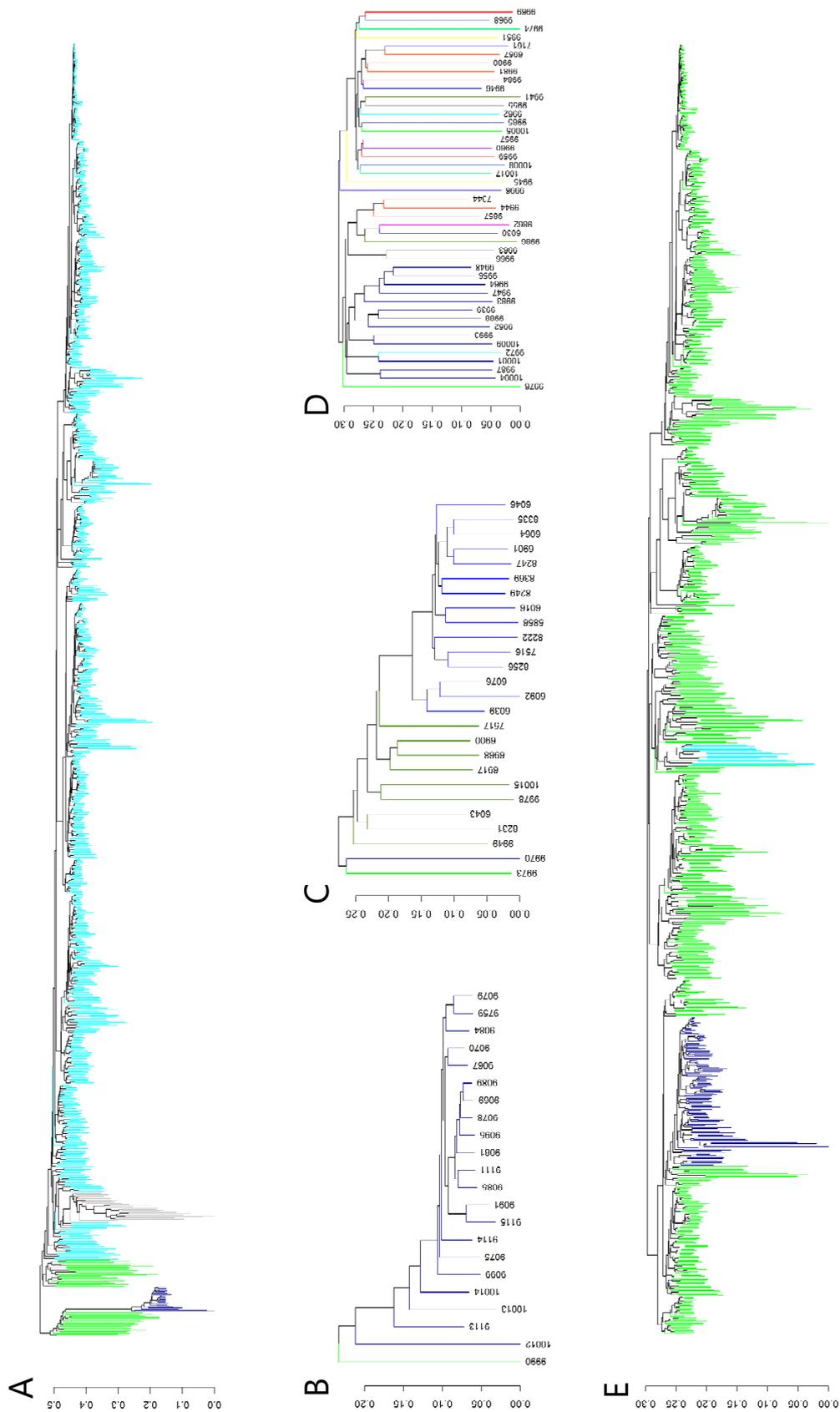


Extended Data Fig. 1

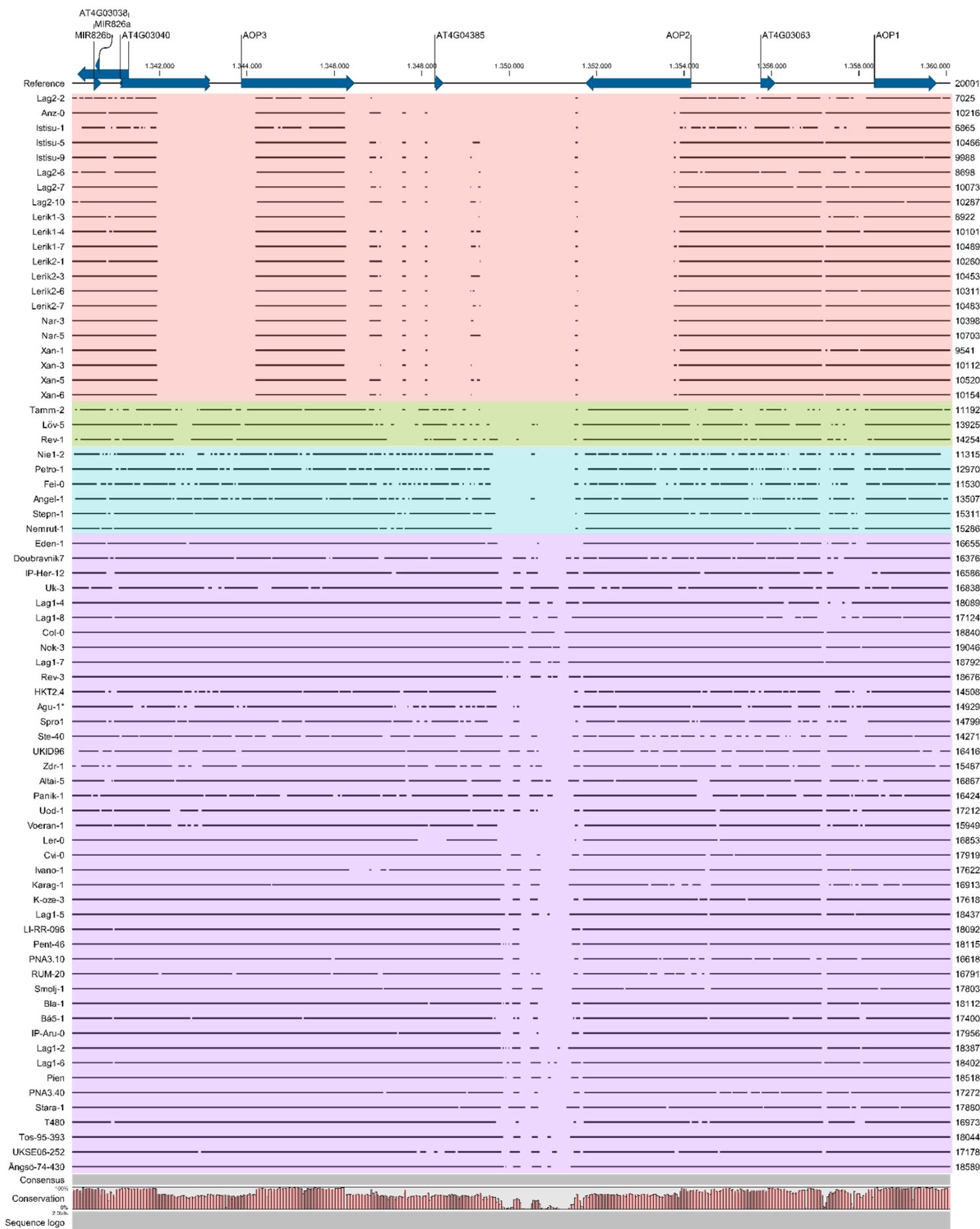
Extended Data Fig. 1

The glucosinolate biosynthesis pathways in *A. thaliana*. GSLs can be grouped into indole GSLs (blue track) and aliphatic GSLs (green track). The three amino acids tryptophan, methionine and phenylalanine are the precursors for indole, aliphatic and benzenic GSL, respectively. Carbon side chain elongation of methionine for aliphatic GSL takes place in the chloroplast and is determined by the *GSL-ELONG* locus genes *MAM1-3*. The subsequent core structure biosynthesis is not shown. Secondary modifications: The indole GSL indolyl-3-methyl GSL (I3M) can be further modified by cytochrome P450 enzymes of the CYP81F family and indole GSL methyltransferases (IGM). Aliphatic methylthioalkyl GSL (C3-C8) can be S-oxygenated to methylsulfinylalkyl GSL by flavin-containing monooxygenases encoded in the *GSL-OX* loci. AOP enzymes build a branch point and catalyze the conversion of C3-C5 methylsulfinylalkyl to alkenyl GSLs (AOP2) or of C3-C4 methylsulfinylalkyl hydroxyalkyl GSLs (AOP3). Hydroxyalkyl GSLs can further be benzyolated by an unknown enzyme (?) using benzoyl-CoA provided by BZO1 (embryo specific). Alkenyl GSL can be hydroxylated by *GSL-OH* to 2-(S/R)-hydroxybut-3-enyl GSL. Turnover of GSL in intact tissue and GSL activation (upon attack) are initiated by hydrolysis by β -glucosidases (classical myrosinases, other BGLUs). Dependent on the physiological condition, the resulting aglucone either spontaneously rearranges to an isothiocyanate (ITC) or the formation of nitriles is promoted by epithiospecifier protein (ESP) or nitrile-specifier proteins (NSPs). Enzymes are shown in rectangular boxes, rounded boxes indicate metabolites. Dashed box margins indicate metabolites not analyzed in the targeted approach used in this study. Dashed arrow lines indicate intermediate steps not shown. Question mark (?) indicates unknown genes. 1OH-I3M, 1-hydroxyindol-3-ylmethyl; 4MOI3M, 4-hydroxyindol-3-ylmethyl; I3M, indol-3-ylmethyl; NMOI3M, N-methoxyindol-3-ylmethyl; 4MOI3M, 4-methoxyindol-3-ylmethyl; 3mtp, 3-methylthiopropyl; 3ohp, 3-hydroxypropyl; 3bzo, 3-benzoyloxypropyl; 4mtb, 4-methylthiobutyl; 4msb, 4-methylsulfinylbutyl; 4ohb, 4-hydroxybutyl; 4bzo, 4-benzoyloxybutyl; 5mtp, 5-methylthiopentyl; 5msp, 5-methylsulfinylpentyl; 6mth, 6-methylthiohexyl; 6msh, 6-methylsulfinylhexyl; 7mth, 7-methylthioheptyl; 7msh, 7-methylsulfinylheptyl; 8mto, 8-methylthiooctyl; 8mso, 8-methylsulfinyloctyl; 2PE, 2-phenylethyl.



Extended Data Fig. 2

GSL-AOP locus dendrogram and main haplotype groups. Dissimilarity based Neighbor-Joining calculated dendrograms of the *GSL-AOP* locus (20kb, Chr4. 1.340.000-1.360.000). **A**: 1135 accessions, split in four main groups. **B-E**: Realignment and sub clustering of the four main groups H1 (B), H2 (C), H3 (D) and H4 (E), colors indicate haplotype sub clusters.

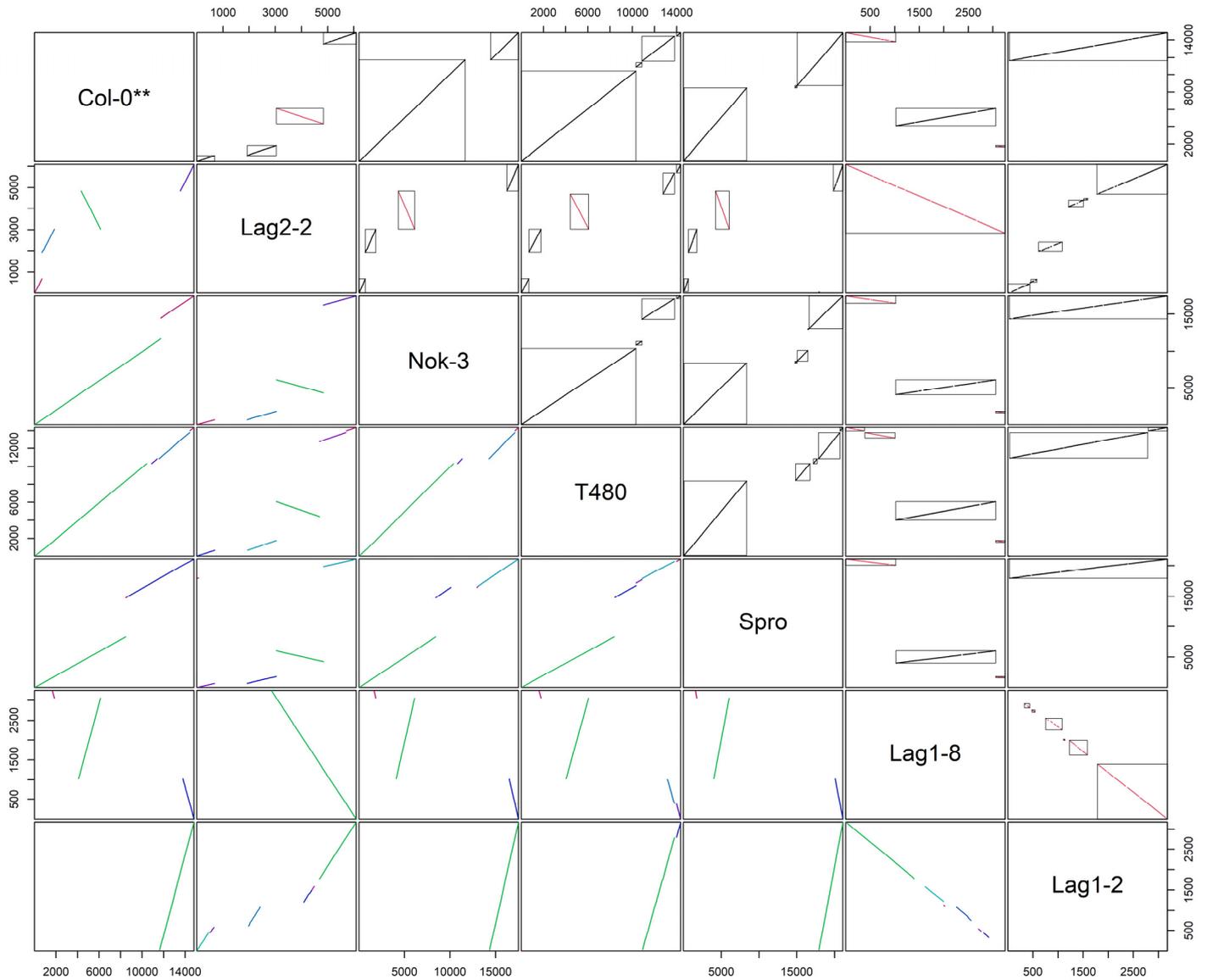


Extended Data Fig. 3

GSL-AOP locus multiple sequences alignment. Arrows show gene position and directionality's according to Col-0 reference (NC_003075) Chromosome 4 including position. Color shows main haplotype groups. Number of nucleotides given on the right of each sequence.

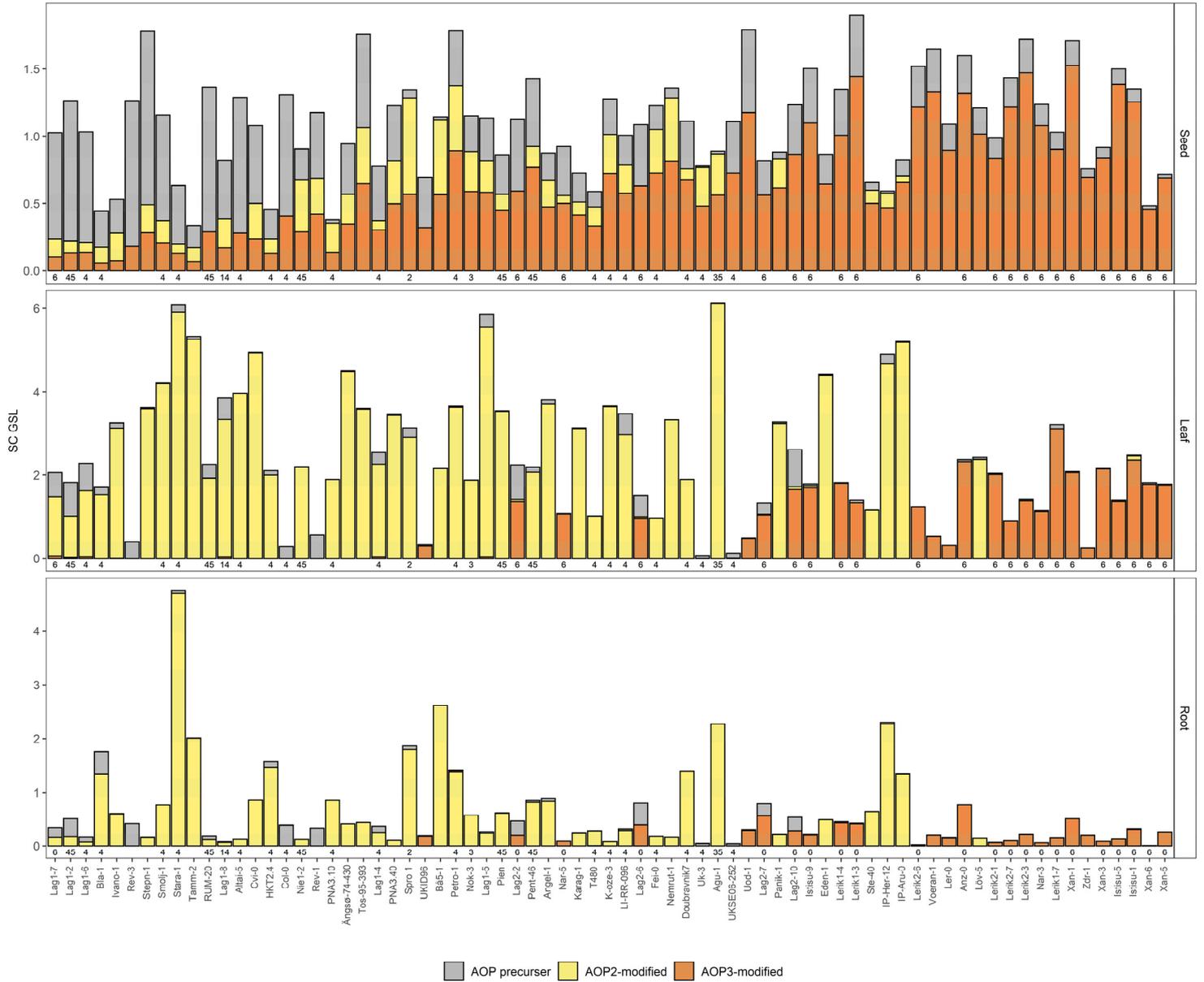
Extended Data Fig. 4

Dendrogram and main haplotype groups for seven ONT sequenced loci (left) and the according Dendrogram for the public available 1001 Data (right). Dissimilarity based Neighbor-Joining calculated dendrograms for *GSL-AOP*: 14.8kb Chr.4 pos. 12340082-1354901, *ESP*: 9.1kb Chr.1 pos. 20169245-20178381, *FLC*: 10.6kb Chr.5 pos. 3171361-3181999, *FRI*: 6.9kb Chr.4 pos. 266932-273864, *FT*: 4.7kb Chr. 1 pos. 24329391-24334102, *GS-OH*: 5.7kb Chr.2 pos. 10828201-10833967, *VRN1*: 7.8kb Chr.3 pos. 6546205-6554036. Col-0** represents the Reference sequence Chr.4 pos. 12340082-1354901 (CP002687), accessions from which additional amplicons were obtained are marked by (*), both amplicons are shown in the tree.



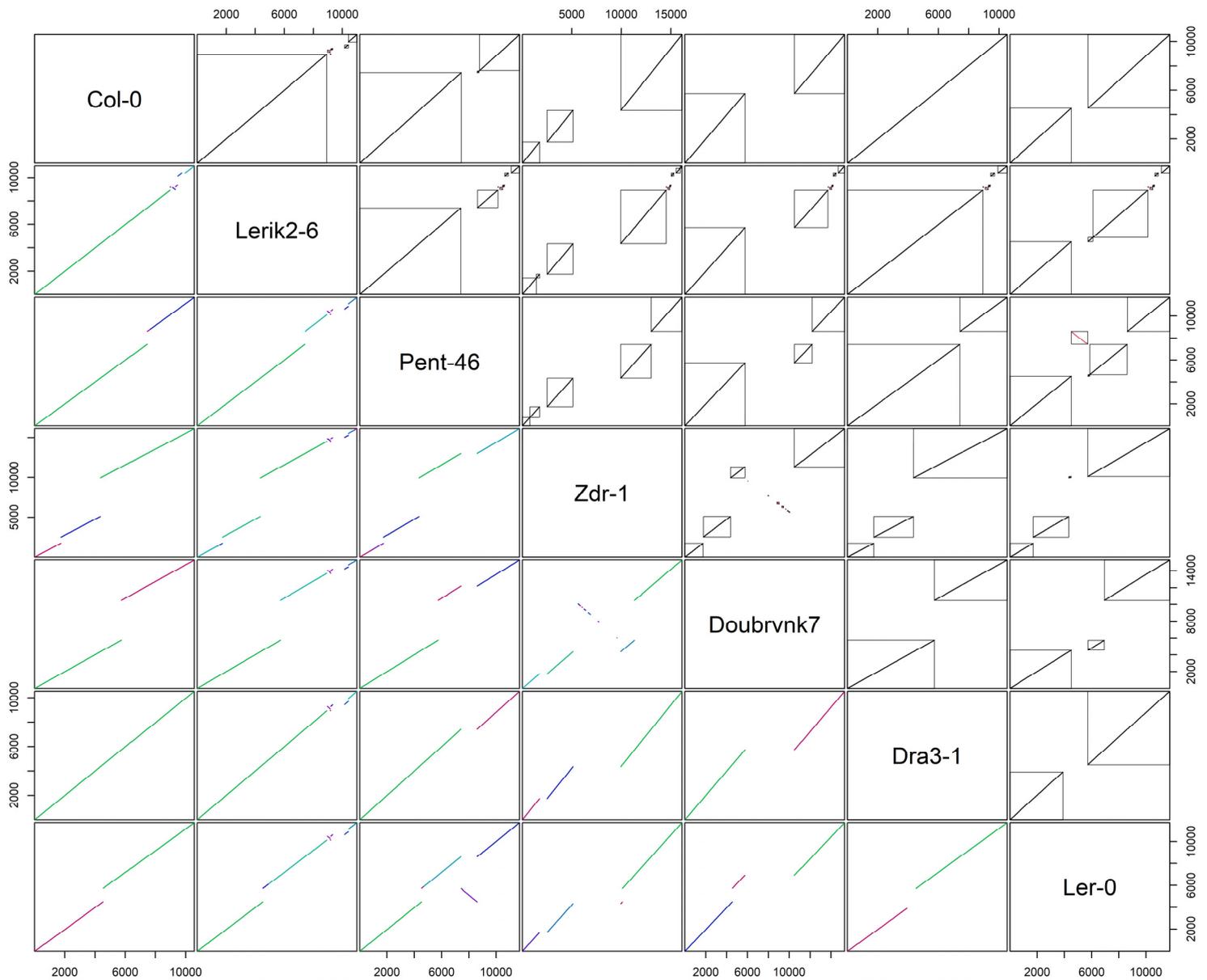
Extended Data Fig. 5

Syntenic pair comparison of the *GSL-AOP* locus between *A. thaliana* accessions. Shown are representative accessions see Figure 2. Diagonal boxes indicate compared accession by name. Lower diagonal boxes: Continuous diagonal lines in same colour show syntenic regions sharing the same orientation. Upper diagonal boxes: Red anti-diagonal lines represent syteny in opposite direction, boxes surround syntenic regions. Numbers indicate cumulative nucleotide position in bp. Col-0**= Reference sequence Chr.4 pos. 12340082-1354901 (CP002687).



Extended Data Fig. 6

Barplot of AOP-modified GSL distribution across tissues. Mean of the different groups of SC GSLs in three different tissues, seed, leaf and root, shown according to their contribution to total SC GSL are shown in nmol/seed (seeds) or nmol/mg fresh weight (leaf, root). Accessions are sorted according to Figure 3 and GSL levels are given in percent of total SC GSLs.



Extended Data Fig. 7

Syntenic pairwise comparison of *FLC* locus among *A. thaliana* accessions. Shown are representative accessions (see Figure 5). Diagonal boxes indicate compared accession by name. Lower diagonal boxes: Continuous diagonal lines in same colour show syntenic regions sharing the same orientation. Upper diagonal boxes: Red anti-diagonal lines represent syntenies in opposite direction, boxes surround syntenic regions. Numbers indicate cumulative nucleotide position in bp. Col-0 Reference sequence Chr.5 pos. 3171361-3181999 (CP002687).