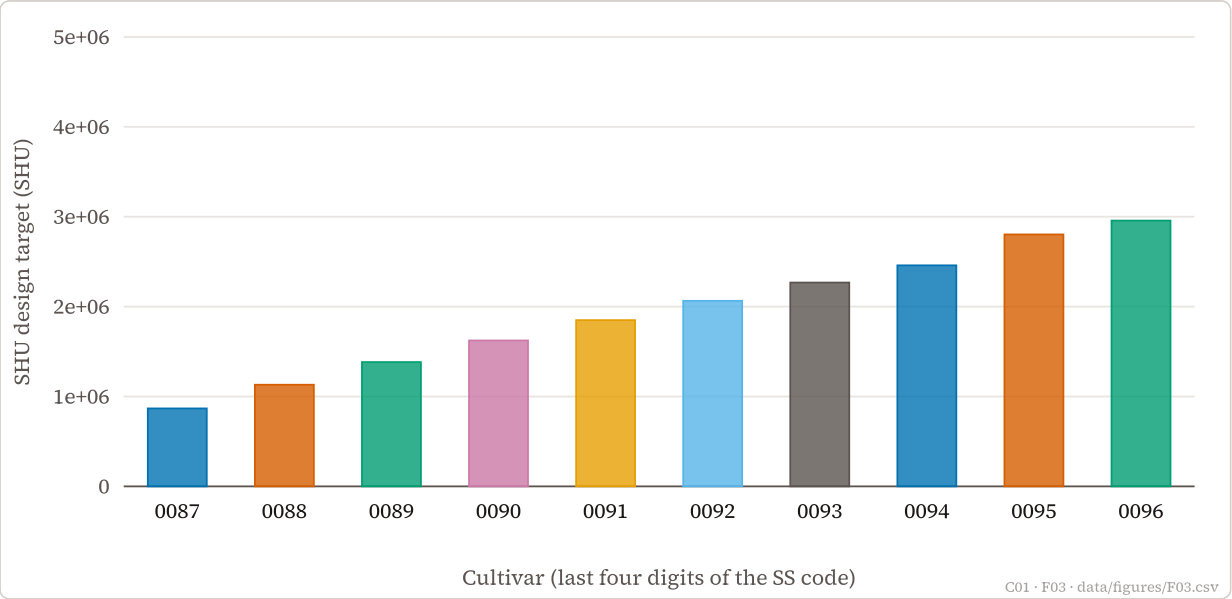


Figure 3. SHU design targets of the ten-cultivar series. The ten cultivars of the series ranked by SHU design target. Values from the series table. Data: `data/figures/F03.csv`; script: `code/01_blend.py`.

3.2 Placenta expression

Table 3 and Figure 4 give the placenta expression of the knockout target and of the pathway transcripts. LOC107856092 (XP_016556569.1) is at 185.6 TPM at 4 DPA, 14.7 at 24 DPA and 1.6 at 34 DPA. It ranks 13 of the 94 peroxidase-family records by mean TPM at 24 and 34 DPA (8.1 TPM); the highest record is LOC107859584 (peroxidase 42, mean 392.3 TPM) (Table S3). In the same series PAL (NP_001311532.1) is at 253.3 TPM and 4CL (NP_001311759.1) at 37.2 TPM at 34 DPA. The CCoAOMT-type transcript NP_001311511.1 is at 746.5 to 752.1 TPM from 24 to 34 DPA, and the COMT-class transcript NP_001311774.1 falls from 721.6 to 280.2 TPM between 24 and 34 DPA. The pAMT-step transcript is at 3,133.2 TPM at 24 DPA. The Pun1 transcript NP_001311698.1 rises from 1.6 TPM at 4 DPA to 2,647.1 at 24 DPA, and the second Pun1 locus (XP_016560262.2) is at 1,580.3 TPM at 24 DPA; read as a group, the two loci carry 4,227.4 TPM at 24 DPA.

Table 3. Placenta expression. Transcripts per million (TPM) at five stages in the public placenta series (one library per stage) for the knockout target and the pathway transcripts. Data: `data/tables/T03.csv`; script: `code/02_expression.py`.

Transcript	Role in this paper	RefSeq product	4 DPA	24 DPA	34 DPA	44 DPA	54 DPA
NP_001311635.1	pAMT step transcript	gamma aminobutyrate transaminase 2	1.8	3133.2	2238.7	2252.2	1260.1
NP_001311869.1	BCAT (valine branch)	branched-chain-amino-acid aminotransferase 2, chloroplastic-like	0.4	639.8	439.1	494.6	263.9
NP_001311532.1	PAL	phenylalanine ammonia-lyase-like	38.8	230.0	253.3	246.6	96.1
XP_016573925.1	PAL	phenylalanine ammonia-lyase 1	7.7	85.2	79.2	46.5	38.1
NP_001311774.1	COMT-class methyltransferase	caffeic acid 3-O-methyltransferase	59.3	721.6	280.2	153.2	99.2
XP_047270260.1	COMT-class methyltransferase	caffeic acid 3-O-methyltransferase	0.3	535.2	166.3	76.1	29.8
NP_001311511.1	CCoAOMT-type methyltransferase	caffeoyl-CoA O-methyltransferase 6	26.7	746.5	752.1	652.3	415.0
NP_001311759.1	4CL	4-coumarate--CoA ligase 1	52.6	63.2	37.2	30.0	19.8
NP_001311698.1	Pun1 acyltransferase (locus read with XP_016560262.2)	acyltransferase Pun1	1.6	2647.1	791.1	48.4	8.9

Transcript	Role in this paper	RefSeq product	4 DPA	24 DPA	34 DPA	44 DPA	54 DPA
XP_016560262.2	Pun1 acyltransferase (locus read with NP_001311698.1)	acyltransferase Pun1	1.0	1580.3	1148.7	855.3	423.3
NP_001311587.1	KAS (acyl branch)	3-oxoacyl-[acyl-carrier-protein] synthase 3 A, chloroplastic-like	2.5	172.4	138.2	155.9	90.3
XP_016579552.2	Acyl-branch acyltransferase	probable long-chain-alcohol O-fatty-acyltransferase 3	0.2	1.1	0.9	0.8	0.4
XP_016556569.1	peroxidase 45 (knockout target)	peroxidase 45	185.6	14.7	1.6	0.8	0.6

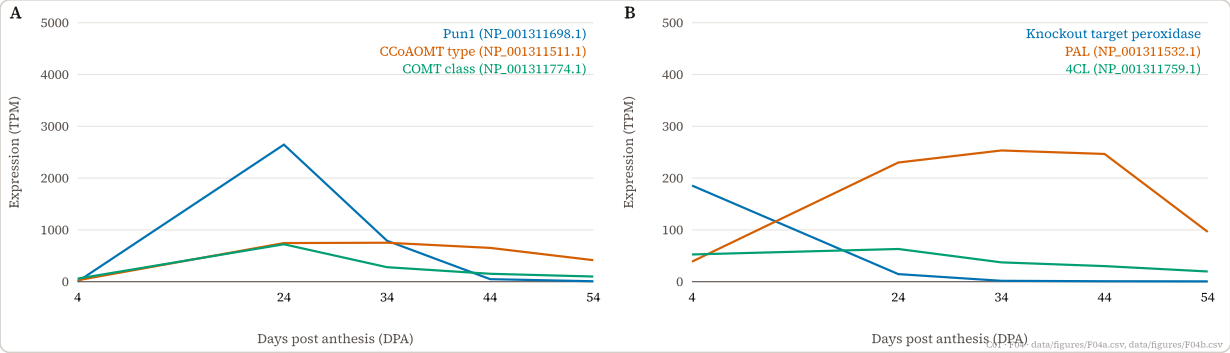


Figure 4. Placenta expression of pathway transcripts and the knockout target. A, Pun1, CCoAOMT-type and COMT-class transcripts. B, the knockout target peroxidase, PAL and 4CL. Transcripts per million in the public placenta series, one library per stage. Values from the stored TPM tables. Data: `data/figures/F04a.csv`, `data/figures/F04b.csv`; script: `code/02_expression.py`.

3.3 Flux model

Table 4 and Figure 5 give the capacities of the steps. With one transcript per step the output is 56.42, set by 4CL; with the family sum the output is 67.28, set by HCT. The capacity control coefficient is 1.00 for the limiting step and 0.00 for every other step. The Pun1/AT3 pool is filled to 0.0776 (representative) and 0.0924 (family sum). Setting any single step to zero gives an output of 0. Halving PAL, C4H, the methylation step, VR, pAMT or the Pun1/AT3 pool leaves the output at 56.42; halving 4CL gives 28.21 and halving HCT gives 31.56. Doubling 4CL gives 63.12, after which HCT limits, and doubling HCT alone leaves the output unchanged. In the family-sum setting doubling HCT gives 112.22 and halving 4CL gives 56.11 (Table S6). Figure 6 and Table S7 scale the Pun1/AT3 pool: the output stays at 56.42 until the pool falls below 7.76% of its capacity (9.24% in the family-sum setting). Table 5 summarises the perturbation runs: an early step limits in 0.9332, 0.9916, 0.9298 and 0.9838 of the draws, and the chance that Pun1/AT3 limits the output is 0.0 in every setting. The median fill of the Pun1/AT3 pool is 0.0607 to 0.0844. In the one-at-a-time analysis only 4CL and HCT move the output (Figure 7). The model ranks transcript-proportional capacities; the peroxidase step is outside it.

Table 4. Capacities of the pathway steps. Step capacity (arbitrary flux units) and saturation (output divided by capacity) with one transcript per step and with the sum of the family; methylation step as a CCoAOMT-type transcript. Data: `data/tables/T04.csv`*; script:* `code/03_flux.py`*.*

Step	Capacity, representative transcript	Saturation	Capacity, family sum	Saturation
PAL	241.6	0.23	580.2	0.12
C4H	214.8	0.26	214.8	0.31
4CL	56.4	1.00	112.2	0.60
HCT	63.1	0.89	67.3	1.00
Methylation	475.9	0.12	763.9	0.09
VR	118.8	0.47	287.2	0.23
pAMT	1416.6	0.04	1416.6	0.05
BCAT (valine)	277.8	0.20	388.3	0.17
KAS (C9)	813.8	0.07	1223.5	0.05
Pun1/AT3 pool	726.9	0.08	728.1	0.09
Output (smallest of the eight chain capacities)	56.42		67.28	

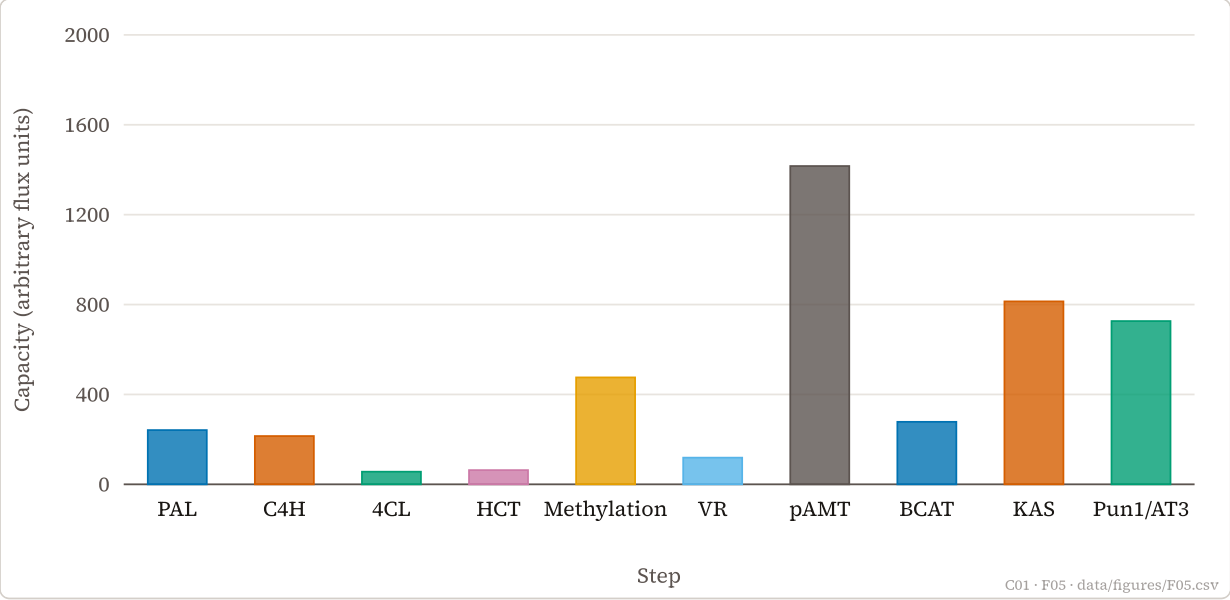


Figure 5. Capacity of each pathway step. Step capacities with one transcript per step and the methylation step as a CCoAOMT-type transcript; the output of the model equals the smallest of the first eight. Capacity proportional to the TPM of the step transcript at 34 DPA. Data: `data/figures/F05.csv`; script: `code/03_flux.py`.

Table 5. Perturbation runs. Summary of 5,000 draws per setting in which every step capacity is drawn from a log-normal distribution with log standard deviation 0.5. Data: `data/tables/T05.csv` *; script:* `code/03_flux.py` *.*

Methylation assignment	Capacity accounting	Chance an early step limits	Chance Pun1/AT3 limits	Pun1/AT3 filled, median	Most common limiting step (share of draws)
CCoAOMT-type	representative	0.9332	0.0	0.0619	4CL (0.5396)
CCoAOMT-type	family-sum	0.9916	0.0	0.0841	HCT (0.7434)
COMT-class	representative	0.9298	0.0	0.0607	4CL (0.5162)
COMT-class	family-sum	0.9838	0.0	0.0844	HCT (0.726)

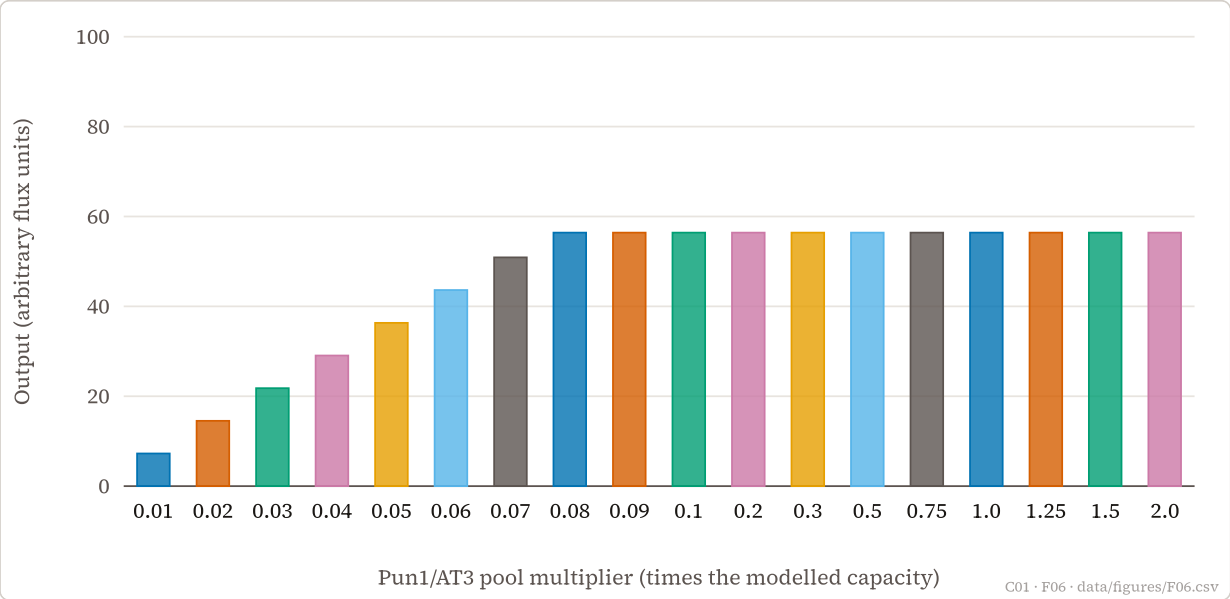


Figure 6. Output as the Pun1/AT3 pool is scaled. Output of the flux model when the capacity of the shared Pun1/AT3 pool is multiplied by the factor on the horizontal axis. Chain rule: output equals the smallest step capacity. Data: `data/figures/F06.csv` ; script: `code/03_flux.py` .

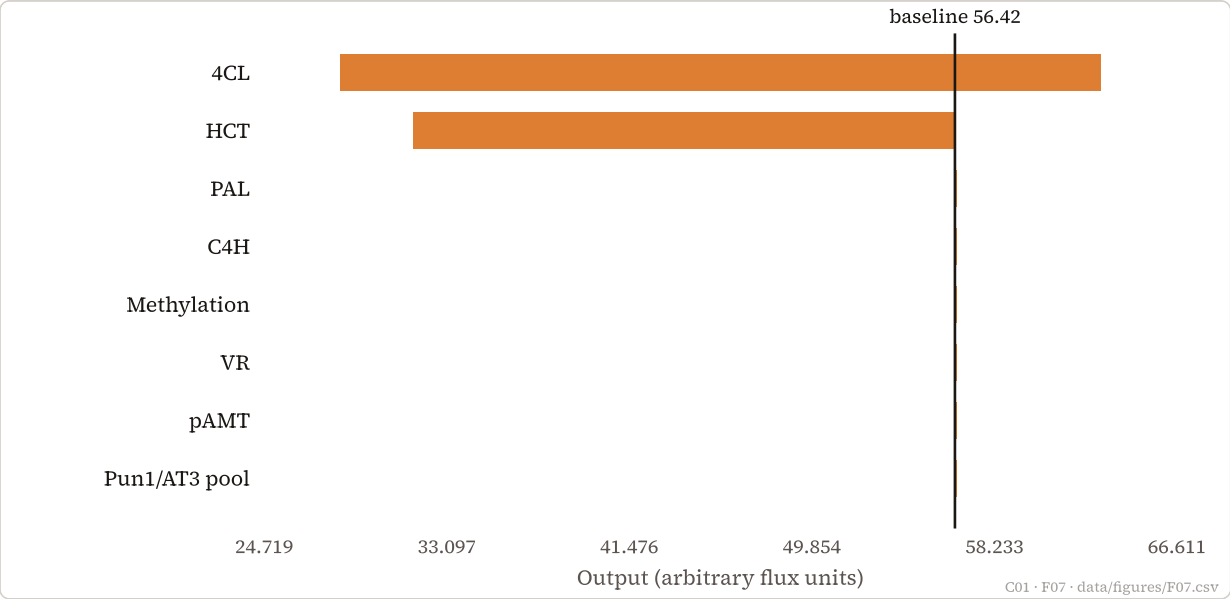


Figure 7. Sensitivity of the output to each step capacity. Each bar spans the output when one step capacity moves to half and to double its value. One-at-a-time analysis of the chain rule. Data: `data/figures/F07.csv`; script: `code/03_flux.py`.

3.4 Guides for the peroxidase gene

LOC107856092 (XP_016556569.1, peroxidase 45) lies on chromosome 1. Table 6 and Figure 8 list the two guides. Guide POX-107856092-g1 is `TCGAAATTCGTGTCCGAACG` with PAM TGG; it cuts after coding base 124 (codon 42 of 333, exon 1) and has a MIT score of 99.3. Guide POX-107856092-g2 is `CATGGCCAAGATGTCAGCAC` with PAM AGG; it cuts after coding base 388 (codon 130, exon 2) and has a MIT score of 94.7. The two cuts are 264 coding bases apart. Each guide has exactly one perfect site in the genome. The sites with one to four mismatches number 1 / 0 / 0 / 0 / 7 for the first guide and 1 / 0 / 0 / 2 / 40 for the second (sites at zero, one, two, three and four mismatches; Table S9). The second guide has 2 sites with three mismatches, both outside genes (Table S10). The recount of every site and score from the scan hits equals the stored table (2 of 2 guides). The best strict-NGG match of the first guide has 1 mismatch in two *C. chinense* assemblies and 0 mismatches in the *C. baccatum* assembly (Table S12).

Table 6. Guides for the peroxidase gene. Each guide has one perfect genomic site. MIT score is the Hsu et al. (2013) aggregate specificity score. The cut position is the last coding base before the cut. Sites are counted with 0 to 4 mismatches. Data: `data/tables/T06.csv`*; script:* `code/04_guides.py`*.*

Guide	Protospacer and PAM	Guide strand	Genomic position	Cut (coding base / codon / exon)	GC (%)	MIT score	Sites at 0 / 1 / 2 / 3 / 4 mismatches
POX-107856092-g1	TCGAAATTCGTGTCCGAACG TGG	-	chr1:32,673,231-32,673,250	124 / 42 / 1	50.0	99.3	1 / 0 / 0 / 0 / 7
POX-107856092-g2	CATGGCCAAGATGTCAGCAC AGG	+	chr1:32,672,639-32,672,658	388 / 130 / 2	55.0	94.7	1 / 0 / 0 / 2 / 40

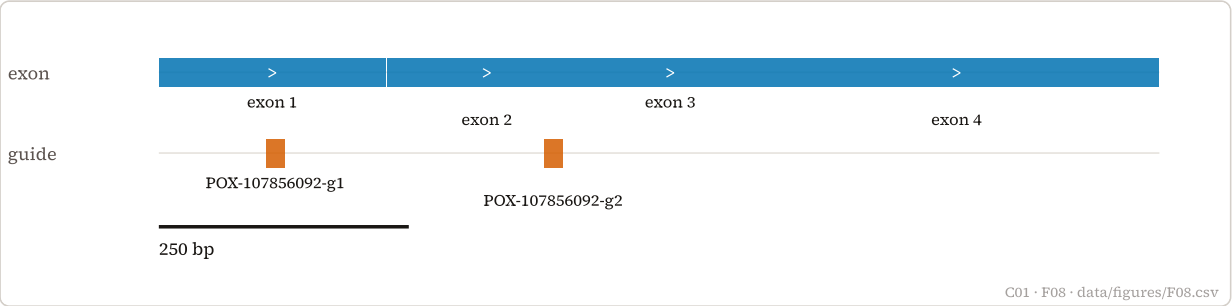


Figure 8. Guides in the coding sequence of the peroxidase gene. Exons and the two guides along the coding sequence; the scale bar is in base pairs. Positions are coding-sequence coordinates computed from the exon coordinates. Data: `data/figures/F08.csv`; script: `code/04_guides.py`.

3.5 Complete knockout cassette

Table 7 and Figure 9 give the cassette: 5,446 bp in total, from a guide unit of 552 bp and a Cas9 unit of 4,894 bp. The Cas9 coding sequence has 4,203 bases, which encode 1,400 residues: 17 + 1,367 + 16 (signal residues at the amino end, SpCas9 residues 2 to 1,368, signal residues at the carboxyl end). The translation equals the GenBank translation, with one start codon and a stop only at the end. The coding sequence is 54.2% GC, and 78.6% GC at the third codon position. The 35S block lies forward at base 553, its reverse complement does not occur in the cassette, and 433 of 435 bases equal the cauliflower mosaic virus genome (219 of 221 in the promoter part and 214 of 214 in the leader). The block holds 7 ATG codons ahead of Cas9, at block positions 25, 52, 125, 151, 273, 326 and 355; 0 of their reading frames reach the Cas9 start. Of the 13 enzymes counted, 7 have zero recognition sites in the cassette (BsaI, EcoRI, XbaI, SphI, BamHI, PstI/SbfI and XmaI/SmaI); it has 6 BsmBI sites, 3 BpiI/BbsI sites and 5 HindIII sites (Table S14). All 19 checks of Table S13 pass. The second script finds 1,400 codons at bases 988 to 5190 as the longest open reading frame in the six frames.

Table 7. Parts of the complete knockout cassette. Positions on the 5,446-bp cassette, one strand, in order: guide unit, 35S block, SpCas9 with nuclear localisation signals and NOS terminator. Data: `data/tables/T07.csv`*; script:* `code/05_cassette.py`.

Part	Start	End	Length (bp)	Source
AtU6-26 promoter (X52528.1, bases 1-448)	1	448	448	GenBank X52528.1, Arabidopsis thaliana U6-26 snRNA gene
Extra 5' G	449	449	1	Native U6-26 start base; the guide begins with T
Guide (20 nt)	450	469	20	Knockout guide of the locus
sgRNA scaffold	470	545	76	Standard SpCas9 scaffold
Pol III stop (T7)	546	552	7	Run of seven T
CaMV 35S block, forward (promoter part 221 bp + leader 214 bp)	553	987	435	CaMV V00141.1 bases 7213-7647
SpCas9 with nuclear localisation signals (KR029112.1)	988	5190	4,203	GenBank KR029112.1 CDS (AKE81020.1)
NOS terminator	5191	5446	256	pBI121 NOS terminator (AF485783 4022-4277)
Complete knockout cassette	1	5446	5,446	

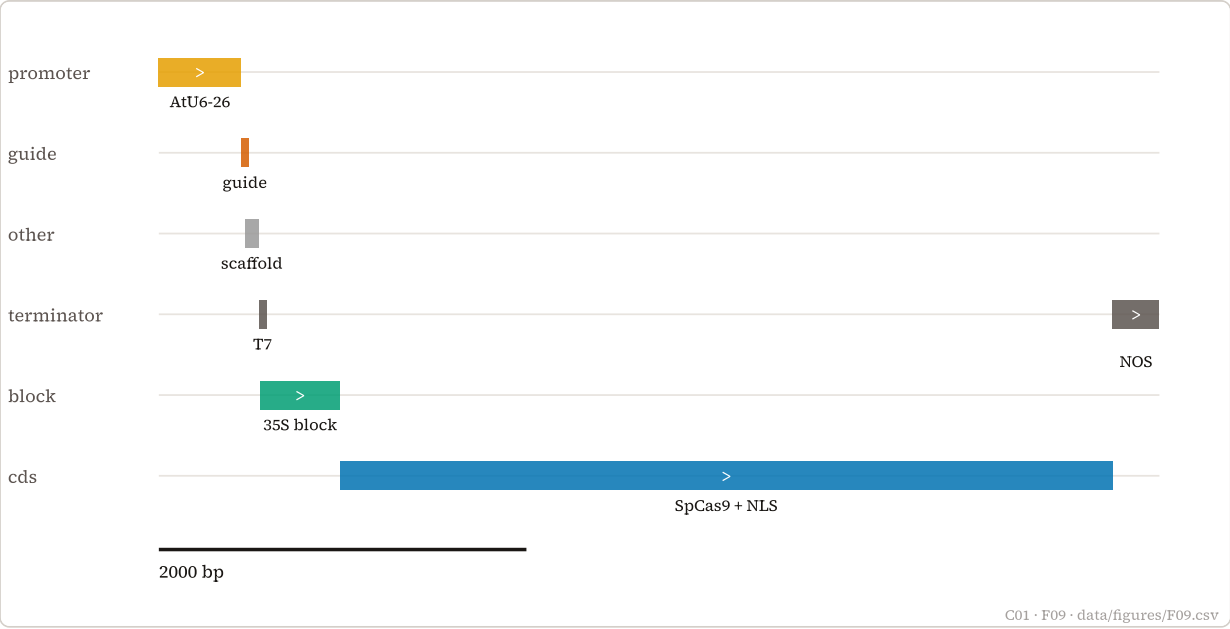


Figure 9. Map of the complete knockout cassette. Parts along the 5,446-bp cassette: AtU6-26 promoter, guide, scaffold, Pol III stop, 35S block, SpCas9 with nuclear localisation signals and NOS terminator, all on one strand. Coordinates from the assembly script. Data: `data/figures/F09.csv`; script: `code/05_cassette.py`.

3.6 Docking into TRPV1

Table 8 and Figure 10 give the scores. In the vanilloid pocket the top pose of 13 of 15 runs places the vanillyl ring within 1.1 Å of the crystal vanillyl ring (0.1 Å at the closest), and the nearest ligand O or N lies 2.8 to 3.2 Å from Y511, S512, T550 or R557 in all 15 runs. The pocket scores run from -6.21 for capsaicin to -5.54 for nonivamide. The mean scores of the five ligands span 0.67 kcal/mol, and the scores of one ligand across seeds span up to 0.53, so the scores are similar in size to the seed spread. In the pore-facing box the docked ring lies 16.5 to 20.7 Å from the crystal vanillyl ring, and the scores are 0.40 to 0.95 kcal/mol lower than in the pocket for every ligand. Re-scoring 30 top poses with Vina differs from the reported scores by at most 0.007 kcal/mol (Table S15).

Table 8. Docking scores. Best Vina score of three seeds per ligand and box (mean and range). The vanilloid pocket is a 24 Angstrom box on the crystal resiniferatoxin; the pore-facing box is a 30 Angstrom box on the channel axis. Data: `data/tables/T08.csv`*; script:* `code/06_docking.py`*.*

Ligand	Vanilloid pocket: mean (range) of three seeds (kcal/mol)	Pore-facing box: mean (range) of three seeds (kcal/mol)	Pocket minus pore-facing (kcal/mol)
capsaicin	-6.21 (-6.40 to -6.04)	-6.66 (-6.84 to -6.46)	0.45
dihydrocapsaicin	-6.03 (-6.11 to -5.97)	-6.58 (-6.70 to -6.44)	0.56
nordihydrocapsaicin	-5.65 (-5.83 to -5.45)	-6.60 (-6.69 to -6.51)	0.95
homodihydrocapsaicin	-5.73 (-5.81 to -5.68)	-6.13 (-6.56 to -5.60)	0.40
nonivamide	-5.54 (-5.80 to -5.27)	-6.43 (-6.67 to -6.20)	0.89

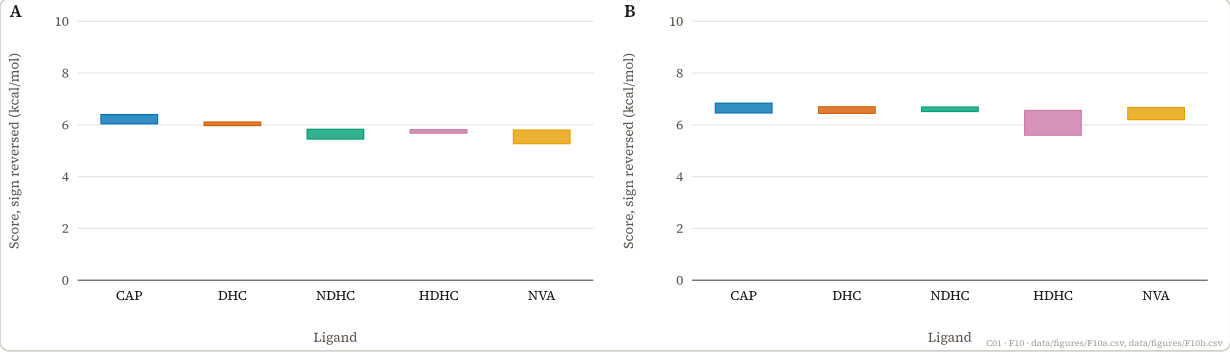


Figure 10. Vina scores of the five capsaicinoids. A, vanilloid-pocket box. B, pore-facing box. Bars span the three seeds; a higher bar is a lower (more negative) Vina score. CAP capsaicin, DHC dihydrocapsaicin, NDHC nordihydrocapsaicin, HDHC homodihydrocapsaicin, NVA nonivamide. Best pose of each run, AutoDock Vina 1.2.5, exhaustiveness 32. Data: `data/figures/F10a.csv`, `data/figures/F10b.csv`; script: `code/06_docking.py`.

4. Limitations

The flux model assumes that step capacity is proportional to transcript abundance and that the perturbation spread is 0.5; it ranks capacities and contains no peroxidase step, so the effect of the knockout on content is a design assumption. The target gene LOC107856092 is expressed mostly at 4 DPA in the placenta series, which has one library per stage. The blend, the SHU coefficients and the comparison value are design inputs. Docking uses one receptor subunit, and Vina scores of single poses do not rank the capsaicinoids. The cassette and the AtU6-26 promoter are designs; their activity in pepper is not computed.

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. The cultivar specification `inputs/cultivar_spec.json` holds every cultivar-specific value. The cassette sequence is in `data/results/knockout_cassette.fasta` and the guides are in `data/results/guides.fasta`. The paper and data are licensed under CC BY 4.0 and the code under the MIT licence. The public input datasets and records are listed in `inputs/input_sources.csv`.

Public records of this paper: Zenodo 10.5281/zenodo.24511362; Figshare: figshare/35015830; OSF: osf.io/4xwcd; Hugging Face: capsaicinoid-data; GitHub: cultivar-ss-0087; Website: ember.

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.

Active experiments for this design include a capsaicinoid time course with SHU measured by ASTA 21.3 in the baseline cultivars, a pure-standard check of the SHU coefficients and the ten blend recipes, resequencing of every target locus in the lab lines, sensory panel screening, training and taster-status genotyping, a release-assay performance check (HPLC-MS) and comparability between the two sites, and capsaicinoid stability in dried tissue, powder and extract under storage and light.

6. References

1. Ebrahim, A., Lerman, J. A., Palsson, B. O., & Hyduke, D. R. (2013). COBRApy: COnstraints-Based Reconstruction and Analysis for Python. *BMC Systems Biology, 7*, 74. doi:10.1186/1752-0509-7-74

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. Gao, Y., Cao, E., Julius, D., & Cheng, Y. (2016). TRPV1 structures in nanodiscs reveal mechanisms of ligand and lipid action. *Nature, 534*(7607), 347–351. doi:10.1038/nature17964

4. 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

5. 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

6. Scoville, W. L. (1912). Note on capsicums. *Journal of the American Pharmaceutical Association, 1*(5), 453–454. doi:10.1002/jps.3080010520

7. Stewart, C., Jr, Kang, B.-C., Liu, K., et al. (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

8. Todd, P. H., Jr, Bensinger, M. G., & Biftu, T. (1977). Determination of pungency due to capsicum by gas-liquid chromatography. *Journal of Food Science, 42*(3), 660–665. doi:10.1111/j.1365-2621.1977.tb12573.x

9. Trott, O., & Olson, A. J. (2010). AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *Journal of Computational Chemistry, 31*(2), 455–461. doi:10.1002/jcc.21334

Supplementary data

The supplementary tables hold the full data behind the paper. Table S1 lists the quantities of the blend arithmetic and Table S2 the sensitivity of the content to the blend inputs. Table S3 gives the placenta TPM of all 94 peroxidase-family records and Table S4 the five libraries. Table S5 gives the capacity, saturation and control coefficient of every step in the four flux settings, Table S6 the knockout and scaling runs, Table S7 the Pun1/AT3 pool sweep and Table S8 the limiting steps of the perturbation runs. Table S9 lists every genome site with up to four mismatches to the guides, Table S10 the annotated sites with up to three mismatches, Table S11 the MIT position weights and Table S12 the guide matches in other assemblies. Table S13 gives the checks of the cassette, Table S14 its restriction-site counts and Table S15 all docking runs. 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. Quantities of the blend arithmetic. Definitions and values of the quantities used in the SHU arithmetic. Data: `data/tables/S01.csv`; script: `code/01_blend.py`.

Quantity	Value	Definition
SHU target	866,611	Design target of the cultivar (SHU)
Sum of the blend fractions (%)	100	Fractions are percentages of the capsaicinoid mass
Blend-weighted coefficient (SHU per g)	13,885,000	Sum of fraction x coefficient divided by the sum of the fractions

Quantity	Value	Definition
Coefficient ratio to capsaicin	0.868	Blend-weighted coefficient divided by the capsaicin coefficient
Content (mg/g dry weight)	62.4	SHU target divided by the blend-weighted coefficient, times 1,000
Content (% dry weight)	6.24	Content in mg/g divided by 10
Comparison value (mg/g dry weight)	106	Design assumption used as a scale for the content
Content divided by the comparison value	0.59	Computed from the unrounded content

Table S2. Sensitivity of the content to the blend inputs. Content (mg/g dry weight) when one fraction moves by five points (other fractions rescaled) or one coefficient moves by ten percent. Data: `data/tables/S02.csv`*; script:* `code/01_blend.py`*.*

Input	Low	High	Content at low (mg/g)	Content at high (mg/g)
capsaicin coefficient	14,400,000	17,600,000	65.43	59.66
dihydrocapsaicin coefficient	14,400,000	17,600,000	64.65	60.33
homodihydrocapsaicin fraction	5.0	15.0	61.12	63.76
nordihydrocapsaicin fraction	10.0	20.0	61.17	63.70
nonivamide fraction	0.0	10.0	61.32	63.54
capsaicin fraction	35.0	45.0	63.22	61.63
dihydrocapsaicin fraction	25.0	35.0	63.10	61.74
nordihydrocapsaicin coefficient	8,190,000	10,010,000	63.03	61.81
homodihydrocapsaicin coefficient	7,740,000	9,460,000	62.80	62.03
nonivamide coefficient	8,280,000	10,120,000	62.62	62.21

Table S3. Class III peroxidase family in the placenta. TPM at five stages of every record of the peroxidase family, ranked by mean TPM at 24 and 34 DPA. Data: `data/tables/S03.csv`*; script:* `code/02_expression.py`*.*

Rank	Locus	Protein	Product	4 DPA	24 DPA	34 DPA	44 DPA	54 DPA	Mean of 24 and 34 DPA
1	LOC107859584	XP_016560122.2	peroxidase 42	1308.7	480.2	304.4	89.4	147.7	392.3
2	LOC107868170	XP_016570262.1	peroxidase 72	0.0	7.1	171.1	15.4	17.3	89.1
3	LOC107860206	XP_016560965.1	peroxidase 51	109.1	68.0	6.5	16.6	18.2	37.3

Rank	Locus	Protein	Product	4 DPA	24 DPA	34 DPA	44 DPA	54 DPA	Mean of 24 and 34 DPA
4	LOC107838944	XP_016537721.1	protein OVEREXPRESSOR OF CATIONIC PEROXIDASE 3	28.4	36.0	30.6	31.9	31.9	33.3
5	LOC124897629	XP_047266625.1	suberization-associated anionic peroxidase-like	0.2	40.1	23.1	0.7	1.8	31.6
6	LOC107860013	XP_016560694.2	peroxidase 49	4.0	3.4	51.4	4.4	26.9	27.4
7	LOC107878262	XP_016580668.1	peroxidase 3	400.4	47.6	0.8	0.1	0.4	24.2
8	LOC107853405	XP_016553885.1	peroxidase 9	0.9	27.5	2.3	0.1	0.1	14.9
9	LOC107844720	XP_016544563.1	peroxidase A2	16.4	16.0	10.6	0.8	1.5	13.3
10	LOC107860325	XP_016561130.1	peroxidase 10	3.1	10.6	8.9	8.2	5.6	9.8
11	LOC107867709	XP_016569561.1	peroxidase 43 isoform X1	18.5	14.3	4.6	1.5	2.8	9.4
12	LOC107851630	XP_016552197.1	peroxidase 17	3.2	10.8	7.9	1.3	1.1	9.3
13	LOC107856092	XP_016556569.1	peroxidase 45	185.6	14.7	1.6	0.8	0.6	8.1
14	LOC107860606	XP_016561495.1	peroxidase 63	97.5	9.6	6.5	69.8	73.1	8.1
15	LOC107867709	XP_016569563.1	peroxidase 43 isoform X3	2.1	5.1	8.2	0.0	0.0	6.6
16	LOC107867617	XP_016569419.1	peroxidase 12	25.9	9.9	2.4	1.1	0.6	6.1
17	LOC124896280	XP_047263783.1	suberization-associated anionic peroxidase 2-like	0.6	9.8	2.0	1.9	2.0	5.9
18	LOC107867212	XP_016568845.1	peroxidase 47	1.9	5.0	3.6	1.6	1.1	4.3
19	LOC107867619	XP_016569421.1	peroxidase 12	11.3	4.0	3.5	1.1	1.2	3.8
20	LOC107860773	XP_016561733.2	peroxidase 16	1.5	0.9	5.8	5.0	9.2	3.4
21	LOC107840010	XP_016539187.2	peroxidase 43	2.9	4.4	1.3	0.9	1.0	2.9
22	LOC107871140	NP_001311926.1	peroxidase P7 precursor	12.4	3.8	0.2	0.4	0.2	2.0
23	LOC107854094	XP_016554572.2	suberization-associated anionic peroxidase 2	0.9	1.6	1.9	1.7	2.0	1.8
24	LOC107852745	XP_016553271.2	peroxidase 11	0.0	0.0	3.2	17.1	23.6	1.6
25	LOC107877246	XP_016579421.1	peroxidase 27	0.0	2.6	0.0	0.0	0.0	1.3
26	LOC107861303	XP_016562131.1	peroxidase 64	0.5	2.4	0.0	0.0	0.0	1.2
27	LOC107855938	XP_016556426.1	peroxidase 5	12.4	2.0	0.0	0.1	0.1	1.0
28	LOC107853865	XP_016554339.1	lignin-forming anionic peroxidase	0.0	0.8	1.1	3.7	2.4	1.0
29	LOC107840027	NP_001311508.1	peroxidase N1-like	0.4	1.4	0.3	0.3	0.1	0.9
30	LOC107859813	XP_016560411.1	peroxidase 64 isoform X1	10.5	1.6	0.0	0.0	0.0	0.8
31	LOC107860014	XP_047263156.1	peroxidase 72-like	127.9	1.5	0.1	0.2	0.2	0.8
32	LOC107861195	XP_016562044.1	peroxidase 19	5.4	0.8	0.1	0.0	0.0	0.4
33	LOC107842218	XP_047252040.1	peroxidase 21	6.1	0.6	0.2	0.2	0.1	0.4
34	LOC107860012	XP_016560693.2	peroxidase 72	0.3	0.4	0.2	0.3	0.2	0.3
35	LOC107860974	XP_047262346.1	LOW QUALITY PROTEIN: peroxidase 25-like	0.3	0.3	0.3	0.2	0.2	0.3
36	LOC107860047	XP_016560733.1	putative Peroxidase 48 isoform X1	0.5	0.3	0.1	0.2	0.0	0.2
37	LOC124892742	XP_047259915.1	peroxidase 60-like	0.0	0.4	0.0	0.0	0.0	0.2
38	LOC107849622	XP_016549663.1	peroxidase 46	0.9	0.0	0.3	0.2	0.1	0.2
39	LOC107863356	XP_016564726.2	peroxidase 66	1.2	0.2	0.1	0.0	0.0	0.1
40	LOC107874883	XP_016577077.1	peroxidase 41	0.0	0.2	0.0	0.0	0.0	0.1
41	LOC107861812	XP_016562667.1	putative Peroxidase 48	0.1	0.1	0.1	0.3	0.1	0.1

Rank	Locus	Protein	Product	4 DPA	24 DPA	34 DPA	44 DPA	54 DPA	Mean of 24 and 34 DPA
42	LOC107843551	XP_016543351.2	peroxidase 21	0.0	0.2	0.0	0.2	0.4	0.1
43	LOC107863246	XP_016564563.2	peroxidase P7	0.0	0.0	0.1	0.5	0.5	0.1
44	LOC107875418	XP_016577619.1	cationic peroxidase 1	0.0	0.1	0.0	0.0	0.0	0.1
45	LOC107860938	XP_016561858.1	lignin-forming anionic peroxidase	0.0	0.1	0.0	0.0	0.0	0.1
46	LOC107840318	XP_016539637.1	peroxidase 10	0.0	0.0	0.1	0.0	0.0	0.0
47	LOC107850917	XP_016551241.1	peroxidase 41	0.0	0.0	0.1	0.1	0.0	0.0
48	LOC107866095	XP_016567764.2	lignin-forming anionic peroxidase-like	0.0	0.1	0.0	0.0	0.0	0.0
49	LOC107854529	XP_047253315.1	peroxidase 60	0.0	0.1	0.0	0.1	0.1	0.0
50	LOC107868171	XP_047267412.1	peroxidase 72-like	0.0	0.0	0.1	0.8	1.7	0.0
51	LOC107874082	XP_016576451.1	lignin-forming anionic peroxidase	0.0	0.0	0.0	0.0	0.0	0.0
52	LOC107877245	XP_016579420.2	peroxidase 27	0.2	0.0	0.0	0.0	0.0	0.0
53	LOC107861283	XP_016562113.1	peroxidase 7	0.0	0.0	0.0	0.0	0.0	0.0
54	LOC107866096	XP_016567765.1	lignin-forming anionic peroxidase	0.1	0.0	0.0	0.0	0.0	0.0
55	LOC107869386	XP_016571405.1	peroxidase 64	0.0	0.0	0.0	0.0	0.0	0.0
56	LOC107847025	XP_016546761.1	peroxidase 6	0.0	0.0	0.0	0.0	0.0	0.0
57	LOC107847017	XP_016546759.1	peroxidase 44-like	0.0	0.0	0.0	0.0	0.0	0.0
58	LOC107869901	XP_016571788.2	peroxidase 40	0.0	0.0	0.0	0.0	0.0	0.0
59	LOC107848684	XP_016548945.2	peroxidase 28	0.0	0.0	0.0	0.0	0.0	0.0
60	LOC107874093	XP_016576460.1	lignin-forming anionic peroxidase	0.0	0.0	0.0	0.0	0.0	0.0
61	LOC107874527	XP_016576787.1	peroxidase N	0.0	0.0	0.0	0.0	0.0	0.0
62	LOC124897880	XP_047267423.1	peroxidase 29-like	0.0	0.0	0.0	0.0	0.0	0.0
63	LOC124898652	XP_047268295.1	LOW QUALITY PROTEIN: peroxidase 24-like	0.0	0.0	0.0	0.0	0.0	0.0
64	LOC107857270	XP_016557678.1	probable peroxidase 26	0.0	0.0	0.0	0.0	0.0	0.0
65	LOC107859813	XP_016560412.1	peroxidase 64 isoform X2	0.4	0.0	0.0	0.0	0.0	0.0
66	LOC107860047	XP_016560734.1	putative Peroxidase 48 isoform X2	0.0	0.0	0.0	0.0	0.1	0.0
67	LOC107860772	XP_016561732.2	peroxidase 55	0.0	0.0	0.0	0.0	0.0	0.0
68	LOC107863896	XP_016565574.1	peroxidase 44	39.3	0.0	0.0	0.0	0.0	0.0
69	LOC107852267	XP_016552813.1	peroxidase 5-like	0.0	0.0	0.0	0.0	0.0	0.0
70	LOC107867709	XP_016569562.1	peroxidase 43 isoform X2	0.0	0.0	0.0	0.0	0.0	0.0
71	LOC107869409	XP_016571432.1	peroxidase 9	0.0	0.0	0.0	0.0	0.0	0.0
72	LOC107868538	XP_016570720.1	peroxidase 9-like	0.0	0.0	0.0	0.0	0.0	0.0
73	LOC107853869	XP_047268512.1	lignin-forming anionic peroxidase-like isoform X1	0.0	0.0	0.0	0.0	0.0	0.0
74	LOC107853869	XP_047268513.1	lignin-forming anionic peroxidase-like isoform X2	0.0	0.0	0.0	0.0	0.0	0.0
75	LOC107866093	XP_016567762.2	lignin-forming anionic peroxidase-like	0.0	0.0	0.0	0.0	0.0	0.0
76	LOC107866094	XP_016567763.1	lignin-forming anionic peroxidase-like	0.0	0.0	0.0	0.0	0.0	0.0
77	LOC107865010	XP_016566867.2	lignin-forming anionic peroxidase-like	0.0	0.0	0.0	0.0	0.0	0.0

Rank	Locus	Protein	Product	4 DPA	24 DPA	34 DPA	44 DPA	54 DPA	Mean of 24 and 34 DPA
78	LOC107866097	NP_001311841.1	lignin-forming anionic peroxidase-like precursor	0.0	0.0	0.0	0.0	0.0	0.0
79	LOC107865011	XP_047268035.1	lignin-forming anionic peroxidase-like	0.0	0.0	0.0	0.1	0.0	0.0
80	LOC107872526	XP_016574676.1	cationic peroxidase 1	0.4	0.0	0.0	0.0	0.0	0.0
81	LOC107871376	XP_016573785.2	peroxidase 5	0.1	0.0	0.0	0.0	0.0	0.0
82	LOC107877215	XP_016579397.1	peroxidase 3	0.0	0.0	0.0	0.0	0.0	0.0
83	LOC107838724	XP_016537380.2	peroxidase 57	0.0	0.0	0.0	0.0	0.0	0.0
84	LOC107872642	XP_016574759.2	peroxidase 5	0.0	0.0	0.0	0.0	0.0	0.0
85	LOC107841908	XP_016541245.1	peroxidase 27	0.1	0.0	0.0	0.0	0.0	0.0
86	LOC107841822	XP_016541157.1	peroxidase 44	0.0	0.0	0.0	0.0	0.0	0.0
87	LOC107844910	XP_047253817.1	peroxidase 2-like	0.0	0.0	0.0	0.0	0.0	0.0
88	LOC107844773	XP_016544608.2	cationic peroxidase 1	0.0	0.0	0.0	0.0	0.0	0.0
89	LOC107844563	XP_047253737.1	peroxidase 21-like	0.0	0.0	0.0	0.0	0.1	0.0
90	LOC124887872	XP_047253747.1	peroxidase 63-like	0.3	0.0	0.0	0.3	1.3	0.0
91	LOC107852613	XP_016553130.2	peroxidase 15	0.1	0.0	0.0	0.0	0.0	0.0
92	LOC107849754	XP_047256950.1	peroxidase 64	0.0	0.0	0.0	0.0	0.0	0.0
93	LOC124885353	XP_047257150.1	peroxidase 64-like	0.0	0.0	0.0	0.0	0.0	0.0
94	LOC107877264	XP_016579433.1	peroxidase 3	0.0	0.0	0.0	0.0	0.0	0.0

Table S4. Placenta libraries. The five public libraries: run, sample, species record, stage, read pairs, mapping rate and transcripts with TPM above zero. Data: `data/tables/S04.csv`*; script:* `code/02_expression.py`*.*

Run	Sample	Species (SRA record)	DPA	Read pairs	Mapped (%)	Transcripts with TPM > 0
SRR17234805	SAMN24045716	Capsicum chinense	4	27.2M	83.6	29,114
SRR17234799	SAMN24045720	Capsicum chinense	24	25.3M	83.8	29,712
SRR17234797	SAMN24045722	Capsicum chinense	34	28.7M	81.8	29,020
SRR17234795	SAMN24045724	Capsicum chinense	44	26.1M	80.8	28,584
SRR17234803	SAMN24045726	Capsicum chinense	54	26.7M	84.0	28,036

Table S5. Capacity, saturation and control of every step. Four settings (methylation assignment and capacity accounting): capacity, output, saturation and capacity control coefficient of every step. Data: `data/tables/S05.csv`*; script:* `code/03_flux.py`*.*

Methylation assignment	Capacity accounting	Step	Capacity	Output	Saturation	Control coefficient	Limiting step
CCoAOMT-type	representative	PAL	241.6213	56.4205	0.2335	0.00	
CCoAOMT-type	representative	C4H	214.7814	56.4205	0.2627	0.00	
CCoAOMT-type	representative	4CL	56.4205	56.4205	1.0000	1.00	yes
CCoAOMT-type	representative	HCT	63.1230	56.4205	0.8938	0.00	
CCoAOMT-type	representative	Methylation	475.9295	56.4205	0.1185	0.00	
CCoAOMT-type	representative	VR (reduction)	118.8183	56.4205	0.4748	0.00	
CCoAOMT-type	representative	pAMT	1416.6469	56.4205	0.0398	0.00	
CCoAOMT-type	representative	BCAT (valine)	277.8304	56.4205	0.2031	0.00	

Methylation assignment	Capacity accounting	Step	Capacity	Output	Saturation	Control coefficient	Limiting step
CCoAOMT-type	representative	KAS (C9)	813.8303	56.4205	0.0693	0.00	
CCoAOMT-type	representative	Pun1/AT3 (shared pool)	726.9181	56.4205	0.0776	0.00	
CCoAOMT-type	family-sum	PAL	580.1543	67.2792	0.1160	0.00	
CCoAOMT-type	family-sum	C4H	214.7814	67.2792	0.3132	0.00	
CCoAOMT-type	family-sum	4CL	112.2183	67.2792	0.5995	0.00	
CCoAOMT-type	family-sum	HCT	67.2792	67.2792	1.0000	1.00	yes
CCoAOMT-type	family-sum	Methylation	763.8764	67.2792	0.0881	0.00	
CCoAOMT-type	family-sum	VR (reduction)	287.1913	67.2792	0.2343	0.00	
CCoAOMT-type	family-sum	pAMT	1416.6469	67.2792	0.0475	0.00	
CCoAOMT-type	family-sum	BCAT (valine)	388.2914	67.2792	0.1733	0.00	
CCoAOMT-type	family-sum	KAS (C9)	1223.4899	67.2792	0.0550	0.00	
CCoAOMT-type	family-sum	Pun1/AT3 (shared pool)	728.1172	67.2792	0.0924	0.00	
COMT-class	representative	PAL	241.6213	56.4205	0.2335	0.00	
COMT-class	representative	C4H	214.7814	56.4205	0.2627	0.00	
COMT-class	representative	4CL	56.4205	56.4205	1.0000	1.00	yes
COMT-class	representative	HCT	63.1230	56.4205	0.8938	0.00	
COMT-class	representative	Methylation	177.2928	56.4205	0.3182	0.00	
COMT-class	representative	VR (reduction)	118.8183	56.4205	0.4748	0.00	
COMT-class	representative	pAMT	1416.6469	56.4205	0.0398	0.00	
COMT-class	representative	BCAT (valine)	277.8304	56.4205	0.2031	0.00	
COMT-class	representative	KAS (C9)	813.8303	56.4205	0.0693	0.00	
COMT-class	representative	Pun1/AT3 (shared pool)	726.9181	56.4205	0.0776	0.00	
COMT-class	family-sum	PAL	580.1543	67.2792	0.1160	0.00	
COMT-class	family-sum	C4H	214.7814	67.2792	0.3132	0.00	
COMT-class	family-sum	4CL	112.2183	67.2792	0.5995	0.00	
COMT-class	family-sum	HCT	67.2792	67.2792	1.0000	1.00	yes
COMT-class	family-sum	Methylation	299.1625	67.2792	0.2249	0.00	
COMT-class	family-sum	VR (reduction)	287.1913	67.2792	0.2343	0.00	
COMT-class	family-sum	pAMT	1416.6469	67.2792	0.0475	0.00	
COMT-class	family-sum	BCAT (valine)	388.2914	67.2792	0.1733	0.00	
COMT-class	family-sum	KAS (C9)	1223.4899	67.2792	0.0550	0.00	
COMT-class	family-sum	Pun1/AT3 (shared pool)	728.1172	67.2792	0.0924	0.00	

Table S6. Knockout and scaling runs of the flux model. Output when one step is set to 0, half and double its capacity; the limiting step is listed. Data: `data/tables/S06.csv`*; script:* `code/03_flux.py`*.*

Capacity accounting	Step changed	Factor	Output	Output over base	Limiting step
representative	PAL	0	0.00	0.0000	PAL
representative	PAL	0.5	56.42	1.0000	4CL
representative	PAL	2	56.42	1.0000	4CL
representative	C4H	0	0.00	0.0000	C4H
representative	C4H	0.5	56.42	1.0000	4CL
representative	C4H	2	56.42	1.0000	4CL
representative	4CL	0	0.00	0.0000	4CL
representative	4CL	0.5	28.21	0.5000	4CL
representative	4CL	2	63.12	1.1188	HCT
representative	HCT	0	0.00	0.0000	HCT
representative	HCT	0.5	31.56	0.5594	HCT
representative	HCT	2	56.42	1.0000	4CL
representative	Methylation	0	0.00	0.0000	Methylation
representative	Methylation	0.5	56.42	1.0000	4CL
representative	Methylation	2	56.42	1.0000	4CL
representative	VR	0	0.00	0.0000	VR (reduction)
representative	VR	0.5	56.42	1.0000	4CL
representative	VR	2	56.42	1.0000	4CL
representative	pAMT	0	0.00	0.0000	pAMT
representative	pAMT	0.5	56.42	1.0000	4CL
representative	pAMT	2	56.42	1.0000	4CL
representative	Pun1/AT3 pool	0	0.00	0.0000	Pun1/AT3 (shared pool)
representative	Pun1/AT3 pool	0.5	56.42	1.0000	4CL
representative	Pun1/AT3 pool	2	56.42	1.0000	4CL
family-sum	PAL	0	0.00	0.0000	PAL

Capacity accounting	Step changed	Factor	Output	Output over base	Limiting step
family-sum	PAL	0.5	67.28	1.0000	HCT
family-sum	PAL	2	67.28	1.0000	HCT
family-sum	C4H	0	0.00	0.0000	C4H
family-sum	C4H	0.5	67.28	1.0000	HCT
family-sum	C4H	2	67.28	1.0000	HCT
family-sum	4CL	0	0.00	0.0000	4CL
family-sum	4CL	0.5	56.11	0.8340	4CL
family-sum	4CL	2	67.28	1.0000	HCT
family-sum	HCT	0	0.00	0.0000	HCT
family-sum	HCT	0.5	33.64	0.5000	HCT
family-sum	HCT	2	112.22	1.6679	4CL
family-sum	Methylation	0	0.00	0.0000	Methylation
family-sum	Methylation	0.5	67.28	1.0000	HCT
family-sum	Methylation	2	67.28	1.0000	HCT
family-sum	VR	0	0.00	0.0000	VR (reduction)
family-sum	VR	0.5	67.28	1.0000	HCT
family-sum	VR	2	67.28	1.0000	HCT
family-sum	pAMT	0	0.00	0.0000	pAMT
family-sum	pAMT	0.5	67.28	1.0000	HCT
family-sum	pAMT	2	67.28	1.0000	HCT
family-sum	Pun1/AT3 pool	0	0.00	0.0000	Pun1/AT3 (shared pool)
family-sum	Pun1/AT3 pool	0.5	67.28	1.0000	HCT
family-sum	Pun1/AT3 pool	2	67.28	1.0000	HCT

Table S7. Pun1/AT3 pool sweep. Output when the Pun1/AT3 pool is scaled; the chain rule is compared with the stored solver values. Data: `data/tables/S07.csv`*; script:* `code/03_flux.py`*.*

Pool multiplier	Pool capacity	Output (chain rule)	Output (stored solver)	Absolute difference
0.01	7.2692	7.2692	7.269181	0.000000
0.02	14.5384	14.5384	14.538362	0.000000
0.03	21.8075	21.8075	21.807542	0.000001
0.04	29.0767	29.0767	29.076723	0.000001
0.05	36.3459	36.3459	36.345904	0.000001
0.06	43.6151	43.6151	43.615085	0.000001
0.07	50.8843	50.8843	50.884265	0.000002
0.08	58.1534	56.4205	56.420478	0.000022
0.09	65.4226	56.4205	56.420478	0.000022
0.1	72.6918	56.4205	56.420478	0.000022
0.2	145.3836	56.4205	56.420478	0.000022
0.3	218.0754	56.4205	56.420478	0.000022
0.5	363.4590	56.4205	56.420478	0.000022
0.75	545.1886	56.4205	56.420478	0.000022
1.0	726.9181	56.4205	56.420478	0.000022
1.25	908.6476	56.4205	56.420478	0.000022
1.5	1090.3772	56.4205	56.420478	0.000022
2.0	1453.8362	56.4205	56.420478	0.000022

Table S8. Limiting steps of the perturbation runs. Share of the draws in which each step sets the output, in four settings. Data: `data/tables/S08.csv`*; script:* `code/03_flux.py`*.*

Methylation assignment	Capacity accounting	Limiting step	Share of draws
CCoAOMT-type	representative	4CL	0.5396
CCoAOMT-type	representative	HCT	0.3832
CCoAOMT-type	representative	VR	0.066
CCoAOMT-type	representative	C4H	0.007

Methylation assignment	Capacity accounting	Limiting step	Share of draws
CCoAOMT-type	representative	PAL	0.0034
CCoAOMT-type	representative	BCAT_val	0.0012
CCoAOMT-type	family-sum	HCT	0.7434
CCoAOMT-type	family-sum	4CL	0.2196
CCoAOMT-type	family-sum	C4H	0.0286
CCoAOMT-type	family-sum	VR	0.0082
CCoAOMT-type	family-sum	BCAT_val	0.0006
COMT-class	representative	4CL	0.5162
COMT-class	representative	HCT	0.4054
COMT-class	representative	VR	0.0588
COMT-class	representative	COMT	0.0102
COMT-class	representative	C4H	0.0056
COMT-class	representative	PAL	0.0026
COMT-class	representative	BCAT_val	0.0018
COMT-class	family-sum	HCT	0.726
COMT-class	family-sum	4CL	0.228
COMT-class	family-sum	C4H	0.0298
COMT-class	family-sum	VR	0.0096
COMT-class	family-sum	COMT	0.0062
COMT-class	family-sum	BCAT_val	0.0018

Table S9. Genome sites with up to four mismatches. Every site found by the whole-genome scan (strict NGG PAM, both strands). Data: `data/tables/S09.csv`*; script:* `code/04_guides.py`*.*

Guide	Chromosome accession	Position of the 23-mer	Strand	Mismatches	Mismatch positions (5' end = 1)
POX-107856092-g1	NC_061111.1	32673228	-	0	none
POX-107856092-g1	NC_061111.1	86344263	+	4	2+8+12+14
POX-107856092-g1	NC_061115.1	78959003	+	4	3+9+13+14
POX-107856092-g1	NC_061116.1	182137776	-	4	2+12+16+19
POX-107856092-g1	NC_061120.1	19565716	+	4	11+12+13+14
POX-107856092-g1	NC_061120.1	174570161	+	4	6+12+16+20
POX-107856092-g1	NC_061122.1	182792299	-	4	2+10+11+20
POX-107856092-g1	NW_025845299.1	1230	-	4	2+8+12+14
POX-107856092-g2	NC_061111.1	32672639	+	0	none
POX-107856092-g2	NC_061116.1	213893413	+	3	9+15+20
POX-107856092-g2	NC_061121.1	191709795	-	3	10+13+20
POX-107856092-g2	NC_061111.1	148614837	-	4	1+2+17+20
POX-107856092-g2	NC_061111.1	158817893	+	4	1+15+17+20
POX-107856092-g2	NC_061112.1	17031346	+	4	6+14+15+19
POX-107856092-g2	NC_061112.1	83297680	-	4	5+6+17+20
POX-107856092-g2	NC_061112.1	149606781	+	4	3+13+14+16

Guide	Chromosome accession	Position of the 23-mer	Strand	Mismatches	Mismatch positions (5' end = 1)
POX-107856092-g2	NC_061113.1	52799390	+	4	2+3+4+17
POX-107856092-g2	NC_061113.1	223048636	-	4	7+11+13+19
POX-107856092-g2	NC_061114.1	42686331	+	4	1+5+6+15
POX-107856092-g2	NC_061115.1	76248949	+	4	11+12+13+19
POX-107856092-g2	NC_061116.1	199335734	-	4	9+11+13+19
POX-107856092-g2	NC_061116.1	212731004	-	4	5+9+15+20
POX-107856092-g2	NC_061116.1	213237087	+	4	5+9+15+20
POX-107856092-g2	NC_061116.1	213487048	+	4	5+9+15+20
POX-107856092-g2	NC_061116.1	213605931	+	4	5+9+15+20
POX-107856092-g2	NC_061116.1	213662364	+	4	5+9+15+20
POX-107856092-g2	NC_061116.1	213707228	+	4	5+9+15+20
POX-107856092-g2	NC_061116.1	221005434	-	4	11+14+15+16
POX-107856092-g2	NC_061118.1	97267372	+	4	4+10+16+17
POX-107856092-g2	NC_061119.1	205209780	-	4	9+11+13+19
POX-107856092-g2	NC_061120.1	101636003	-	4	7+8+15+18
POX-107856092-g2	NC_061120.1	133748111	+	4	5+7+12+20
POX-107856092-g2	NC_061120.1	144320566	-	4	10+13+15+19

Guide	Chromosome accession	Position of the 23-mer	Strand	Mismatches	Mismatch positions (5' end = 1)
POX-107856092-g2	NC_061120.1	182307598	-	4	3+11+13+19
POX-107856092-g2	NC_061120.1	189080887	+	4	9+11+12+19
POX-107856092-g2	NC_061120.1	194271466	-	4	1+7+11+20
POX-107856092-g2	NC_061120.1	202391527	+	4	3+4+7+19
POX-107856092-g2	NC_061121.1	18917354	+	4	2+6+7+15
POX-107856092-g2	NC_061121.1	50348159	-	4	5+10+14+15
POX-107856092-g2	NC_061121.1	170783536	-	4	4+7+10+18
POX-107856092-g2	NC_061122.1	8017233	+	4	11+13+14+19
POX-107856092-g2	NC_061122.1	49643914	-	4	1+8+17+19
POX-107856092-g2	NC_061122.1	125727356	-	4	4+6+11+20
POX-107856092-g2	NC_061122.1	138755916	-	4	1+6+14+15
POX-107856092-g2	NW_025838913.1	10469	-	4	11+13+14+15
POX-107856092-g2	NW_025853076.1	23568	+	4	5+9+15+20
POX-107856092-g2	NW_025853076.1	66023	+	4	5+9+15+20
POX-107856092-g2	NW_025853076.1	142993	+	4	5+9+15+20
POX-107856092-g2	NW_025862766.1	1224	+	4	5+9+15+20

Guide	Chromosome accession	Position of the 23-mer	Strand	Mismatches	Mismatch positions (5' end = 1)
POX-107856092-g2	NW_025879081.1	105	-	4	5+9+15+20
POX-107856092-g2	NW_025888827.1	372472	+	4	6+12+13+17

Table S10. Annotated sites with up to three mismatches. Sites with up to three mismatches and the gene each falls in. Data: `data/tables/S10.csv`*; script:* `code/04_guides.py`*.*

Guide	Mismatches	Mismatch positions	Site (plus-strand coordinates)	Strand	Site sequence (PAM after the space)	Overlapping gene
POX-107856092-g2	3	9+15+20	NC_061116.1:213893413-213893435	+	CATGGCCATGATGTAAGCAA AGG	intergenic
POX-107856092-g2	3	10+13+20	NC_061121.1:191709795-191709817	-	CATGGCCAATATTTCAGCAG CGG	intergenic

Table S11. MIT position weights. Position weights of the Hsu et al. (2013) score; position 1 is the PAM-distal end. Data: `data/tables/S11.csv`*; script:* `code/04_guides.py`*.*

Guide position (5' end = 1)	Weight (w)
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

Guide position (5' end = 1)	Weight (w)
15	0.732
16	0.828
17	0.615
18	0.804
19	0.685
20	0.583

Table S12. Guide matches in other assemblies. Best strict-NGG match of each guide in three other Capsicum assemblies (up to two mismatches). Data: `data/tables/S12.csv`*; script:* `code/04_guides.py`*.*

Guide	Assembly	Fewest mismatches (strict NGG)	Sites with 0 mismatches	Sites with 1 mismatch	Sites with 2 mismatches
POX-107856092-g1	C_chinense_PI159236_GCA_002271895.2	1	0	1	0
POX-107856092-g2	C_chinense_PI159236_GCA_002271895.2	0	1	0	0
POX-107856092-g1	C_baccatum_PBC81_GCA_002271885.2	0	1	0	0
POX-107856092-g2	C_baccatum_PBC81_GCA_002271885.2	0	1	0	0
POX-107856092-g1	C_chinense_Superhot_GCA_054135625.1	1	0	1	0
POX-107856092-g2	C_chinense_Superhot_GCA_054135625.1	0	1	0	0

Table S13. Checks on the knockout cassette. Every check with its value and result. Data: `data/tables/S13.csv`*; script:* `code/05_cassette.py`.

Check	Value	Result
Cassette length (bp)	5446	pass
Length arithmetic: 448 + 1 + 20 + 76 + 7 = 552 (guide unit); 435 + 4203 + 256 = 4894 (Cas9 unit)	552; 4894; total 5446	pass
Guide in the cassette equals the specified guide, found once at base	450	pass
Scaffold equals the standard SpCas9 sgRNA scaffold (76 nt)	76	pass
Base after the AtU6-26 promoter in X52528.1 (start of the U6 transcript)	G	pass
Cas9 reading frame: start codon, length in bases, divisible by 3	ATG; 4203; True	pass
Stop codons in the Cas9 frame (only the last codon)	TAG; internal stops: 0	pass
Translation equals the GenBank translation (AKE81020.1), residues	1400	pass
Protein composition: N-terminal residues + SpCas9 residues 2-1368 + C-terminal residues	17 + 1367 + 16	pass
SV40 NLS PKKKRKV in the N-terminal residues; nucleoplasmin-type KRPAATKKAGQAKKKK in the C-terminal residues	MAPKKKRKVGIHGVPAA; KRPAATKKAGQAKKKK	pass
Cas9 coding sequence GC (%); GC at third codon positions (%)	54.2; 78.6	pass
35S block (forward) found at base; reverse complement of the block present	553; no	pass
35S block bases identical to CaMV V00141.1 7213-7647 (forward); reverse complement	433 of 435; 111 of 435	pass
35S block: bases identical to CaMV in the promoter part (block 1-221, CaMV 7213-7433) and in the leader part (block 222-435, CaMV 7434-7647)	219 of 221; 214 of 214	pass
ATG codons in the 35S block ahead of Cas9 (block positions); reading frames reaching the Cas9 start	25,52,125,151,273,326,355; 0	pass
Promoter faces Cas9: block upstream of the Cas9 start on the same strand; Cas9 start base	988	pass
Junction guide unit | 35S block (bases 541-558)	GGTGCTTTTTTTGATAAA	pass
Junction 35S block | Cas9 start (bases 975-995)	TAAAATCCAGATCATGGCTCC	pass
Junction Cas9 stop | NOS (bases 5180-5200)	AAAAGAAGTAGGATCGTTCAA	pass

Table S14. Restriction-site counts of the cassette. Sites on both strands and the first eight positions. Data: `data/tables/S14.csv`*; script:* `code/05_cassette.py`*.*

Enzyme	Recognition site	Sites in the cassette	Positions (first eight)
BsaI	GGTCTC	0	
BsmBI/Esp3I	CGTCTC	6	1204,1489,1696,3802,3921,4225
BpiI/BbsI	GAAGAC	3	640,657,4323
HindIII	AAGCTT	5	1,1456,1822,2209,4042
EcoRI	GAATTC	0	
XbaI	TCTAGA	0	
SphI	GCATGC	0	
BamHI	GGATCC	0	
PstI/SbfI	CTGCAG	0	
XmaI/SmaI	CCCGGG	0	
SalI	GTCGAC	4	1312,1402,1651,3547
SacI	GAGCTC	5	2602,3514,4519,4648,4690
KpnI	GGTACC	2	2809,4015

Table S15. All docking runs. Reported and re-scored Vina scores, distance of the pose centroid to the crystal ligand, nearest contact distance and ring distance for every run. Data: `data/tables/S15.csv`*; script:* `code/06_docking.py`*.*

Run	Reported score	Re-scored (score-only)	Centroid to crystal ligand (A)	Nearest ligand O/N to Y511, S512, T550, R557 (A)	Docked ring to crystal vanillyl ring (A)
capsaicin_old_box_s1	-6.841	-6.838	14.8	13.4	20.0
capsaicin_old_box_s2	-6.462	-6.461	14.0	15.0	20.5
capsaicin_old_box_s3	-6.678	-6.678	14.6	14.4	20.2
capsaicin_vanilloid_s1	-6.036	-6.029	3.0	3.0	0.1
capsaicin_vanilloid_s2	-6.397	-6.392	3.0	2.9	0.6
capsaicin_vanilloid_s3	-6.207	-6.206	2.8	3.0	1.1
dihydrocapsaicin_old_box_s1	-6.614	-6.613	14.9	13.5	20.0

Run	Reported score	Re-scored (score-only)	Centroid to crystal ligand (A)	Nearest ligand O/N to Y511, S512, T550, R557 (A)	Docked ring to crystal vanillyl ring (A)
dihydrocapsaicin_old_box_s2	-6.439	-6.439	14.4	13.8	20.6
dihydrocapsaicin_old_box_s3	-6.701	-6.7	14.8	13.3	20.0
dihydrocapsaicin_vanilloid_s1	-5.966	-5.962	3.2	3.0	0.1
dihydrocapsaicin_vanilloid_s2	-6.112	-6.113	3.1	2.9	0.8
dihydrocapsaicin_vanilloid_s3	-6.009	-6.009	3.0	2.8	0.7
homodihydrocapsaicin_old_box_s1	-5.595	-5.594	13.1	12.9	16.5
homodihydrocapsaicin_old_box_s2	-6.559	-6.562	14.7	13.0	20.0
homodihydrocapsaicin_old_box_s3	-6.247	-6.247	14.3	14.1	20.2
homodihydrocapsaicin_vanilloid_s1	-5.677	-5.678	2.8	3.0	0.8
homodihydrocapsaicin_vanilloid_s2	-5.806	-5.806	2.9	2.9	0.7
homodihydrocapsaicin_vanilloid_s3	-5.715	-5.711	2.7	2.8	0.6
nonivamide_old_box_s1	-6.674	-6.673	14.8	12.9	20.0
nonivamide_old_box_s2	-6.196	-6.197	14.8	15.5	20.7
nonivamide_old_box_s3	-6.428	-6.43	14.7	12.9	20.1
nonivamide_vanilloid_s1	-5.553	-5.553	3.0	2.9	0.2
nonivamide_vanilloid_s2	-5.271	-5.27	1.6	3.2	7.5
nonivamide_vanilloid_s3	-5.802	-5.8	3.1	3.1	0.3
nordihydrocapsaicin_old_box_s1	-6.51	-6.509	14.7	14.0	20.0
nordihydrocapsaicin_old_box_s2	-6.694	-6.695	14.8	13.1	20.0
nordihydrocapsaicin_old_box_s3	-6.603	-6.602	14.6	13.6	20.5
nordihydrocapsaicin_vanilloid_s1	-5.447	-5.449	3.0	3.0	8.3
nordihydrocapsaicin_vanilloid_s2	-5.667	-5.666	3.3	3.0	0.3
nordihydrocapsaicin_vanilloid_s3	-5.831	-5.83	2.9	3.1	1.1

Supplementary data: file map

Path	Contents
`data/tables/T01.csv`	Table 1: Blend and SHU contributions
`data/tables/T02.csv`	Table 2: Series of ten cultivars
`data/tables/T03.csv`	Table 3: Placenta expression
`data/tables/T04.csv`	Table 4: Capacities of the pathway steps
`data/tables/T05.csv`	Table 5: Perturbation runs
`data/tables/T06.csv`	Table 6: Guides for the peroxidase gene
`data/tables/T07.csv`	Table 7: Parts of the complete knockout cassette
`data/tables/T08.csv`	Table 8: Docking scores
`data/tables/S01.csv`	Table S1: Quantities of the blend arithmetic
`data/tables/S02.csv`	Table S2: Sensitivity of the content to the blend inputs
`data/tables/S03.csv`	Table S3: Class III peroxidase family in the placenta
`data/tables/S04.csv`	Table S4: Placenta libraries
`data/tables/S05.csv`	Table S5: Capacity, saturation and control of every step
`data/tables/S06.csv`	Table S6: Knockout and scaling runs of the flux model
`data/tables/S07.csv`	Table S7: Pun1/AT3 pool sweep
`data/tables/S08.csv`	Table S8: Limiting steps of the perturbation runs
`data/tables/S09.csv`	Table S9: Genome sites with up to four mismatches
`data/tables/S10.csv`	Table S10: Annotated sites with up to three mismatches
`data/tables/S11.csv`	Table S11: MIT position weights
`data/tables/S12.csv`	Table S12: Guide matches in other assemblies
`data/tables/S13.csv`	Table S13: Checks on the knockout cassette
`data/tables/S14.csv`	Table S14: Restriction-site counts of the cassette
`data/tables/S15.csv`	Table S15: All docking runs
`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/F04a.csv`	Data behind Figure 4
`data/figures/F04b.csv`	Data behind Figure 4
`data/figures/F05.csv`	Data behind Figure 5
`data/figures/F06.csv`	Data behind Figure 6
`data/figures/F07.csv`	Data behind Figure 7
`data/figures/F08.csv`	Data behind Figure 8
`data/figures/F09.csv`	Data behind Figure 9
`data/figures/F10a.csv`	Data behind Figure 10
`data/figures/F10b.csv`	Data behind Figure 10

Path	Contents
`data/results/blend_calculation.csv`	SHU contribution and content (mg/g dry weight) of each capsaicinoid, with the blend total
`data/results/blend_summary.csv`	SHU target, blend-weighted coefficient, content (mg/g and % dry weight) and ratio to the comparison value
`data/results/series_position.csv`	The ten cultivars of the series ranked by SHU target
`data/results/blend_sensitivity.csv`	Content when one blend fraction or one SHU coefficient moves over its range
`data/results/expression_summary.csv`	Placenta TPM at five stages of 13 transcripts, to three decimals, with the peak stage
`data/results/peroxidase_rank.csv`	Rank and TPM of the knockout target among the class III peroxidase-family records
`data/results/pun1_group.csv`	TPM of the two Pun1 loci and their sum at five stages
`data/results/flux_model_summary.csv`	Size of the pathway model and the capacity scale
`data/results/flux_steps.csv`	Capacity, output, saturation and capacity control coefficient of each step in four settings
`data/results/flux_knockout_runs.csv`	Output when one step is set to 0, half and double capacity
`data/results/pun1_capacity_sweep.csv`	Output when the Pun1/AT3 pool is scaled from 0.01 to 2 times
`data/results/pun1_limit_fraction.csv`	Share of the pool capacity at which Pun1/AT3 would limit the output
`data/results/sweep_agreement.csv`	Sweep output from the chain rule against the stored solver values
`data/results/flux_monte_carlo.csv`	Summary of the stored perturbation runs (5,000 draws per setting)
`data/results/flux_sensitivity.csv`	Output when one step capacity moves to half and to double
`data/results/guide_table.csv`	The guides: gene, position, protospacer, PAM, GC, cut position in the coding sequence, MIT score and site counts
`data/results/guides.fasta`	The guide sequences with gene, position, strand and PAM
`data/results/scan_hits.tsv`	Every genome site with up to four mismatches to the guides
`data/results/gene_model_summary.csv`	Gene model of the knockout target: exons, coding bases, residues, translation check
`data/results/guide_agreement.csv`	Recount of sites and MIT scores from the scan hits against the stored table
`data/results/guide_pair.csv`	Coding bases between the two cuts
`data/results/knockout_cassette.fasta`	The complete knockout cassette (5,446 bp), one record per knockout construct
`data/results/cassette_parts.csv`	Start, end and length of each part of the cassette
`data/results/cassette_checks.csv`	Checks on the cassette: frame, translation, orientation, identity with CaMV, junctions
`data/results/cassette_sites.csv`	Restriction-site counts and positions in the cassette

Path	Contents
`data/results/cassette_summary.csv`	Cassette quantities used in the text
`data/results/overexpression_cassettes.*`	Stored overexpression cassettes of the cultivar, when listed in the specification
`data/results/docking_summary.csv`	Mean, lowest and highest Vina score per ligand and box over three seeds, with centroid distance to the crystal ligand
`data/results/docking_runs.csv`	Best score, centroid distance, contact residues and hydrogen-bond candidates of every run
`data/results/docking_geometry.csv`	Ring overlap, nearest contact distances, box distance and re-scoring agreement
`data/results/docking_receptor.csv`	Receptor chain, atom and residue counts and run settings
`data/results/full_scan_comparison.csv`	Full mode only: comparison of the repeated genome scan with the stored scan
`data/wet_lab_active.csv`	Experiments active in the authors' laboratories for this paper: register identifier (WL-01 to WL-84), experiment name and stage code (S0 paperwork, S1 calibration and cheap decisive tests, S2 transient and in vitro mechanism tests, S3 stable lines, S4 tests on engineered material, S5 optional and long-lead)
`inputs/cultivar_spec.json`	The cultivar specification: name, code, tier, SHU target, blend fractions, SHU coefficients, comparison value, knockout target and guides, constructs (the only cultivar-specific input)
`inputs/input_sources.csv`	Public inputs: genome, RNA-seq series, GenBank and UniProt records, TRPV1 structure
`inputs/series_shu_targets.csv`	SHU design targets and capsaicinoid content of the ten cultivars of the series
`inputs/expression/*`	Placenta TPM tables (selected transcripts, target loci, peroxidase family), library table, run table, transcript labels and a note on the Pun1 loci
`inputs/flux/*`	Stoichiometric pathway model (37 metabolites, 44 reactions), step capacities in four settings and the stored perturbation summary
`inputs/guides/*`	Guide table, gene model of the target locus, stored genome scan hits (up to four mismatches), annotated sites (up to three mismatches), matches in other assemblies
`inputs/cassette/*`	Stored cassette parts and public records: construct sequences, NOS terminator, X52528.1, KR029112.1, SpCas9 protein Q99ZW2, CaMV V00141.1
`inputs/docking/*`	Stored Vina results, re-scoring and geometry table, docking parameters, receptor chain A of 5IRX
`code/cultivar_lib.py`	Reads inputs/cultivar_spec.json and formats numbers (round half up)
`code/01_blend.py`	Blend arithmetic and sensitivity
`code/02_expression.py`	Expression tables
`code/03_flux.py`	Flux model as a ranking of capacities
`code/04_guides.py`	Guide table from the stored genome scan
`code/05_cassette.py`	Knockout cassette assembly and checks
`code/06_docking.py`	Docking tables and geometry
`code/07_text.py`	Claims table and paper source
`code/08_full_guide_scan.py`	Full-mode genome scan of the guides
`code/text/paper_template.html`	Text template of the paper with claim markers; the paper source is written from it
`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
`research_record/scripts/flux/*`	Stored pathway-model scripts: model library, perturbation run, checks, knockout runs and the second perturbation run with a different solver

Path	Contents
`research_record/scripts/cassette/*`	First route of the cassette assembly and checks (own codon table, reading frame, translation, orientation)
`research_record/scripts/docking/*`	Docking run script (Vina 1.2.5, Meeko, RDKit)
`research_record/scripts/guides/*`	Recount of the guide sites and MIT scores from the scan hits
`research_record/outputs/*`	Raw outputs of the scripts above
`research_record/second_script_rerun/*`	Independent second scripts and their outputs: linear-programme check of the chain rule, six-frame check of the cassette, re-scoring and geometry of the docking poses, independent guide verification

Cite this paper

Ford, P.I. (2026). Ember: In Silico Design of a Peroxidase-Knockout Capsicum annuum Cultivar Targeting 866,611 Scoville Heat Units. Scoville Splice Research Series. https://doi.org/10.5281/zenodo.24511362

```
@misc{Ford2026_C01,
  author = {Ford, P.I.},
  title = {Ember: In Silico Design of a Peroxidase-Knockout Capsicum annuum Cultivar
Targeting 866,611 Scoville Heat Units},
  year = {2026},
  publisher = {Scoville Splice LLC},
  series = {Scoville Splice Research Series},
  doi = {10.5281/zenodo.24511362},
  url = {https://doi.org/10.5281/zenodo.24511362}
}
```

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
P04	The Iyengar Flavor Index (IFI): A Receptor-Normalized Standard for Cross-Modal Flavor Measurement	10.5281/zenodo.24511320
P11	The Capsaicin Design Platform: Architecture, Sixteen Design Tools and the Flavor Index Framework of a Subscription Service	10.5281/zenodo.24511346
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
C01-C001	SHU design target of the cultivar	866,611	design target	`inputs/cultivar_spec.json`	field shu_target
C01-C002	Capsaicinoid content implied by the SHU target and the blend (mg/g dry weight)	62.4 mg/g	computed	`data/results/blend_summary.csv`	row content_mg_per_g_dw
C01-C003	Capsaicinoid content implied by the SHU target and the blend (% dry weight)	6.24%	computed	`data/results/blend_summary.csv`	row content_pct_dw
C01-C004	Content divided by the comparison value	0.59	computed	`data/results/blend_summary.csv`	row content_over_comparison
C01-C005	Comparison value used as a scale for the content	106 mg/g	design assumption	`inputs/cultivar_spec.json`	field comparison_mg_per_g
C01-C006	Share of the Pun1/AT3 pool capacity below which Pun1/AT3 limits the output, representative	7.76%	computed	`data/results/pun1_limit_fraction.csv`	row CCoAOMT-type / representative
C01-C007	Share of the Pun1/AT3 pool capacity below which Pun1/AT3 limits the output, family sum	9.24%	computed	`data/results/pun1_limit_fraction.csv`	row CCoAOMT-type / family-sum
C01-C008	MIT score of guide 1	99.3	computed	`data/results/guide_table.csv`	row 1, column mit_score
C01-C009	MIT score of guide 2	94.7	computed	`data/results/guide_table.csv`	row 2, column mit_score
C01-C010	Length of the complete knockout cassette (bp)	5,446	computed	`data/results/cassette_summary.csv`	row knockout_cassette_bp
C01-C011	Ring distance bound for those runs	1.1 Å	computed	`data/results/docking_geometry.csv`	row pocket_ring_distance_largest_within_1.1_A
C01-C012	Pocket runs whose docked vanillyl ring lies within 1.1 A of the crystal vanillyl ring	13 of 15	computed	`data/results/docking_geometry.csv`	rows pocket_runs_ring_within_1.1_A_of_crystal_vanillyl_ring and pocket_runs

Claim ID	Claim	Value	Kind	Source file	Location
C01-C013	Chance Pun1/AT3 limits the output in the perturbation settings	0.0	computed	`data/results/flux_monte_carlo.csv`	column probability_pun1_carries_control
C01-C014	Perfect genomic sites of each guide	exactly one	computed	`data/results/guide_table.csv`	column sites_0_mismatch
C01-C015	Codon of the cut of guide 1	42	computed	`data/results/guide_table.csv`	row 1, column cut_codon
C01-C016	Rank of the SHU target among the ten cultivars of the series	1 of 10	computed	`data/results/series_position.csv`	row of this cultivar, column rank_by_shu_target
C01-C017	Reference genome assembly	GCF_002878395.1	definition	`inputs/input_sources.csv`	row reference genome
C01-C018	Public placenta RNA-seq series	PRJNA789050	definition	`inputs/input_sources.csv`	row RNA-seq series
C01-C019	Python version for the analysis steps	3.9	definition	`code/LIBRARIES.txt`	line Python
C01-C020	numpy version used in the research record	2.0.2	definition	`code/LIBRARIES.txt`	line numpy
C01-C021	scipy version used in the research record	1.13.1	definition	`code/LIBRARIES.txt`	line scipy
C01-C022	Sum of the five blend fractions	100%	design target	`data/results/blend_summary.csv`	row sum_of_blend_fractions_pct
C01-C023	Metabolites of the pathway model	37	computed	`data/results/flux_model_summary.csv`	row metabolites
C01-C024	Reactions of the pathway model	44	computed	`data/results/flux_model_summary.csv`	row reactions
C01-C025	Steps of the capsaicinoid chain	8	definition	`data/results/flux_model_summary.csv`	row chain_steps
C01-C026	Stage (DPA) whose transcript counts set the step capacities	34	definition	`research_record/scripts/flux/knockout_run.py`	argument stage=34dpa
C01-C027	TPM of the Pun1/AT3 paralog XP_016560262.2 that sets the capacity scale	1580.3	definition	`data/results/flux_model_summary.csv`	row pun1_scale_transcript_tpm

Claim ID	Claim	Value	Kind	Source file	Location
C01-C028	Capacity assigned to that transcript (arbitrary units)	1,000	definition	`data/results/flux_model_summary.csv`	row pun1_scale_capacity
C01-C029	COBRApy version of the stored model runs	0.32.1	definition	`research_record/scripts/flux/model/flux_model.py`	header, libraries
C01-C030	Largest difference between the second linear-programme solver and the chain rule	5e-09	computed	`research_record/second_script_rerun/flux_chain_rule_check_output.txt`	line max |LP - minrule|
C01-C031	Log standard deviation of the capacity draws	0.5	design assumption	`data/results/flux_monte_carlo.csv`	column log_sigma
C01-C032	Draws per setting	5,000	definition	`data/results/flux_monte_carlo.csv`	column draws
C01-C033	Random seed of the perturbation run	20261111	definition	`research_record/scripts/flux/model/flux_monte_carlo.py`	header
C01-C034	Days post anthesis of the five placenta libraries	4, 24, 34, 44 and 54	definition	`inputs/expression/placenta_libraries.csv`	column DPA
C01-C035	Mapping rates of the five libraries	80.8 to 84.0%	computed	`inputs/expression/placenta_libraries.csv`	column Mapped (%)
C01-C036	Records in the class III peroxidase family table	94	computed	`data/results/peroxidase_rank.csv`	row records_in_family
C01-C037	Exons of the target gene model	4	computed	`data/results/gene_model_summary.csv`	column exons
C01-C038	Coding bases of the target gene model	1,002	computed	`data/results/gene_model_summary.csv`	column cds_bases
C01-C039	Residues of the target protein	333	computed	`data/results/gene_model_summary.csv`	column protein_residues

Claim ID	Claim	Value	Kind	Source file	Location
C01-C040	Identity of the translation from the genome to the RefSeq protein	1.00	computed	`data/results/gene_model_summary.csv`	column translation_identity_to_refseq_protein
C01-C041	GenBank record of the Arabidopsis U6-26 gene	X52528.1	definition	`inputs/input_sources.csv`	row U6-26
C01-C042	Length of the 35S block (bp)	435	computed	`data/results/cassette_parts.csv`	row of the 35S block
C01-C043	GenBank record of the cauliflower mosaic virus genome	V00141.1	definition	`inputs/input_sources.csv`	row CaMV
C01-C044	GenBank record of the plant-optimised SpCas9 coding sequence	KR029112.1	definition	`inputs/input_sources.csv`	row SpCas9
C01-C045	UniProt entry of SpCas9	Q99ZW2	definition	`inputs/input_sources.csv`	row SpCas9 protein
C01-C046	Length of the NOS terminator (bp)	256	computed	`data/results/cassette_parts.csv`	row NOS terminator
C01-C047	AutoDock Vina version	1.2.5	definition	`inputs/docking/docking_setup.csv`	row program
C01-C048	TRPV1 structure used for docking	5IRX	definition	`inputs/input_sources.csv`	row TRPV1 structure
C01-C049	First residue of the receptor chain	335	computed	`data/results/docking_receptor.csv`	row residue_first
C01-C050	Last residue of the receptor chain	751	computed	`data/results/docking_receptor.csv`	row residue_last
C01-C051	Meeko version	0.8	definition	`inputs/docking/docking_setup.csv`	row receptor
C01-C052	Edge of the vanilloid-pocket box	24 Å	definition	`inputs/docking/docking_setup.csv`	row vanilloid_pocket_box_size_angstrom
C01-C053	Edge of the pore-facing box	30 Å	definition	`inputs/docking/docking_setup.csv`	row pore_facing_box_size_angstrom

Claim ID	Claim	Value	Kind	Source file	Location
C01-C054	Distance between the centres of the two boxes	24.6 Å	computed	data/results/docking_geometry.csv	row box_centre_to_box_centre_A
C01-C055	Seeds per ligand per box	3	definition	data/results/docking_receptor.csv	row seeds_per_ligand_per_box
C01-C056	Vina exhaustiveness	32	definition	data/results/docking_receptor.csv	row exhaustiveness
C01-C057	Blend-weighted SHU coefficient (SHU per gram of capsaicinoid)	13,885,000	computed	data/results/blend_summary.csv	row blend_weighted_coefficient_shu_per_g
C01-C058	Blend-weighted coefficient divided by the capsaicin coefficient	0.868	computed	data/results/blend_summary.csv	row coefficient_ratio_to_reference_compound
C01-C059	SHU contribution of each of the five capsaicinoids	399,446, 299,585, 85,194, 53,676 and 28,710	computed	data/results/blend_calculation.csv	column shu_contribution, five compound rows
C01-C060	Sum of the five SHU contributions	866,611	computed	data/results/blend_calculation.csv	row total, column shu_contribution
C01-C061	Content when the capsaicin coefficient moves by ten percent	59.66 to 65.43	computed	data/results/blend_sensitivity.csv	row capsaicin coefficient
C01-C062	Content when the capsaicin fraction moves by five points	61.63 to 63.22	computed	data/results/blend_sensitivity.csv	row capsaicin fraction
C01-C063	Lowest SHU design target of the series	866,611	design target	data/results/series_position.csv	first row
C01-C064	Highest SHU design target of the series	2,956,310	design target	data/results/series_position.csv	last row
C01-C065	TPM of the target gene at 4 DPA	185.6	computed	data/tables/T03.csv	row XP_016556569.1, column 4 DPA
C01-C066	TPM of the target gene at 24 DPA	14.7	computed	data/tables/T03.csv	row XP_016556569.1, column 24 DPA
C01-C067	TPM of the target gene at 34 DPA	1.6	computed	data/tables/T03.csv	row XP_016556569.1, column 34 DPA

Claim ID	Claim	Value	Kind	Source file	Location
C01-C068	Rank of the target gene among the peroxidase-family records by mean TPM at 24 and 34 DPA	13	computed	`data/results/peroxidase_rank.csv`	row target_rank_by_mean_24_34_dpa
C01-C069	Mean TPM of the target gene at 24 and 34 DPA	8.1	computed	`data/results/peroxidase_rank.csv`	row target_mean_24_34_dpa_tpm
C01-C070	Locus with the highest mean TPM in the peroxidase family	LOC107859584	computed	`data/results/peroxidase_rank.csv`	row top_locus
C01-C071	Product of that locus	peroxidase 42	computed	`data/results/peroxidase_rank.csv`	row top_product
C01-C072	Mean TPM of that locus at 24 and 34 DPA	392.3	computed	`data/results/peroxidase_rank.csv`	row top_mean_24_34_dpa_tpm
C01-C073	TPM of PAL NP_001311532.1 at 34 DPA	253.3	computed	`data/tables/T03.csv`	row NP_001311532.1, column 34 DPA
C01-C074	TPM of 4CL NP_001311759.1 at 34 DPA	37.2	computed	`data/tables/T03.csv`	row NP_001311759.1, column 34 DPA
C01-C075	TPM of the CCoAOMT-type transcript NP_001311511.1 at 24 and 34 DPA	746.5 to 752.1	computed	`data/tables/T03.csv`	row NP_001311511.1, columns 24 and 34 DPA
C01-C076	TPM of the COMT-class transcript NP_001311774.1 at 24 and 34 DPA	721.6 to 280.2	computed	`data/tables/T03.csv`	row NP_001311774.1, columns 24 and 34 DPA
C01-C077	TPM of the pAMT-step transcript NP_001311635.1 at 24 DPA	3,133.2	computed	`data/tables/T03.csv`	row NP_001311635.1, column 24 DPA
C01-C078	TPM of the Pun1 transcript NP_001311698.1 at 4 DPA	1.6	computed	`data/tables/T03.csv`	row NP_001311698.1, column 4 DPA
C01-C079	TPM of the Pun1 transcript NP_001311698.1 at 24 DPA	2,647.1	computed	`data/tables/T03.csv`	row NP_001311698.1, column 24 DPA
C01-C080	TPM of the second Pun1 locus XP_016560262.2 at 24 DPA	1,580.3	computed	`data/tables/T03.csv`	row XP_016560262.2, column 24 DPA
C01-C081	Sum of the TPM of the two Pun1 loci at 24 DPA	4,227.4	computed	`data/results/pun1_group.csv`	row 24, column sum_tpm

Claim ID	Claim	Value	Kind	Source file	Location
C01-C082	Output of the flux model, representative transcripts (arbitrary units)	56.42	computed	data/tables/T04.csv	output row, column representative
C01-C083	Limiting step, representative transcripts	4CL	computed	data/results/flux_steps.csv	rows with limiting = yes, representative
C01-C084	Output of the flux model, family sum (arbitrary units)	67.28	computed	data/tables/T04.csv	output row, column family sum
C01-C085	Limiting step, family sum	HCT	computed	data/results/flux_steps.csv	rows with limiting = yes, family-sum
C01-C086	Capacity control coefficient of the limiting step	1.00	computed	data/results/flux_steps.csv	column capacity_control_coefficient, limiting step
C01-C087	Capacity control coefficient of every other step	0.00	computed	data/results/flux_steps.csv	column capacity_control_coefficient, other steps
C01-C088	Fill level of the Pun1/AT3 pool, representative transcripts	0.0776	computed	data/results/flux_steps.csv	row Pun1/AT3 pool, representative
C01-C089	Fill level of the Pun1/AT3 pool, family sum	0.0924	computed	data/results/flux_steps.csv	row Pun1/AT3 pool, family-sum
C01-C090	Output when any single step is set to zero	0	computed	data/results/flux_knockout_runs.csv	rows with factor 0
C01-C091	Output with 4CL at half capacity, representative	28.21	computed	data/results/flux_knockout_runs.csv	row representative, 4CL, 0.5
C01-C092	Output with HCT at half capacity, representative	31.56	computed	data/results/flux_knockout_runs.csv	row representative, HCT, 0.5
C01-C093	Output with 4CL at double capacity, representative	63.12	computed	data/results/flux_knockout_runs.csv	row representative, 4CL, 2
C01-C094	Output with HCT at double capacity, family sum	112.22	computed	data/results/flux_knockout_runs.csv	row family-sum, HCT, 2
C01-C095	Output with 4CL at half capacity, family sum	56.11	computed	data/results/flux_knockout_runs.csv	row family-sum, 4CL, 0.5
C01-C096	Chance an early step limits, in the four perturbation settings	0.9332, 0.9916, 0.9298 and 0.9838	computed	data/results/flux_monte_carlo.csv	column probability_early_step_limits

Claim ID	Claim	Value	Kind	Source file	Location
C01-C097	Median fill of the Pun1/AT3 pool across the four settings	0.0607 to 0.0844	computed	`data/results/flux_monte_carlo.csv`	column pun1_saturation_median
C01-C098	RefSeq product name of the target	peroxidase 45	computed	`data/results/gene_model_summary.csv`	column product
C01-C099	Chromosome of the target gene	1	computed	`data/results/guide_table.csv`	column chromosome
C01-C100	First guide	POX-107856092-g1	computed	`data/results/guide_table.csv`	row 1, column guide_id
C01-C101	Protospacer of the first guide	TCGAAATTCGTGTCCGAACG	computed	`data/results/guide_table.csv`	row 1, column protospacer_20nt
C01-C102	PAM of the first guide	TGG	computed	`data/results/guide_table.csv`	row 1, column pam
C01-C103	Coding base after which guide 1 cuts	124	computed	`data/results/guide_table.csv`	row 1, column cut_after_coding_base
C01-C104	Exon of the cut of guide 1	1	computed	`data/results/guide_table.csv`	row 1, column exon
C01-C105	Second guide	POX-107856092-g2	computed	`data/results/guide_table.csv`	row 2, column guide_id
C01-C106	Protospacer of the second guide	CATGGCCAAGATGTCAGCAC	computed	`data/results/guide_table.csv`	row 2, column protospacer_20nt
C01-C107	PAM of the second guide	AGG	computed	`data/results/guide_table.csv`	row 2, column pam
C01-C108	Coding base after which guide 2 cuts	388	computed	`data/results/guide_table.csv`	row 2, column cut_after_coding_base
C01-C109	Codon of the cut of guide 2	130	computed	`data/results/guide_table.csv`	row 2, column cut_codon
C01-C110	Exon of the cut of guide 2	2	computed	`data/results/guide_table.csv`	row 2, column exon
C01-C111	Coding bases between the two cuts	264	computed	`data/results/guide_pair.csv`	row coding_bases_between_cuts
C01-C112	Genome sites of guide 1 at 0 / 1 / 2 / 3 / 4 mismatches	1 / 0 / 0 / 0 / 7	computed	`data/results/guide_table.csv`	row 1, columns sites_0 to sites_4
C01-C113	Genome sites of guide 2 at 0 / 1 / 2 / 3 / 4 mismatches	1 / 0 / 0 / 2 / 40	computed	`data/results/guide_table.csv`	row 2, columns sites_0 to sites_4

Claim ID	Claim	Value	Kind	Source file	Location
C01-C114	Sites of guide 2 with three mismatches	2	computed	data/results/guide_table.csv	row 2, column sites_3_mismatches
C01-C115	Guides whose recounted sites and scores equal the stored table	2 of 2 guides	computed	data/results/guide_agreement.csv	column agrees
C01-C116	Fewest mismatches of the first guide in the two C. chinense assemblies	1	computed	inputs/guides/guide_transfer_other_assemblies.csv	rows of the first guide, C. chinense
C01-C117	Fewest mismatches of the first guide in the C. baccatum assembly	0	computed	inputs/guides/guide_transfer_other_assemblies.csv	row of the first guide, C. baccatum
C01-C118	Length of the guide unit (bp)	552	computed	data/results/cassette_summary.csv	row guide_unit_bp
C01-C119	Length of the Cas9 unit (bp)	4,894	computed	data/results/cassette_summary.csv	row cas9_unit_bp
C01-C120	Length of the Cas9 coding sequence (bases)	4,203	computed	data/results/cassette_summary.csv	row cas9_bp
C01-C121	Residues encoded by the Cas9 coding sequence	1,400	computed	data/results/cassette_summary.csv	row cas9_residues
C01-C122	Composition of the Cas9 protein (N-terminal residues + SpCas9 residues + C-terminal residues)	17 + 1,367 + 16	computed	data/results/cassette_checks.csv	row Protein composition
C01-C123	GC content of the Cas9 coding sequence	54.2%	computed	data/results/cassette_summary.csv	row cas9_gc_percent
C01-C124	GC content at third codon positions of the Cas9 coding sequence	78.6%	computed	data/results/cassette_summary.csv	row cas9_gc3_percent
C01-C125	Base at which the forward 35S block starts	553	computed	data/results/cassette_checks.csv	row 35S block (forward) found
C01-C126	Bases of the 35S block identical to the cauliflower mosaic virus genome	433 of 435	computed	data/results/cassette_checks.csv	row 35S block bases identical to CaMV V00141.1

Claim ID	Claim	Value	Kind	Source file	Location
C01-C127	Identical bases in the promoter part of the block	219 of 221	computed	`data/results/cassette_checks.csv`	row 35S block: bases identical
C01-C128	Identical bases in the leader part of the block	214 of 214	computed	`data/results/cassette_checks.csv`	row 35S block: bases identical
C01-C129	ATG codons in the 35S block ahead of Cas9	7	computed	`data/results/cassette_summary.csv`	row block_atg_count
C01-C130	Block positions of those ATG codons	25, 52, 125, 151, 273, 326 and 355	computed	`data/results/cassette_checks.csv`	row ATG codons in the 35S block
C01-C131	Reading frames of those ATG codons that reach the Cas9 start	0	computed	`data/results/cassette_checks.csv`	row ATG codons in the 35S block
C01-C132	Restriction enzymes counted	13	computed	`data/results/cassette_summary.csv`	row recognition_sites_counted
C01-C133	Counted recognition sites with zero occurrences	7	computed	`data/results/cassette_summary.csv`	row recognition_sites_at_zero
C01-C134	Enzymes with zero sites	BsaI, EcoRI, XbaI, SphI, BamHI, PstI/SbfI and XmaI/SmaI	computed	`data/results/cassette_sites.csv`	rows with sites = 0
C01-C135	Sites of BsmBI/Esp3I in the cassette	6	computed	`data/results/cassette_sites.csv`	row BsmBI/Esp3I
C01-C136	Sites of BpiI/BbsI in the cassette	3	computed	`data/results/cassette_sites.csv`	row BpiI/BbsI
C01-C137	Sites of HindIII in the cassette	5	computed	`data/results/cassette_sites.csv`	row HindIII
C01-C138	Checks on the cassette that pass	19	computed	`data/results/cassette_summary.csv`	row checks_passed
C01-C139	Longest open reading frame found by the second script, which reads only the FASTA file (codons and bases)	1,400 codons at bases 988 to 5190	computed	`research_record/second_script_rerun/cassette_check_second_route_output.txt`	line of ORFs of 100 codons or more

Claim ID	Claim	Value	Kind	Source file	Location
C01-C140	Smallest ring distance in the pocket runs	0.1 Å	computed	`data/results/docking_geometry.csv`	row pocket_ring_distance_min_A
C01-C141	Nearest ligand O or N to Y511, S512, T550 or R557 in the pocket runs	2.8 to 3.2 Å	computed	`data/results/docking_geometry.csv`	rows pocket_nearest_ON_to_Y511_S512_T550_R557
C01-C142	Pocket runs	15	computed	`data/results/docking_geometry.csv`	row pocket_runs
C01-C143	Mean pocket score of capsaicin (kcal/mol)	-6.21	computed	`data/results/docking_summary.csv`	row capsaicin, vanilloid pocket
C01-C144	Mean pocket score of nonivamide (kcal/mol)	-5.54	computed	`data/results/docking_summary.csv`	row nonivamide, vanilloid pocket
C01-C145	Span of the mean pocket scores of the five ligands (kcal/mol)	0.67	computed	`data/results/docking_geometry.csv`	row ligand_mean_score_range_in_pocket
C01-C146	Largest span of the pocket scores of one ligand across seeds (kcal/mol)	0.53	computed	`data/results/docking_geometry.csv`	row seed_spread_largest_in_pocket
C01-C147	Distance of the docked ring from the crystal vanillyl ring in the pore-facing box	16.5 to 20.7 Å	computed	`data/results/docking_geometry.csv`	rows pore_box_ring_distance
C01-C148	Pore-facing scores minus pocket scores, range over the five ligands (kcal/mol)	0.40 to 0.95	computed	`data/results/docking_geometry.csv`	rows pore_box_better_than_pocket
C01-C149	Top poses re-scored with Vina	30	computed	`data/results/docking_geometry.csv`	row runs_rescored
C01-C150	Largest difference between reported and re-scored values (kcal/mol)	0.007	computed	`data/results/docking_geometry.csv`	row rescoring_largest_difference_kcal_per_mol

Appendix B. Glossary, abbreviations and units

Term	Meaning	Type
SHU	Scoville Heat Unit: a rating of pungency based on a dilution scale	abbreviation
DPA	Days post anthesis: the age of the fruit counted from flowering	abbreviation
TPM	Transcripts per million: a measure of transcript abundance in an RNA-seq library	abbreviation
PAL	Phenylalanine ammonia-lyase: first step of the phenylpropanoid pathway	abbreviation
C4H	Cinnamate 4-hydroxylase	abbreviation
4CL	4-coumarate-CoA ligase	abbreviation
HCT	Hydroxycinnamoyl-CoA shikimate/quinate hydroxycinnamoyl transferase	abbreviation
COMT	Caffeic acid O-methyltransferase	abbreviation
CCoAOMT	Caffeoyl-CoA O-methyltransferase	abbreviation
VR	Vanillin reductase step that reduces vanillin on the way to vanillylamine	abbreviation
pAMT	Putative aminotransferase step that forms vanillylamine	abbreviation
BCAT	Branched-chain amino acid aminotransferase	abbreviation
KAS	3-oxoacyl-acyl-carrier-protein synthase	abbreviation
Pun1/AT3	Acyltransferase that joins vanillylamine and a fatty acid chain to make a capsaicinoid	abbreviation
Chain rule	In a one-chain model the output equals the smallest step capacity	term
Capacity control coefficient	Relative change of the output per relative change of one step capacity	term
Saturation	Output divided by the capacity of a step	term
Class III peroxidase	A family of secreted plant peroxidases	term
Knockout	Loss of a gene function by a cut and repair in the gene	term
sgRNA	Single guide RNA: a guide of 20 bases joined to a scaffold	abbreviation
PAM	Protospacer adjacent motif: the NGG next to the protospacer that SpCas9 needs	abbreviation

Term	Meaning	Type
SpCas9	Cas9 nuclease of Streptococcus pyogenes	abbreviation
NLS	Nuclear localisation signal: a short peptide that sends a protein into the nucleus	abbreviation
MIT score	Aggregate specificity score of a guide (Hsu et al., 2013), 0 to 100	term
Pol III	RNA polymerase III, which transcribes U6 genes	abbreviation
AtU6-26	Arabidopsis thaliana U6-26 snRNA gene promoter	term
CaMV	Cauliflower mosaic virus	abbreviation
35S block	The 435-bp piece of CaMV (promoter part and leader) used as the promoter of SpCas9	term
NOS	Nopaline synthase terminator	abbreviation
FASTA	Plain-text format for sequences	term
TRPV1	Transient receptor potential vanilloid 1: the channel activated by capsaicin	abbreviation
RTX	Resiniferatoxin: the vanilloid bound in the crystal structure 5IRX	abbreviation
Vina	AutoDock Vina, a docking program	term
Vanilloid pocket	The binding site of vanilloids in TRPV1	term
Pore-facing box	A docking box on the channel axis, away from the vanilloid pocket	term
CAP	Capsaicin	abbreviation
DHC	Dihydrocapsaicin	abbreviation
NDHC	Nordihydrocapsaicin	abbreviation
HDHC	Homodihydrocapsaicin	abbreviation
NVA	Nonivamide	abbreviation
RefSeq	NCBI reference sequence database	term
LOC	Prefix of a RefSeq gene identifier	abbreviation
GC	Share of G and C bases	abbreviation
bp	Base pairs	unit
nt	Nucleotides	unit
mg/g	Milligrams per gram of dry weight	unit

Term	Meaning	Type
kcal/mol	Kilocalories per mole	unit
Å	Angstrom, 0.1 nanometre	unit
kcal	Kilocalorie	unit
Design target	A value chosen as a goal of the design	term
Design assumption	An input with no source, stated so that its effect can be read	term

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_blend.py: 0.1 s; 2 code/02_expression.py: 0.1 s; 3 code/03_flux.py: 0.1 s; 4 code/04_guides.py: 0.1 s; 5 code/05_cassette.py: 0.1 s; 6 code/06_docking.py: 0.1 s; 7 code/07_text.py: 0.1 s
Input files	29 files, SHA-256 in `code/environment.json`
Data fingerprint	`d3a0165533e3e9e09055c5a6b9238964317f47dea3edade242a6bc593344cf46`

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.24511362 · 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