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.

Zero-Bitter Brassica: In Silico Design of Myrosinase Knockout in Brassica oleracea

Priyanca Ford

scovillesplice@gmail.com

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

Research Article

Keywords: Brassica oleracea, myrosinase, isothiocyanate, glucosinolate, bitter taste, CRISPR-Cas9 guide design, sulforaphane, Iyengar Flavor Index

Posted Date: November 26th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11875882/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

Zero-Bitter Brassica: In Silico Design of Myrosinase Knockout in Brassica oleracea

A computed design for a five-gene myrosinase knockout in Brassica oleracea, with the guide designs, cut positions and a sulforaphane delivery model.

P.I. Ford

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

DOI: 10.5281/zenodo.24511326 License: CC BY 4.0 **Patents:** US 64/172,827; US 64/188,197; US 64/192,062; US 64/199,105

Harvard Dataverse: DVN/LC54R5 OSF: osf.io/hjm74 Hugging Face: capsaicinoid-data GitHub: paper-06

Website: zero-bitter-brassica

Package record. 122 claims traced to a file or a cited source (93 computed, 12 definition, 13 design assumption, 2 design target, 2 literature value). 18 tables, 7 figures, 64 data files. Data fingerprint `1c93c01d8bf7`. Rerun every result with `bash run_all.sh`.



Broccoli cross-section. Scoville Splice LLC image library.

ABSTRACT

Isothiocyanates formed by the enzyme myrosinase when Brassica tissue is cut give broccoli, cabbage and kale a bitter and sharp taste. This paper describes a computed design that removes myrosinase by knocking out every classical myrosinase gene of *Brassica oleracea*. The reference assembly carries five such genes: *BoMYR1* (LOC106326798), *BoMYR2* (LOC106319184), *BoMYR3* (LOC106309978), *BoMYR4* (LOC106310472) and *BoMYR5* (LOC106310801). Their proteins are 67.5 to 72.6% identical to Arabidopsis TGG1. Two guide RNAs give every gene at least one perfect-match guide; each has perfect sites in three of the genes, with MIT scores of 69.0 and 84.0. In *BoMYR1* (LOC106326798) the two cuts lie 761 bp apart and flank the nucleophile codon. The design target for the knockout is an isothiocyanate content below 0.1 umol/g DW for each compound. A delivery model built from stated assumptions gives a change in absorbed sulforaphane of −14% to +36%,

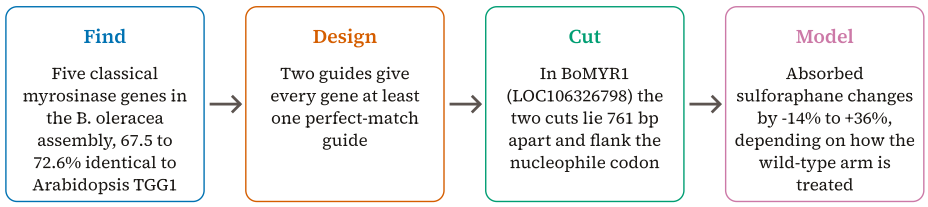
depending on how the wild-type arm is treated. All results are computed. No experimental data are reported in this paper.

Keywords: Brassica oleracea · myrosinase · isothiocyanate · glucosinolate · bitter taste · CRISPR-Cas9 guide design · sulforaphane · Iyengar Flavor Index

KEY POINTS

- The *B. oleracea* assembly carries five classical myrosinase genes, 67.5 to 72.6% identical to Arabidopsis TGG1.
- Two guides give every gene at least one perfect-match guide; their MIT scores are 69.0 and 84.0.
- In *BoMYR1* (LOC106326798) the two cuts lie 761 bp apart and flank the nucleophile codon.
- The design target is below 0.1 umol/g DW for each isothiocyanate in the knockout.
- A delivery model gives a change in absorbed sulforaphane of −14% to +36%, depending on how the wild-type arm is treated.

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



P06 · GA · data/figures/GA.csv

Graphical abstract. The design in four steps: five myrosinase genes are found, two guides are designed, the cuts are placed, and a delivery model gives the change in absorbed sulforaphane. Drawn from the steps file. Data: `data/figures/GA.csv`; script: `code/04_supplementary.py`.

1. Introduction

Broccoli, cabbage, kale and other Brassica vegetables store glucosinolates in their cells. When the tissue is cut or chewed, the enzyme myrosinase splits the glucosinolates and isothiocyanates (ITCs) are released. Bitterness limits the acceptance of Brassica vegetables (Drewnowski and Gomez-Carneros, 2000). Allyl isothiocyanate activates the bitter taste receptor TAS2R38 (Meyerhof et al., 2010; Tran et al., 2021), which responds to compounds that carry an N–C=S group (Bufe et al., 2005).

In Arabidopsis, the genes *TGG1* and *TGG2* encode the main leaf myrosinases, and leaf extracts of the *tgg1 tgg2* double mutant have undetectable myrosinase activity in vitro (Barth and Jander, 2006). *Brassica oleracea* is diploid and has a reference assembly with RefSeq annotation (Parkin et al., 2014). This paper designs a knockout of the whole classical myrosinase family in that species. It reports the myrosinase genes of the assembly, two guide RNAs that give every gene a perfect-match guide, the positions of the cuts, a design target for isothiocyanate content, and a delivery model for glucoraphanin and sulforaphane. Delivery of Cas9 ribonucleoprotein into *B. oleracea* protoplasts with polyethylene glycol has been reported (Murovec et al., 2018). Gene names in this paper are our own, and each is given with its RefSeq locus.

The IFI-B axis of the Iyengar Flavor Index scores a compound as a bitter-taste blocker at one receptor (Ford, 2026; 10.5281/zenodo.24511320). The knockout acts at the source of the isothiocyanates and adds no blocker, so this paper uses IFI-B only to state its reference line (Section 2.9).

2. Methods

2.1 Computational basis

Gene models are the RefSeq annotation of the *Brassica oleracea* var. *oleracea* assembly GCF_000695525.1 (BOL, cultivar TO1000, annotation release 100; Parkin et al., 2014), used directly. The assembly has nine chromosomes, C1 to C9, and unplaced scaffolds. NCBI Gene IDs equal the digits of the LOC symbols. Scripts run on Python 3.9 or newer with the standard library only. The whole-assembly scanner is a C program built with `cc -O2` (`code/guide_design/scan.c`). The default run uses the stored gene models and the stored scan in `inputs/`; the full run repeats both from the RefSeq files, whose checksums are in `inputs/input_sources.csv`.

2.2 Gene identification

Annotated proteins whose product name contains myrosinase, thioglucosidase or glucosidase and that have at least 300 residues were aligned locally to Arabidopsis TGG1 (UniProt P37702, 541 residues) and TGG2 (UniProt Q9C5C2, 547 residues) with Biopython (BLOSUM62, gap open −10, gap extend −0.5); the 30 closest are in Table S4. For the five genes, identity to each Arabidopsis protein is the number of identical residues divided by the length of the Arabidopsis protein in a global alignment with the scoring BLOSUM62, gap open −10, gap extend −0.5. The aligner (`code/guide_design/align.py`) is an affine-gap implementation that gives the same values for all ten gene and Arabidopsis pairs. The nucleophile glutamate is the third residue of the (V/I)TENG motif. A gene model is complete when its coding bases equal three times the protein length plus three. The exons are in Table S11.

As a second route, every 24-mer of the five coding sequences (6,761 sequences) was searched in the whole assembly on both strands. Hits closer than 3,000 bp were joined into one cluster, and clusters with at least 90 hits were annotated with the RefSeq loci under them (Table S5). The share of a gene is the number of distinct 24-mer sequences of that gene found in the cluster divided by the number of distinct 24-mer sequences of that gene; a 24-mer shared by two genes counts for both.

2.3 Guide design

Candidates were listed in the three genes with complete models. A candidate is a strict-NGG SpCas9 site inside one exon, on either strand, whose cut lies within 250 coding bases either side of the nucleophile codon, with GC content of 40 to 70%, no TTTT and no run of five identical bases. The whole assembly was scanned with our own C scanner for sites with up to four mismatches (strict NGG PAM, both strands; sites with an N in the protospacer are skipped); it found 1,058 sites for the 61 candidate protospacers. The MIT score of a guide is 100 / (100 + the sum of its site scores) × 100, summed over the sites with one to four mismatches. The score of one site is the product of $(1 - w)$ over its mismatch positions, where w is the position weight (20 weights, 0 to 0.851; Table S10), times $1 / (((19 - d) / 19) \times 4 + 1)$ for sites with more than one mismatch, where d is the mean distance between mismatches, times $1 / n^2$ for n mismatches, times 100 (Hsu et al., 2013; `code/guide_design/design_lib.py`). On-target activity scores are not computed. From the pairs of candidates that give every gene a perfect site, the pair with the most perfect gene sites was chosen.

2.4 Cut positions

A site's position is the leftmost plus-strand base of the 20-nucleotide protospacer. The cut lies three bases from the PAM, between two plus-strand bases; the cut position is the first plus-strand base after the cut. The coding position after the cut is counted from the first coding base of the gene model in transcript orientation. The offset is the coding position after the cut minus the first base of the nucleophile codon, found in the coding sequence as the codon of the (V/I)TENG motif (negative = upstream). The distance between the two cuts of a gene is the difference of the two genomic cut positions.

2.5 Isothiocyanate design target and assumed pools

The knockout design target is an isothiocyanate content below 0.1 umol/g DW for each of two compounds, allyl isothiocyanate (from sinigrin) and sulforaphane (from glucoraphanin), a limit chosen by us. The wild-type values are design assumptions with no source: a sinigrin pool of 14.2 umol/g DW, which sets the ceiling for allyl isothiocyanate, a glucoraphanin pool of 28.4 umol/g DW and a sulforaphane content of 9.6 umol/g DW, which is 33.8% of the glucoraphanin pool and lies inside the assumed in-plant hydrolysis

range of 0.30 to 0.45. The pools of Table 4 are design assumptions with no source; the knockout acts after glucosinolate synthesis, so its pools are set close to the wild-type pools.

2.6 Delivery model

The model is a chain of fractions applied to the glucoraphanin pool in umol/g DW. The assumed factors are a cooking loss of 0.35 (wild-type tissue only), a gastric loss of 0.15, a microbial conversion of 0.88 (range 0.85 to 0.95 in reading 3) and an absorption fraction of 0.62. Bacteria of the human gut have been reported to metabolise glucoraphanin (Luang-In et al., 2014); the conversion factor itself is an assumption. Knockout tissue is taken to reach the gut with its glucosinolates intact and has no cooking loss. Three readings treat the wild-type arm differently. Reading 1 uses the stated factors only: wild type = pool × (1 − cooking loss) × absorption, and knockout = pool × (1 − gastric loss) × microbial conversion × absorption. Reading 2 also applies the gastric loss to the wild-type arm. Reading 3 uses one delivery chain for both arms: a fraction h of the wild-type pool (0.30 to 0.45) is hydrolysed in the plant and absorbed, and the rest reaches the gut and follows the knockout chain. The same cooking loss on both arms and equal glucoraphanin pools are run as sensitivity cases. A second route (exact fractions and the algebraic form in which the absorption fraction cancels) recomputes every quantity (`code/04_supplementary.py`; Table S12).

2.7 Sensitivity

For each reading the model was run with one input moved at a time while the others stay at their assumed values; the step is 0.10 and 10% (`code/03_delivery_model.py`, `common.one_at_a_time`). The baseline in-plant hydrolysis for reading 3 is 0.375, the midpoint of 0.30 to 0.45. The output is the change in absorbed sulforaphane of knockout tissue relative to wild type, in percent.

2.8 Allele count

Brassica oleracea is diploid, so each of the five genes has two alleles and a homozygous knockout edits ten alleles. With an assumed editing probability of 0.60 per allele, the probability that all alleles are edited in one cell is the per-allele probability raised to the number of alleles.

2.9 IFI-B reference line

IFI-B scores one compound as a blocker at TAS2R31 against the reference blocker GIV3727 (Ford, 2026; 10.5281/zenodo.24511320). The reference value is the mean of two reported IC50 values of GIV3727 at TAS2R31, 6.4 micromolar against acesulfame K and 7.9 micromolar against saccharin (Slack et al., 2010): 7.15 micromolar. This paper assigns no IFI-B score to any compound, because the knockout removes isothiocyanates at their source and adds no blocker.

3. Results

3.1 Five classical myrosinase genes

The reference assembly carries five classical myrosinase genes (Table 1, Figure 1, Figure 2): *BoMYR1* (LOC106326798) on chromosome C2, *BoMYR2* (LOC106319184) on chromosome C9, and *BoMYR3* (LOC106309978), *BoMYR4* (LOC106310472) and *BoMYR5* (LOC106310801), three genes on chromosome C8. Their proteins are 543 to 548 residues long, 67.5 to 72.6% identical to AtTGG1 and 66.2 to 68.9% identical to AtTGG2. All five carry the (V/I)TENG motif (three VTENG and two ITENG). The nucleophile glutamate is Glu429, Glu427, Glu427, Glu428 and Glu427 in *BoMYR1* (LOC106326798), *BoMYR2* (LOC106319184), *BoMYR3* (LOC106309978), *BoMYR4* (LOC106310472) and *BoMYR5* (LOC106310801), against Glu420 in AtTGG1. The genes have 3 to 12 exons. Of the five gene models, three are complete; the model of *BoMYR1* (LOC106326798) has 1,504 of 1,647 coding bases and the model of *BoMYR2* (LOC106319184) has 458 of 1,641 coding bases in the assembly. The five genes are the five annotated proteins closest to AtTGG1 and AtTGG2 (ranks 1 to 5; Table S4): the lowest local identity to AtTGG1 among the five is 68.4%, and the highest among the 25 other proteins compared is 57.5%.

Table 1. The five classical myrosinase genes. Genes of Brassica oleracea (assembly GCF_000695525.1) with the closest identity to Arabidopsis TGG1 and TGG2. Identity is the share of Arabidopsis residues that are identical in a global alignment. The last column gives the coding bases of the gene model and, for incomplete models, the coding bases expected from the protein length. Data: `data/tables/T01.csv` *; script:* `code/01_gene_table.py` *.*

Gene	RefSeq locus (Gene ID)	Chromosome	Protein	Length (aa)	Identity to AtTGG1 (%)	Identity to AtTGG2 (%)	Nucleophile Glu (motif)	Exons	Coding bases in the model
BoMYR1	LOC106326798 (106326798)	C2 (NC_027749.1)	XP_013620174.1	548	72.6	68.9	Glu429 (VTENG)	10	1,504 of 1,647 (incomplete)
BoMYR2	LOC106319184 (106319184)	C9 (NC_027756.1)	XP_013612895.1	546	71.7	66.9	Glu427 (ITENG)	3	458 of 1,641 (incomplete)
BoMYR3	LOC106309978 (106309978)	C8 (NC_027755.1)	XP_013602611.1	545	69.7	68.0	Glu427 (VTENG)	12	1,638 (complete)
BoMYR4	LOC106310472 (106310472)	C8 (NC_027755.1)	XP_013603159.1	547	69.5	66.4	Glu428 (ITENG)	12	1,644 (complete)
BoMYR5	LOC106310801 (106310801)	C8 (NC_027755.1)	XP_013603475.1	543	67.5	66.2	Glu427 (VTENG)	12	1,632 (complete)

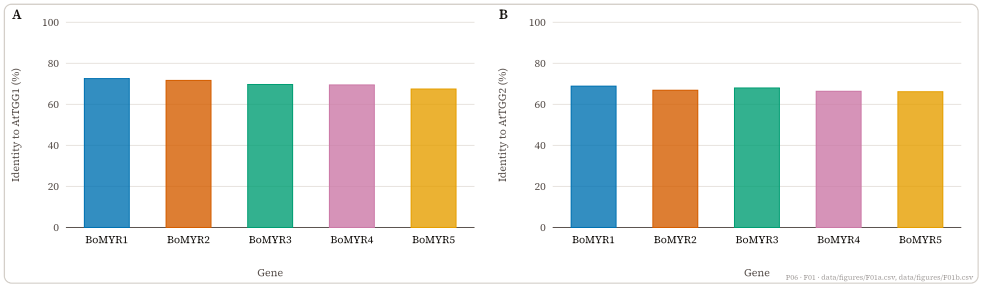


Figure 1. Identity of the five myrosinase proteins to Arabidopsis TGG1 and TGG2. A, identity of each protein to AtTGG1 (P37702). B, identity to AtTGG2 (Q9C5C2). The genes are BoMYR1 (LOC106326798), BoMYR2 (LOC106319184), BoMYR3 (LOC106309978), BoMYR4 (LOC106310472) and BoMYR5 (LOC106310801). Identity is the share of Arabidopsis residues that are identical in a global alignment. Global alignment, BLOSUM62, gap open -10 and extend -0.5. Data: `data/figures/F01a.csv` , `data/figures/F01b.csv` ; script: `code/01_gene_table.py` .

3.2 Guide designs

Of the sites in the complete gene models, 70 pass the filters; they are 61 distinct protospacers, with MIT scores from 18.1 to 100.0 (Table S1, Figure 3). No single candidate has a perfect site in more than three of the five genes. Of the candidates, nine pairs give every gene a perfect site (Table S2); all contain TGG-gRNA2, and TGG-gRNA1 with TGG-gRNA2 is the pair with the most perfect gene sites (six against five for each of the other eight pairs). Table 2 lists the two guides. TGG-gRNA1 is `CACACTGCCATGATGGACGC` with PAM AGG and TGG-gRNA2 is `GGATACTTTGCATGGGCTCT` with PAM CGG. Their MIT scores are 69.0 and 84.0.

TGG-gRNA1 is perfect in *BoMYR1* (LOC106326798), *BoMYR3* (LOC106309978) and *BoMYR5* (LOC106310801), has a site with two mismatches in *BoMYR4* (LOC106310472) and no site within four mismatches in *BoMYR2* (LOC106319184). TGG-gRNA2 is perfect in *BoMYR1* (LOC106326798), *BoMYR2* (LOC106319184) and *BoMYR4* (LOC106310472) and has sites with two mismatches in *BoMYR3* (LOC106309978) and *BoMYR5* (LOC106310801). Each guide therefore has a perfect site in three genes, and every gene has a perfect-match guide. Only one gene, *BoMYR1* (LOC106326798), has a perfect site for both guides. Across the whole assembly, TGG-gRNA1 has 6 / 1 / 2 / 0 / 2 sites at zero to four mismatches and TGG-gRNA2 has 7 / 1 / 2 / 1 / 4 (Figure 4).

Table 2. The two designed guides. TGG-gRNA1 and TGG-gRNA2 with their MIT specificity score, the assembly sites at 0 to 4 mismatches, the perfect sites outside the five genes, and the match in each gene (perfect, 2 mismatches or no site within 4 mismatches). Data: `data/tables/T02.csv` *; script:* `code/02_guides.py` *.*

Property	TGG-gRNA1	TGG-gRNA2
Protospacer (20 nt)	CACACTGCCATGATGGACGC	GGATACTTTGCATGGGCTCT
PAM	AGG	CGG
MIT score	69.0	84.0

Property	TGG-gRNA1	TGG-gRNA2
Genes with a perfect site	3	3
Perfect sites outside the five genes	3	4
Assembly sites at 0 / 1 / 2 / 3 / 4 mismatches	6 / 1 / 2 / 0 / 2	7 / 1 / 2 / 1 / 4
Match in BoMYR1 (LOC106326798)	perfect	perfect
Match in BoMYR2 (LOC106319184)	none	perfect
Match in BoMYR3 (LOC106309978)	perfect	2 mm
Match in BoMYR4 (LOC106310472)	2 mm	perfect
Match in BoMYR5 (LOC106310801)	perfect	2 mm

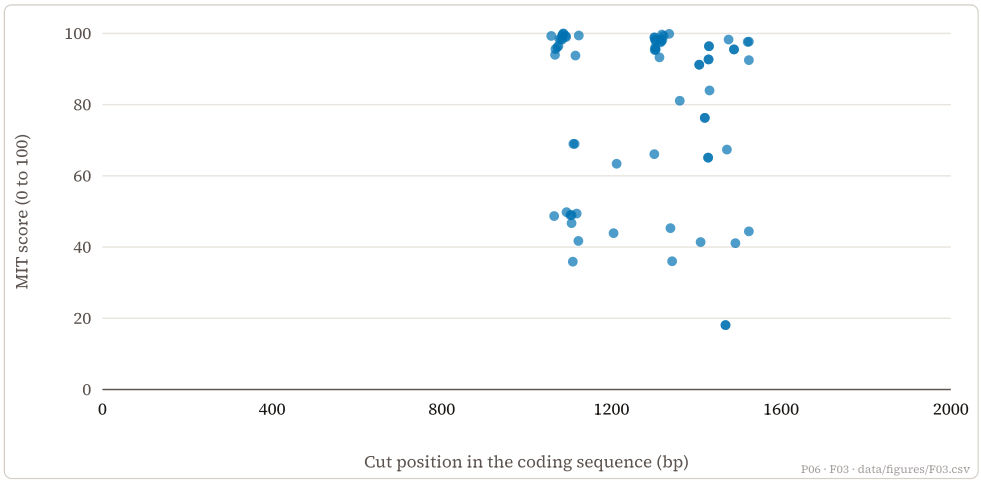


Figure 3. Specificity of the 70 candidate guide sites. Each point is one candidate site in the three complete gene models: its cut position along the coding sequence and the MIT score of its protospacer. Scores from the whole-assembly scan; the same protospacer in different genes gives one point per gene. Data: `data/figures/F03.csv`; script: `code/02_guides.py`.

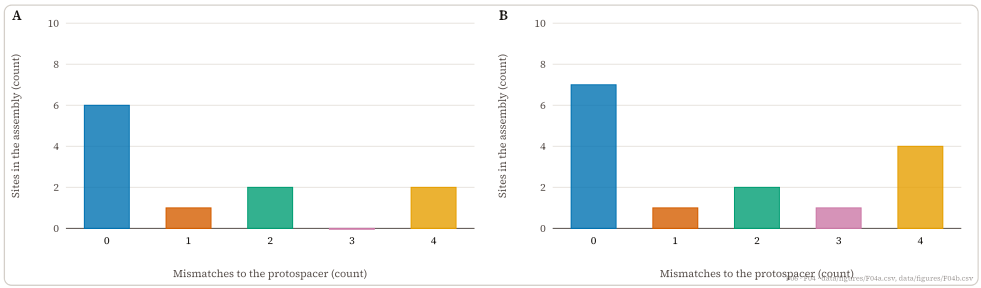


Figure 4. Assembly sites of the two chosen guides by number of mismatches. A, TGG-gRNA1. B, TGG-gRNA2. Every strict-NGG site of the assembly with 0 to 4 mismatches to the protospacer. Whole-assembly scan, both strands. Data: `data/figures/F04a.csv`, `data/figures/F04b.csv`; script: `code/02_guides.py`.

3.3 Sites outside the five genes and further copies

The two guides have 13 perfect sites in the assembly; Table S3 lists them with their RefSeq loci. Of these, seven lie outside the five genes: three for TGG-gRNA1 and four for TGG-gRNA2. Of the seven, four lie in pseudogene loci (LOC106344637, LOC106326800 with two sites, and LOC106330038), one lies in the coding sequence of the protein-coding locus LOC106320470, annotated as myrosinase MB1 with a protein of 136 residues on the unplaced scaffold NW_013617532.1, and two lie in sequence with no annotated locus.

The 24-mer search finds 13 clusters (Table S5, Figure 7). Of these, five are the five genes and eight are further copies of the myrosinase sequence: five lie under pseudogene loci, one under the protein-coding locus LOC106320470 and two in sequence with no annotated locus. The largest share of a gene's 24-mers in

a further cluster is 73.4%, under the transcribed pseudogene LOC106326800 on the same chromosome as *BoMYR1* (LOC106326798).

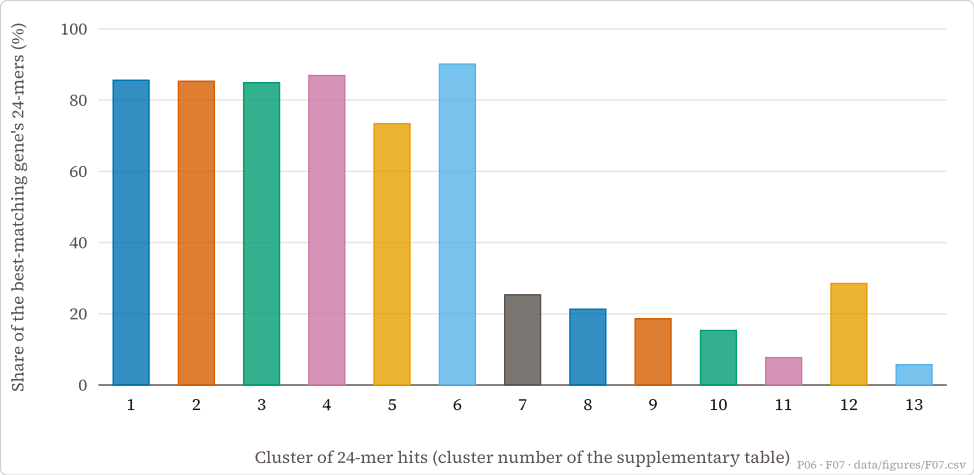


Figure 7. Copies of the myrosinase sequence in the assembly. Each bar is one cluster of hits of the 24-mers of the five coding sequences; clusters 1 to 4 and 6 are the five genes, clusters 5 and 7 to 13 are further copies. Every 24-mer of the five coding sequences was searched in the whole assembly; clusters have at least 90 hits. Data: `data/figures/F07.csv`; script: `code/04_supplementary.py`.

3.4 Cut positions

Table S6 lists every cut inside the five genes. TGG-gRNA1 cuts 163, 166 and 169 coding bases upstream of the nucleophile codon in *BoMYR1* (LOC106326798), *BoMYR3* (LOC106309978) and *BoMYR5* (LOC106310801), and TGG-gRNA2 cuts 149 coding bases downstream in *BoMYR1* (LOC106326798), *BoMYR2* (LOC106319184) and *BoMYR4* (LOC106310472) (Figure 2). *BoMYR1* (LOC106326798) is cut by both guides, at NC_027749.1:48,767,157 and NC_027749.1:48,766,396, 761 bp apart. The two cuts flank the nucleophile codon, with 312 coding bases of the assembled model between them (a multiple of three). *BoMYR2* (LOC106319184) and *BoMYR4* (LOC106310472) are cut once by TGG-gRNA2, and *BoMYR3* (LOC106309978) and *BoMYR5* (LOC106310801) are cut once by TGG-gRNA1.



Figure 2. Coding-sequence maps of the five myrosinase genes. A to E, BoMYR1 (LOC106326798), BoMYR2 (LOC106319184), BoMYR3 (LOC106309978), BoMYR4 (LOC106310472) and BoMYR5 (LOC106310801). Exons, the nucleophile codon and the perfect sites of TGG-gRNA1 and TGG-gRNA2 along the coding sequence; the scale bar is in base pairs. Positions are coding-sequence coordinates computed from the exon coordinates; BoMYR1 and BoMYR2 are drawn on the coding bases present in the assembly. Data: `data/figures/F02a.csv`, `data/figures/F02b.csv`, `data/figures/F02c.csv`, `data/figures/F02d.csv`, `data/figures/F02e.csv`; script: `code/02_guides.py`.

3.5 Isothiocyanate design target and assumed pools

Table 3 sets the wild-type isothiocyanate pools beside the design target. The wild-type sum is at most 23.8 umol/g DW: at most 14.2 umol/g DW as allyl isothiocyanate and 9.6 umol/g DW as sulforaphane. The knockout design target is below 0.1 umol/g DW for each compound and below 0.2 umol/g DW for the sum. Table 4 gives the assumed pools: glucoraphanin 28.4 umol/g DW in wild type and 28.6 umol/g DW in the knockout, and a change of −1.4% to +1.9% for the four listed glucosinolates. The four listed glucosinolates sum to 56.4 and 56.6 umol/g DW, and the total pools are 63.8 and 64.2 umol/g DW.

Table 3. Isothiocyanate design target. Wild-type isothiocyanate pools and the design target for the knockout. All wild-type values are design assumptions with no source; the knockout limits are design targets chosen by the authors. Data: `data/tables/T03.csv`; script: `code/03_delivery_model.py`.

Isothiocyanate	Wild type (umol/g DW)	Five-gene knockout (umol/g DW)	Basis
Allyl isothiocyanate (from sinigrin)	<= 14.2	< 0.1	Wild type: ceiling set by the whole sinigrin pool (design assumption). Knockout: design target.
Sulforaphane (from glucoraphanin)	9.6	< 0.1	Wild type: 33.8% of the glucoraphanin pool (design assumption). Knockout: design target.
Sum of the two	<= 23.8	< 0.2	Sum of the two rows above.

Table 4. Assumed glucosinolate pools. Assumed pool sizes of wild-type and knockout tissue. Every value is a design assumption with no source, set so that the knockout leaves glucosinolate synthesis unchanged; the change column is the arithmetic difference between the two assumed columns. Data: `data/tables/T04.csv`*; script:* `code/03_delivery_model.py`*.*

Glucosinolate	Wild type, assumed (umol/g DW)	Knockout, assumed (umol/g DW)	Change (%)
Glucoraphanin	28.4	28.6	+0.7
Sinigrin	14.2	14.0	-1.4
Glucobrassicin	8.6	8.7	+1.2
Gluconapin	5.2	5.3	+1.9
Sum of the four listed	56.4	56.6	+0.4
Other glucosinolates (total minus the four)	7.4	7.6	+2.7
Total glucosinolates	63.8	64.2	+0.6

3.6 Delivery model

Table 5 follows the glucoraphanin pool through the chain of reading 1. The wild-type pool is 18.46 umol/g DW after cooking. The knockout pool is 24.31 after gastric loss and 21.39 after microbial conversion. Absorbed sulforaphane is 13.26 umol/g DW for the knockout and 11.45 for wild type. Table 6 and Figure 5 give the knockout to wild-type ratio under each reading: reading 1 gives 1.159 (+15.9%), reading 2 gives 1.363 (+36.3%) and reading 3 gives 0.859 to 0.940 (−14.1% to −6.0%). The modelled change in absorbed sulforaphane is therefore −14% to +36%, depending on how the wild-type arm is treated. With the same cooking loss on both arms, reading 1 gives −24.7% and reading 2 gives −11.4%; with equal glucoraphanin pools the three readings give +15.1%, +35.4% and −14.7%. The main route and a second route of the model agree for 34 quantities (Table S12); the largest relative difference is 2.9e-06.

Table 5. Delivery chain, reading 1. Glucoraphanin carried through cooking, gastric loss, microbial conversion and absorption. Every factor is a design assumption with no source. Data: `data/tables/T05.csv`*; script:* `code/03_delivery_model.py`*.*

Step	Wild type (umol/g DW)	Knockout (umol/g DW)	Note
Glucoraphanin pool	28.4	28.6	Design assumptions
After cooking loss (35%, wild type only)	18.46	28.6	Knockout tissue has no cooking loss in readings 1 to 3
After gastric loss (15%, knockout; wild type in reading 2)	18.46	24.31	Wild type in reading 1 passes unchanged
After microbial conversion (88%, knockout)	18.46	21.39	Wild type in reading 1 passes unchanged
Absorbed (62%)	11.45	13.26	Reading 1; ratio KO / WT = 1.159

Table 6. Modelled change in absorbed sulforaphane. Ratio of absorbed sulforaphane in knockout to wild-type tissue under three readings of the wild-type arm and six sensitivity cases. Data: `data/tables/T06.csv`*; script:* `code/03_delivery_model.py`*.*

Reading	Treatment of the wild-type arm	Absorbed sulforaphane, KO / WT	Change, KO vs WT (%)
Reading 1	Stated factors only	1.159	+15.9
Reading 2	Gastric loss also applied to the wild-type arm	1.363	+36.3
Reading 3, lowest	One delivery chain for both arms; hydrolysis 0.45, microbial conversion 0.85	0.859	-14.1

Reading	Treatment of the wild-type arm	Absorbed sulforaphane, KO / WT	Change, KO vs WT (%)
Reading 3, highest	One delivery chain for both arms; hydrolysis 0.30, microbial conversion 0.95	0.940	-6.0
Reading 1, same cooking loss on both arms	Sensitivity case	0.753	-24.7
Reading 2, same cooking loss on both arms	Sensitivity case	0.886	-11.4
Reading 1, equal glucoraphanin pools	Sensitivity case: KO pool = WT pool	1.151	+15.1
Reading 2, equal glucoraphanin pools	Sensitivity case: KO pool = WT pool	1.354	+35.4
Reading 3, equal glucoraphanin pools, lowest	Sensitivity case: KO pool = WT pool	0.853	-14.7

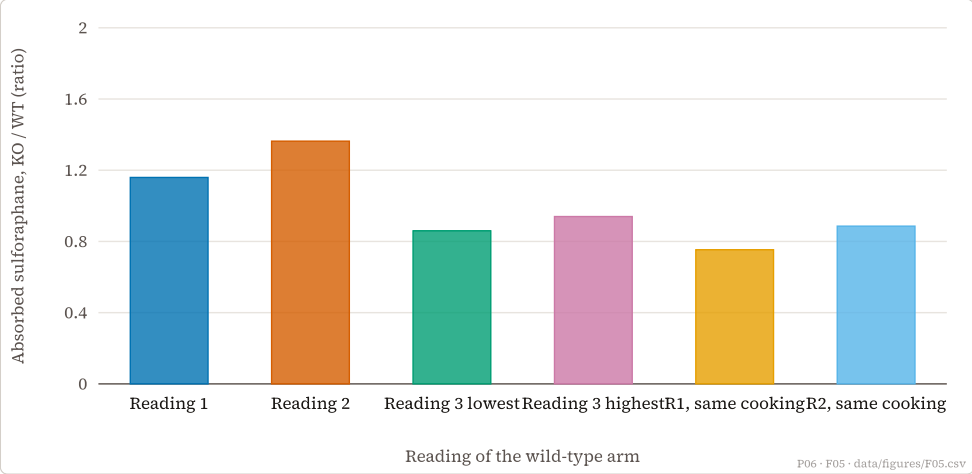


Figure 5. Modelled ratio of absorbed sulforaphane, knockout to wild type. Ratio of absorbed sulforaphane in knockout tissue to wild-type tissue for the three readings (reading 3 at the lowest and highest values of its ranges) and for two readings with the same cooking loss on both arms. A ratio of 1 means no change. Chain of fractions applied to the glucoraphanin pool; every input is a design assumption. Data: `data/figures/F05.csv`; script: `code/03_delivery_model.py`.

3.7 Sensitivity

Figure 6 and Table S9 show the effect of moving one input at a time. In reading 1 the change runs from +0.4% to +37.0%; in reading 2 it runs from +18.2% to +61.1%; in reading 3 it runs from −19.5% to −0.7%. The sign of the change stays the same in each reading over all steps. The cooking loss of wild-type tissue has the largest span in readings 1 and 2 (36.5 and 43.0 percentage points), and the wild-type glucoraphanin pool has the largest span in reading 3 (18.1 percentage points). The absorption fraction has no effect on the ratio in any reading, because it appears in both arms.

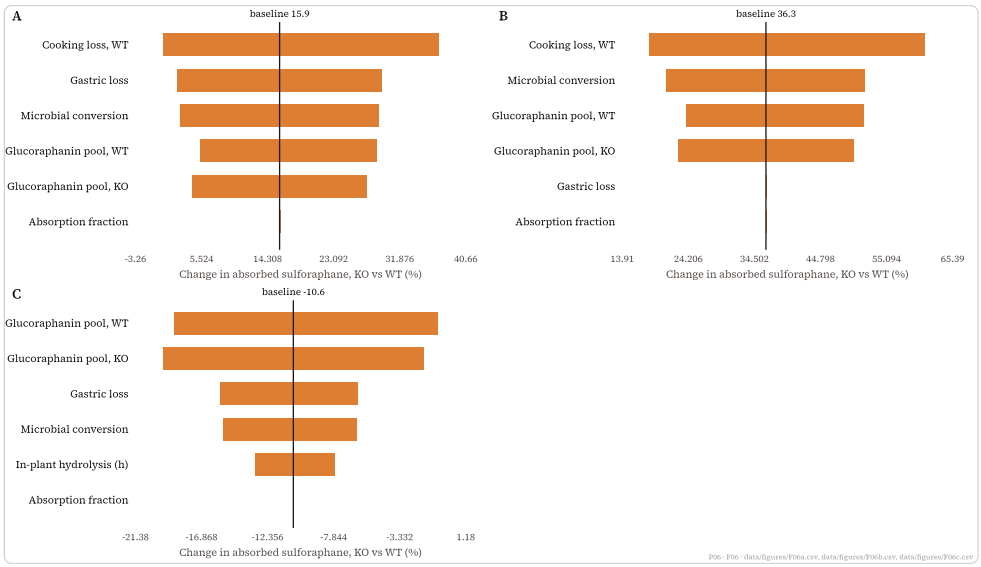


Figure 6. Sensitivity of the modelled change to each input. A, reading 1. B, reading 2. C, reading 3. Each bar spans the result when one input moves by 0.10 (fractions) or by 10% (pools) and the others stay at their assumed values. One-at-a-time analysis of the delivery model. Data: `data/figures/F06a.csv`, `data/figures/F06b.csv`, `data/figures/F06c.csv`; script: `code/03_delivery_model.py`.

3.8 Allele count

Five genes with two alleles each give ten alleles. At an assumed editing probability of 0.60 per allele, the probability that all alleles are edited in one cell is 0.0060 (0.6%).

4. Limitations

Every input of the delivery model is a design assumption with no source, so the modelled change depends on those assumptions and on how the wild-type arm is treated (Table 6, Figure 6). The isothiocyanate limits are design targets. Two of the five gene models are incomplete in the assembly, and TGG-gRNA2 also has a perfect site in the coding sequence of LOC106320470 on an unplaced scaffold. Guide specificity is the MIT score; on-target activity scores are not computed. Breakdown of indole glucosinolates is reported to be slowed, not abolished, in the Arabidopsis double mutant (Barth and Jander, 2006).

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 paper and data are licensed under CC BY 4.0 and the code under the MIT licence. The public input datasets are listed in `inputs/input_sources.csv`.

Public records of this paper: Zenodo 10.5281/zenodo.24511326; Harvard Dataverse: DVN/LC54R5; OSF: osf.io/hjm74; Hugging Face: capsaicinoid-data; GitHub: paper-06; Website: zero-bitter-brassica.

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 Brassica glucosinolate and isothiocyanate baseline, resequencing of every target locus in the lab lines, protoplast RNP editing with the TGG1 and TGG2 myrosinase guides across the B. oleracea genes, TRPA1 cell assays with allyl isothiocyanate, cinnamaldehyde and related compounds, sensory panel screening, training and taster-status genotyping, and the human-subject approvals for sensory panels and for saliva and gut-microbiota donors.

6. References

1. Barth, C., & Jander, G. (2006). Arabidopsis myrosinases TGG1 and TGG2 have redundant function in glucosinolate breakdown and insect defense. *The Plant Journal, 46*(4), 549–562. doi:10.1111/j.1365-313X.2006.02716.x

2. Bufe, B., Breslin, P. A. S., Kuhn, C., Reed, D. R., Tharp, C. D., Slack, J. P., ... & Meyerhof, W. (2005). The molecular basis of individual differences in phenylthiocarbamide and propylthiouracil bitterness perception. *Current Biology, 15*(4), 322–327. doi:10.1016/j.cub.2005.01.047

3. Drewnowski, A., & Gomez-Carneros, C. (2000). Bitter taste, phytonutrients, and the consumer: a review. *American Journal of Clinical Nutrition, 72*(6), 1424–1435. doi:10.1093/ajcn/72.6.1424

4. Ford, P.I. (2026). The Iyengar Flavor Index (IFI): A Receptor-Normalized Standard for Cross-Modal Flavor Measurement. *Scoville Splice Research Series*. doi:10.5281/zenodo.24511320

5. Hsu, P. D., Scott, D. A., Weinstein, J. A., Ran, F. A., Konermann, S., Agarwala, V., ... & Zhang, F. (2013). DNA targeting specificity of RNA-guided Cas9 nucleases. *Nature Biotechnology, 31*(9), 827–832. doi:10.1038/nbt.2647

6. Luang-In, V., Narbad, A., Nueno-Palop, C., Mithen, R., Bennett, M., & Rossiter, J. T. (2014). The metabolism of methylsulfinylalkyl- and methylthioalkyl-glucosinolates by a selection of human gut bacteria. *Molecular Nutrition & Food Research, 58*(4), 875–883. doi:10.1002/mnfr.201300377

7. Meyerhof, W., Batram, C., Kuhn, C., Brockhoff, A., Chudoba, E., Bufe, B., ... & Behrens, M. (2010). The molecular receptive ranges of human TAS2R bitter taste receptors. *Chemical Senses, 35*(2), 157–170. doi:10.1093/chemse/bjp092

8. Murovec, J., Guček, K., Bohanec, B., Avbelj, M., & Jerala, R. (2018). DNA-free genome editing of *Brassica oleracea* and *B. rapa* protoplasts using CRISPR-Cas9 ribonucleoprotein complexes. *Frontiers in Plant Science, 9*, 1594. doi:10.3389/fpls.2018.01594

9. Parkin, I. A. P., Koh, C., Tang, H., Robinson, S. J., Kagale, S., Clarke, W. E., ... & Sharpe, A. G. (2014). Transcriptome and methylome profiling reveals relics of genome dominance in the mesopolyploid *Brassica oleracea*. *Genome Biology, 15*(6), R77. doi:10.1186/gb-2014-15-6-r77

10. Slack, J. P., Brockhoff, A., Batram, C., Menzel, S., Sonnabend, C., Born, S., ... & Meyerhof, W. (2010). Modulation of bitter taste perception by a small molecule hTAS2R antagonist. *Current Biology, 20*(12), 1104–1109. doi:10.1016/j.cub.2010.04.043

11. Tran, H. T. T., Stetter, R., Herz, C., Spöttel, J., Krell, M., Hanschen, F. S., ... & Lamy, E. (2021). Allyl isothiocyanate: a TAS2R38 receptor-dependent immune modulator at the interface between personalized medicine and nutrition. *Frontiers in Immunology, 12*, 669005. doi:10.3389/fimmu.2021.669005

Supplementary data

The supplementary tables hold the full data behind the paper. Table S1 lists all 61 candidate protospacers with their assembly sites, MIT scores and match in each gene. Table S2 lists the nine pairs of candidates that give every gene a perfect site. Table S3 lists the 13 perfect sites of the two guides with strand, PAM, cut position and the RefSeq loci under each site. Table S4 lists the 30 annotated proteins closest to the two Arabidopsis myrosinases. Table S5 lists the 13 clusters of the 24-mer search. Table S6 gives the cut positions and offsets inside the five genes. Table S7 and Table S8 give every input and every result of the delivery model, Table S9 the sensitivity analysis and Table S12 the agreement of its two routes. Table S10 gives the MIT position weights and Table S11 the exon coordinates. Every table is stored as a CSV file in `data/tables/`, and the scripts that write them are in `code/`. The file map below lists every data, input, code and research-record file.

Table S1. All 61 candidate guides. Every protospacer that passed the filters, with the assembly sites at 0 to 4 mismatches, the MIT score, the genes in which it is a candidate and its match in each of the five genes. Data: `data/tables/S01.csv`*; script:* `code/02_guides.py`.

Guide	Protospacer	PAM	Candidate in	Sites 0 mm	Sites 1 mm	Sites 2 mm	Sites 3 mm	Sites 4 mm	MIT score	Genes with a perfect site	BoMYR1 (LOC106326798)	BoMYR2 (LOC106319184)	BoMYR3 (LOC10630
G01	CAAGCCATAATCCAGTTCAT	TGG	BoMYR3 (LOC106309978)	1	0	0	0	22	98.3	1	none	none	perfect
G02	GGCTAATCACACTGCCATGA	TGG	BoMYR3 (LOC106309978); BoMYR5 (LOC106310801)	2	1	0	1	14	49.0	2	none	none	perfect
TGG-gRNA1	CACACTGCCATGATGGACGC	AGG	BoMYR3 (LOC106309978); BoMYR5 (LOC106310801)	6	1	2	0	2	69.0	3	perfect	none	perfect
G04	AACTGGATTATGGCTTGGTT	GGG	BoMYR3 (LOC106309978)	1	0	0	2	27	94.0	1	none	none	perfect
G05	GAACTGGATTATGGCTTGGT	TGG	BoMYR3 (LOC106309978)	1	0	0	0	21	95.6	1	none	none	perfect
G06	CAATGAACTGGATTATGGCT	TGG	BoMYR3 (LOC106309978)	1	0	0	2	15	96.1	1	none	none	perfect
G07	TAGCCCAATGAACTGGATTA	TGG	BoMYR3 (LOC106309978)	1	0	0	0	8	98.4	1	none	none	perfect
G08	GTGTGATTAGCCCAATGAAC	TGG	BoMYR3 (LOC106309978)	1	0	0	1	2	99.5	1	none	none	perfect
G09	CTTTGCGCCTGCGTCCATCA	TGG	BoMYR3 (LOC106309978)	1	2	3	1	63	35.9	1	2 mm	none	perfect
G10	ATCACGCGTTTCATTACCGG	GGG	BoMYR3 (LOC106309978)	1	0	0	1	6	98.9	1	none	none	perfect
G11	CATCACGCGTTTCATTACCG	GGG	BoMYR3 (LOC106309978)	1	0	0	0	7	98.8	1	none	none	perfect
G12	TCATCACGCGTTTCATTACC	GGG	BoMYR3 (LOC106309978)	1	0	0	0	17	96.1	1	none	none	perfect
G13	CTCATCACGCGTTTCATTAC	CGG	BoMYR3 (LOC106309978)	1	0	0	1	21	95.4	1	none	none	perfect
G14	GAGAAGCATGTCAACGTGAA	AGG	BoMYR3 (LOC106309978); BoMYR5 (LOC106310801)	2	0	1	6	17	91.2	2	3 mm	4 mm	perfect
G15	ACGTGAAAGGATACTTTGCG	TGG	BoMYR3 (LOC106309978); BoMYR5 (LOC106310801)	2	0	5	11	25	76.3	2	2 mm	3 mm	perfect
G16	GGATACTTTGCGTGGTCTCT	TGG	BoMYR3 (LOC106309978); BoMYR5 (LOC106310801)	2	1	7	1	3	65.1	2	2 mm	2 mm	perfect
G17	GATACTTTGCGTGGTCTCTT	GGG	BoMYR3 (LOC106309978); BoMYR5 (LOC106310801)	2	0	0	0	31	92.7	2	none	none	perfect
G18	ATACTTTGCGTGGTCTCTTG	GGG	BoMYR3 (LOC106309978); BoMYR5 (LOC106310801)	2	0	0	0	12	96.4	2	none	none	perfect
G19	TCGGACTCAGTTACATTGAT	TGG	BoMYR3 (LOC106309978); BoMYR5 (LOC106310801)	2	0	1	6	5	95.5	2	3 mm	4 mm	perfect
G20	GACAGAGACCTCAAACTTTC	TGG	BoMYR3 (LOC106309978)	1	0	0	1	29	92.5	1	4 mm	none	perfect
G21	GTAACTGAGTCCGAATCTGA	CGG	BoMYR3 (LOC106309978); BoMYR5 (LOC106310801)	2	8	2	1	10	18.1	2	1 mm	1 mm	perfect
G22	TAACCCACTCCCCGTTACAT	CGG	BoMYR4 (LOC106310472)	1	1	0	0	7	49.8	1	none	none	none
G23	GGAAAGATACACTGCCATGA	TGG	BoMYR4 (LOC106310472)	2	0	0	1	20	93.8	1	4 mm	none	none
G24	TACACTGCCATGATGGACCC	AGG	BoMYR4 (LOC106310472)	2	1	6	0	7	41.7	1	2 mm	none	2 mm

Guide	Protospacer	PAM	Candidate in	Sites 0 mm	Sites 1 mm	Sites 2 mm	Sites 3 mm	Sites 4 mm	MIT score	Genes with a perfect site	BoMYR1 (LOC106326798)	BoMYR2 (LOC106319184)	BoMYR3 (LOC10630
G25	ACACTGCCATGATGGACCCA	GGG	BoMYR4 (LOC106310472)	2	0	0	1	4	99.4	1	none	none	none
G26	AGTTGGCTGTACGAACTGAG	TGG	BoMYR4 (LOC106310472)	1	0	0	0	5	99.3	1	none	none	none
G27	ACGGGGAGTGGGTTAGGAGT	TGG	BoMYR4 (LOC106310472)	1	0	1	2	12	96.5	1	none	none	none
G28	GATGTAACGGGGAGTGGGTT	AGG	BoMYR4 (LOC106310472)	2	0	0	1	4	98.4	1	none	none	none
G29	TTTCCGATGTAACGGGGAGT	GGG	BoMYR4 (LOC106310472)	2	0	0	0	4	99.9	1	none	none	none
G30	CTTTCCGATGTAACGGGGAG	TGG	BoMYR4 (LOC106310472)	2	0	0	0	0	100.0	1	none	none	none
G31	TGTATCTTTCCGATGTAACG	GGG	BoMYR4 (LOC106310472)	2	0	0	0	4	99.3	1	none	none	none
G32	GTGTATCTTTCCGATGTAAC	GGG	BoMYR4 (LOC106310472)	2	0	0	2	5	99.0	1	none	none	none
G33	TCTTGTCCCTGGGTCCATCA	TGG	BoMYR4 (LOC106310472)	1	1	0	0	19	49.4	1	none	none	none
G34	CTCTGGTGATCAAACCCGCC	AGG	BoMYR4 (LOC106310472)	1	0	0	0	11	97.5	1	none	none	none
G35	AAGCTGTTGCCGATTCCAAG	CGG	BoMYR4 (LOC106310472)	1	2	1	2	17	45.3	1	none	1 mm	none
G36	CTGGCGGGTTTGATCACCAG	AGG	BoMYR4 (LOC106310472)	1	0	0	1	5	97.9	1	none	none	none
G37	ATCGGCAACAGCTTCCTGGC	GGG	BoMYR4 (LOC106310472)	1	0	0	1	3	99.7	1	none	none	none
G38	AATCGGCAACAGCTTCCTGG	CGG	BoMYR4 (LOC106310472)	1	0	0	1	11	98.1	1	4 mm	none	none
G39	TGGAATCGGCAACAGCTTCC	TGG	BoMYR4 (LOC106310472)	1	0	0	0	6	99.3	1	none	none	none
G40	GCAGAGATAATCAATCCGCT	TGG	BoMYR4 (LOC106310472)	1	3	0	1	1	36.0	1	none	1 mm	none
G41	GAGAAGCGTGTCAACATAAA	AGG	BoMYR4 (LOC106310472)	2	2	0	4	16	41.4	1	4 mm	1 mm	3 mm
TGG-gRNA2	GGATACTTTGCATGGGCTCT	CGG	BoMYR4 (LOC106310472)	7	1	2	1	4	84.0	3	perfect	perfect	2 mm
G43	TCGGACTCAGTTACGTTAAC	TGG	BoMYR4 (LOC106310472)	1	1	6	4	9	41.1	1	2 mm	1 mm	3 mm
G44	GTAACTGAGTCCGAATCGGA	CGG	BoMYR4 (LOC106310472)	1	2	7	2	7	67.4	1	2 mm	2 mm	1 mm
G45	TAACGTAACTGAGTCCGAAT	CGG	BoMYR4 (LOC106310472)	1	0	0	0	8	98.3	1	none	none	none
G46	ACCATTTGCCAGAATCTTTG	AGG	BoMYR4 (LOC106310472)	1	5	3	4	14	44.4	1	1 mm	none	3 mm
G47	CAAGATGATTAGGACTTGGC	TGG	BoMYR5 (LOC106310801)	1	1	0	1	13	48.7	1	none	none	none
G48	CGTTGCGCCTGCGTCCATCA	TGG	BoMYR5 (LOC106310801)	1	1	1	4	5	46.7	1	3 mm	none	1 mm
G49	AACACCTATTACTACCCAGA	AGG	BoMYR5 (LOC106310801)	1	1	6	3	12	63.4	1	2 mm	4 mm	3 mm
G50	AATGCCTTCTGGGTAGTAAT	AGG	BoMYR5 (LOC106310801)	1	2	1	0	9	43.9	1	none	1 mm	2 mm
G51	CCCCGGTGATCAAACCCGTG	AGG	BoMYR5 (LOC106310801)	1	0	1	5	8	93.3	1	none	none	none
G52	ACGCTAAGCTCGATTGCAAG	CGG	BoMYR5 (LOC106310801)	1	0	0	0	2	99.9	1	none	none	none
G53	CTCACGGGTTTGATCACCGG	GGG	BoMYR5 (LOC106310801)	1	1	0	1	8	66.1	1	none	none	4 mm
G54	CCTCACGGGTTTGATCACCG	GGG	BoMYR5 (LOC106310801)	1	0	0	1	14	95.3	1	none	none	4 mm
G55	TCCTCACGGGTTTGATCACC	GGG	BoMYR5 (LOC106310801)	1	0	1	0	9	98.2	1	none	none	4 mm
G56	GTCCTCACGGGTTTGATCAC	CGG	BoMYR5 (LOC106310801)	1	0	1	1	6	95.7	1	none	none	none

Guide	Protospacer	PAM	Candidate in	Sites 0 mm	Sites 1 mm	Sites 2 mm	Sites 3 mm	Sites 4 mm	MIT score	Genes with a perfect site	BoMYR1 (LOC106326798)	BoMYR2 (LOC106319184)	BoMYR3 (LOC10630
G57	ATCGAGCTTAGCGTCCTCAC	GGG	BoMYR5 (LOC106310801)	1	0	0	1	7	98.4	1	none	none	none
G58	AATCGAGCTTAGCGTCCTCA	CGG	BoMYR5 (LOC106310801)	1	0	0	2	7	97.8	1	none	none	none
G59	ACTGAGAAAAGAGAGATGAC	TGG	BoMYR5 (LOC106310801)	1	0	0	5	91	81.1	1	none	none	none
G60	GATAGAGACCTCAAACAGTC	TGG	BoMYR5 (LOC106310801)	1	0	1	4	12	97.7	1	4 mm	4 mm	3 mm
G61	ACCATTTGCCAGACTGTTTG	AGG	BoMYR5 (LOC106310801)	1	0	5	3	10	97.6	1	2 mm	none	3 mm

Table S2. Guide pairs that give every gene a perfect site. The nine pairs of candidates in which at least one guide has a perfect site in each of the five genes, with MIT scores and perfect sites elsewhere in the assembly. Data: `data/tables/S02.csv`*; script:* `code/02_guides.py`*.*

Guide A	Guide B	Protospacer A	Protospacer B	MIT A	MIT B	Genes with a perfect site, A	Genes with a perfect site, B	Perfect sites elsewhere, A	Perfect sites elsewhere, B	Perfect gene sites, A + B
TGG-gRNA1	TGG-gRNA2	CACACTGCCATGATGGACGC	GGATACTTTGCATGGGCTCT	69.0	84.0	BoMYR1, BoMYR3, BoMYR5	BoMYR1, BoMYR2, BoMYR4	3	4	6
G02	TGG-gRNA2	GGCTAATCACACTGCCATGA	GGATACTTTGCATGGGCTCT	49.0	84.0	BoMYR3, BoMYR5	BoMYR1, BoMYR2, BoMYR4	0	4	5
G14	TGG-gRNA2	GAGAAGCATGTCAACGTGAA	GGATACTTTGCATGGGCTCT	91.2	84.0	BoMYR3, BoMYR5	BoMYR1, BoMYR2, BoMYR4	0	4	5
G15	TGG-gRNA2	ACGTGAAAGGATACTTTGCG	GGATACTTTGCATGGGCTCT	76.3	84.0	BoMYR3, BoMYR5	BoMYR1, BoMYR2, BoMYR4	0	4	5
G16	TGG-gRNA2	GGATACTTTGCGTGGTCTCT	GGATACTTTGCATGGGCTCT	65.1	84.0	BoMYR3, BoMYR5	BoMYR1, BoMYR2, BoMYR4	0	4	5
G17	TGG-gRNA2	GATACTTTGCGTGGTCTCTT	GGATACTTTGCATGGGCTCT	92.7	84.0	BoMYR3, BoMYR5	BoMYR1, BoMYR2, BoMYR4	0	4	5
G18	TGG-gRNA2	ATACTTTGCGTGGTCTCTTG	GGATACTTTGCATGGGCTCT	96.4	84.0	BoMYR3, BoMYR5	BoMYR1, BoMYR2, BoMYR4	0	4	5
G19	TGG-gRNA2	TCGGACTCAGTTACATTGAT	GGATACTTTGCATGGGCTCT	95.5	84.0	BoMYR3, BoMYR5	BoMYR1, BoMYR2, BoMYR4	0	4	5
G21	TGG-gRNA2	GTAACTGAGTCCGAATCTGA	GGATACTTTGCATGGGCTCT	18.1	84.0	BoMYR3, BoMYR5	BoMYR1, BoMYR2, BoMYR4	0	4	5

Table S3. Perfect sites of the two guides. All 13 assembly sites with a perfect match to TGG-gRNA1 or TGG-gRNA2, with strand, PAM, cut position and the RefSeq loci under the site. Position is the leftmost plus-strand base of the protospacer. Data: `data/tables/S03.csv`*; script:* `code/02_guides.py`*.*

Guide	Chromosome	Position	Strand	PAM	Cut position	Inside	Overlapping RefSeq loci	Coding sequence under the site
TGG-gRNA1	NC_027748.1	32116203	+	TGG	32116220	outside the five genes	LOC106344637 (pseudogene)	
TGG-gRNA1	NC_027749.1	48751948	-	TGG	48751951	outside the five genes	LOC106326800 (transcribed_pseudogene)	
TGG-gRNA1	NC_027749.1	48767154	-	TGG	48767157	BoMYR1 (LOC106326798)	LOC106326798 (protein_coding)	coding sequence of LOC106326798, myrosinase, 548 aa
TGG-gRNA1	NC_027750.1	18923691	+	TGG	18923708	outside the five genes	LOC106330038 (pseudogene)	
TGG-gRNA1	NC_027755.1	13655460	-	AGG	13655463	BoMYR3 (LOC106309978)	LOC106309978 (protein_coding)	coding sequence of LOC106309978, myrosinase MA1, 545 aa

Guide	Chromosome	Position	Strand	PAM	Cut position	Inside	Overlapping RefSeq loci	Coding sequence under the site
TGG-gRNA1	NC_027755.1	13784755	+	AGG	13784772	BoMYR5 (LOC106310801)	LOC106310801 (protein_coding)	coding sequence of LOC106310801, myrosinase MA1-like, 543 aa
TGG-gRNA2	NC_027749.1	48751183	-	TGG	48751186	outside the five genes	LOC106326800 (transcribed_pseudogene)	
TGG-gRNA2	NC_027749.1	48766393	-	TGG	48766396	BoMYR1 (LOC106326798)	LOC106326798 (protein_coding)	coding sequence of LOC106326798, myrosinase, 548 aa
TGG-gRNA2	NC_027750.1	18925982	+	TGG	18925999	outside the five genes	none	
TGG-gRNA2	NC_027755.1	20282846	+	CGG	20282863	BoMYR4 (LOC106310472)	LOC106310472 (protein_coding)	coding sequence of LOC106310472, myrosinase-like, 547 aa
TGG-gRNA2	NC_027756.1	7226811	+	TGG	7226828	BoMYR2 (LOC106319184)	LOC106319184 (protein_coding)	coding sequence of LOC106319184, myrosinase-like, 546 aa
TGG-gRNA2	NC_027756.1	7252363	+	TGG	7252380	outside the five genes	none	
TGG-gRNA2	NW_013617532.1	3960	+	TGG	3977	outside the five genes	LOC106320470 (protein_coding)	coding sequence of LOC106320470, myrosinase MB1, 136 aa

Table S4. Closest annotated proteins. The 30 annotated glucosidase-family proteins of the assembly most similar to Arabidopsis TGG1 or TGG2 by local alignment; ranks 1 to 5 are the five genes. Data: `data/tables/S04.csv`*; script:* `code/01_gene_table.py`*.*

Rank	Protein	RefSeq locus	Product	Length (aa)	Identity to AtTGG1 (%)	Aligned residues (AtTGG1)	Identity to AtTGG2 (%)	Aligned residues (AtTGG2)	One of the five
1	XP_013620174.1	LOC106326798	myrosinase	548	73.0	538	72.3	519	yes
2	XP_013612895.1	LOC106319184	myrosinase-like	546	72.1	538	68.9	531	yes
3	XP_013602611.1	LOC106309978	myrosinase MA1	545	70.3	536	70.2	530	yes
4	XP_013603159.1	LOC106310472	myrosinase-like	547	70.4	530	68.5	530	yes
5	XP_013603475.1	LOC106310801	myrosinase MA1-like	543	68.4	534	69.6	520	yes
6	XP_013627388.1	LOC106333491	myrosinase 2-like	522	57.5	464	62.1	451	no
7	XP_013584584.1	LOC106293453	myrosinase 4	515	52.3	491	52.6	489	no
8	XP_013626777.1	LOC106332836	beta-glucosidase 13-like	507	49.5	479	46.5	492	no
9	XP_013599775.1	LOC106307383	beta-glucosidase 16 isoform X1	514	47.0	489	47.8	487	no
10	XP_013631280.1	LOC106336864	beta-glucosidase 15	525	48.8	477	46.1	492	no
11	XP_013609671.1	LOC106316338	beta-glucosidase 15-like	507	48.6	477	46.4	487	no
12	XP_013635999.1	LOC106341817	beta-glucosidase 15-like	505	48.0	481	46.2	487	no
13	XP_013636000.1	LOC106341818	beta-glucosidase 15-like	502	49.5	461	46.9	490	no
14	XP_013602564.1	LOC106309949	beta-glucosidase 16-like isoform X1	514	45.5	497	46.7	490	no
15	XP_013631179.1	LOC106336786	beta-glucosidase 15-like isoform X1	517	47.4	481	47.0	485	no
16	XP_013587431.1	LOC106295962	beta-glucosidase 40	508	47.0	483	45.2	489	no
17	XP_013634663.1	LOC106340326	beta-glucosidase 16-like isoform X1	524	47.7	459	48.5	468	no
18	XP_013634664.1	LOC106340326	beta-glucosidase 16-like isoform X2	514	48.3	458	48.7	466	no
19	XP_013634904.1	LOC106340598	beta-glucosidase 32-like	533	43.1	508	45.0	505	no
20	XP_013630284.1	LOC106336105	beta-glucosidase 40 isoform X1	512	46.4	483	46.7	480	no
21	XP_013612042.1	LOC106318562	putative beta-glucosidase 41 isoform X1	549	44.7	499	45.4	491	no
22	XP_013612043.1	LOC106318562	putative beta-glucosidase 41 isoform X2	522	43.8	502	44.8	495	no
23	XP_013634136.1	LOC106339807	beta-glucosidase 27-like	540	45.4	485	42.5	482	no

Rank	Protein	RefSeq locus	Product	Length (aa)	Identity to AtTGG1 (%)	Aligned residues (AtTGG1)	Identity to AtTGG2 (%)	Aligned residues (AtTGG2)	One of the five
24	XP_013635997.1	LOC106341816	beta-glucosidase 32-like	533	44.6	493	43.7	497	no
25	XP_013592374.1	LOC106300718	beta-D-glucopyranosyl abscisate beta-glucosidase-like	526	42.9	492	43.2	505	no
26	XP_013609009.1	LOC106315739	beta-glucosidase 44	514	45.3	481	42.3	487	no
27	XP_013634359.1	LOC106340036	beta-glucosidase 27-like isoform X2	510	45.7	475	43.5	476	no
28	XP_013630157.1	LOC106335968	probable inactive beta-glucosidase 25	521	44.4	487	43.2	491	no
29	XP_013631103.1	LOC106336723	beta-glucosidase 33-like	602	45.8	467	46.1	469	no
30	XP_013600411.1	LOC106307874	beta-glucosidase 30	570	44.0	473	42.9	501	no

Table S5. Clusters of the myrosinase sequence. Clusters of at least 90 hits of the 24-mers of the five coding sequences, with the best-matching gene, the share of that gene's 24-mers found in the cluster and the RefSeq locus under it. Data: `data/tables/S05.csv`*; script:* `code/04_supplementary.py`*.*

Cluster	Chromosome	Start	End	24-mer hits	Best-matching gene	Share of that gene's 24-mers (%)	RefSeq locus under the cluster	One of the five genes
1	NC_027755.1	20280369	20283076	1388	BoMYR4 (LOC106310472)	85.6	LOC106310472 (protein_coding)	yes
2	NC_027755.1	13654551	13657259	1379	BoMYR3 (LOC106309978)	85.3	LOC106309978 (protein_coding)	yes
3	NC_027755.1	13782909	13785665	1366	BoMYR5 (LOC106310801)	84.9	LOC106310801 (protein_coding)	yes
4	NC_027749.1	48766182	48768912	1288	BoMYR1 (LOC106326798)	86.9	LOC106326798 (protein_coding)	yes
5	NC_027749.1	48750972	48753378	1098	BoMYR1 (LOC106326798)	73.4	LOC106326800 (transcribed_pseudogene)	no
6	NC_027756.1	7226366	7227041	392	BoMYR2 (LOC106319184)	90.1	LOC106319184 (protein_coding)	yes
7	NW_013617532.1	3336	4190	374	BoMYR1 (LOC106326798)	25.3	LOC106320470 (protein_coding)	no
8	NC_027750.1	18922983	18926169	356	BoMYR1 (LOC106326798)	21.3	LOC106330038 (pseudogene)	no
9	NC_027753.1	30133320	30135459	306	BoMYR4 (LOC106310472)	18.6	LOC106299074 (pseudogene)	no
10	NC_027749.1	48727971	48729612	274	BoMYR1 (LOC106326798)	15.3	LOC106326801 (pseudogene)	no
11	NC_027748.1	32115246	32117116	179	BoMYR3 (LOC106309978)	7.7	LOC106344637 (pseudogene)	no
12	NC_027756.1	7252141	7252539	129	BoMYR2 (LOC106319184)	28.5	no annotated locus	no
13	NC_027749.1	48743310	48744354	90	BoMYR1 (LOC106326798)	5.7	no annotated locus	no

Table S6. Cut positions and offsets. Genomic and coding position of every cut inside the five genes, the exon, and the distance from the cut to the first base of the nucleophile codon (negative = upstream). Data: `data/tables/S06.csv`*; script:* `code/02_guides.py`*.*

Guide	Gene	RefSeq locus	Chromosome	Site position	Strand	PAM	Genomic cut position	Coding position after the cut	Exon	Nucleophile codon (first base)	Cut minus nucleophile codon (bp)	Gene model	Guide on gene strand
TGG-gRNA1	BoMYR1	LOC106326798	NC_027749.1	48767154	-	TGG	48767157	979	6	1142	-163	incomplete	yes
TGG-gRNA1	BoMYR3	LOC106309978	NC_027755.1	13655460	-	AGG	13655463	1113	8	1279	-166	complete	yes
TGG-gRNA1	BoMYR5	LOC106310801	NC_027755.1	13784755	+	AGG	13784772	1110	8	1279	-169	complete	yes
TGG-gRNA2	BoMYR1	LOC106326798	NC_027749.1	48766393	-	TGG	48766396	1291	10	1142	149	incomplete	yes
TGG-gRNA2	BoMYR2	LOC106319184	NC_027756.1	7226811	+	TGG	7226828	245	3	96	149	incomplete	yes
TGG-gRNA2	BoMYR4	LOC106310472	NC_027755.1	20282846	+	CGG	20282863	1431	12	1282	149	complete	yes

Table S7. Inputs of the delivery model. Every input with its unit and kind: design assumption (no source), design target or definition. Data: `data/tables/S07.csv`*; script:* `code/03_delivery_model.py`*.*

Parameter	Value	Unit	Kind	Where used
glucoraphanin_WT	28.4	umol per g dry weight	design assumption (no source)	Tables 3 4 5
glucoraphanin_KO	28.6	umol per g dry weight	design assumption (no source)	Tables 4 5
sinigrin_WT	14.2	umol per g dry weight	design assumption (no source)	Tables 3 4
sinigrin_KO	14.0	umol per g dry weight	design assumption (no source)	Table 4
glucobrassicin_WT	8.6	umol per g dry weight	design assumption (no source)	Table 4
glucobrassicin_KO	8.7	umol per g dry weight	design assumption (no source)	Table 4
gluconapin_WT	5.2	umol per g dry weight	design assumption (no source)	Table 4
gluconapin_KO	5.3	umol per g dry weight	design assumption (no source)	Table 4
total_glucosinolate_WT	63.8	umol per g dry weight	design assumption (no source)	Table 4
total_glucosinolate_KO	64.2	umol per g dry weight	design assumption (no source)	Table 4
sulforaphane_itc_WT	9.6	umol per g dry weight	design assumption (no source)	Table 3
itc_limit_KO_each	0.1	umol per g dry weight	design target (limit chosen by the authors)	Table 3
cooking_loss_WT	0.35	fraction	design assumption (no source)	Table 5
gastric_loss	0.15	fraction	design assumption (no source)	Table 5
microbial_conversion_mid	0.88	fraction	design assumption (no source)	Table 5
microbial_conversion_low	0.85	fraction	design assumption (no source)	Section 3.6 reading 3
microbial_conversion_high	0.95	fraction	design assumption (no source)	Section 3.6 reading 3
absorption_fraction	0.62	fraction	design assumption (no source)	Table 5
inplant_hydrolysis_low	0.30	fraction of glucoraphanin	design assumption (no source)	Section 3.6 reading 3
inplant_hydrolysis_high	0.45	fraction of glucoraphanin	design assumption (no source)	Section 3.6 reading 3
editing_probability_per_allele	0.60	fraction	design assumption (no source)	Section 3.8
genes_targeted	5	count	definition	Section 3.8
alleles_per_gene	2	count	definition (diploid)	Section 3.8

Parameter	Value	Unit	Kind	Where used
sensitivity_step_fraction	0.10	fraction (absolute)	definition (step applied to each fraction in the sensitivity analysis)	Section 3.7
sensitivity_step_pool	0.10	fraction of the pool (relative)	definition (step applied to each pool in the sensitivity analysis)	Section 3.7

Table S8. Results of the delivery model. Every computed quantity of the model. Data: `data/tables/S08.csv`*; script:* `code/03_delivery_model.py`.

Id	Quantity	Value	Unit
T3_sfn_itc_WT_share_of_glucoraphanin	Table 3 sulforaphane-ITC / glucoraphanin pool	0.338028	fraction
T3_total_itc_WT	Table 3 allyl-ITC ceiling (sinigrin pool) + sulforaphane-ITC	23.8	umol per g dry weight
T3_total_itc_limit_KO	Table 3 limit with all five genes knocked out, two ITCs	0.2	umol per g dry weight
T4_sum_listed_WT	Sum of the four listed glucosinolates, WT	56.4	umol per g dry weight
T4_sum_listed_KO	Sum of the four listed glucosinolates, KO	56.6	umol per g dry weight
T4_total_minus_listed_WT	Total row minus the four listed, WT	7.4	umol per g dry weight
T4_total_minus_listed_KO	Total row minus the four listed, KO	7.6	umol per g dry weight
T5_WT_after_cooking	Table 5 wildtype after cooking	18.46	umol per g dry weight
T5_WT_absorbed_reading1	Table 5 wildtype absorbed (reading 1)	11.4452	umol per g dry weight
T5_KO_after_gastric	Table 5 knockout after gastric loss	24.31	umol per g dry weight
T5_KO_after_microbial	Table 5 knockout after microbial conversion	21.3928	umol per g dry weight
T5_KO_absorbed	Table 5 knockout absorbed	13.2635	umol per g dry weight
R1_ratio_KO_to_WT	Reading 1: KO absorbed / WT absorbed	1.15887	ratio
R1_change_percent	Reading 1: change KO vs WT	15.8873	percent
R2_WT_absorbed	Reading 2: wildtype absorbed	9.72842	umol per g dry weight
R2_ratio_KO_to_WT	Reading 2: KO absorbed / WT absorbed	1.36338	ratio
R2_change_percent	Reading 2: change KO vs WT	36.338	percent
R3_ratio_h030_m085	Reading 3: KO/WT at hydrolysis 0.3 and microbial conversion 0.85	0.902995	ratio
R3_change_percent_h030_m085	Reading 3: change KO vs WT at hydrolysis 0.3 and microbial conversion 0.85	-9.70052	percent
R3_ratio_h030_m095	Reading 3: KO/WT at hydrolysis 0.3 and microbial conversion 0.95	0.939829	ratio

Id	Quantity	Value	Unit
R3_change_percent_h030_m095	Reading 3: change KO vs WT at hydrolysis 0.3 and microbial conversion 0.95	-6.01715	percent
R3_ratio_h045_m085	Reading 3: KO/WT at hydrolysis 0.45 and microbial conversion 0.85	0.858638	ratio
R3_change_percent_h045_m085	Reading 3: change KO vs WT at hydrolysis 0.45 and microbial conversion 0.85	-14.1362	percent
R3_ratio_h045_m095	Reading 3: KO/WT at hydrolysis 0.45 and microbial conversion 0.95	0.909478	ratio
R3_change_percent_h045_m095	Reading 3: change KO vs WT at hydrolysis 0.45 and microbial conversion 0.95	-9.05224	percent
R3_ratio_min	Reading 3: smallest KO/WT over the stated ranges	0.858638	ratio
R3_ratio_max	Reading 3: largest KO/WT over the stated ranges	0.939829	ratio
range_change_percent_min	Lowest change over the three readings	-14.1362	percent
range_change_percent_max	Highest change over the three readings	36.338	percent
S_equalpool_R1_change_percent	Sensitivity, KO pool = WT pool, reading 1 change	15.0769	percent
S_equalpool_R2_change_percent	Sensitivity, KO pool = WT pool, reading 2 change	35.3846	percent
S_equalpool_R3_change_percent_min	Sensitivity, KO pool = WT pool, reading 3 lowest change	-14.7367	percent
S_equalcooking_R1_change_percent	Sensitivity, cooking loss on both arms, reading 1 change	-24.6732	percent
S_equalcooking_R2_change_percent	Sensitivity, cooking loss on both arms, reading 2 change	-11.3803	percent

Table S9. Sensitivity of the delivery model. Change in absorbed sulforaphane (knockout vs wild type) when one input moves over its step and the others stay at their assumed values. Data: `data/tables/S09.csv`*; script:* `code/03_delivery_model.py`*.*

Reading	Input	Low value	High value	Change at low (%)	Change at high (%)	Baseline change (%)	Span (percentage points)
Reading 1	Cooking loss, WT	0.250	0.450	0.4	37.0	15.9	36.5
Reading 1	Gastric loss	0.050	0.250	29.5	2.3	15.9	27.3
Reading 1	Microbial conversion	0.780	0.980	2.7	29.1	15.9	26.3
Reading 1	Glucoraphanin pool, WT	25.560	31.240	28.8	5.4	15.9	23.4
Reading 1	Glucoraphanin pool, KO	25.740	31.460	4.3	27.5	15.9	23.2
Reading 1	Absorption fraction	0.520	0.720	15.9	15.9	15.9	0.0
Reading 2	Cooking loss, WT	0.250	0.450	18.2	61.1	36.3	43.0
Reading 2	Microbial conversion	0.780	0.980	20.8	51.8	36.3	31.0
Reading 2	Glucoraphanin pool, WT	25.560	31.240	51.5	23.9	36.3	27.5
Reading 2	Glucoraphanin pool, KO	25.740	31.460	22.7	50.0	36.3	27.3
Reading 2	Gastric loss	0.050	0.250	36.3	36.3	36.3	0.0
Reading 2	Absorption fraction	0.520	0.720	36.3	36.3	36.3	0.0
Reading 3	Glucoraphanin pool, WT	25.560	31.240	-0.7	-18.7	-10.6	18.1
Reading 3	Glucoraphanin pool, KO	25.740	31.460	-19.5	-1.7	-10.6	17.9
Reading 3	Gastric loss	0.050	0.250	-6.2	-15.6	-10.6	9.4

Reading	Input	Low value	High value	Change at low (%)	Change at high (%)	Baseline change (%)	Span (percentage points)
Reading 3	Microbial conversion	0.780	0.980	-15.4	-6.3	-10.6	9.1
Reading 3	In-plant hydrolysis (h)	0.275	0.475	-7.8	-13.2	-10.6	5.4
Reading 3	Absorption fraction	0.520	0.720	-10.6	-10.6	-10.6	0.0

Table S10. MIT position weights. Position weights of the Hsu et al. (2013) score; position 1 is the PAM-distal base. Data: `data/tables/S10.csv`*; script:* `code/02_guides.py`*.*

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

Table S11. Exon coordinates of the five genes. Genomic and coding coordinates of every coding exon, in transcript order. Data: `data/tables/S11.csv`*; script:* `code/01_gene_table.py`*.*

Gene	RefSeq locus	Exon (transcript order)	Chromosome	Genomic start	Genomic end	Strand	Coding start	Coding end	Length (bp)
BoMYR1	LOC106326798	1	NC_027749.1	48768736	48768912	-	1	177	177
BoMYR1	LOC106326798	2	NC_027749.1	48768590	48768645	-	178	233	56
BoMYR1	LOC106326798	3	NC_027749.1	48768008	48768173	-	234	399	166
BoMYR1	LOC106326798	4	NC_027749.1	48767641	48767902	-	400	661	262
BoMYR1	LOC106326798	5	NC_027749.1	48767438	48767553	-	662	777	116
BoMYR1	LOC106326798	6	NC_027749.1	48767137	48767357	-	778	998	221
BoMYR1	LOC106326798	7	NC_027749.1	48766999	48767033	-	999	1033	35
BoMYR1	LOC106326798	8	NC_027749.1	48766762	48766876	-	1034	1148	115
BoMYR1	LOC106326798	9	NC_027749.1	48766526	48766628	-	1149	1251	103
BoMYR1	LOC106326798	10	NC_027749.1	48766182	48766434	-	1252	1504	253

Gene	RefSeq locus	Exon (transcript order)	Chromosome	Genomic start	Genomic end	Strand	Coding start	Coding end	Length (bp)
BoMYR2	LOC106319184	1	NC_027756.1	7226366	7226467	+	1	102	102
BoMYR2	LOC106319184	2	NC_027756.1	7226579	7226681	+	103	205	103
BoMYR2	LOC106319184	3	NC_027756.1	7226789	7227041	+	206	458	253
BoMYR3	LOC106309978	1	NC_027755.1	13657086	13657259	-	1	174	174
BoMYR3	LOC106309978	2	NC_027755.1	13656912	13656972	-	175	235	61
BoMYR3	LOC106309978	3	NC_027755.1	13656769	13656830	-	236	297	62
BoMYR3	LOC106309978	4	NC_027755.1	13656592	13656667	-	298	373	76
BoMYR3	LOC106309978	5	NC_027755.1	13656340	13656505	-	374	539	166
BoMYR3	LOC106309978	6	NC_027755.1	13655984	13656239	-	540	795	256
BoMYR3	LOC106309978	7	NC_027755.1	13655740	13655855	-	796	911	116
BoMYR3	LOC106309978	8	NC_027755.1	13655443	13655663	-	912	1132	221
BoMYR3	LOC106309978	9	NC_027755.1	13655301	13655335	-	1133	1167	35
BoMYR3	LOC106309978	10	NC_027755.1	13655096	13655213	-	1168	1285	118
BoMYR3	LOC106309978	11	NC_027755.1	13654876	13654978	-	1286	1388	103
BoMYR3	LOC106309978	12	NC_027755.1	13654551	13654800	-	1389	1638	250
BoMYR4	LOC106310472	1	NC_027755.1	20280369	20280545	+	1	177	177
BoMYR4	LOC106310472	2	NC_027755.1	20280632	20280692	+	178	238	61
BoMYR4	LOC106310472	3	NC_027755.1	20280767	20280828	+	239	300	62
BoMYR4	LOC106310472	4	NC_027755.1	20280921	20280996	+	301	376	76
BoMYR4	LOC106310472	5	NC_027755.1	20281094	20281259	+	377	542	166
BoMYR4	LOC106310472	6	NC_027755.1	20281367	20281622	+	543	798	256
BoMYR4	LOC106310472	7	NC_027755.1	20281703	20281818	+	799	914	116
BoMYR4	LOC106310472	8	NC_027755.1	20281953	20282179	+	915	1141	227
BoMYR4	LOC106310472	9	NC_027755.1	20282290	20282324	+	1142	1176	35
BoMYR4	LOC106310472	10	NC_027755.1	20282426	20282537	+	1177	1288	112
BoMYR4	LOC106310472	11	NC_027755.1	20282642	20282744	+	1289	1391	103
BoMYR4	LOC106310472	12	NC_027755.1	20282824	20283076	+	1392	1644	253
BoMYR5	LOC106310801	1	NC_027755.1	13782909	13783079	+	1	171	171
BoMYR5	LOC106310801	2	NC_027755.1	13783202	13783262	+	172	232	61
BoMYR5	LOC106310801	3	NC_027755.1	13783344	13783401	+	233	290	58
BoMYR5	LOC106310801	4	NC_027755.1	13783546	13783625	+	291	370	80
BoMYR5	LOC106310801	5	NC_027755.1	13783788	13783953	+	371	536	166
BoMYR5	LOC106310801	6	NC_027755.1	13784041	13784296	+	537	792	256
BoMYR5	LOC106310801	7	NC_027755.1	13784378	13784496	+	793	911	119
BoMYR5	LOC106310801	8	NC_027755.1	13784574	13784791	+	912	1129	218
BoMYR5	LOC106310801	9	NC_027755.1	13784877	13784911	+	1130	1164	35
BoMYR5	LOC106310801	10	NC_027755.1	13785006	13785126	+	1165	1285	121
BoMYR5	LOC106310801	11	NC_027755.1	13785237	13785339	+	1286	1388	103
BoMYR5	LOC106310801	12	NC_027755.1	13785422	13785665	+	1389	1632	244

Table S12. Agreement of two routes of the delivery model. Every model quantity computed by the main script and by a second route (exact fractions and the algebraic form). Data: `data/tables/S12.csv`*; script:* `code/04_supplementary.py`*.*

Quantity	Main route	Second route (exact fractions)	Relative difference	Within 1%
T3_sfn_itc_WT_share_of_glucoraphanin	0.338028	0.338028	5.000e-07	yes
T3_total_itc_WT	23.8	23.8	0.000e+00	yes
T3_total_itc_limit_KO	0.2	0.2	0.000e+00	yes
T4_sum_listed_WT	56.4	56.4	0.000e+00	yes
T4_sum_listed_KO	56.6	56.6	0.000e+00	yes
T4_total_minus_listed_WT	7.4	7.4	0.000e+00	yes

Quantity	Main route	Second route (exact fractions)	Relative difference	Within 1%
T4_total_minus_listed_KO	7.6	7.6	0.000e+00	yes
T5_WT_after_cooking	18.46	18.46	0.000e+00	yes
T5_WT_absorbed_reading1	11.4452	11.4452	0.000e+00	yes
T5_KO_after_gastric	24.31	24.31	0.000e+00	yes
T5_KO_after_microbial	21.3928	21.3928	0.000e+00	yes
T5_KO_absorbed	13.2635	13.2635	2.714e-06	yes
R1_ratio_KO_to_WT	1.15887	1.15887	2.795e-06	yes
R1_change_percent	15.8873	15.8873	1.507e-06	yes
R2_WT_absorbed	9.72842	9.72842	0.000e+00	yes
R2_ratio_KO_to_WT	1.36338	1.36338	2.066e-07	yes
R2_change_percent	36.338	36.338	7.752e-07	yes
R3_ratio_h030_m085	0.902995	0.902995	2.654e-07	yes
R3_change_percent_h030_m085	-9.70052	-9.70052	4.085e-07	yes
R3_ratio_h030_m095	0.939829	0.939829	5.196e-07	yes
R3_change_percent_h030_m095	-6.01715	-6.01715	1.939e-07	yes
R3_ratio_h045_m085	0.858638	0.858638	4.770e-07	yes
R3_change_percent_h045_m085	-14.1362	-14.1362	2.897e-06	yes
R3_ratio_h045_m095	0.909478	0.909478	4.882e-07	yes
R3_change_percent_h045_m095	-9.05224	-9.05224	4.866e-07	yes
R3_ratio_min	0.858638	0.858638	4.770e-07	yes
R3_ratio_max	0.939829	0.939829	5.196e-07	yes
range_change_percent_min	-14.1362	-14.1362	2.897e-06	yes
range_change_percent_max	36.338	36.338	7.752e-07	yes
S_equalpool_R1_change_percent	15.0769	15.0769	1.531e-06	yes
S_equalpool_R2_change_percent	35.3846	35.3846	4.348e-07	yes
S_equalpool_R3_change_percent_min	-14.7367	-14.7367	8.939e-07	yes
S_equalcooking_R1_change_percent	-24.6732	-24.6732	1.598e-06	yes
S_equalcooking_R2_change_percent	-11.3803	-11.3803	1.609e-06	yes

Supplementary data: file map

Path	Contents
`data/tables/T01.csv`	Table 1: The five classical myrosinase genes
`data/tables/T02.csv`	Table 2: The two designed guides
`data/tables/T03.csv`	Table 3: Isothiocyanate design target
`data/tables/T04.csv`	Table 4: Assumed glucosinolate pools
`data/tables/T05.csv`	Table 5: Delivery chain, reading 1

Path	Contents
data/tables/T06.csv	Table 6: Modelled change in absorbed sulforaphane
data/tables/S01.csv	Table S1: All 61 candidate guides
data/tables/S02.csv	Table S2: Guide pairs that give every gene a perfect site
data/tables/S03.csv	Table S3: Perfect sites of the two guides
data/tables/S04.csv	Table S4: Closest annotated proteins
data/tables/S05.csv	Table S5: Clusters of the myrosinase sequence
data/tables/S06.csv	Table S6: Cut positions and offsets
data/tables/S07.csv	Table S7: Inputs of the delivery model
data/tables/S08.csv	Table S8: Results of the delivery model
data/tables/S09.csv	Table S9: Sensitivity of the delivery model
data/tables/S10.csv	Table S10: MIT position weights
data/tables/S11.csv	Table S11: Exon coordinates of the five genes
data/tables/S12.csv	Table S12: Agreement of two routes of the delivery model
data/figures/GA.csv	Data behind the graphical abstract
data/figures/F01a.csv	Data behind Figure 1
data/figures/F01b.csv	Data behind Figure 1
data/figures/F02a.csv	Data behind Figure 2
data/figures/F02b.csv	Data behind Figure 2
data/figures/F02c.csv	Data behind Figure 2
data/figures/F02d.csv	Data behind Figure 2
data/figures/F02e.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/F06a.csv	Data behind Figure 6
data/figures/F06b.csv	Data behind Figure 6
data/figures/F06c.csv	Data behind Figure 6
data/figures/F07.csv	Data behind Figure 7
data/results/gene_table.csv	The five genes: name, locus, Gene ID, protein, chromosome, strand, product, length, identity to AtTGG1 and AtTGG2, nucleophile residue and motif, coding bases in the model and expected, exons, model status, gene span
data/results/gene_exons.csv	Every coding exon of the five genes in transcript order with genomic and coding coordinates
data/results/gene_summary.csv	Counts and ranges of the gene table used in the text (metric,value)
data/results/closest_proteins.csv	The 30 annotated proteins closest to AtTGG1 or AtTGG2 by local alignment
data/results/candidate_sites.csv	The 70 candidate guide sites in the three complete gene models: guide, protospacer, PAM, gene, strand, coding cut position, offset from the nucleophile codon, GC and MIT score
data/results/guide_scores.csv	The 61 candidate protospacers: assembly sites at 0 to 4 mismatches, MIT score, genes with a perfect site and the match in each gene
data/results/guides_to_scan.txt	The 61 protospacers given to the scanner, in scan order
data/results/covering_pairs.csv	Pairs of candidates that give every gene a perfect site
data/results/guide_table.csv	The two chosen guides: protospacer, PAM, MIT score, sites by mismatch number, perfect sites outside the five genes and the match in each gene
data/results/guides.fasta	The two chosen guides as FASTA (protospacer, PAM and MIT score in the header)
data/results/perfect_sites.csv	All 13 perfect sites of the two guides with strand, PAM, cut position and RefSeq loci
data/results/cut_sites.csv	Cuts inside the five genes: genomic and coding position, exon, nucleophile codon and offset
data/results/guide_summary.csv	Counts and ranges of the guide design used in the text (metric,value)
data/results/further_copies.csv	The 13 clusters of 24-mer hits with the best-matching gene, its share and the RefSeq locus under the cluster
data/results/further_copies_summary.csv	Counts of the clusters used in the text (metric,value)
data/results/model_results.csv	Every quantity of the delivery model (id, quantity, value, unit)
data/results/model_readings.csv	Ratio and percent change of absorbed sulforaphane for each reading and sensitivity case, rounded as printed

Path	Contents
`data/results/model_sensitivity.csv`	One-at-a-time sensitivity of the three readings
`data/results/model_agreement.csv`	Main route and second route of the model, quantity by quantity
`data/results/model_agreement_summary.csv`	Number of quantities compared and the largest relative difference
`data/results/itc_design_target.csv`	Isothiocyanate pools of wild type and the design target of the knockout
`data/results/glucosinolate_pools.csv`	Assumed glucosinolate pools of wild type and knockout
`data/results/delivery_chain.csv`	The delivery chain of reading 1 step by step
`data/results/allele_probability.csv`	Probability that all alleles of the five genes are edited at an assumed per-allele probability
`data/results/ifi_b_reference.csv`	IFI-B reference line: GIV3727 at TAS2R31, the two IC50 values and their mean
`data/results/sensitivity_summary.csv`	Lowest and highest change over all sensitivity steps for each reading, and the input with the largest span
`data/results/sensitivity_baseline.csv`	Baseline value of each input of the sensitivity analysis (in-plant hydrolysis is the midpoint of its range)
`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/model_inputs.csv`	Every input of the delivery model with unit and kind (design assumption, design target or definition)
`inputs/ifi_reference_values.csv`	IFI-B reference line: GIV3727 at TAS2R31, the two reported IC50 values (6.4 and 7.9 micromolar)
`inputs/gene_names.csv`	Names used in this paper with RefSeq locus, Gene ID, protein, chromosome and product
`inputs/input_sources.csv`	Public inputs: assembly, annotation, proteins and the two Arabidopsis UniProt entries, with accession and checksum
`inputs/gene_models/bol_models.json`	Exon coordinates and protein sequences of the five myrosinase genes (from the RefSeq annotation)
`inputs/gene_models/coding_sequences.fasta`	Coding sequences of the five genes in transcript orientation, exons joined
`inputs/gene_models/P37702.fasta`	Arabidopsis TGG1 (UniProt P37702)
`inputs/gene_models/Q9C5C2.fasta`	Arabidopsis TGG2 (UniProt Q9C5C2)
`inputs/guide_design/hits.tsv`	Every assembly site with up to four mismatches to a candidate guide (stored scan output; columns: contig, position of the 23-mer, strand, guide index, mismatches, mismatch bitmask)
`inputs/guide_design/site_annotation.tsv`	Strand, PAM and RefSeq loci under every perfect site of the two chosen guides (from the stored annotation)
`inputs/guide_design/selected_guides.csv`	The two chosen guides
`inputs/further_copies/annot_route.tsv`	Local-alignment identity of the 30 closest annotated proteins to Arabidopsis TGG1 and TGG2
`inputs/further_copies/kmers.txt`	The 24-mers of the five coding sequences
`inputs/further_copies/kmer_hits.tsv`	Every assembly position of those 24-mers (contig, position, strand, 24-mer index)
`inputs/further_copies/kmer_clusters.tsv`	Clusters of 24-mer hits
`inputs/further_copies/cluster_annotation.tsv`	RefSeq features under each cluster
`inputs/further_copies/five_gene_spans.json`	Genomic span, protein and coding length of the five genes
`code/01_gene_table.py`	Gene table, exons and closest annotated proteins
`code/02_guides.py`	Candidate guides, scores, covering pairs, chosen guides, perfect sites, cut positions and gene maps
`code/03_delivery_model.py`	Isothiocyanate design target, glucosinolate pools, delivery model, sensitivity, allele probability and IFI-B reference line
`code/04_supplementary.py`	Further copies of the myrosinase sequence, second route of the model and graphical abstract
`code/05_full_from_genome.py`	Full-mode guide design from the RefSeq files
`code/guide_design/align.py`	Global protein alignment (Gotoh, BLOSUM62), standard library only
`code/guide_design/design_lib.py`	Coding sequences, candidate guides and MIT scoring shared by the quick and full steps
`code/guide_design/gene_models.py`	Full mode: gene models and coding sequences from the RefSeq files

Path	Contents
`code/guide_design/site_annotation.py`	Full mode: strand, PAM and RefSeq loci under every perfect site
`code/guide_design/scan.c`	C program: scans the whole assembly for sites with up to four mismatches to each candidate guide (strict NGG PAM, both strands)
`code/heavy/NEEDS.txt`	What full mode needs
`research_record/scripts/guide_design/*`	Research scripts for the guide design (gene search, gene models, candidates, scan, scoring, covering pairs, site details) with their outputs and RUN_ORDER.txt
`research_record/scripts/further_copies/*`	Research scripts for the further copies of the myrosinase sequence (alignment route, 24-mer scan, cluster annotation, guide coverage matrix, shares) with their outputs
`research_record/scripts/model/*`	Research scripts of the delivery model (main route and exact-fraction route) with results and agreement
`research_record/second_script_rerun/guide_design/*`	Independent second scripts for gene identity, candidates, scan, scoring, cut positions and site annotation, with their outputs
`research_record/second_script_rerun/model/*`	Independent second script for the delivery-model numbers, with its output
`research_record/outputs/*`	Raw outputs copied from the scripts

Cite this paper

Ford, P.I. (2026). Zero-Bitter Brassica: In Silico Design of Myrosinase Knockout in Brassica oleracea. Scoville Splice Research Series. https://doi.org/10.5281/zenodo.24511326

```
@misc{Ford2026_P06,
  author = {Ford, P.I.},
  title = {Zero-Bitter Brassica: In Silico Design of Myrosinase Knockout in Brassica oleracea},
  year = {2026},
  publisher = {Scoville Splice LLC},
  series = {Scoville Splice Research Series},
  doi = {10.5281/zenodo.24511326},
  url = {https://doi.org/10.5281/zenodo.24511326}
}
```

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

Related papers in this series

ID	Title	DOI
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

Appendix A. Evidence table

Claim ID	Claim	Value	Kind	Source file	Location
P06-C001	The assembly carries five classical myrosinase genes	five	computed	`data/results/gene_summary.csv`	row genes
P06-C002	The five proteins are 67.5 to 72.6% identical to AtTGG1	67.5 to 72.6%	computed	`data/results/gene_summary.csv`	rows identity_AtTGG1_min and identity_AtTGG1_max
P06-C003	Each guide has a perfect site in three of the five genes	three	computed	`data/results/guide_table.csv`	column genes_with_perfect_site

Claim ID	Claim	Value	Kind	Source file	Location
P06-C004	TGG-gRNA1 has MIT score 69.0	69.0	computed	data/results/guide_table.csv	row TGG-gRNA1, column mit_score
P06-C005	TGG-gRNA2 has MIT score 84.0	84.0	computed	data/results/guide_table.csv	row TGG-gRNA2, column mit_score
P06-C006	The two cuts in BoMYR1 (LOC106326798) lie 761 bp apart	761 bp	computed	data/results/guide_summary.csv	row distance_between_cuts_BoMYR1_genomic_bp
P06-C007	The nucleophile codon of BoMYR1 (LOC106326798) lies between the two cuts	flank the nucleophile codon	computed	data/results/guide_summary.csv	row nucleophile_codon_between_cuts_BoMYR1 (yes)
P06-C008	The design target for the knockout is below 0.1 umol/g DW for each isothiocyanate	0.1 umol/g DW	design target	inputs/model_inputs.csv	row itc_limit_KO_each
P06-C009	Modelled change in absorbed sulforaphane of -14% to +36%, depending on how the wild-type arm is treated	−14% to +36%	computed	data/results/model_results.csv	rows range_change_percent_min and range_change_percent_max
P06-C010	The assembly is RefSeq GCF_000695525.1 (BOL)	GCF_000695525.1	definition	inputs/input_sources.csv	row assembly
P06-C011	The RefSeq annotation is release 100	100	definition	inputs/input_sources.csv	row assembly
P06-C012	Scripts run on Python 3.9 or newer	3.9 or newer	definition	code/LIBRARIES.txt	line 1
P06-C013	Proteins of at least 300 residues were compared	300	definition	research_record/scripts/further_copies/cl_gene_set.py	filter len(s)<300
P06-C014	AtTGG1 (UniProt P37702) has 541 residues	541	computed	data/results/gene_summary.csv	row AtTGG1_length_aa
P06-C015	AtTGG2 (UniProt Q9C5C2) has 547 residues	547	computed	data/results/gene_summary.csv	row AtTGG2_length_aa
P06-C016	The 30 closest annotated proteins are listed	30	computed	data/results/gene_summary.csv	row proteins_compared
P06-C017	Alignments use BLOSUM62, gap open -10 and gap extend -0.5	BLOSUM62, gap open −10, gap extend −0.5	definition	code/guide_design/align.py	defaults open_gap and extend_gap
P06-C018	The standard-library aligner and the Biopython aligner give the same identity for all ten gene and Arabidopsis pairs	the same values for all ten gene and Arabidopsis pairs	computed	data/results/gene_table.csv	columns pct_identity_AtTGG1 and pct_identity_AtTGG2 against research_record/second_script_rerun/guide_design/step1_output.txt

Claim ID	Claim	Value	Kind	Source file	Location
P06-C019	The k-mers are 24 bases long	24	definition	`data/results/further_copies_summary.csv`	row kmer_length_bp
P06-C020	6,761 24-mers were searched	6,761	computed	`data/results/further_copies_summary.csv`	row kmers_searched
P06-C021	Hits closer than 3,000 bp are joined into one cluster	3,000 bp	definition	`research_record/scripts/further_copies/cl_gene_set.py`	cluster rule x[0]-cur[-1][0]<3000
P06-C022	Clusters with at least 90 hits are reported	90	definition	`research_record/scripts/further_copies/c4_plain_shares.py`	filter n < 90
P06-C023	Candidates were listed in the three genes with complete models	three	computed	`data/results/guide_summary.csv`	row genes_with_complete_models
P06-C024	The cut window is 250 coding bases either side of the nucleophile codon	250 coding bases	definition	`code/guide_design/design_lib.py`	WINDOW
P06-C025	Candidate GC content is 40 to 70%	40 to 70%	definition	`code/guide_design/design_lib.py`	GC_RANGE
P06-C026	The scan found 1,058 sites with up to four mismatches for the 61 candidates	1,058	computed	`data/results/guide_summary.csv`	row sites_scanned_total
P06-C027	The 70 sites correspond to 61 distinct protospacers	61	computed	`data/results/guide_summary.csv`	row unique_protospacers
P06-C028	The MIT score uses 20 position weights from 0 to 0.851	20 weights, 0 to 0.851	definition	`data/tables/S10.csv`	column Weight
P06-C029	The wild-type sinigrin pool is 14.2 umol/g DW and sets the ceiling for allyl isothiocyanate	14.2 umol/g DW	design assumption	`inputs/model_inputs.csv`	row sinigrin_WT
P06-C030	The wild-type glucoraphanin pool is 28.4 umol/g DW	28.4 umol/g DW	design assumption	`inputs/model_inputs.csv`	row glucoraphanin_WT
P06-C031	The wild-type sulforaphane isothiocyanate is 9.6 umol/g DW	9.6 umol/g DW	design assumption	`inputs/model_inputs.csv`	row sulforaphane_itc_WT
P06-C032	The wild-type sulforaphane isothiocyanate is 33.8% of the glucoraphanin pool	33.8%	design assumption	`data/tables/T03.csv`	row Sulforaphane, column Basis
P06-C033	In-plant hydrolysis of glucoraphanin is assumed to be 0.30 to 0.45 of the pool	0.30 to 0.45	design assumption	`inputs/model_inputs.csv`	rows inplant_hydrolysis_low and inplant_hydrolysis_high

Claim ID	Claim	Value	Kind	Source file	Location
P06-C034	The cooking loss of wild-type tissue is assumed to be 0.35	0.35	design assumption	inputs/model_inputs.csv	row cooking_loss_WT
P06-C035	The gastric loss is assumed to be 0.15	0.15	design assumption	inputs/model_inputs.csv	row gastric_loss
P06-C036	The microbial conversion is assumed to be 0.88	0.88	design assumption	inputs/model_inputs.csv	row microbial_conversion_mid
P06-C037	The microbial conversion range for reading 3 is 0.85 to 0.95	0.85 to 0.95	design assumption	inputs/model_inputs.csv	rows microbial_conversion_low and microbial_conversion_high
P06-C038	The absorption fraction is assumed to be 0.62	0.62	design assumption	inputs/model_inputs.csv	row absorption_fraction
P06-C039	Each fraction moves by 0.10 and each pool by 10% in the sensitivity analysis	0.10 and 10%	definition	inputs/model_inputs.csv	rows sensitivity_step_fraction and sensitivity_step_pool
P06-C040	The baseline in-plant hydrolysis of the sensitivity analysis is 0.375, the midpoint of 0.30 to 0.45	0.375	computed	data/results/sensitivity_baseline.csv	row In-plant hydrolysis (h)
P06-C041	Five genes with two alleles each give ten alleles	ten	computed	data/results/allele_probability.csv	column alleles (10)
P06-C042	The editing probability per allele is assumed to be 0.60	0.60	design assumption	inputs/model_inputs.csv	row editing_probability_per_allele
P06-C043	The reported IC50 of GIV3727 at TAS2R31 against acesulfame K is 6.4 micromolar	6.4	literature value	literature	Slack et al. 2010, doi:10.1016/j.cub.2010.04.043
P06-C044	The reported IC50 of GIV3727 at TAS2R31 against saccharin is 7.9 micromolar	7.9	literature value	literature	Slack et al. 2010, doi:10.1016/j.cub.2010.04.043
P06-C045	The IFI-B reference value is 7.15 micromolar, the mean of 6.4 and 7.9	7.15	computed	data/results/ifi_b_reference.csv	column reference_um
P06-C046	No IFI-B score is assigned to any compound in this paper	assigns no IFI-B score to any compound	computed	data/results/ifi_b_reference.csv	column compounds_scored (0)
P06-C047	BoMYR1 (LOC106326798) lies on chromosome C2	C2	computed	data/results/gene_table.csv	row BoMYR1, column chromosome
P06-C048	BoMYR2 (LOC106319184) lies on chromosome C9	C9	computed	data/results/gene_table.csv	row BoMYR2, column chromosome
P06-C049	Three of the five genes lie on one chromosome	three	computed	data/results/gene_summary.csv	row genes_on_chromosome_C8

Claim ID	Claim	Value	Kind	Source file	Location
P06-C050	BoMYR3, BoMYR4 and BoMYR5 lie on chromosome C8	C8	computed	data/results/gene_table.csv	rows BoMYR3 to BoMYR5, column chromosome
P06-C051	The five proteins are 543 to 548 residues long	543 to 548 residues	computed	data/results/gene_summary.csv	rows protein_length_min_aa and protein_length_max_aa
P06-C052	The five proteins are 66.2 to 68.9% identical to AtTGG2	66.2 to 68.9%	computed	data/results/gene_summary.csv	rows identity_AtTGG2_min and identity_AtTGG2_max
P06-C053	Three genes carry the motif VTENG and two carry ITENG	three VTENG and two ITENG	computed	data/results/gene_summary.csv	rows nucleophile_motif_VTENG_genes and nucleophile_motif_ITENG_genes
P06-C054	The nucleophile glutamate is Glu429, Glu427, Glu427, Glu428 and Glu427 in BoMYR1 to BoMYR5	Glu429, Glu427, Glu427, Glu428 and Glu427	computed	data/tables/T01.csv	column Nucleophile Glu (motif)
P06-C055	The nucleophile glutamate of AtTGG1 is Glu420	Glu420	computed	data/results/gene_summary.csv	row AtTGG1_nucleophile_residue
P06-C056	The five genes have 3 to 12 exons in the assembly	3 to 12 exons	computed	data/results/gene_summary.csv	rows exons_min and exons_max
P06-C057	Three gene models are complete	three	computed	data/results/gene_summary.csv	row models_complete
P06-C058	The BoMYR1 (LOC106326798) model has 1,504 of 1,647 expected coding bases	1,504 of 1,647 coding bases	computed	data/tables/T01.csv	row BoMYR1
P06-C059	The BoMYR2 (LOC106319184) model has 458 of 1,641 expected coding bases	458 of 1,641 coding bases	computed	data/tables/T01.csv	row BoMYR2
P06-C060	The five genes are the five annotated proteins closest to AtTGG1 and AtTGG2 (ranks 1 to 5)	ranks 1 to 5	computed	data/results/closest_proteins.csv	column one_of_the_five, ranks 1 to 5
P06-C061	The lowest local identity to AtTGG1 among the five is 68.4%	68.4%	computed	data/results/gene_summary.csv	row lowest_local_identity_AtTGG1_of_the_five
P06-C062	25 other annotated proteins were compared	25	computed	data/results/gene_summary.csv	row other_proteins_compared
P06-C063	The highest local identity to AtTGG1 among the other proteins is 57.5%	57.5%	computed	data/results/gene_summary.csv	row highest_local_identity_AtTGG1_of_the_rest
P06-C064	70 candidate sites pass the filters in the three complete gene models	70	computed	data/results/guide_summary.csv	row candidate_sites

Claim ID	Claim	Value	Kind	Source file	Location
P06-C065	The MIT scores of the 61 candidates run from 18.1 to 100.0	18.1 to 100.0	computed	`data/results/guide_summary.csv`	rows mit_min and mit_max
P06-C066	No single candidate has a perfect site in more than three of the five genes	three	computed	`data/results/guide_summary.csv`	row max_genes_with_perfect_site_single_guide
P06-C067	Nine pairs of candidates give every gene a perfect site	nine	computed	`data/results/guide_summary.csv`	row guide_pairs_covering_all_five_genes
P06-C068	All nine pairs contain TGG-gRNA2	all contain TGG-gRNA2	computed	`data/results/guide_summary.csv`	row covering_pairs_containing_TGG-gRNA2
P06-C069	TGG-gRNA1 with TGG-gRNA2 has six perfect gene sites; each of the other eight pairs has five	six	computed	`data/results/covering_pairs.csv`	column perfect_gene_sites_total
P06-C070	TGG-gRNA1 is CACACTGCCATGATGGACGC with PAM AGG	CACACTGCCATGATGGACGC	computed	`data/results/guides.fasta`	record TGG-gRNA1
P06-C071	TGG-gRNA2 is GGATACTTTGCATGGGCTCT with PAM CGG	GGATACTTTGCATGGGCTCT	computed	`data/results/guides.fasta`	record TGG-gRNA2
P06-C072	Only one gene, BoMYR1 (LOC106326798), has a perfect site for both guides	one	computed	`data/results/guide_summary.csv`	row genes_with_perfect_site_both_guides (1)
P06-C073	TGG-gRNA1 has 6, 1, 2, 0 and 2 assembly sites at 0, 1, 2, 3 and 4 mismatches	6 / 1 / 2 / 0 / 2	computed	`data/tables/T02.csv`	row Assembly sites at 0 / 1 / 2 / 3 / 4 mismatches, column TGG-gRNA1
P06-C074	TGG-gRNA2 has 7, 1, 2, 1 and 4 assembly sites at 0, 1, 2, 3 and 4 mismatches	7 / 1 / 2 / 1 / 4	computed	`data/tables/T02.csv`	row Assembly sites at 0 / 1 / 2 / 3 / 4 mismatches, column TGG-gRNA2
P06-C075	The two guides have 13 perfect sites in the assembly	13	computed	`data/results/guide_summary.csv`	row perfect_sites_both_guides
P06-C076	Seven perfect sites lie outside the five genes	seven	computed	`data/results/guide_summary.csv`	row perfect_sites_outside_five_genes
P06-C077	TGG-gRNA1 has three perfect sites outside the five genes	three	computed	`data/results/guide_table.csv`	row TGG-gRNA1, column perfect_sites_outside_the_five_genes
P06-C078	TGG-gRNA2 has four perfect sites outside the five genes	four	computed	`data/results/guide_table.csv`	row TGG-gRNA2, column perfect_sites_outside_the_five_genes
P06-C079	Four of the seven outside sites lie in pseudogene loci	four	computed	`data/results/guide_summary.csv`	row perfect_sites_outside_in_pseudogene_loci
P06-C080	One outside site lies in the coding sequence of the protein-coding locus LOC106320470 (myrosinase MB1, 136 residues, scaffold NW_013617532.1)	one	computed	`data/results/perfect_sites.csv`	row TGG-gRNA2, NW_013617532.1:3960

Claim ID	Claim	Value	Kind	Source file	Location
P06-C081	LOC106320470 is annotated as myrosinase MB1 with a 136-residue protein	136 residues	computed	data/results/perfect_sites.csv	row NW_013617532.1:3960, column coding_detail
P06-C082	Two outside sites lie in sequence with no annotated locus	two	computed	data/results/guide_summary.csv	row perfect_sites_outside_no_annotated_locus
P06-C083	The 24-mer search finds 13 clusters with at least 90 hits	13	computed	data/results/further_copies_summary.csv	row clusters
P06-C084	Five clusters are the five genes	five	computed	data/results/further_copies_summary.csv	row clusters_that_are_the_five_genes
P06-C085	Eight clusters are further copies	eight	computed	data/results/further_copies_summary.csv	row further_clusters
P06-C086	Five of the further clusters lie under pseudogene loci	five	computed	data/results/further_copies_summary.csv	row further_clusters_pseudogene_loci
P06-C087	One further cluster lies under a protein-coding locus	one	computed	data/results/further_copies_summary.csv	row further_clusters_protein_coding_locus
P06-C088	Two further clusters have no annotated locus	two	computed	data/results/further_copies_summary.csv	row further_clusters_no_annotated_locus
P06-C089	The largest share of a gene's 24-mers in a further cluster is 73.4%	73.4%	computed	data/results/further_copies_summary.csv	row largest_further_share_percent
P06-C090	TGG-gRNA1 cuts 163, 166 and 169 coding bases upstream of the nucleophile codon in BoMYR1, BoMYR3 and BoMYR5	163, 166 and 169	computed	data/results/cut_sites.csv	rows TGG-gRNA1, column offset_to_nucleophile_codon_bp
P06-C091	TGG-gRNA2 cuts 149 coding bases downstream of the nucleophile codon in BoMYR1, BoMYR2 and BoMYR4	149	computed	data/results/cut_sites.csv	rows TGG-gRNA2, column offset_to_nucleophile_codon_bp
P06-C092	The genomic cut of TGG-gRNA1 in BoMYR1 is at NC_027749.1:48,767,157	48,767,157	computed	data/results/cut_sites.csv	row TGG-gRNA1 BoMYR1
P06-C093	The genomic cut of TGG-gRNA2 in BoMYR1 is at NC_027749.1:48,766,396	48,766,396	computed	data/results/cut_sites.csv	row TGG-gRNA2 BoMYR1
P06-C094	312 coding bases of the assembled BoMYR1 model lie between the two cuts	312 coding bases	computed	data/results/guide_summary.csv	row coding_bases_between_cuts_BoMYR1

Claim ID	Claim	Value	Kind	Source file	Location
P06-C095	312 is a multiple of three	a multiple of three	computed	data/results/guide_summary.csv	row coding_bases_between_cuts_mod_3 (0)
P06-C096	The wild-type isothiocyanate sum is at most 23.8 umol/g DW	23.8 umol/g DW	computed	data/results/model_results.csv	row T3_total_itc_WT
P06-C097	The knockout design target for the sum of the two isothiocyanates is below 0.2 umol/g DW	0.2 umol/g DW	design target	data/results/model_results.csv	row T3_total_itc_limit_KO
P06-C098	The knockout glucoraphanin pool is 28.6 umol/g DW	28.6 umol/g DW	design assumption	inputs/model_inputs.csv	row glucoraphanin_KO
P06-C099	The assumed knockout pools differ from the wild-type pools by -1.4% to +1.9% for the four listed glucosinolates	−1.4% to +1.9%	computed	data/results/glucosinolate_pools.csv	column change_percent, rows 1 to 4
P06-C100	The four listed glucosinolates sum to 56.4 (wild type) and 56.6 (knockout) umol/g DW	56.4 and 56.6 umol/g DW	computed	data/results/model_results.csv	rows T4_sum_listed_WT and T4_sum_listed_KO
P06-C101	The assumed total glucosinolate pools are 63.8 (wild type) and 64.2 (knockout) umol/g DW	63.8 and 64.2 umol/g DW	design assumption	inputs/model_inputs.csv	rows total_glucosinolate_WT and total_glucosinolate_KO
P06-C102	The wild-type pool after cooking is 18.46 umol/g DW	18.46	computed	data/tables/T05.csv	row After cooking loss
P06-C103	The knockout pool after gastric loss is 24.31 umol/g DW	24.31	computed	data/tables/T05.csv	row After gastric loss
P06-C104	The knockout pool after microbial conversion is 21.39 umol/g DW	21.39	computed	data/tables/T05.csv	row After microbial conversion
P06-C105	Absorbed knockout sulforaphane is 13.26 umol/g DW in reading 1	13.26	computed	data/tables/T05.csv	row Absorbed
P06-C106	Absorbed wild-type sulforaphane is 11.45 umol/g DW in reading 1	11.45	computed	data/tables/T05.csv	row Absorbed
P06-C107	Reading 1 gives a knockout to wild-type ratio of 1.159 (+15.9%)	1.159 (+15.9%)	computed	data/results/model_readings.csv	row Reading 1
P06-C108	Reading 2 gives a ratio of 1.363 (+36.3%)	1.363 (+36.3%)	computed	data/results/model_readings.csv	row Reading 2
P06-C109	Reading 3 gives ratios of 0.859 to 0.940 (-14.1% to -6.0%)	0.859 to 0.940 (−14.1% to −6.0%)	computed	data/results/model_readings.csv	rows Reading 3, lowest and Reading 3, highest
P06-C110	With the same cooking loss on both arms, reading 1 gives -24.7%	−24.7%	computed	data/results/model_readings.csv	row Reading 1, same cooking loss on both arms
P06-C111	With the same cooking loss on both arms, reading 2 gives -11.4%	−11.4%	computed	data/results/model_readings.csv	row Reading 2, same cooking loss on both arms

Claim ID	Claim	Value	Kind	Source file	Location
P06-C112	With equal glucoraphanin pools, the three readings give +15.1%, +35.4% and -14.7%	+15.1%, +35.4% and −14.7%	computed	data/results/model_readings.csv	rows with equal glucoraphanin pools
P06-C113	The two routes of the model agree for 34 quantities within 1%	34	computed	data/results/model_agreement_summary.csv	row quantities_compared
P06-C114	The largest relative difference between the two routes is 2.9e-06	2.9e-06	computed	data/results/model_agreement_summary.csv	row largest_relative_difference
P06-C115	In reading 1 the change runs from +0.4% to +37.0% over the sensitivity steps	+0.4% to +37.0%	computed	data/results/sensitivity_summary.csv	row Reading 1
P06-C116	In reading 2 the change runs from +18.2% to +61.1% over the sensitivity steps	+18.2% to +61.1%	computed	data/results/sensitivity_summary.csv	row Reading 2
P06-C117	In reading 3 the change runs from -19.5% to -0.7% over the sensitivity steps	−19.5% to −0.7%	computed	data/results/sensitivity_summary.csv	row Reading 3
P06-C118	The sign of the modelled change is the same over all sensitivity steps in each reading	stays the same in each reading	computed	data/results/sensitivity_summary.csv	column sign_unchanged_over_all_steps (yes for all three readings)
P06-C119	Cooking loss has the largest span in readings 1 and 2 (36.5 and 43.0 percentage points)	36.5 and 43.0 percentage points	computed	data/results/sensitivity_summary.csv	rows Reading 1 and Reading 2, column largest_span_percentage_points
P06-C120	The wild-type glucoraphanin pool has the largest span in reading 3 (18.1 percentage points)	18.1 percentage points	computed	data/results/sensitivity_summary.csv	row Reading 3
P06-C121	The absorption fraction has no effect on the ratio in any reading	no effect	computed	data/results/model_sensitivity.csv	rows Absorption fraction (zero span)
P06-C122	The probability that all ten alleles are edited is 0.0060 (0.6%)	0.0060 (0.6%)	computed	data/results/allele_probability.csv	columns probability_all_alleles_edited and percent

Appendix B. Glossary, abbreviations and units

Term	Meaning	Type
Myrosinase	Enzyme that splits glucosinolates when plant tissue is damaged, releasing isothiocyanates	term
Glucosinolate	Sulfur-containing compound stored in Brassica tissue that myrosinase turns into isothiocyanates	term
ITC	Isothiocyanate: the reactive product of glucosinolate breakdown; allyl isothiocyanate is sharp and sulforaphane is bitter	abbreviation
TGG	Thioglucoside glucohydrolase: the name of the classical myrosinase genes in Arabidopsis (TGG1 and TGG2)	abbreviation
AtTGG1	Arabidopsis thaliana myrosinase 1 (UniProt P37702)	term
AtTGG2	Arabidopsis thaliana myrosinase 2 (UniProt Q9C5C2)	term
BoMYR	The names used in this paper for the five Brassica oleracea myrosinase genes (BoMYR1 to BoMYR5), each given with its RefSeq locus	term
LOC	RefSeq locus symbol of a gene without an established name; the digits equal the NCBI Gene ID	abbreviation
Nucleophile glutamate	The glutamate of the (V/I)TENG motif that attacks the sugar bond in the active site	term
Gene model	The exons and protein of one gene as annotated on the assembly	term
Guide RNA	Short RNA that directs Cas9 to a 20-nucleotide DNA target	term
Protospacer	The 20-nucleotide DNA target matched by the guide	term
PAM	Protospacer adjacent motif: the NGG next to the protospacer that SpCas9 requires	abbreviation
SpCas9	Cas9 nuclease of Streptococcus pyogenes	abbreviation
MIT score	Aggregate specificity score of a guide from 0 to 100 (Hsu et al. 2013): higher means fewer and weaker off-target sites	term
Mismatch	A position where the DNA site differs from the protospacer	term
Perfect site	An assembly site that matches the protospacer at all 20 positions next to an NGG PAM	term
RNP	Ribonucleoprotein: Cas9 protein loaded with its guide RNA	abbreviation
Pseudogene	A gene copy that carries disabling changes and is annotated as non-coding	term
k-mer	A sequence of k nucleotides; here k is 24	term
KO	Knockout: all five myrosinase genes edited	abbreviation
WT	Wild type: unedited tissue	abbreviation
DW	Dry weight	abbreviation
umol/g DW	Micromoles per gram of dry weight	unit
bp	Base pairs	unit
aa	Amino acids (residues)	unit
Mb	Megabases	unit
TAS2R38	Human bitter taste receptor that responds to N-C=S compounds including isothiocyanates	abbreviation

Term	Meaning	Type
TAS2R31	Human bitter taste receptor used for the IFI-B reference line	abbreviation
GIV3727	Small-molecule bitter-taste blocker reported to block TAS2R31	term
IFI	Iyengar Flavor Index: scores one compound against one reference receptor assay	abbreviation
IFI-B	The bitter-blocking axis of the Iyengar Flavor Index	abbreviation
IC50	Concentration of a blocker that reduces a receptor response by half	abbreviation
uM	Micromolar	unit
Reading	One way of treating the wild-type arm in the delivery model (readings 1 to 3)	term
Hydrolysis fraction h	Fraction of the wild-type glucoraphanin pool broken down in the plant, assumed 0.30 to 0.45	term
RefSeq	NCBI reference sequence database	term
Assembly	A genome sequence assembled from reads; here GCF_000695525.1 (BOL)	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_gene_table.py: 2.0 s; 2 code/02_guides.py: 0.1 s; 3 code/03_delivery_model.py: 0.1 s; 4 code/04_supplementary.py: 0.1 s
Input files	18 files, SHA-256 in `code/environment.json`
Data fingerprint	`1c93c01d8bf7d1c087c6e3396848e929ea5bf25c1808f8bfcf79a58758350e9b`

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.24511326 · 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