

Supplementary figures

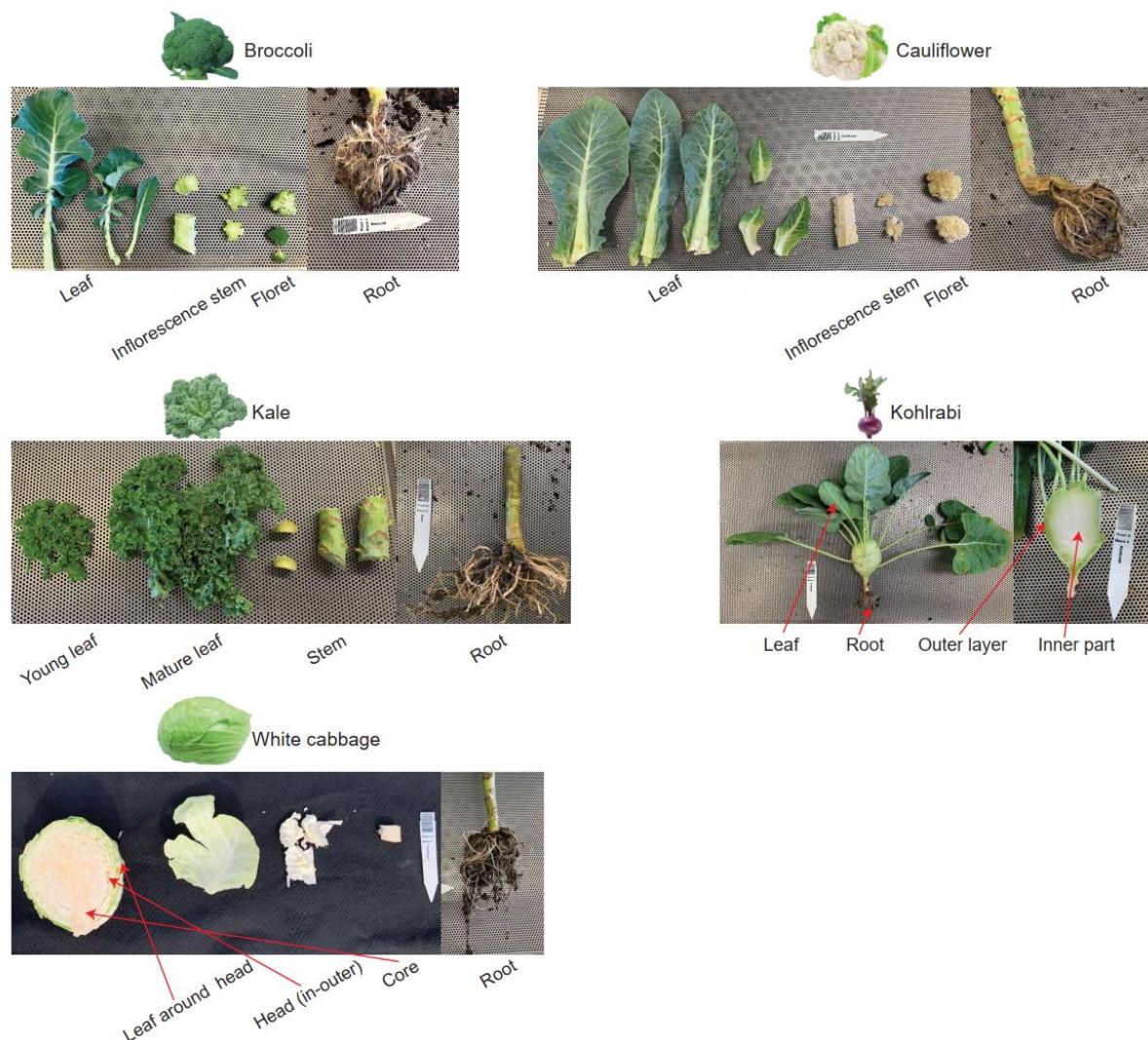


Fig. S1 The five *B. oleracea* morphotypes and collected tissues for GSL extraction and RNA sequencing.

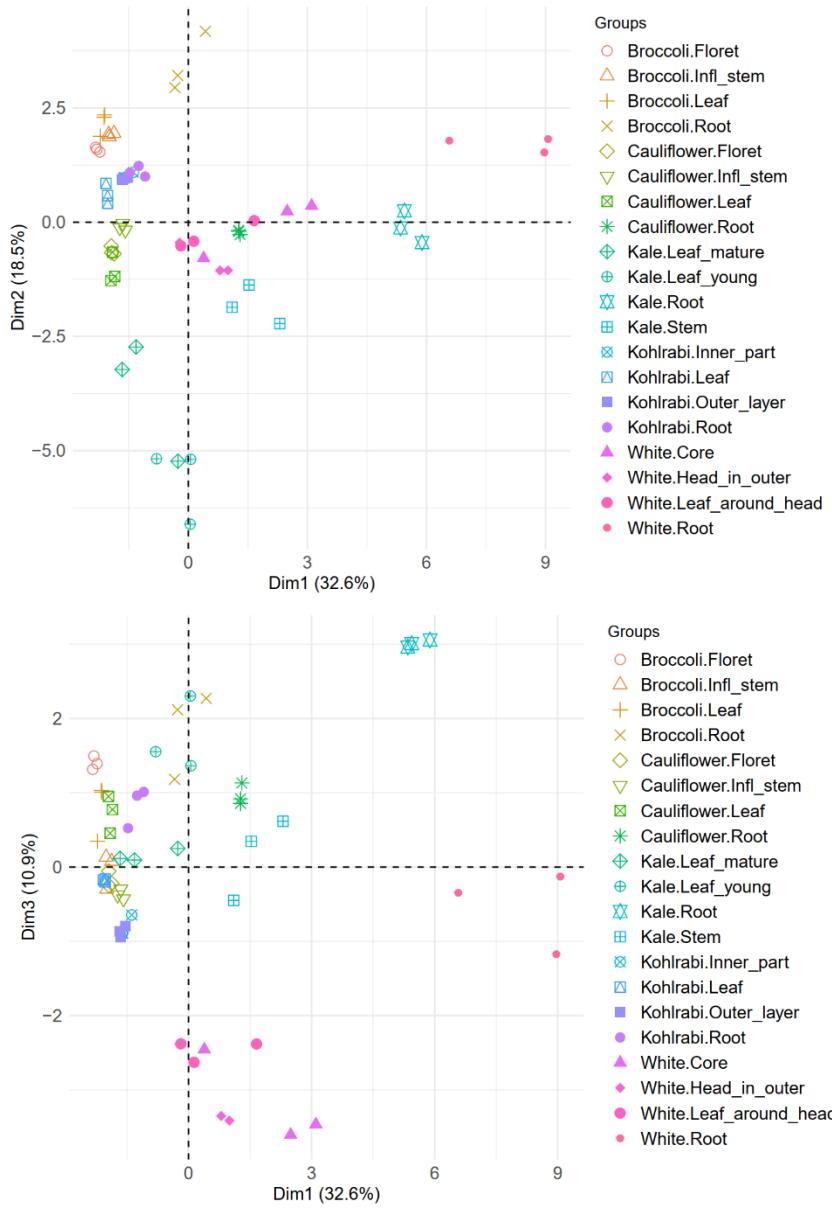
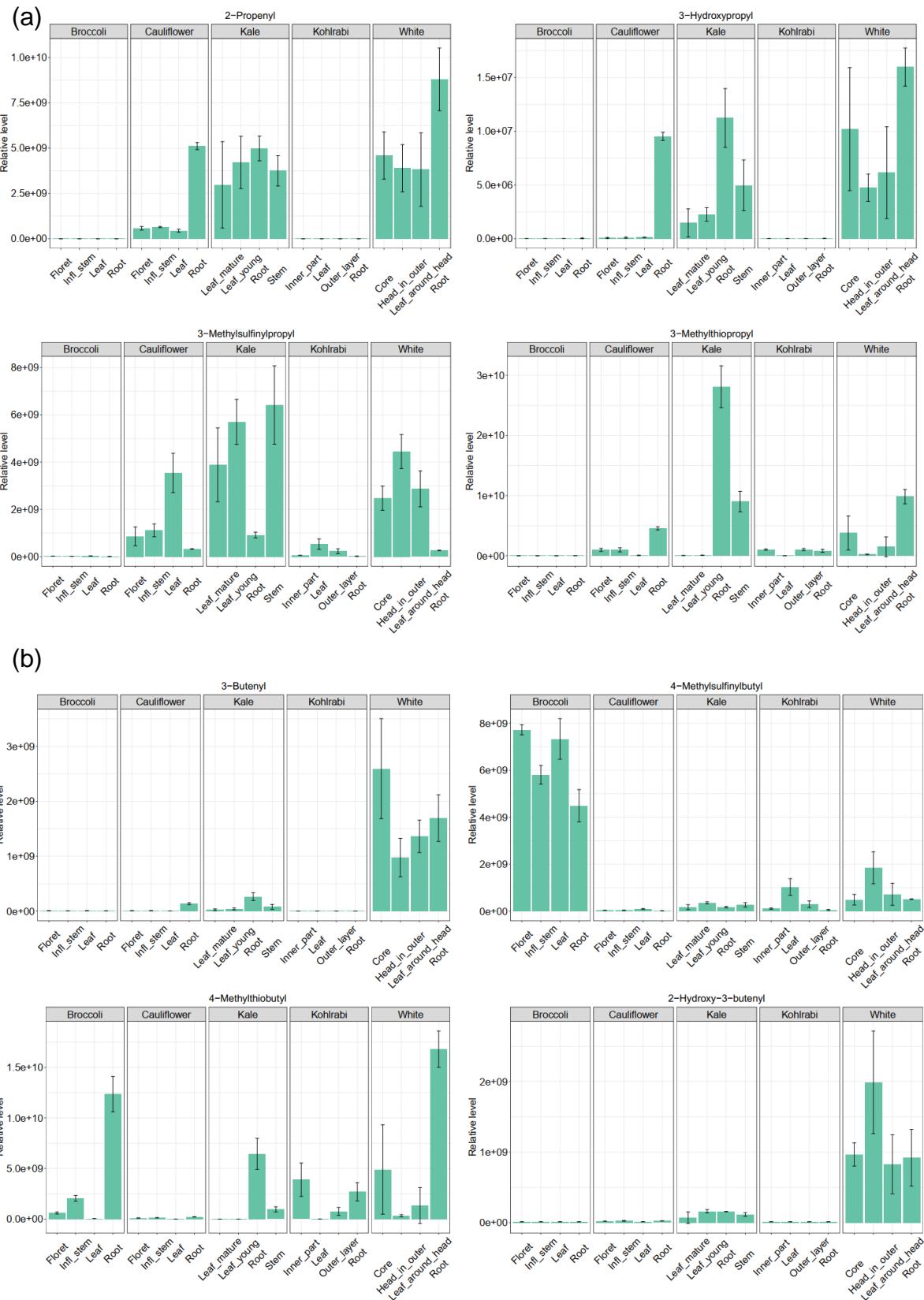
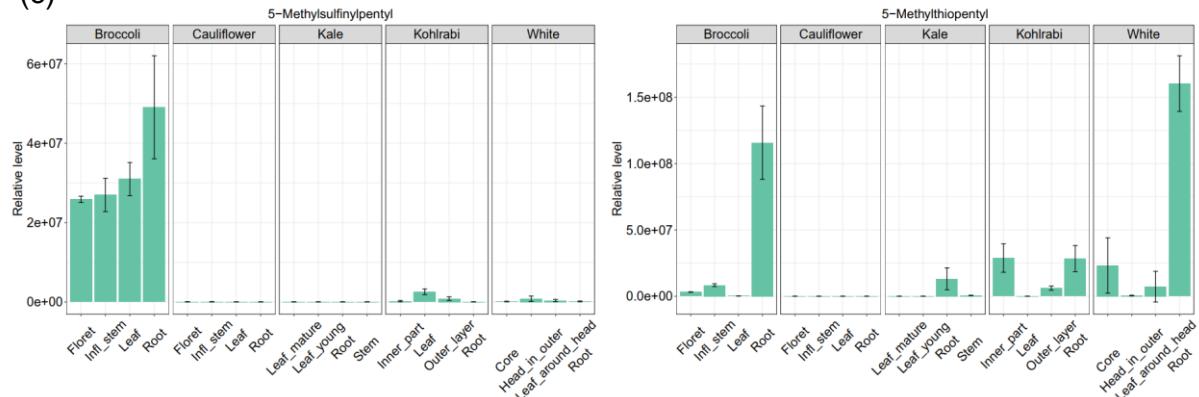


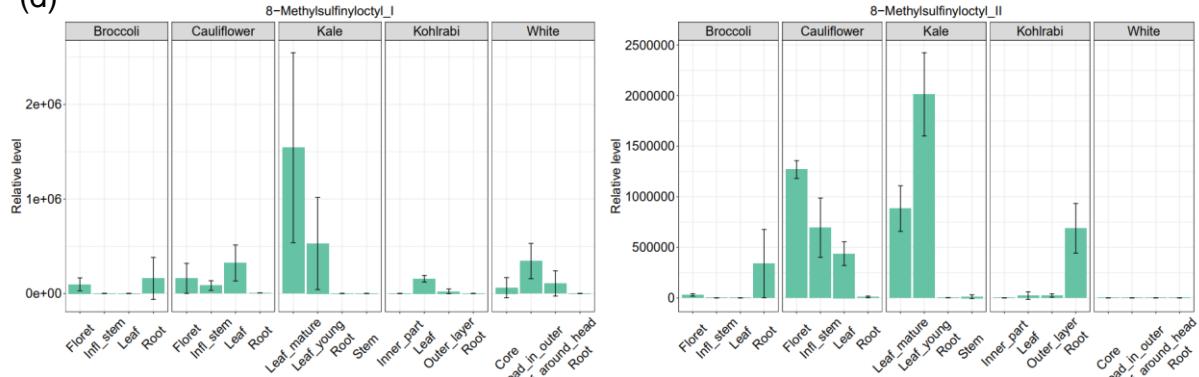
Fig. S2 Principle component analysis (PCA) based on GSL data showing overall variation between the three biological replicates (20 samples \times 3 biological replicates).



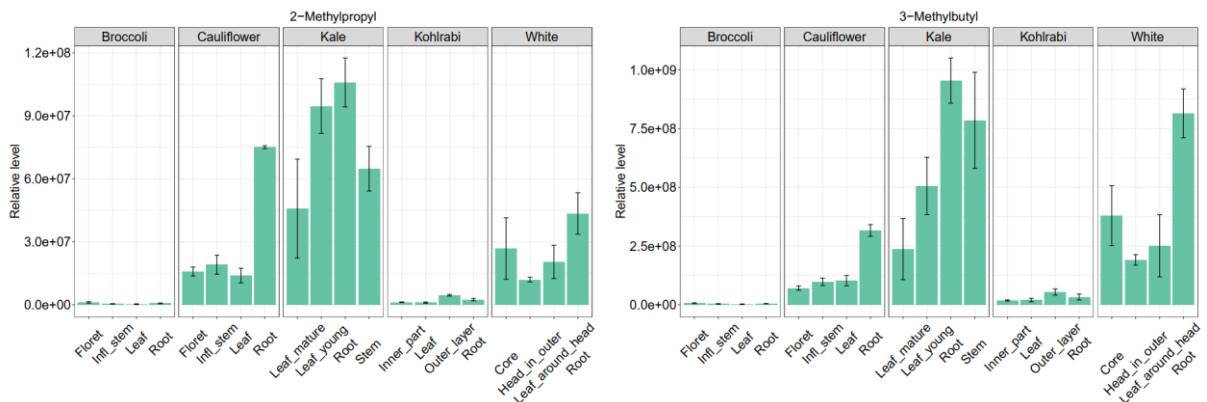
(c)



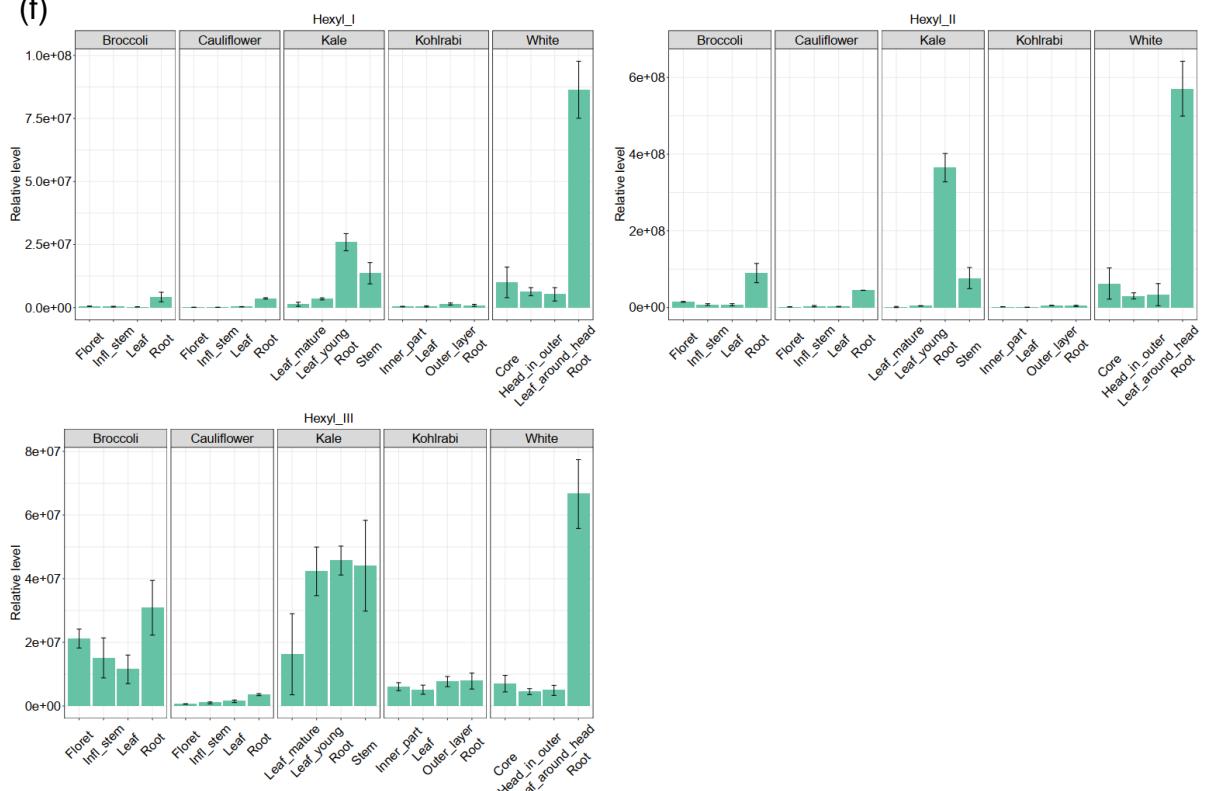
(d)



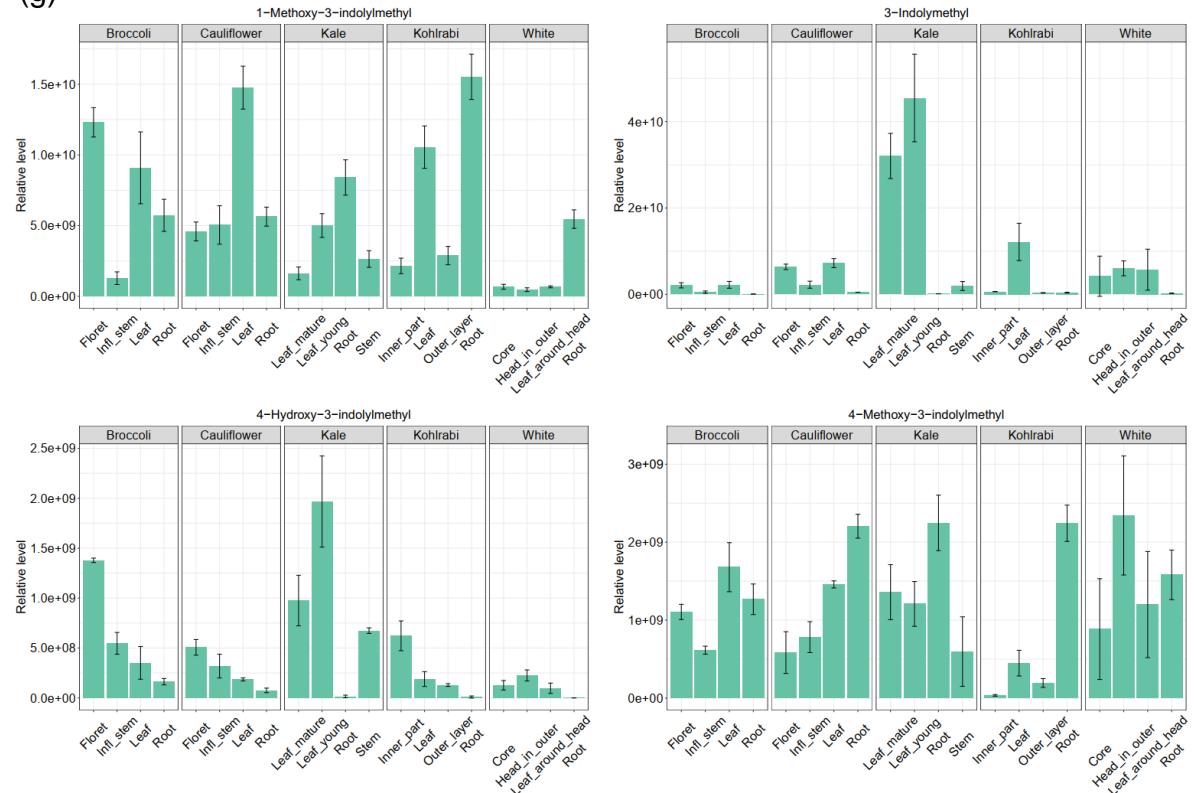
(e)



(f)



(g)



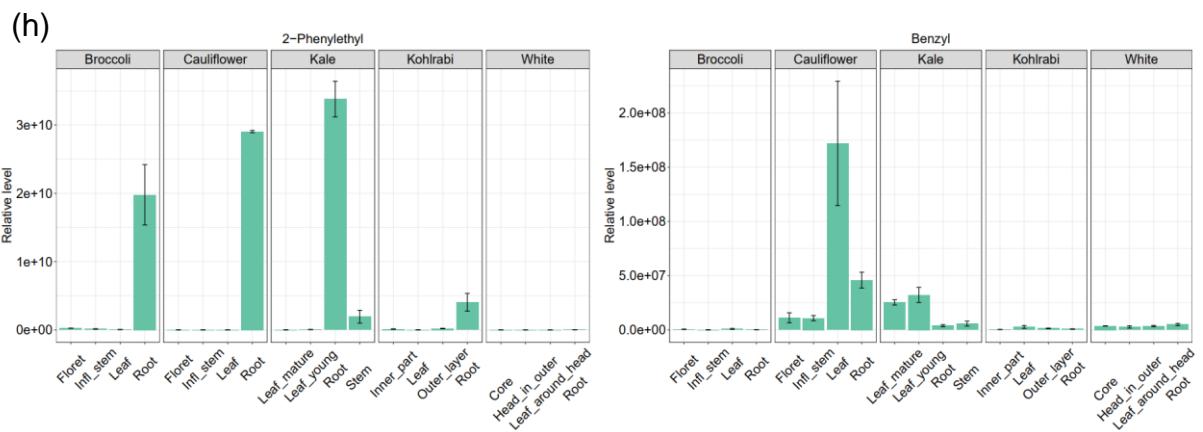
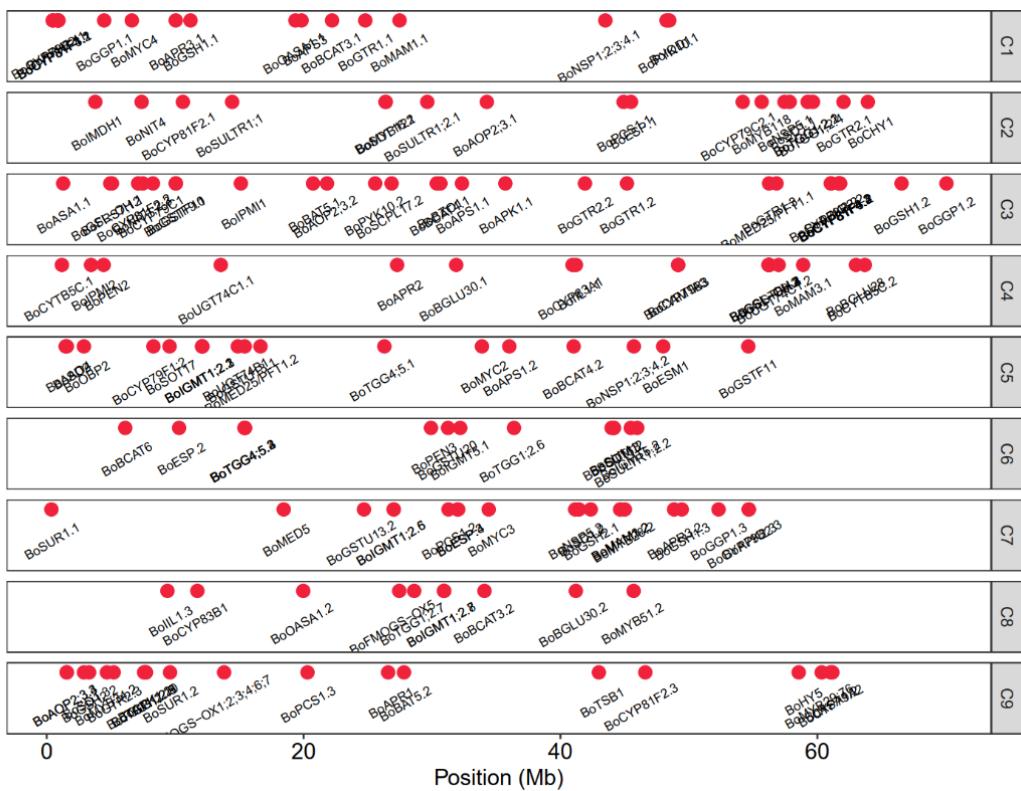
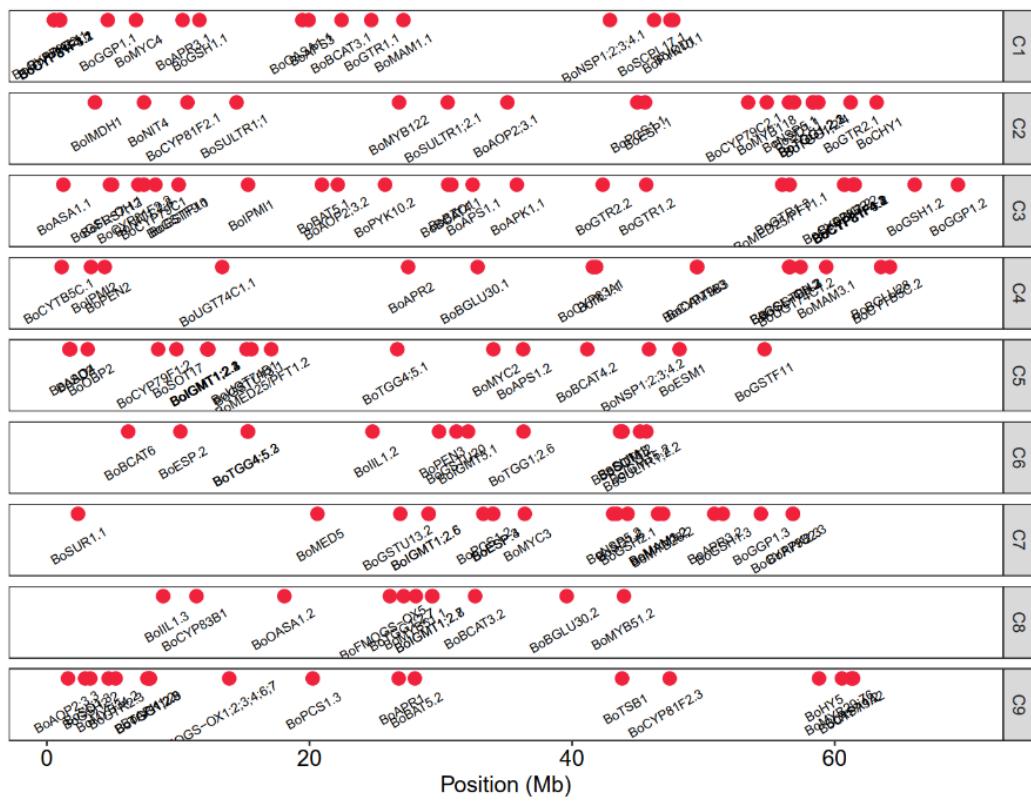


Fig. S3 Relative quantity of individual GSL in four tissues of five *B. oleracea* morphotypes. (a) Methionine derived aliphatic C3 GSLs. (b) Methionine derived aliphatic C4 GSLs. (c) Methionine derived aliphatic C5 GSLs. (d) Methionine derived aliphatic C8 GSLs. (e) Branched-chain amino acid derived aliphatic GSLs. (f) Aliphatic hexyl GSLs. (g) Indolic GSLs. (h) Aromatic GSLs. The Y-axis shows the peak surface area measured in LCMS for the indicated compound. Error bars indicate standard deviation (n = 3).

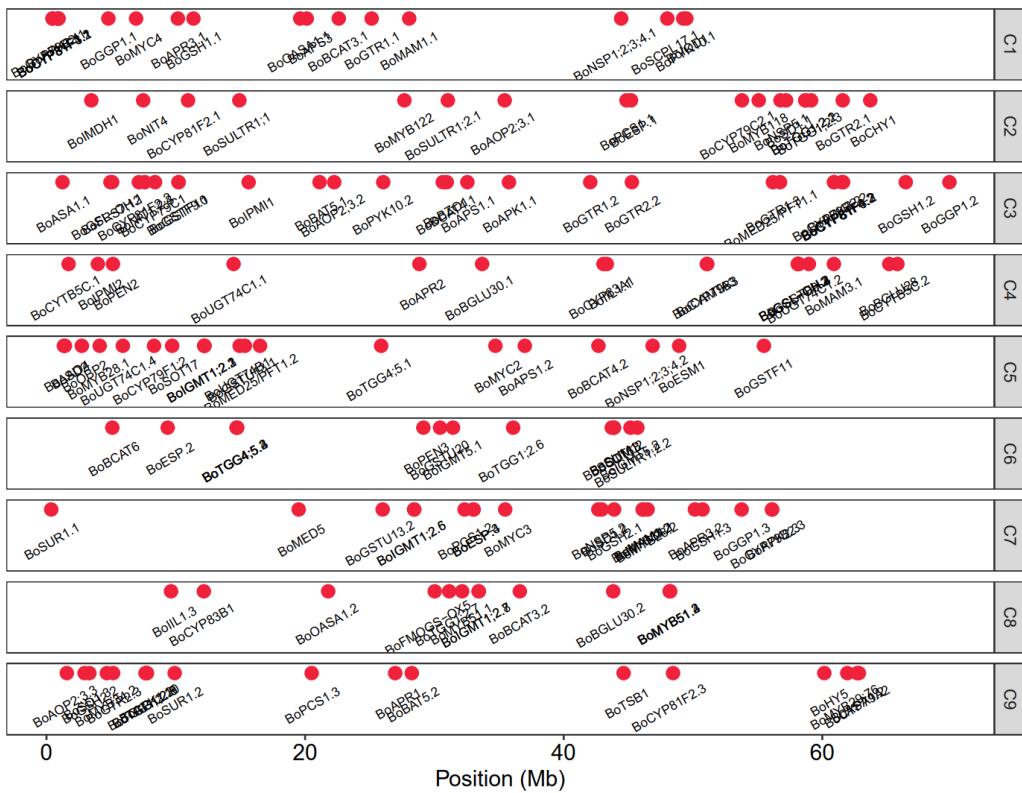
Broccoli



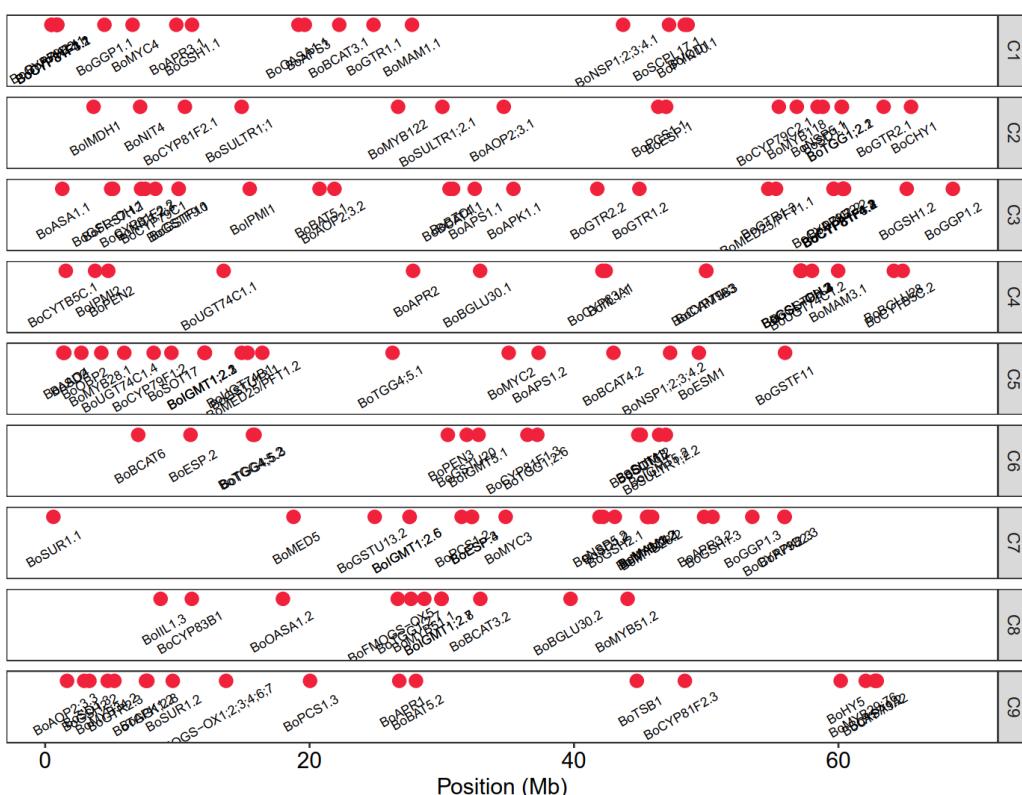
Cauliflower



Kale



Kohlrabi



White Cabbage

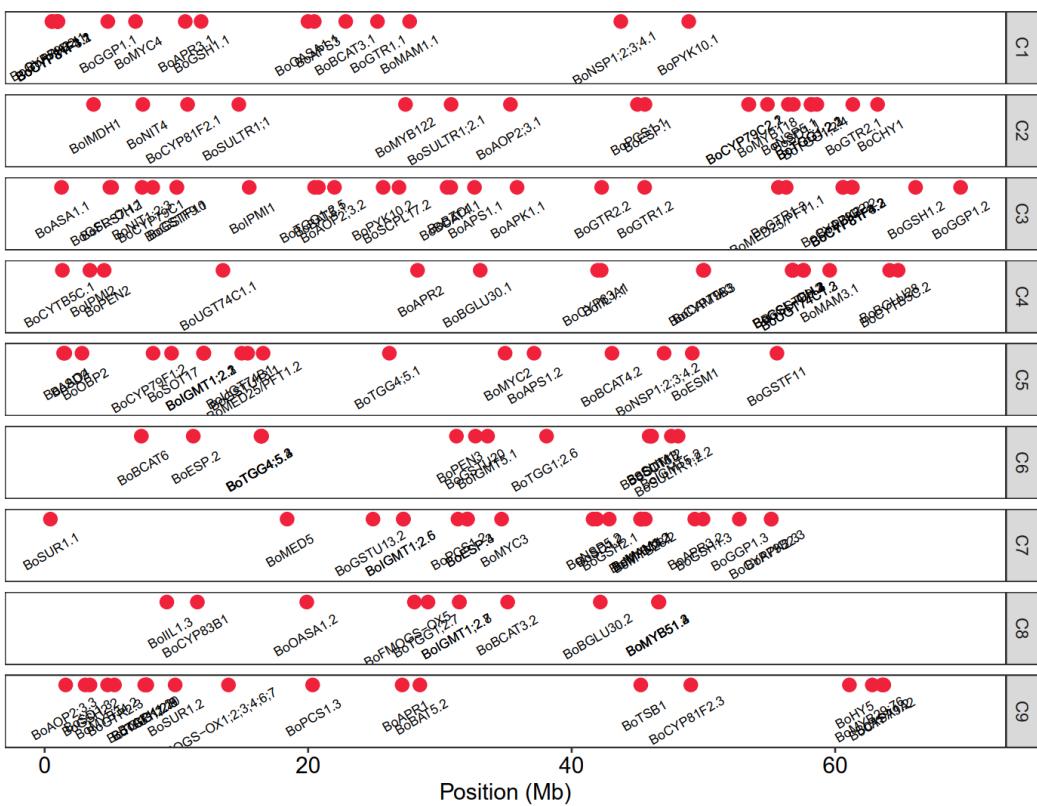
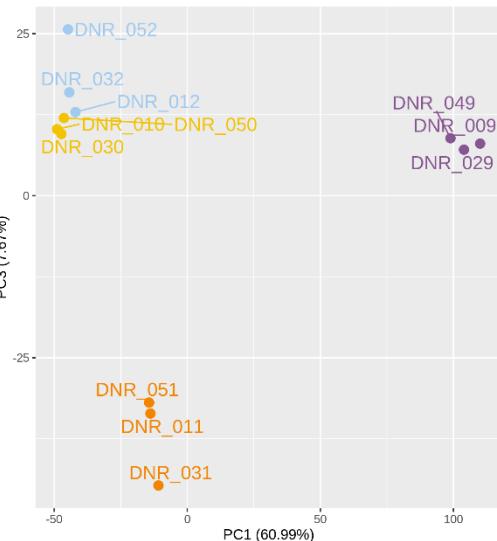
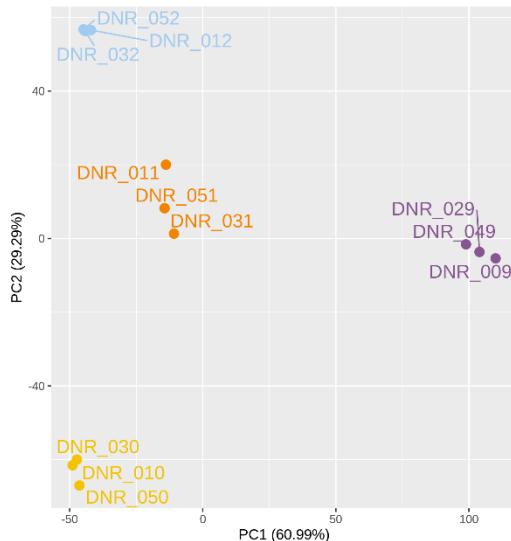
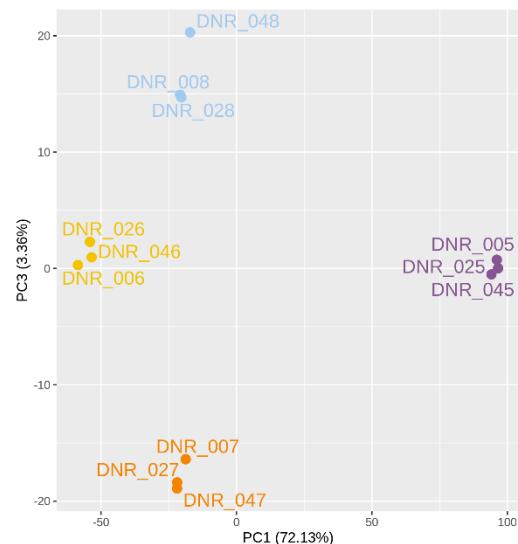
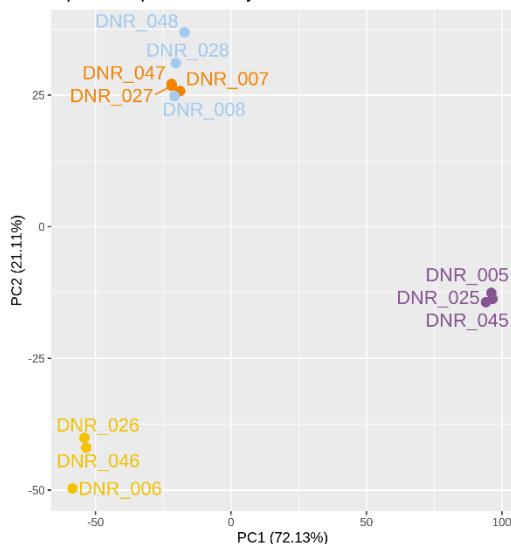


Fig. S4 Distribution of GSL related genes in the five *B. oleracea* genomes. See Table S3 for source data.

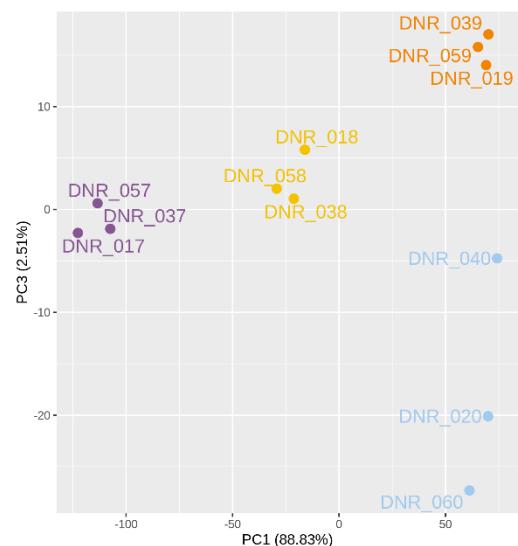
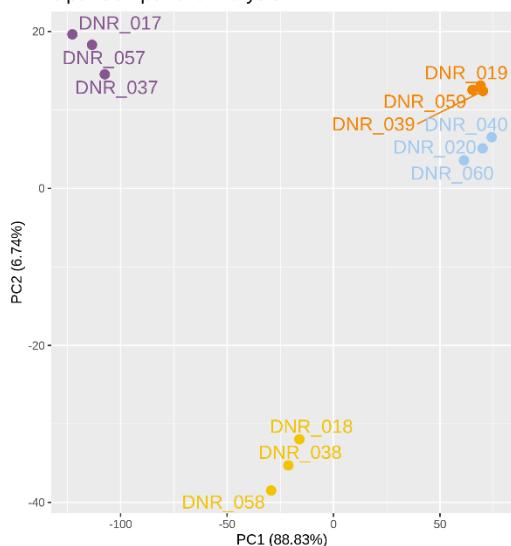
Principal Component Analysis



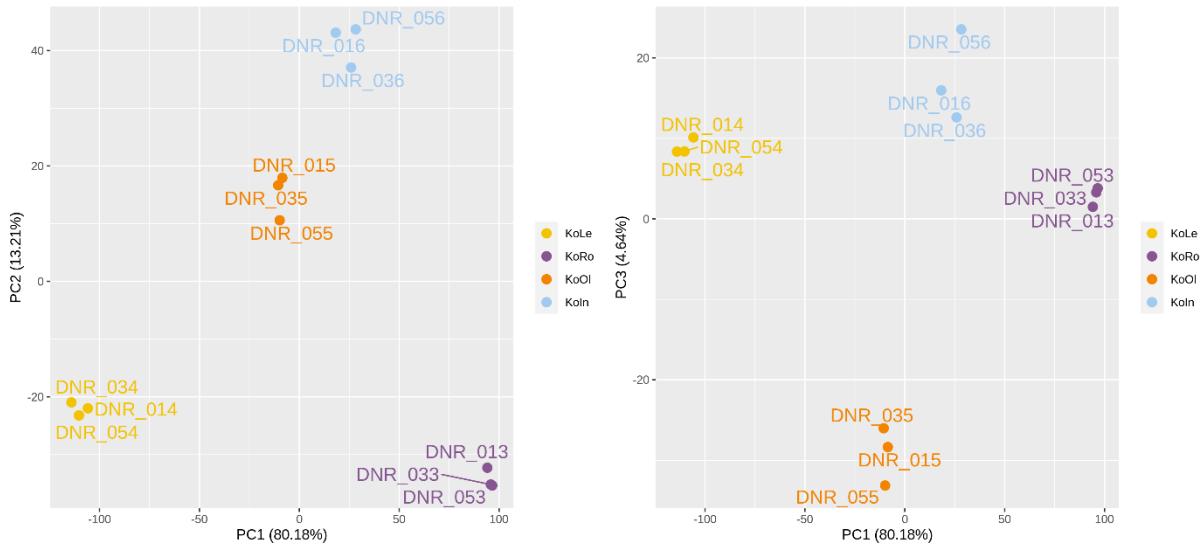
Principal Component Analysis



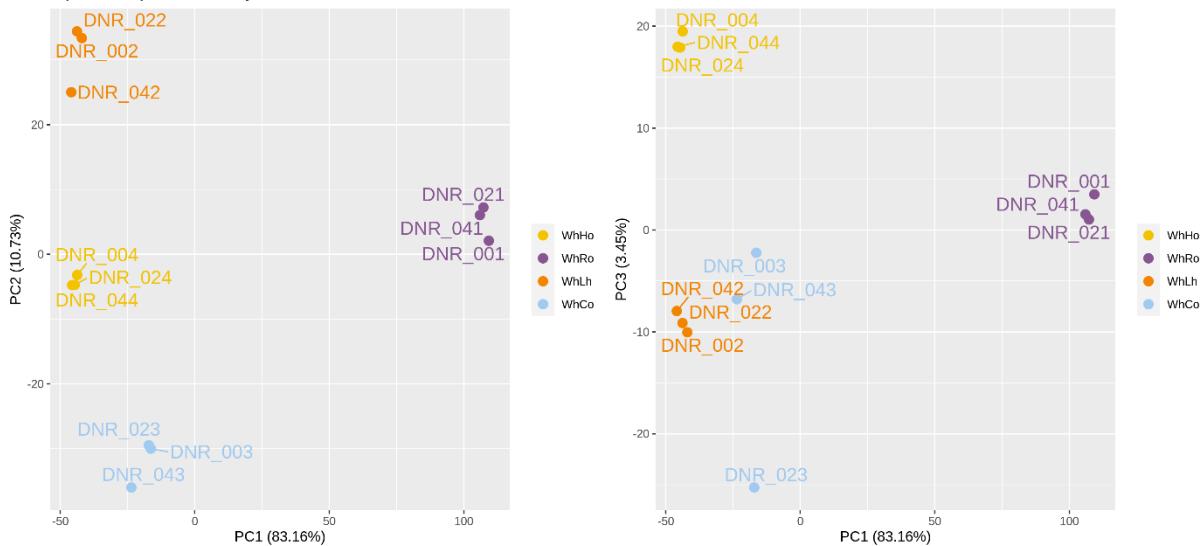
Principal Component Analysis



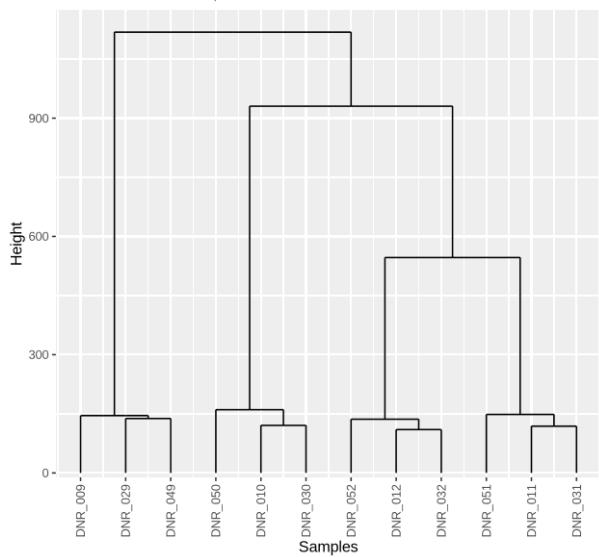
Principal Component Analysis



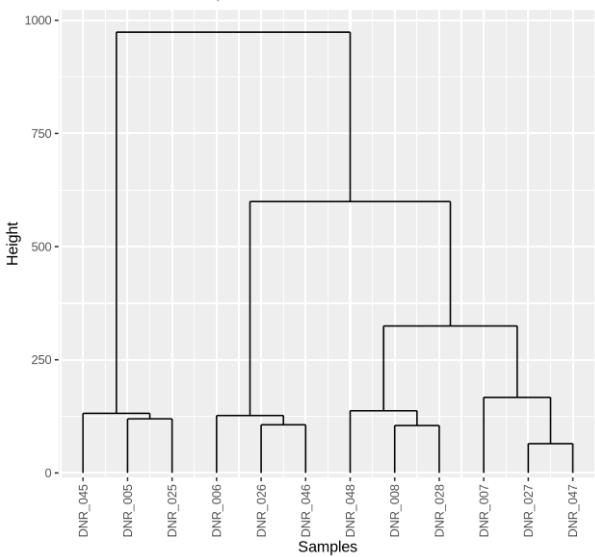
Principal Component Analysis



Cluster dendrogram
Euclidean distance, Ward criterion



Cluster dendrogram
Euclidean distance, Ward criterion



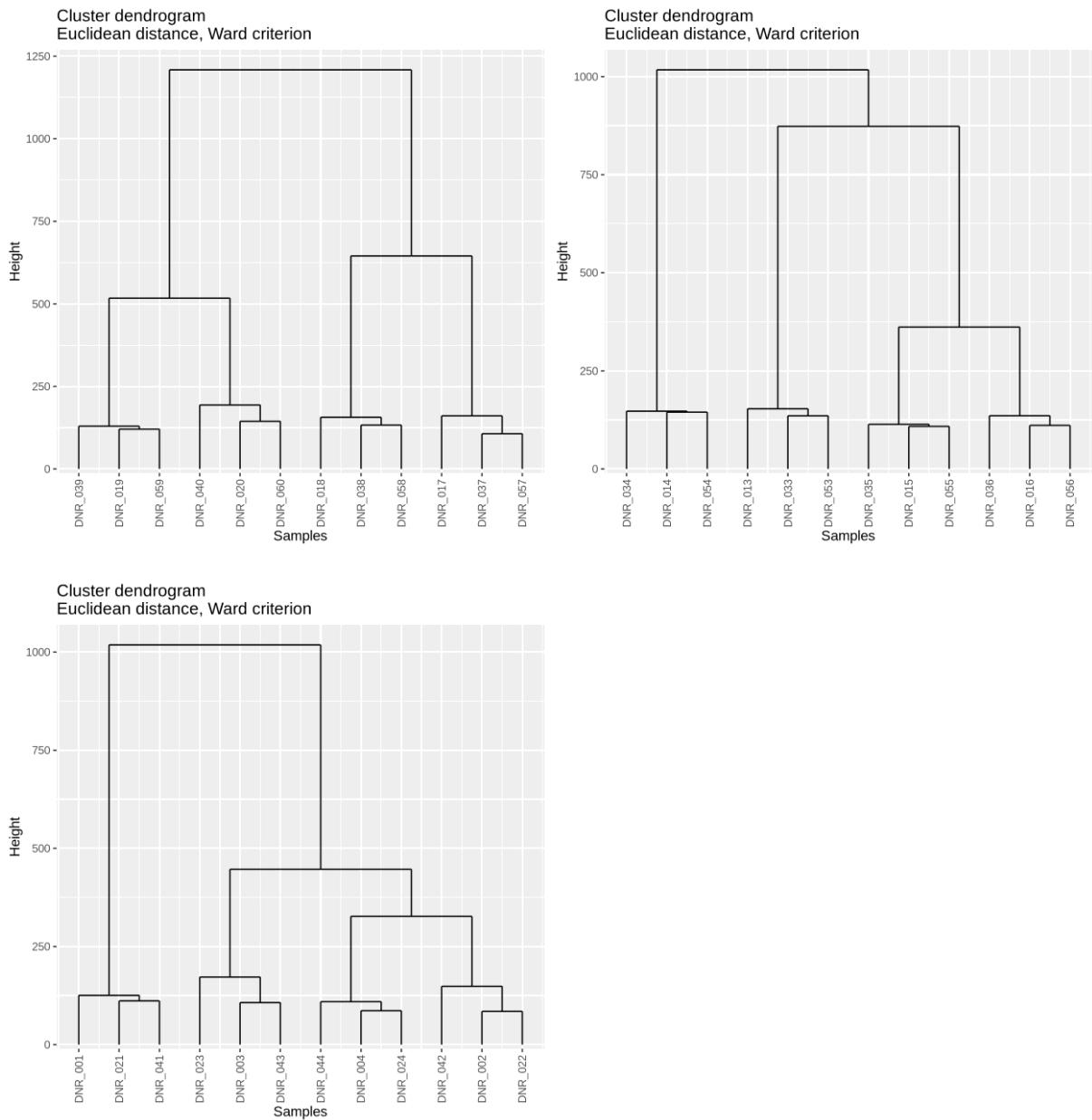


Fig. S5 PCA plots and cluster dendrogram based on mRNA-Seq data. In PCA plots, morphotypes from the top to bottom are: broccoli, cauliflower, kale, kohlrabi and white cabbage. Sample IDs in cluster dendrogram correspond to those used in PCA plots. Abbreviations in the PCA legends: BrLe: Broccoli Leaf, BrRo: Broccoli Root, BrIs: Broccoli infl_stem, BrFl: Broccoli Floret, CaLe: Cauliflower Leaf, CaRo: Cauliflower Root, CaIs: Cauliflower Infl_stem, CaFl: Cauliflower Floret, KaLm: Kale Leaf_mature, KaLy: Kale Leaf_young, KaSt: Kale Stem, KaRo: sssKale Root, KoLe: Kohlrabi Leaf, KoOl: Kohlrabi Outer_layer, KoIn: Kohlrabi Inner_part, KoRo: Kohlrabi Root, WhCo: White Core, WhLh: White Leaf_around_head, WhHo: White Head_in_outer, WhRo: White Root.

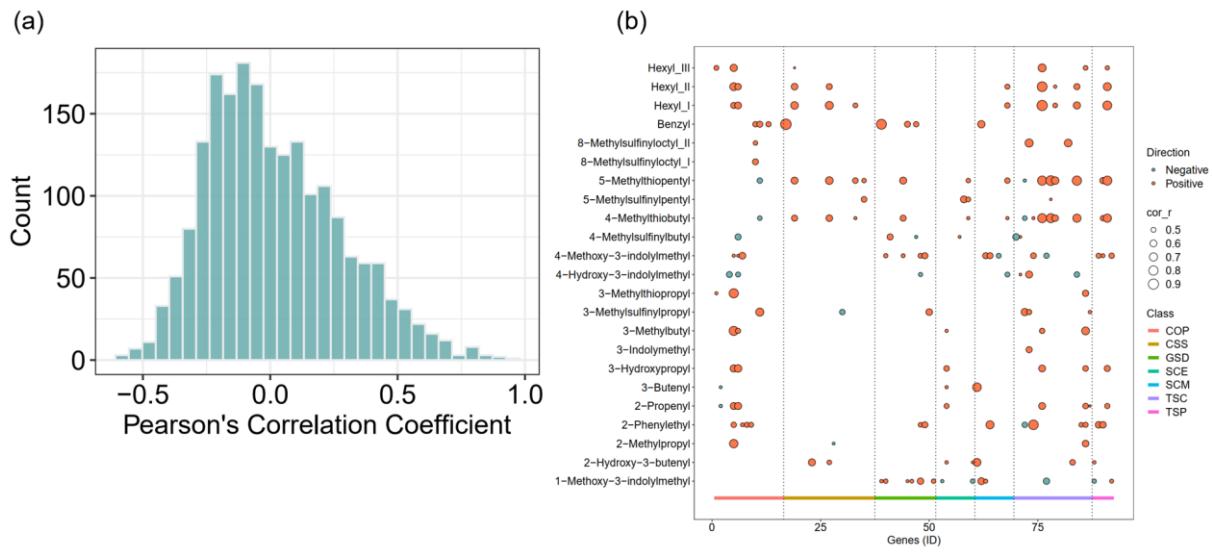


Fig. S6 Pearson's correlation analysis between GSLs and their related genes. Gene expression profiles (TPM values) are pooled for different copies of paralogous genes in *B. oleracea*. (a) Distribution of Pearson's correlation coefficient. (b) Significantly ($P < 0.05$) correlated GSLs and genes. Genes are classified based on their involvement in different processes/phases as shown in Fig. 2. (The abbreviations: COP: Cosubstrate Pathways, CSS: Core Structure Synthesis, GSD: GSL Degradation, SCE: Side-Chain Elongation, SCM: Side-Chain Modification, TSC: Transcriptional Components, TSP: Transporters).

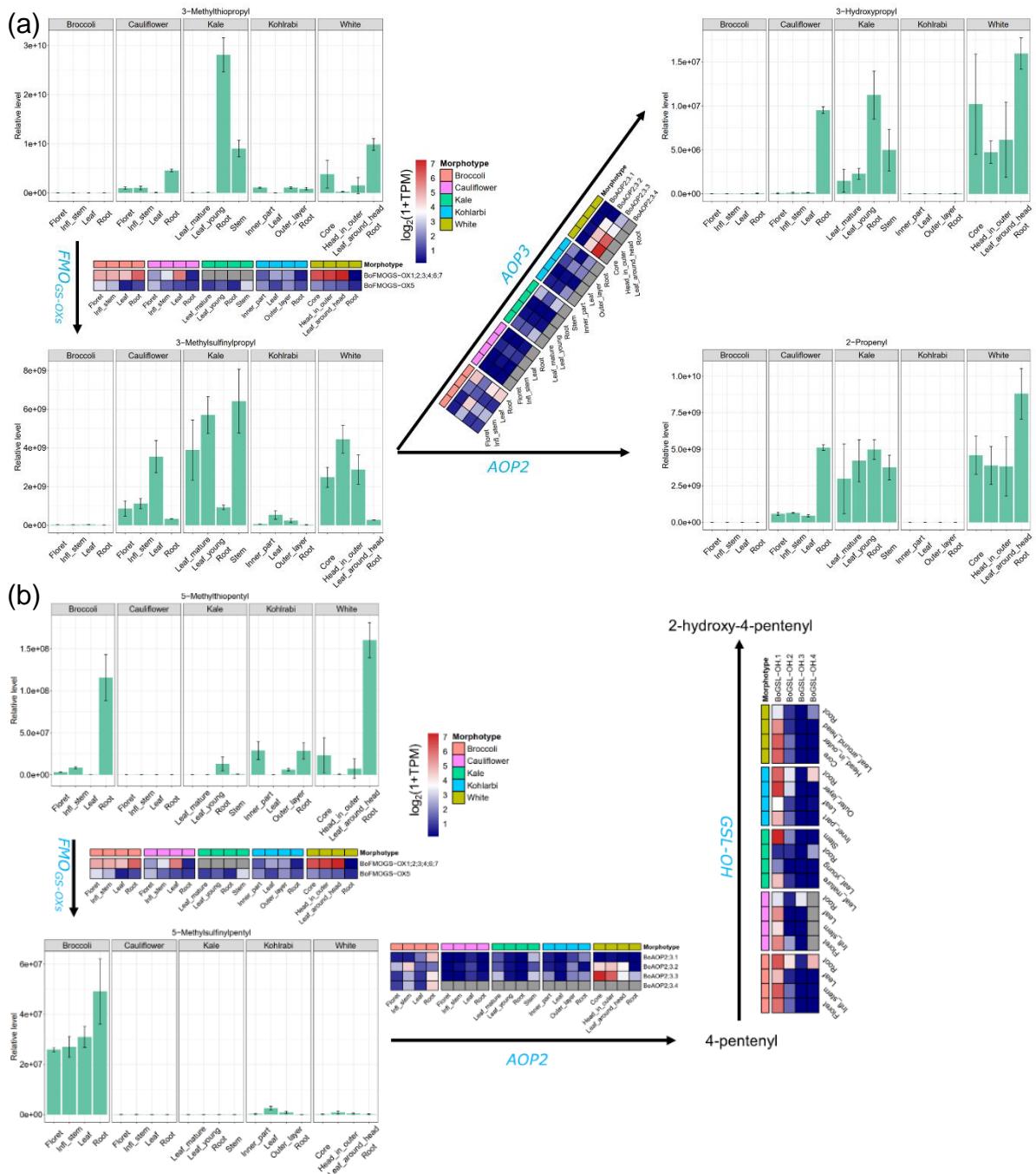
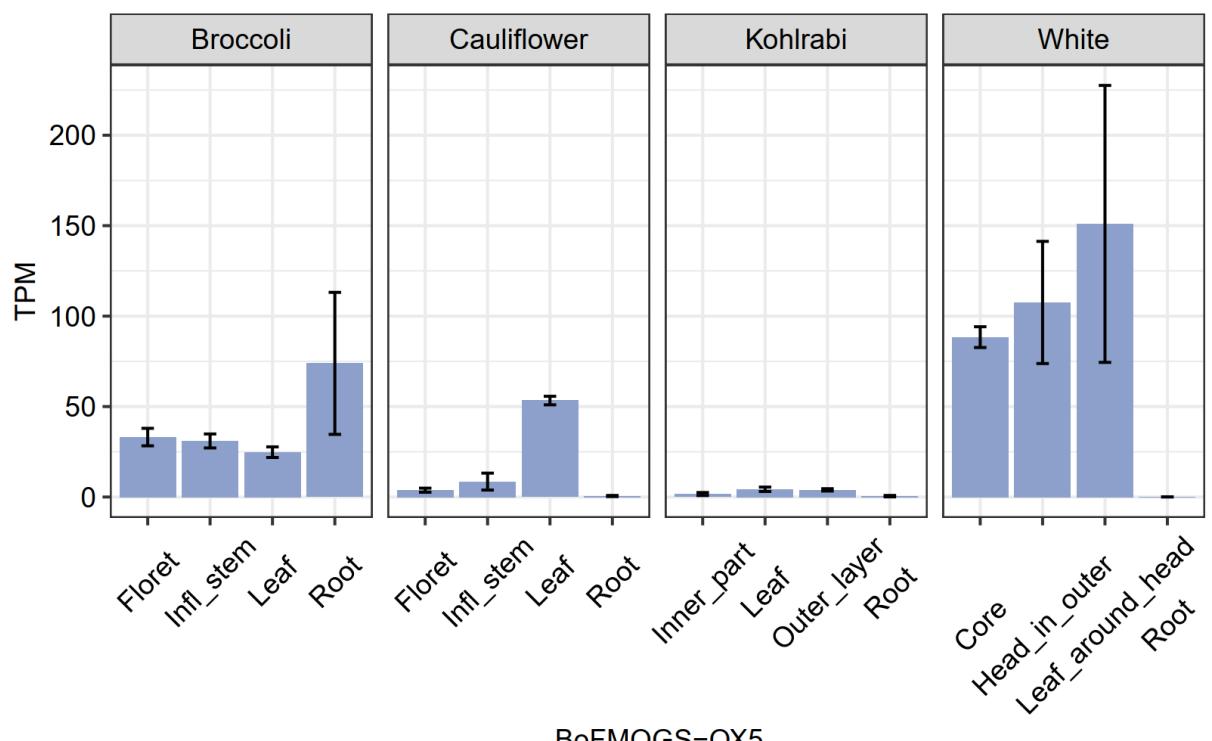


Fig. S7 C3 (a) and C5 (b) aliphatic GSL profiles and expression levels of related genes in different tissues and morphotypes. The bar charts show relative quantity of individual GSLs in respective tissues and morphotypes. Error bars indicate standard deviation ($n = 3$). Heatmaps show gene expression levels. Blue and red colors are used to represent low to high expression levels. Gray color denotes that the gene is not identified in the corresponding morphotype. Note: *BoAOP2;3.3* is *BoAOP2*.

BoFMOGS-OX1;2;3;4;6;7



BoFMOGS-OX5

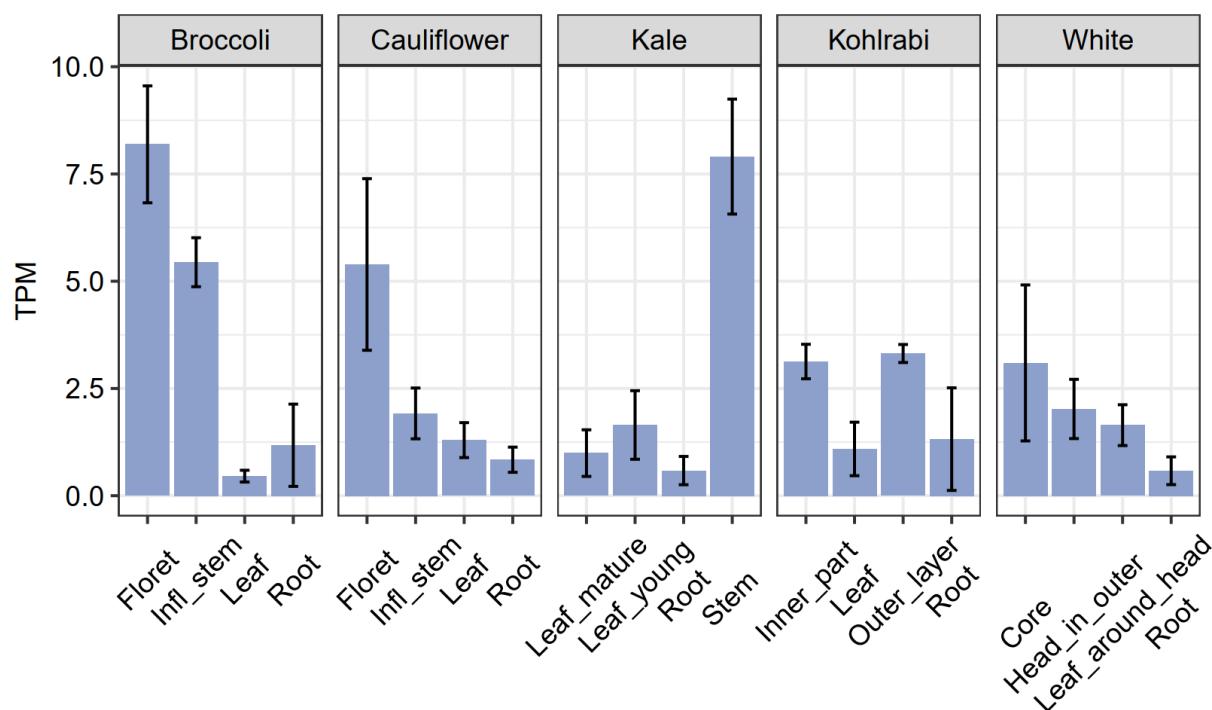


Fig. S8 Gene expression analysis of *FMO_{GS-OX}* paralogues in four tissues in five *B. oleracea* morphotypes. The expression level was estimated using TPM values based on mRNA-Seq data. Error bars indicate standard deviation (n = 3).

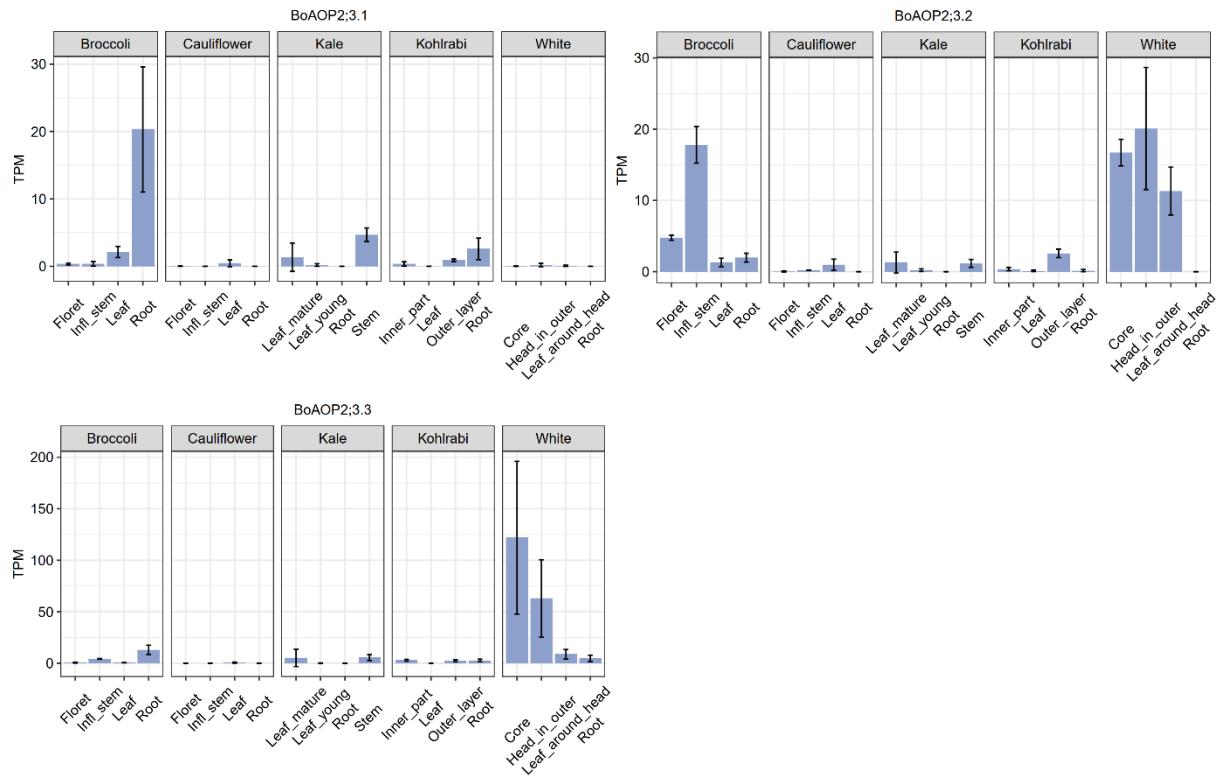


Fig. S9 Gene expression analysis of *AOP* paralogues in four tissues in five *B. oleracea* morphotypes. The expression level was estimated using TPM values based on mRNA-Seq data. Error bars indicate standard deviation (n = 3). Note: *BoAOP2;3.3* is *BoAOP2*.

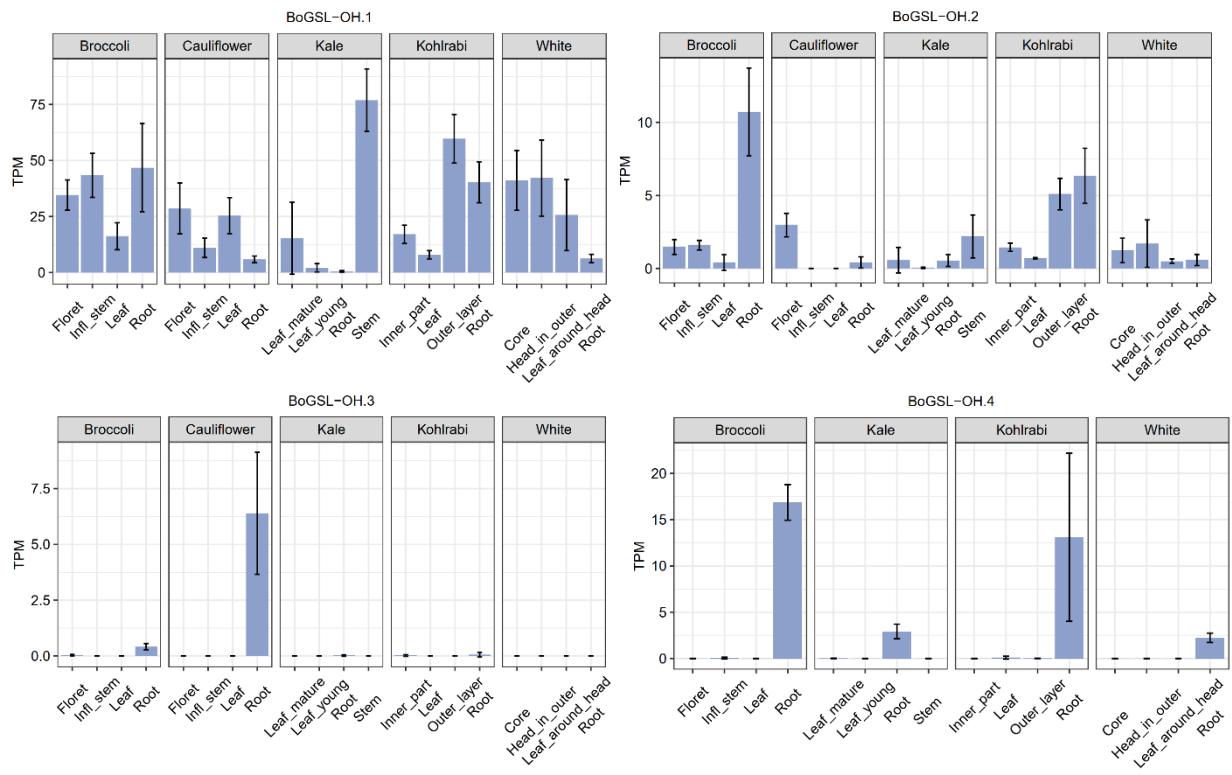


Fig. S10 Gene expression analysis of *GSL-OH* paralogues in four tissues in five *B. oleracea* morphotypes. The expression level was estimated using TPM values based on mRNA-Seq data. Error bars indicate standard deviation (n = 3).

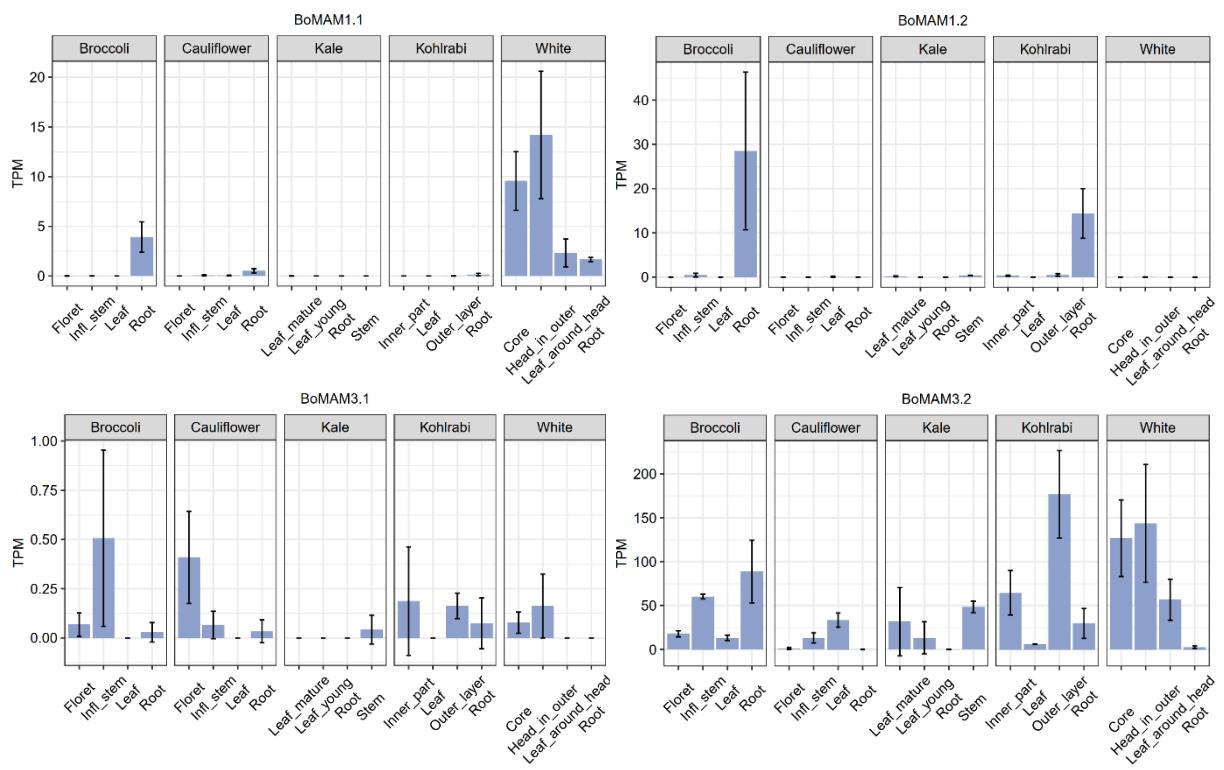
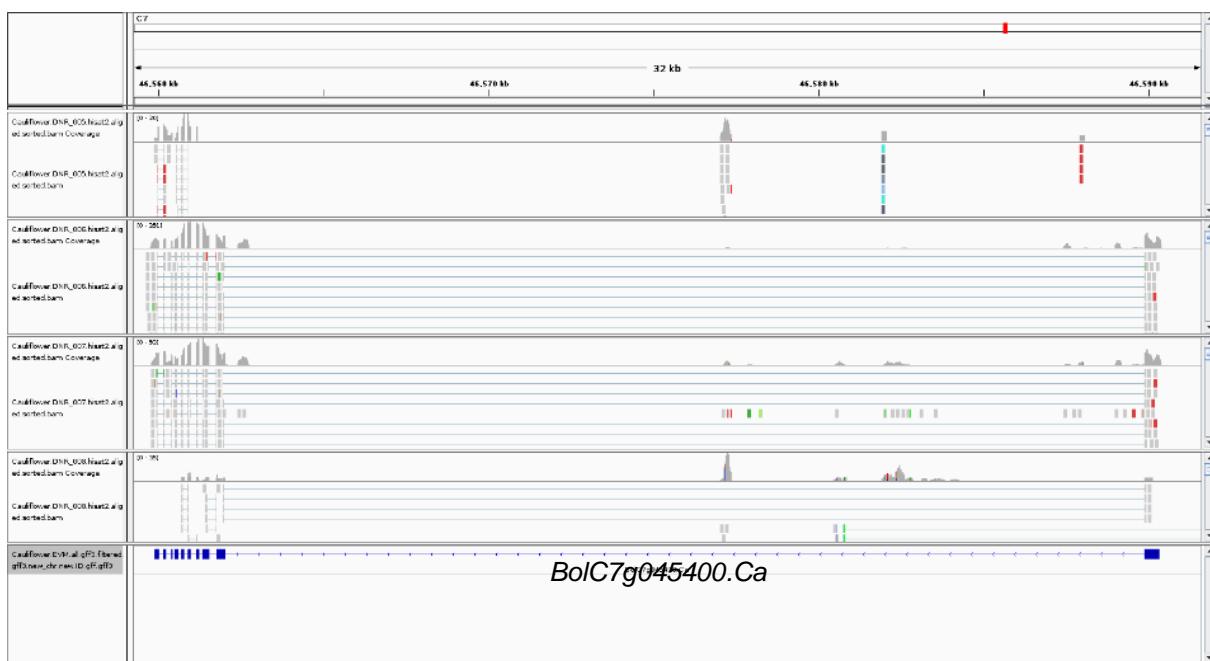


Fig. S11 Gene expression analysis of *MAM* paralogues in four tissues in five *B. oleracea* morphotypes. The expression level was estimated using TPM values based on mRNA-Seq data. Error bars indicate standard deviation ($n = 3$).



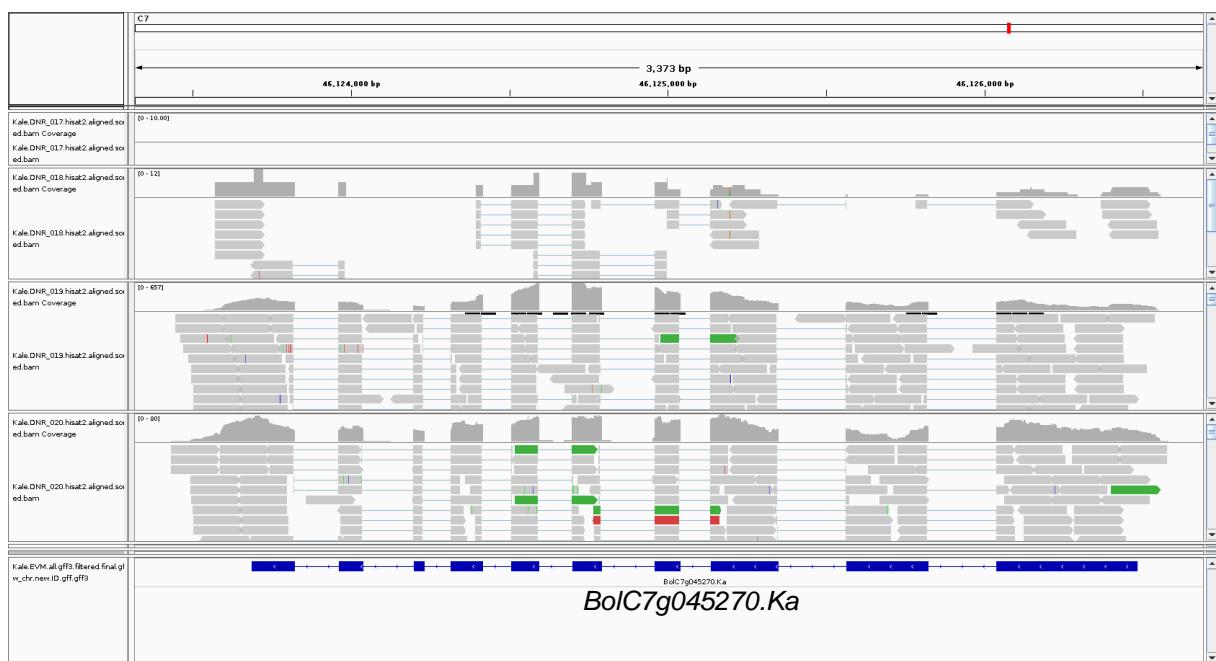
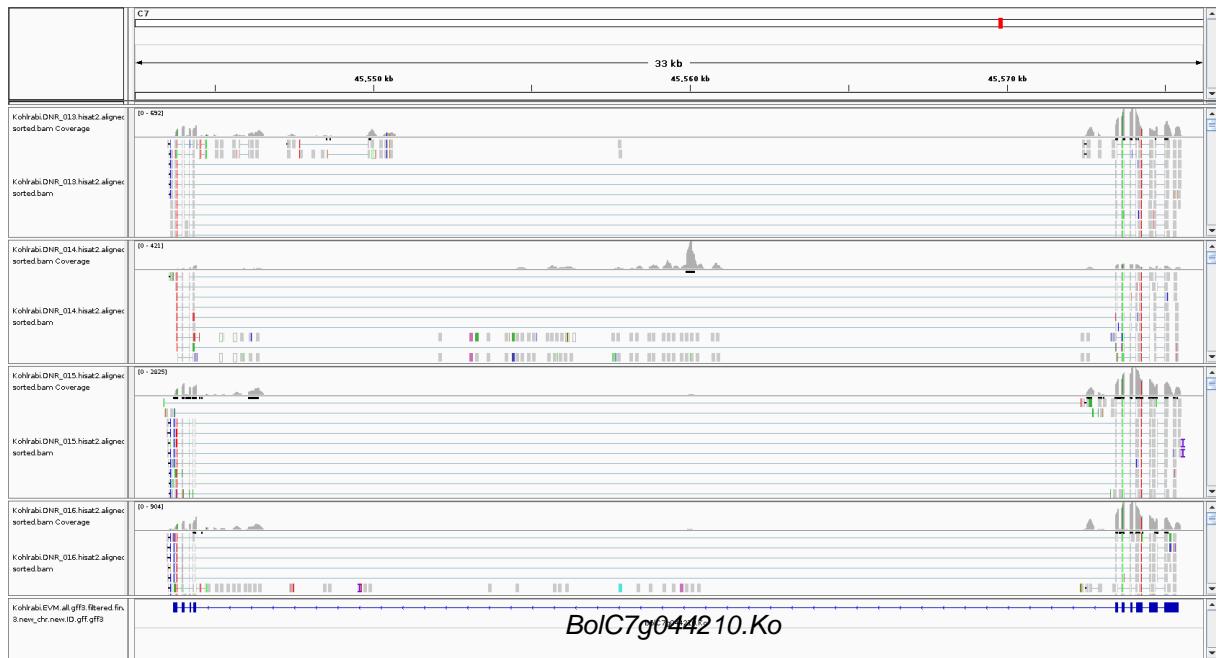




Fig. S12 IGV snapshots showing mRNA-Seq alignments in *BoMAM3.2* genes from five *B. oleracea* morphotypes (Top to bottom: broccoli, cauliflower, kohlrabi, kale and white cabbage). In each snapshot, four tracks from top to bottom represent alignments in four different tissues.