ViroidHeat — Supplementary Information

Additional file 1 to: Citrus Bark Cracking Viroid Infection Leads to a Coordinated Growth Suppression in Hop.

Supplementary Information: Inventory

Coordinated Suppression of Anabolic Growth Programs in Hop During Citrus Bark Cracking Viroid Associated Stunting. Organised as six thematic Supplementary Notes in order of first citation in the main text; figures (Fig S1–S11) are numbered globally within this file and data tables (Table ST1–ST15) are provided as Additional file 2.

Supplementary Notes

Note	Topic	Cited in main text
1	Leaf proline (methods + figure + discussion)	Methods, Physiological measurements
2	Sequencing & mapping QC (Cascade)	Methods, RNA sequencing
3	DEG overview, sample MDS + treatment-main volcano	Methods, Read processing; Results
4	GO & KEGG gene-set enrichment (Cascade)	Methods, Functional annotation; Results
5	qPCR validation + Sanger amplicon confirmation	Methods, qPCR and Sanger validation; Results; Discussion
6	Reference-genome comparison (Cascade, Kew, Apollo)	Methods, Reference cross-validation

Supplementary Figures

Fig	Note	Content
S1	1	Leaf proline time course
S2	2	Per-library mapping QC (Cascade), fate + library size
S3	3	Sample MDS
S4	3	Treatment main-effect volcano (pooled, 353 DEGs)
S5	4	GO enrichment map
S6	4	GO ridge
S7	4	KEGG enrichment map
S8	4	KEGG ridge
S9	5	qPCR per-gene boxplot
S10	6	Kew read mapping fate (cross-validation)
S11	6	Cascade vs Kew logFC concordance

All figures are sized to Springer column widths (85 mm single / 170 mm double; max height 225 mm).

Supplementary Tables (Additional file 2)

Table	Note	Content
ST1	1	Leaf proline per-plant data
ST2	2	Per-library sequencing/mapping QC (raw→trimmed→Cascade & Kew)
ST3	3	DEGs May (169)
ST4	3	DEGs August (42)
ST5	3	Persistent core (11, May ∩ August)
ST6	3	DEGs treatment-overall, season-averaged (353)
ST7	4	GSEA GO/KEGG (Contrast + Database cols)
ST8	4	ORA per timepoint, GO (a) / KEGG (b)
ST9	5	qPCR primers (F/R sequence, length, GC %)
ST10	5	qPCR × RNA-seq correlation (16 points)
ST11	5	Sanger amplicon consensus sequences (FASTA)
ST12	6	Cascade↔Kew reciprocal best hits
ST13	6	DEGs shared Cascade ∩ Kew (20)
ST14	6	GO (a) / KEGG (b) enrichment Cascade vs Kew
ST15	2/3	Raw featureCounts (a) / TMM-normalised (b) counts

Notes on completeness

- **Data & code availability:** raw reads at ENA **PRJEB110290**; all scripts, intermediate tables
  and publication figures at the accompanying repository (Zenodo DOI to be assigned).

- **Supplementary Methods** are distributed into the relevant Notes (proline assay → Note 1; STAR/
  edgeR pipeline → main Methods + Note 2; Sanger → Note 5; reference comparison → Note 6).

- **Optional / deliberately not included:** a per-timepoint GO ridge (May | August) symmetric to the
  main KEGG ridge, data are present (ST7), build only if reviewers request GO/KEGG symmetry.

Supplementary Note 1: Leaf proline

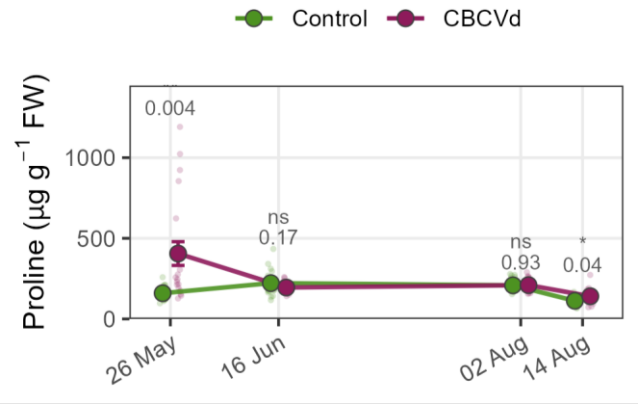
Methods

Free proline was determined in leaf material of the same field/chamber experiment by the acid-ninhydrin colorimetric assay (after Bates et al. [S1]): tissue was extracted, reacted with ninhydrin, and absorbance read at 520 nm against an L-proline standard curve (three technical replicates per sample), and the content expressed per gram fresh weight. Samples spanned the 2023 season (26 May, 16 June, 2 August, 14 August) for CBCVd-infected and healthy control plants across both temperature chambers and irrigation regimes (pooled here). Per-date Welch *t*-tests compared CBCVd vs. control. Raw values: **Table ST1**.

Reference cited in this Note

[S1] Bates LS, Waldren RP, Teare ID. Rapid determination of free proline for water-stress studies. Plant Soil. 1973;39(1):205–7. https://doi.org/10.1007/BF00018060

Result



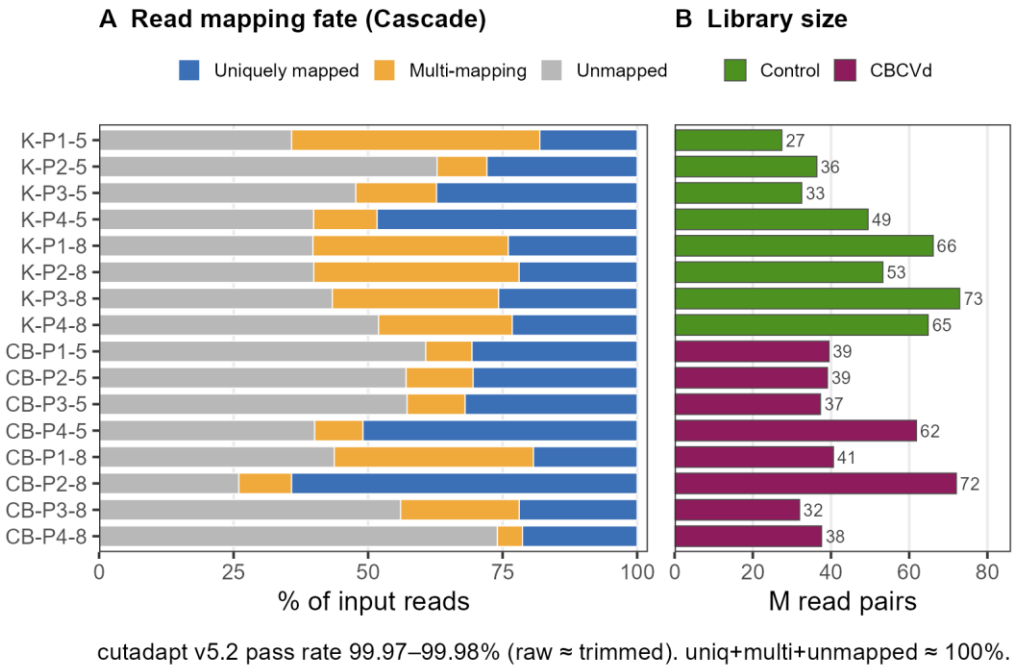
Supplementary Fig S1. Leaf proline over the 2023 season, by infection status. Leaf proline (µg g^{-1} fresh weight); points = individual plants (jittered), symbols and error bars = group mean ± SE; green = healthy control, magenta = CBCVd-infected; per-date Welch *t*-tests annotated (significance code and *p*). Temperature chambers (1 and 2) and irrigation regimes are pooled, and measurement dates differ from the RNA-seq sampling.

Brief discussion

Proline tended to be higher in CBCVd-infected leaves and reached significance on 2 of 4 dates (26 May $p = 0.004$; 14 Aug $p = 0.039$), but the pattern is weak and inconsistent, the large May contrast is inflated by a few extreme values, and the mid-season dates show no difference. Notably, despite the ABA/drought-like transcriptional programme in the DEG set (*ATHB7, DHN1, MFT*), proline , the classic ABA/osmotic-stress osmolyte, does **not** accumulate consistently. This argues that the early response is a transcriptional reprogramming rather than a fully realised osmotic-stress physiology. (Leaf chlorophyll, the other physiological readout, is shown in main **Fig 1B**.)

Supplementary Note 2: Sequencing & mapping quality control

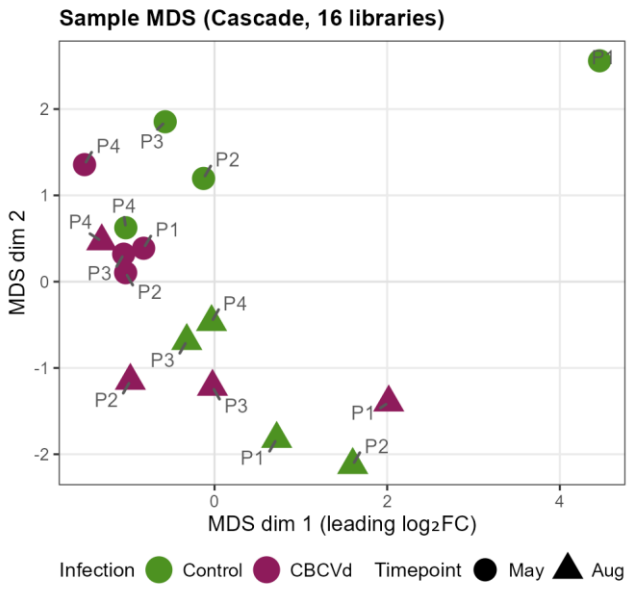
Per-library read and mapping statistics for the 16 mRNA libraries against the **Cascade** reference used for all downstream analyses. (The parallel Kew/drHumLupu mapping is shown as a reference cross-validation in **Supplementary Note 6**.) Full per-library table: **Table ST2**.



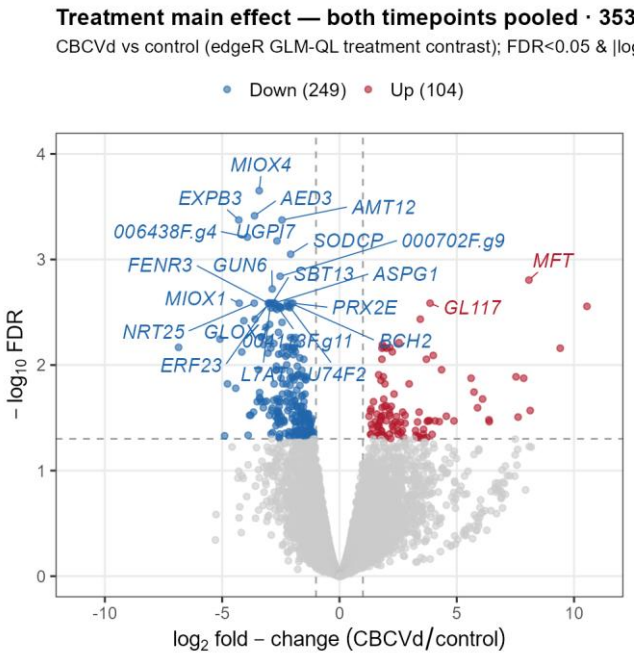
Supplementary Fig S2. Per-library sequencing and mapping QC (Cascade reference). (A) Read mapping fate (STAR 2-pass): uniquely mapped (blue), multi-mapping (amber), unmapped (grey), as % of input reads. **(B)** Library size (million raw read pairs) by infection. cutadapt v5.2 retained **99.97–99.98 %** of read pairs (raw ≈ trimmed). The **unmapped fraction is high** (mean uniquely mapped ≈ 31 %, range 18–64 %), as expected from **Herkules × Cascade cultivar divergence**, hop genomes are highly heterozygous with large structural variation between cultivars. This is *not* a library-quality problem: trimming pass rates are ~100 %, no library is an outlier, and the same pattern reproduces against an independent reference (Note 6), where the DEG calls are concordant, confirming the biology is robust to the high unmapped fraction. Cascade mapping statistics are from the original study STAR run (parameters as in Methods).

Supplementary Note 3: Differential-expression overview

Sample relationships and the season-averaged (treatment main-effect) differential expression that underlie the timepoint-resolved analysis in the main text. DEG lists: **Tables ST3–ST6**.



Supplementary Fig S3. Sample MDS (edgeR leading-log_2FC distances, Cascade, 16 libraries). Colour = infection (green control / magenta CBCVd), symbol = timepoint (circle May / triangle August), labels = biological replicate plant. Dimension 2 separates the two timepoints; the infection effect is subtler, consistent with the modest per-timepoint DEG counts. No library is an extreme outlier.



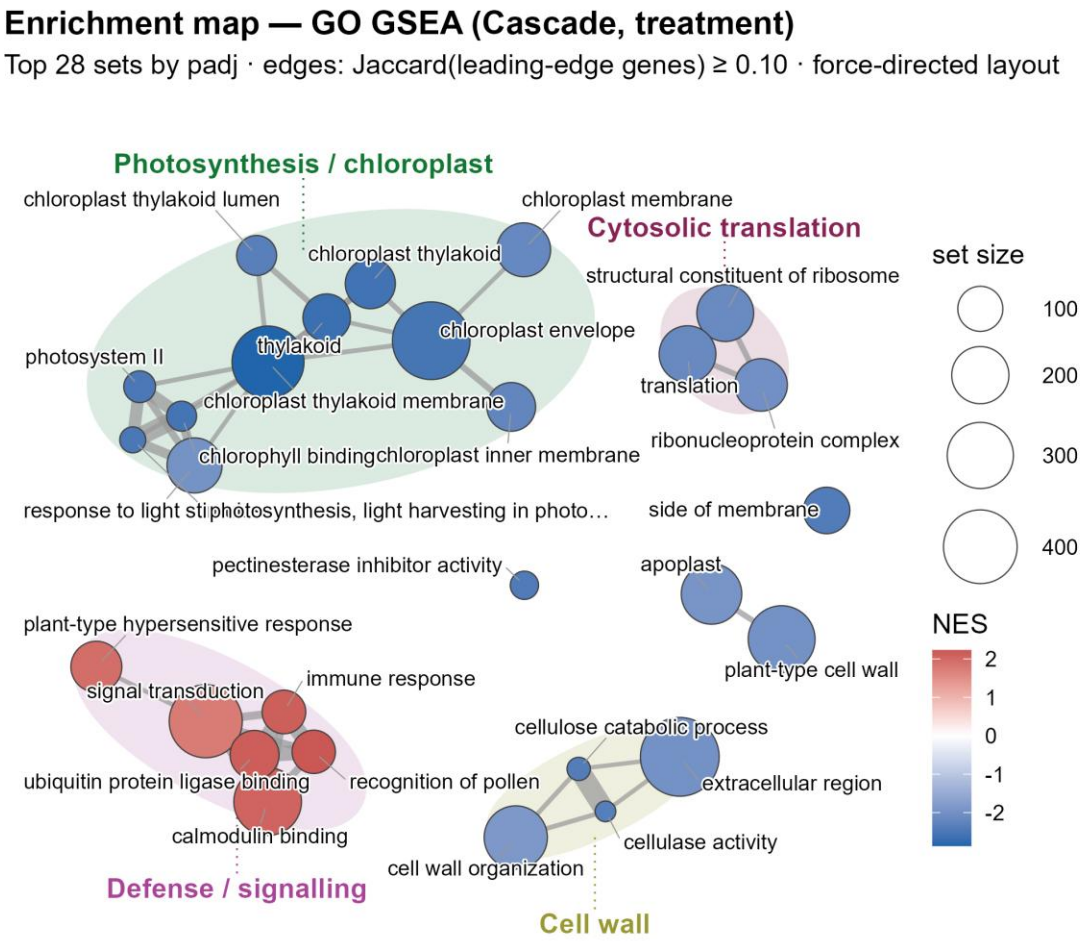
Supplementary Fig S4. Volcano of the treatment main effect (CBCVd vs control, edgeR GLM-QL `~ plant + treatment * timepoint` treatment contrast; both timepoints pooled). Blue = down, red = up (FDR < 0.05 & $|\log_2 FC| \geq 1$); **249 down / 104 up = 353 DEGs**. The response is strongly down-skewed and recovers the same anabolic-suppression core seen per timepoint (MIOX4, NRT25, SODCP, EXPB3, UGPI7 …), with MFT and GL117 among the few induced loci. This season-averaged view is the highest-power contrast but merges the acute (May) and chronic (August) phases; the main text therefore presents the response timepoint-resolved (main Fig 2).

Tables: ST3 DEGs May (169) · ST4 DEGs August (42) · ST5 persistent core (11, May ∩ August) · ST6 treatment-overall season-averaged (353). Normalised/raw counts: ST15.

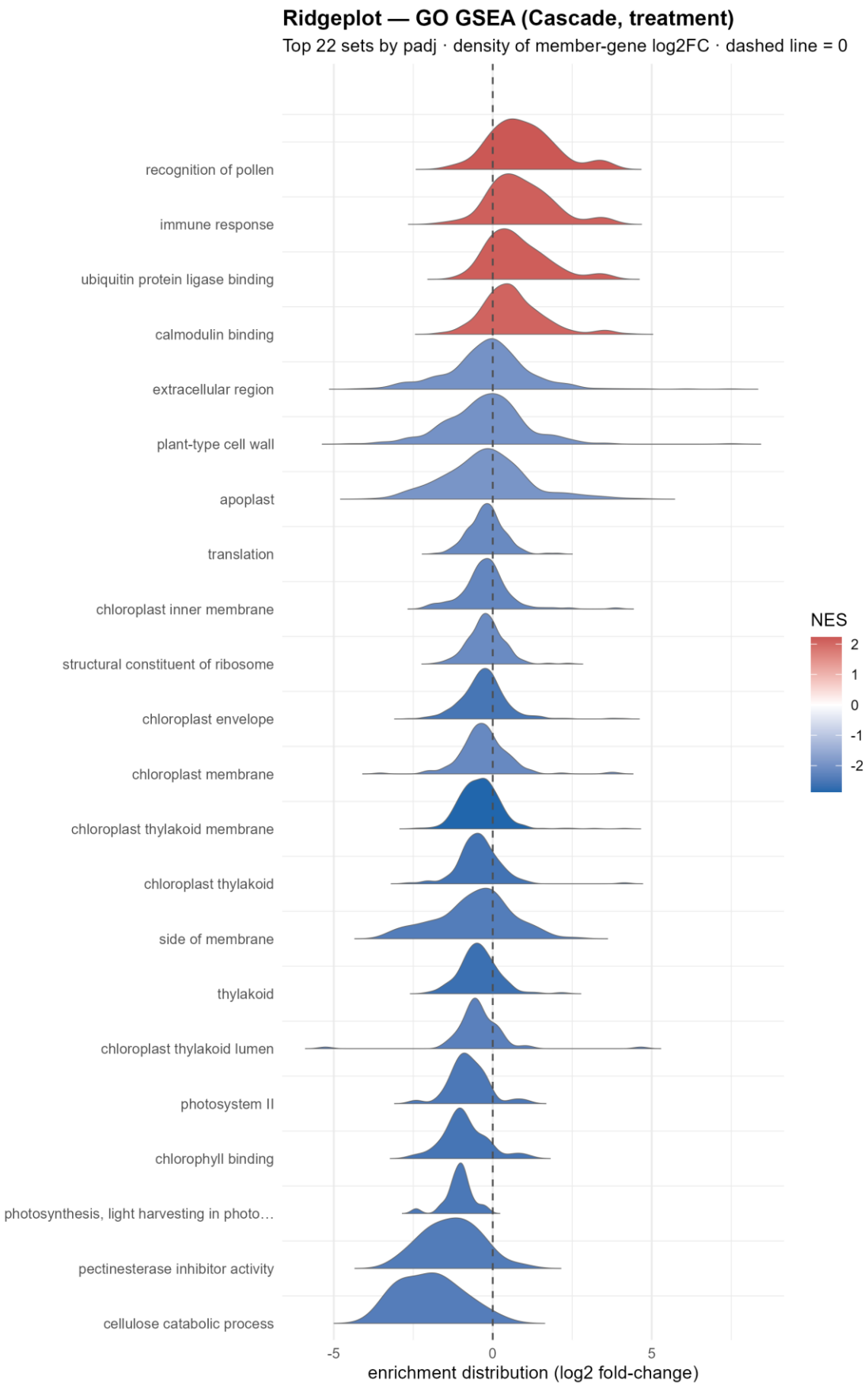
Supplementary Note 4: GO & KEGG gene-set enrichment

Season-averaged GSEA (fgsea) on the full ranked gene lists, complementing the per-timepoint KEGG ridge in the main text (main Fig 3). Enrichment tables: **ST7** (GSEA GO/KEGG, single table with Contrast and Database columns), **ST8a/b** (per-timepoint ORA, GO and KEGG).

GO

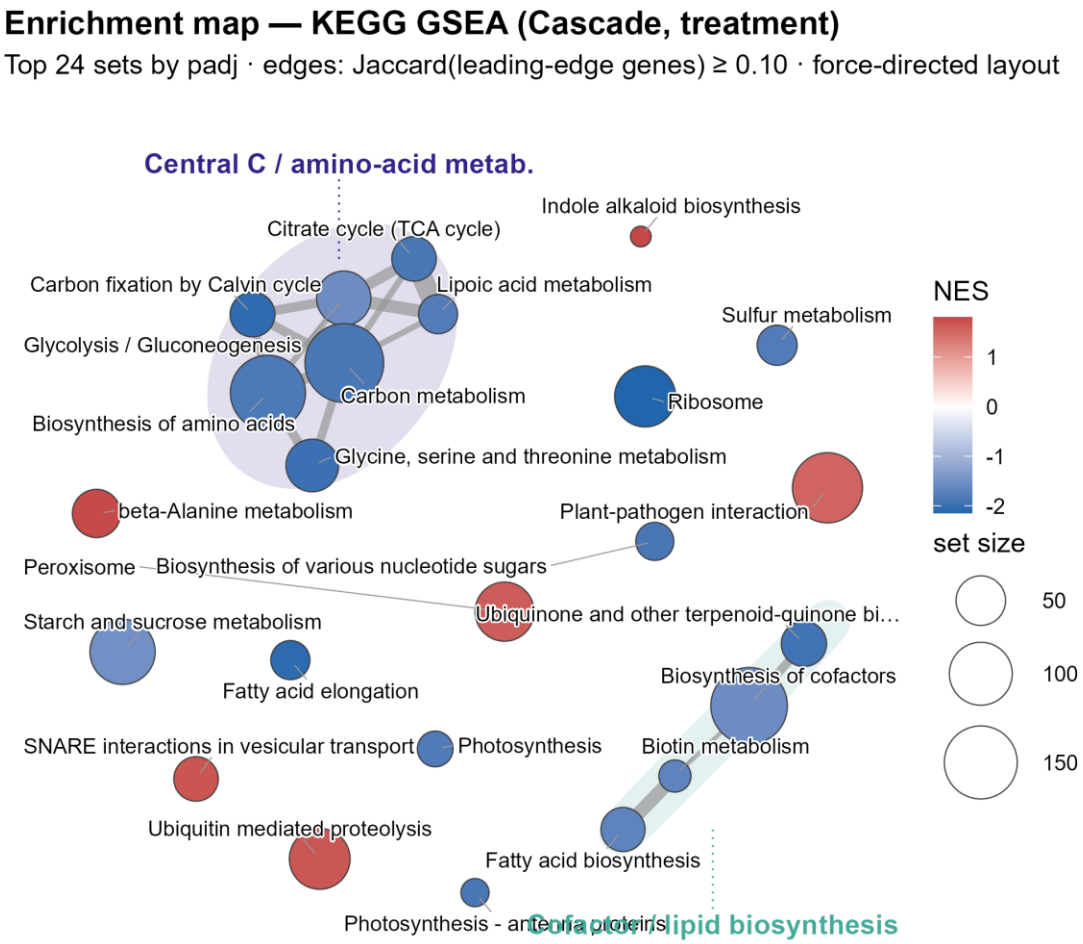


Supplementary Fig S5. GO enrichment map (season-averaged treatment contrast). Nodes = significant GO sets (fill = NES, blue suppressed / red activated, same scale as main Fig 2; size = set size), edges = shared leading-edge genes (Jaccard ≥ 0.10), force-directed layout. Shaded coloured ellipses mark connected-component clusters by functional theme: a large **Photosynthesis / chloroplast** cluster and a **Cell wall** cluster are suppressed (blue), a **Defense / signalling** cluster is activated (red), and **Cytosolic translation** forms its own group.

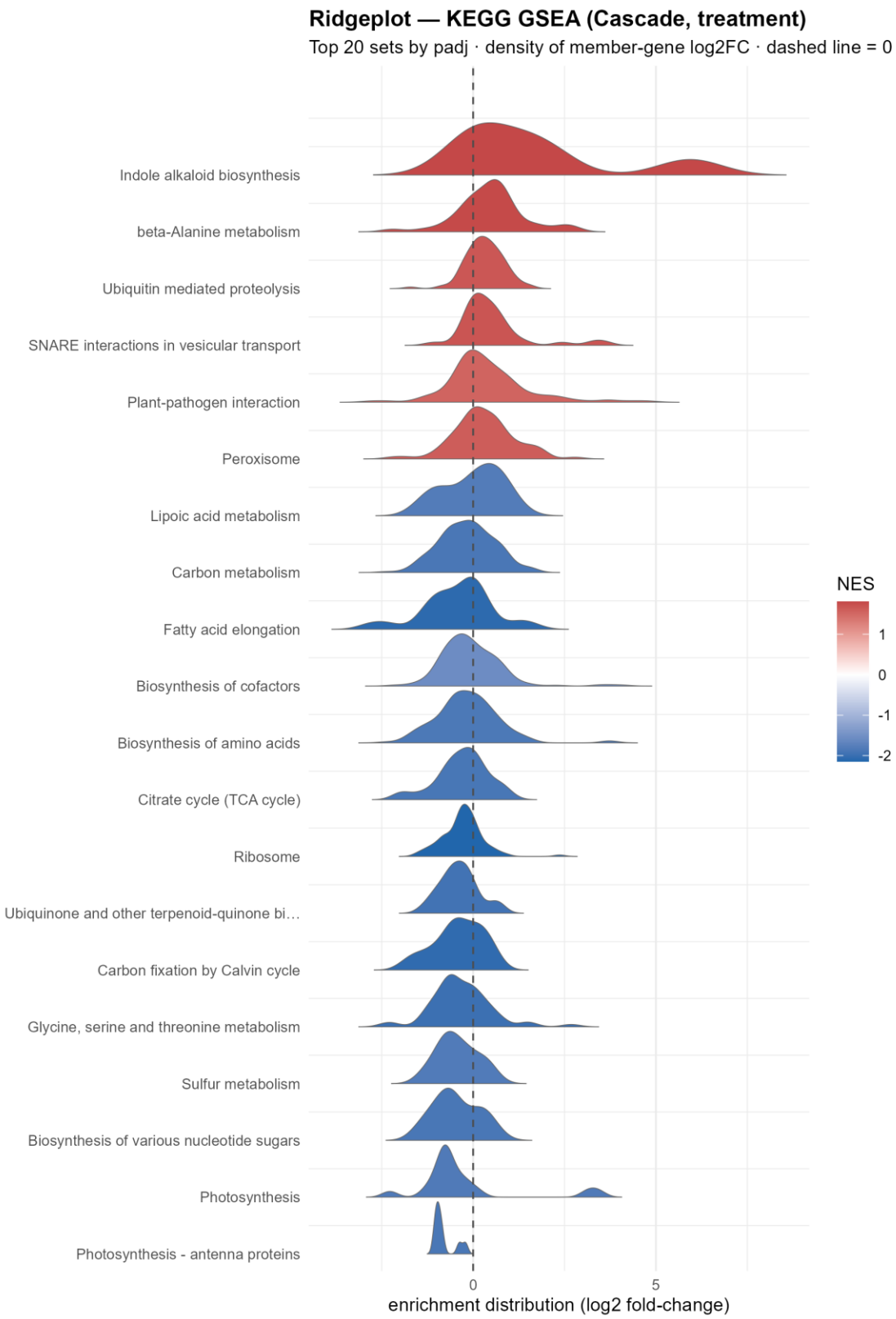

Ridgeplot — GO GSEA (Cascade, treatment)
Top 22 sets by padj · density of member-gene log2FC · dashed line = 0
recognition of pollen
immune response
ubiquitin protein ligase binding
calmodulin binding
extracellular region
plant-type cell wall
apoplast
translation
chloroplast inner membrane
structural constituent of ribosome
chloroplast envelope
chloroplast membrane
chloroplast thylakoid membrane
chloroplast thylakoid
side of membrane
thylakoid
chloroplast thylakoid lumen
photosystem II
chlorophyll binding
photosynthesis, light harvesting in photo…
pectinesterase inhibitor activity
cellulose catabolic process
-5
0
5
enrichment distribution (log2 fold-change)
NES
2
1
0
-1
-2

Supplementary Fig S6. GO ridge plot. Density of member-gene $\log_2$FC per top set; the headline term **chloroplast thylakoid membrane (NES −2.48, padj 7×10^{-23})** sits in the strongly suppressed block.

KEGG



Supplementary Fig S7. KEGG enrichment map (season-averaged). Encoding as Fig S5. A **Central C / amino-acid metabolism** cluster and a **Cofactor / lipid biosynthesis** cluster are suppressed (blue); indole-alkaloid biosynthesis, ubiquitin-mediated proteolysis, SNARE and plant-pathogen interaction are activated (red).



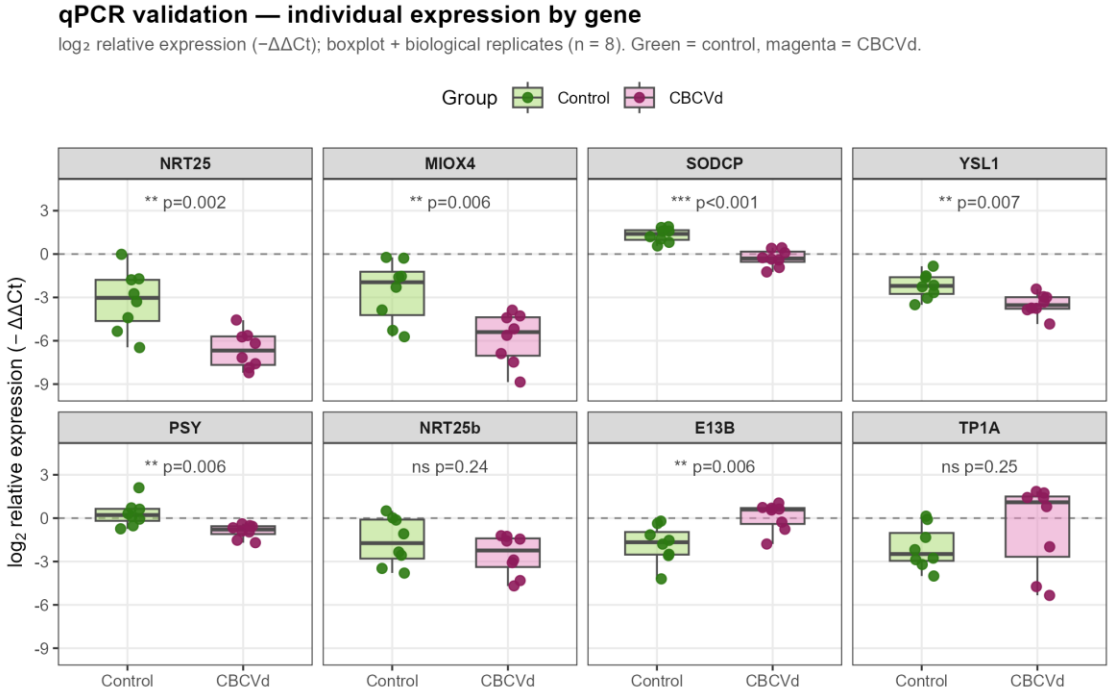
Supplementary Fig S8. KEGG ridge plot (season-averaged) complementing the per-timepoint main Fig 3. KEGG enrichment used an *Arabidopsis thaliana* proxy mapping; pathway-level calls are approximate.

Supplementary Note 5: qPCR validation & Sanger amplicon confirmation

Independent validation of the RNA-seq DEGs by RT-qPCR, with the qPCR amplicons of the significant differentially expressed genes checked for identity by Sanger sequencing. Primers: **Table ST9**; qPCR×RNA-seq values: **Table ST10**; Sanger amplicon consensus sequences: **Table ST11**. (The main-text qPCR×RNA-seq correlation is main Fig 2D.)

The qPCR followed MIQE essentials: amplification efficiencies from LinRegPCR, **7SL** reference gene, the Linker sample as calibrator, and n = 8 biological replicates per group.

Per-gene qPCR detail



Supplementary Fig S9. Per-sample relative expression ($\log_2$, −ΔΔCt vs 7SL, Linker calibrator) by gene and treatment group (n = 8); boxplots with individual points, green = control, magenta = CBCVd. **NRT25b** is excluded from the quantitative correlation (LinRegPCR efficiency 71.9 ± 6.9 %, flagged for redesign); overall Pearson r = 0.89 across the remaining genes (main Fig 2D).

Sanger confirmation of amplicon identity

The qPCR products were Sanger-sequenced from both the forward and reverse primer across two sequencing rounds (Macrogen Europe), base-called and quality-trimmed (Biopython, Mott's algorithm), and identified by BLASTn against the full panel of target transcripts. Accepted reads (≥ 25 bp aligned, ≥ 85 % identity) were assembled into a reference-guided, per-amplicon consensus by orienting each read onto the plus strand of its target transcript and merging overlapping reads by per-position majority vote. Every assembled amplicon returned its intended target transcript as the best BLAST hit, giving a

gapless, multi-read consensus for each of the differentially expressed genes below; the consensus sequences are provided in **Table ST11**.

Gene	Target transcript	Function	Reads	Consensus
NRT25	000404F.g19.t1	High-affinity nitrate transporter 2.5	4	169 bp, gapless
MIOX4	001331F.g30.t3	Inositol oxygenase	4	98 bp, gapless
SODCP	004860F.g4.t1	Superoxide dismutase [Cu-Zn]	3	62 bp, gapless
YSL1	000482F.g43.t1	Metal-nicotianamine transporter	6	189 bp, gapless
PSY	000802F.g35.t2	Phytoene synthase	4	148 bp, gapless
DHN1	000562F.g82.t1	Dehydrin	4	73 bp, gapless
E13B	007625F.g12.t1	Glucan endo-1,3-β-glucosidase	4	128 bp, gapless

Reads were pooled across both sequencing rounds; per-read BLAST metrics and the read-to-gene assignment are in the run tables. # Supplementary Note 6: Reference-genome comparison (Cascade, Kew and Apollo)

Rationale

The primary analysis used the *H. lupulus* cv. Cascade assembly (`combinedMaskedCascadePrimary_v2`), which diverges from the sequenced cultivar (Herkules) and yields a high unmapped fraction (Note 2). To confirm that the CBCVd response is biological rather than a Cascade-assembly artefact, and that Cascade is the best available reference, we benchmarked three public hop genomes against each other on the same data.

Design

All three references were first pre-screened by mapping quality on all 16 libraries (the identical 2 × 2 design, CBCVd vs control × May vs August, four replicates). The differential-expression deep-dive was then run only on the references that passed the pre-screen (Cascade and Kew), using the same edgeR treatment contrast (CBCVd vs control, FDR < 0.05 and $|\log_2 FC| \geq 1$). The European chromosome-level genome Kew serves as the independent cross-validation, and the European × North-American hybrid genome Apollo was tested as a candidate alternative.

Mapping pre-screen (16 libraries)

Mean STAR mapping across the 16 libraries (unique / multiple / unmapped):

Reference	Accession	Type	n libs	Unique %	Multi %	Unmapped %
Cascade	GCA_023660075	draft, contigs (US cultivar)	16	**31.1**	20.4	48.5
Kew / drHumLupu1.1	GCA_963169125	chromosome-level (European)	16	**31.1**	5.6	63.4
Apollo	GCA_963992765	chromosome-level (Hybrid)	16	30.7 [1]	26.6 [1]	42.7 [1]

[1] Apollo was run single-pass without a gene model (no matching annotation available, see below), per-library unique range 12.6–69.8 %. Multi = "multiple" + "too many" loci; unmapped = "too short" + "other". Per-library values in **Table ST2**.

The three genomes are tied on usable (uniquely mapped) signal (Cascade 31.1 %, Kew 31.1 %, Apollo 30.7 %): neither chromosome-scale contiguity (Kew) nor the hybrid assembly (Apollo) raises unique mapping above the Cascade draft. The surplus only shifts elsewhere, into multi-mapping for Cascade and Apollo (20–27 %, from redundant or repetitive content) and into the unmapped fraction for Kew (63 %). The consistently high unmapped fraction across all genome references reflects the divergence between Herkules and every public cultivar assembly (hop genomes are highly heterozygous); it is not a library-quality problem, since adapter and quality trimming passed essentially all reads (Note 2).

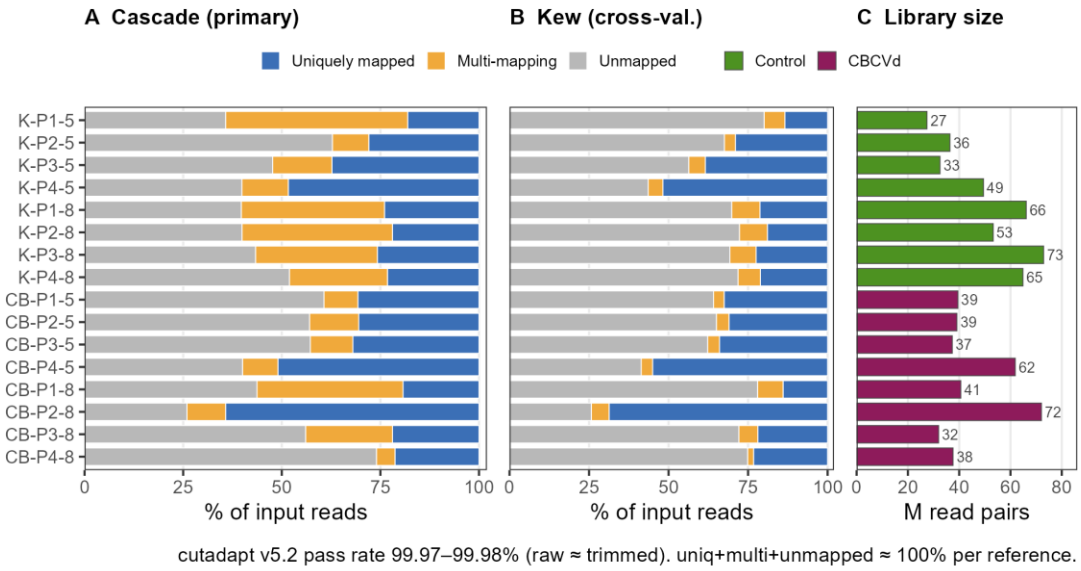
Apollo: assessed by mapping but not taken forward to differential expression

Apollo was evaluated at the mapping stage only and was not carried into the differential-expression analysis, for two reasons. First, it provided no gain in usable signal: its mean unique mapping (30.7 %) was the lowest of the three references, marginally below Cascade and Kew (31.1 %), while its multi-mapping was the highest (mean 26.6 %, up to 61 % in individual libraries), consistent with duplicated and repetitive content in the hybrid assembly. This comparison is conservative in Apollo's favour, as Apollo was mapped single-pass without a gene model whereas Cascade and Kew were mapped two-pass with annotation. Second, because the Apollo genome was published only very recently, no matching gene annotation is yet publicly available for gene-level read counting: the only Apollo annotation (PGSB Augustus models) is on the phased pseudomolecules and does not correspond to the downloadable haploid assembly, and the matching phased FASTA is not publicly deposited. Given the absence of any mapping advantage together with the lack of a usable annotation, we restricted the differential-expression deep-dive to Cascade and Kew.

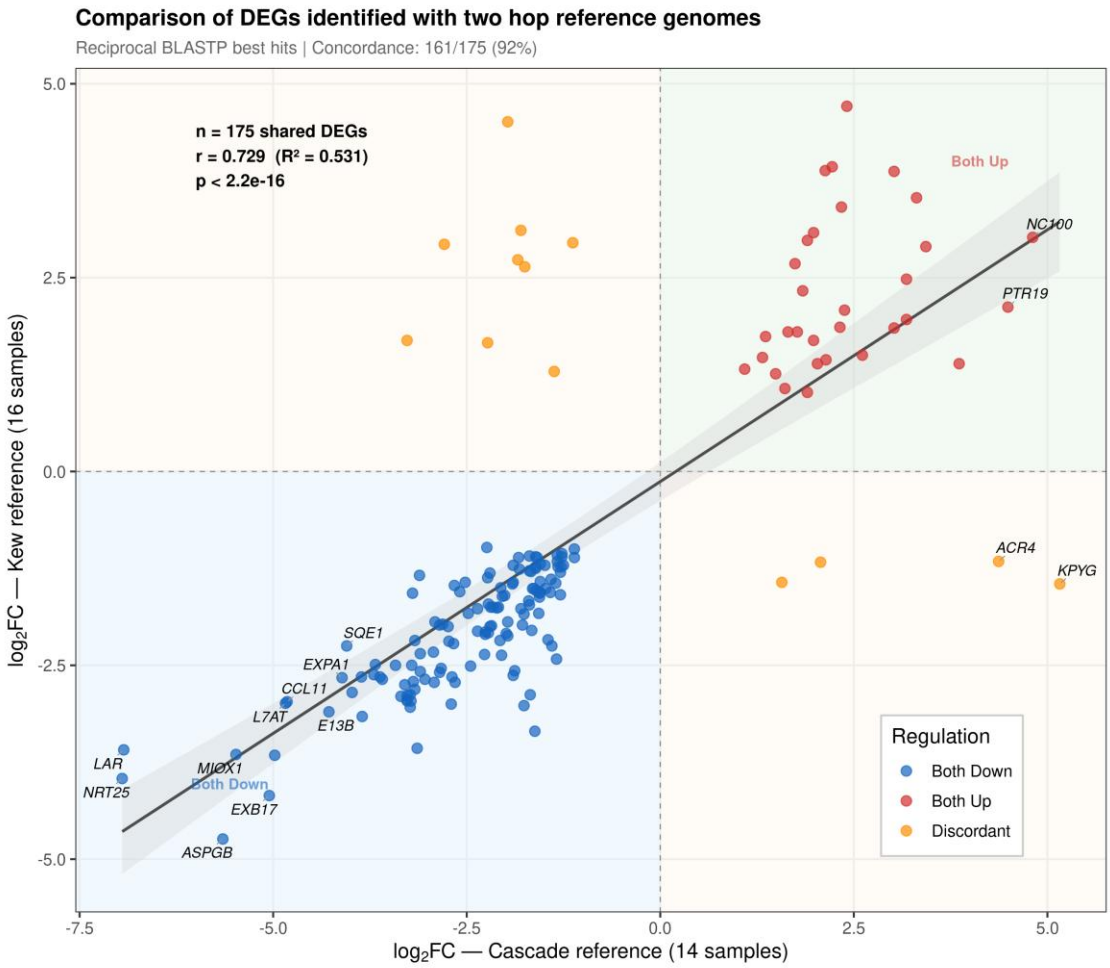
Cascade vs Kew cross-validation

DEGs were independently called against the Kew reference and compared with Cascade. The decisive treatment contrast yields a closely matching number of DEGs (Cascade 353, Kew 371), and the two references agree in direction across shared genes:

- **175 reciprocal-best-hit ortholog pairs** between the Cascade and Kew DEG sets (**Table ST12**).
- **92.0 % directional (sign) agreement** in $\log_2$FC across shared orthologs (161/175).
- **20 DEGs significant in both references** (**Table ST13**); GO/KEGG enrichment of each set in **Table ST14**.
- The same high unmapped fraction appears under both references (cultivar divergence), while multi-mapping is lower against the more contiguous Kew assembly.



Supplementary Fig S10. Per-library read mapping fate, Cascade vs Kew (STAR 2-pass, 16 libraries). (A) Cascade (primary reference) and (B) Kew (cross-validation): uniquely mapped (blue), multi-mapping (orange) and unmapped (grey) as percent of input reads per library; (C) library size in million read pairs, bars coloured by treatment (green control, magenta CBCVd). Mean uniquely mapped is about 31 % against both references, so the high unmapped fraction is reference-independent and reflects Herkules genotype divergence rather than library quality (cutadapt pass rate 99.97 to 99.98 %).



Supplementary Fig S11. $\log_2$FC of shared orthologs (Cascade vs Kew); the strong positive correlation and sign agreement (92.0 % concordant, 161/175) confirm reference-independence.

Conclusion

Tested on the same 16 libraries, Cascade (draft, contigs) gives 353 treatment DEGs and is the best and chosen primary reference; Kew (chromosome-level, European) gives 371 with identical unique mapping (31.1 %) and validates Cascade at 92 % directional concordance, so chromosome-scale contiguity adds no extra usable signal; and Apollo (chromosome-level, hybrid) was assessed but not carried into the differential-expression analysis, showing the lowest unique mapping and the highest multi-mapping of the three and lacking a publicly available annotation matching its downloadable assembly. Usable, uniquely mapped signal, not assembly contiguity, limits DEG detection, because Herkules is a different genotype from every public assembly. The Cascade result is therefore as good as it gets with current references, and the case for a dedicated, chromosome-scale Herkules genome is strengthened: neither a European chromosome-level genome nor a hybrid one beats the Cascade draft.