Reads were trimmed for adaptor sequences using Cutadapt v2.3 [1] and read quality assessed using FastQC[2]. They were then aligned against the *Brassica oleracea* TO1000 (EnsemblPlants 43[3]) genome [4] and uniquely mapped reads for each gene counted using STAR v2.7.1a [5]. Read counts were imported into R v3.6.0 [6] and differential expression determined using DESeq2 v1.24.0 [7]. Specifically, read counts were normalized, with the TO1000 samples as reference, and fitted to a parametric model. We looked for differentially expressed genes for each pairwise comparison (Kale vs Cabbage, Kale vs TO1000, and Cabbage vs TO1000). As each pairwise comparisons increase the potential number of false positives, these were combined into a single table and the False Discovery Rate (FDR) adjusted for the entire table. Genes with an FDR < 0.05 and Log-fold change of 1 were considered as differentially expressed genes.

Gene ontology (GO) terms[8] are not currently available for the TO1000 assembly. In order to annotate TO1000 gene models with GO terms, TO1000 protein sequences were Blast (blastp) against protein sequences from UniProtKB Swiss-Prot (release 2019\_05)[9] using Diamond[10], with an e-value cutoff of 1e-5 and a max of 1 hit per query. UniProt IDs were then used to download GO terms and these were mapped to the T01000 gene models. In total, 35,422 genes (out of 59,225 total) had a GO term assigned. GO term enrichment was performed using topGO[11] and GO.db[12] with the parent-child algorithm[13] and Fisher’s Exact Test.

Plots were made using the R packages ggplot2[14], pheatmap[15], and VennDiagram[16]. All code for bioinformatic analysis and original figure pdfs and tables are available on GitHub ().

1. Martin, M., *Cutadapt removes adapter sequences from high-throughput sequencing reads.* EMBnet.journal, 2011. **17**(1).

2. Andrews, S., *FastQC A Quality Control tool for High Throughput Sequence Data.* 2010.

3. Kersey, P.J., et al., *Ensembl Genomes 2018: an integrated omics infrastructure for non-vertebrate species.* Nucleic Acids Research, 2017. **46**(D1): p. D802-D808.

4. Parkin, I.A., et al., *Transcriptome and methylome profiling reveals relics of genome dominance in the mesopolyploid Brassica oleracea.* Genome Biol, 2014. **15**(6): p. R77.

5. Dobin, A., et al., *STAR: ultrafast universal RNA-seq aligner.* Bioinformatics, 2013. **29**(1): p. 15-21.

6. Team, R.C. *R: A Language and Environment for Statistical Computing*. Available from: <http://www.R-project.org/>.

7. Love, M.I., W. Huber, and S. Anders, *Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2.* Genome Biol, 2014. **15**(12): p. 550.

8. Ashburner, M., et al., *Gene ontology: tool for the unification of biology. The Gene Ontology Consortium.* Nat Genet, 2000. **25**(1): p. 25-9.

9. Consortium, T.U., *UniProt: a worldwide hub of protein knowledge.* Nucleic Acids Research, 2018. **47**(D1): p. D506-D515.

10. Buchfink, B., C. Xie, and D.H. Huson, *Fast and sensitive protein alignment using DIAMOND.* Nature Methods, 2014. **12**: p. 59.

11. Alexa, A. and J. Rahnenfuhrer. *topGO: Enrichment Analysis for Gene Ontology*. R package 2019; Available from: <https://www.bioconductor.org/packages/release/bioc/html/topGO.html>.

12. Carlson, M. *GO.db: A set of annotation maps describing the entire Gene Ontology*. R package 2019; Available from: <https://www.bioconductor.org/packages/release/data/annotation/html/GO.db.html>.

13. Grossmann, S., et al., *Improved detection of overrepresentation of Gene-Ontology annotations with parent–child analysis.* Bioinformatics, 2007. **23**(22): p. 3024-3031.

14. Wickham, H., *ggplot2: Elegant Graphics for Data Analysis*. 2016: Springer-Verlag New York.

15. Kolde, R. *pheatmap: Pretty Heatmaps.* R package 2019; Available from: <https://CRAN.R-project.org/package=pheatmap>.

16. Chen, H. *VennDiagram: Generate High-Resolution Venn and Euler Plots.* R package 2018; Available from: <https://CRAN.R-project.org/package=VennDiagram>.