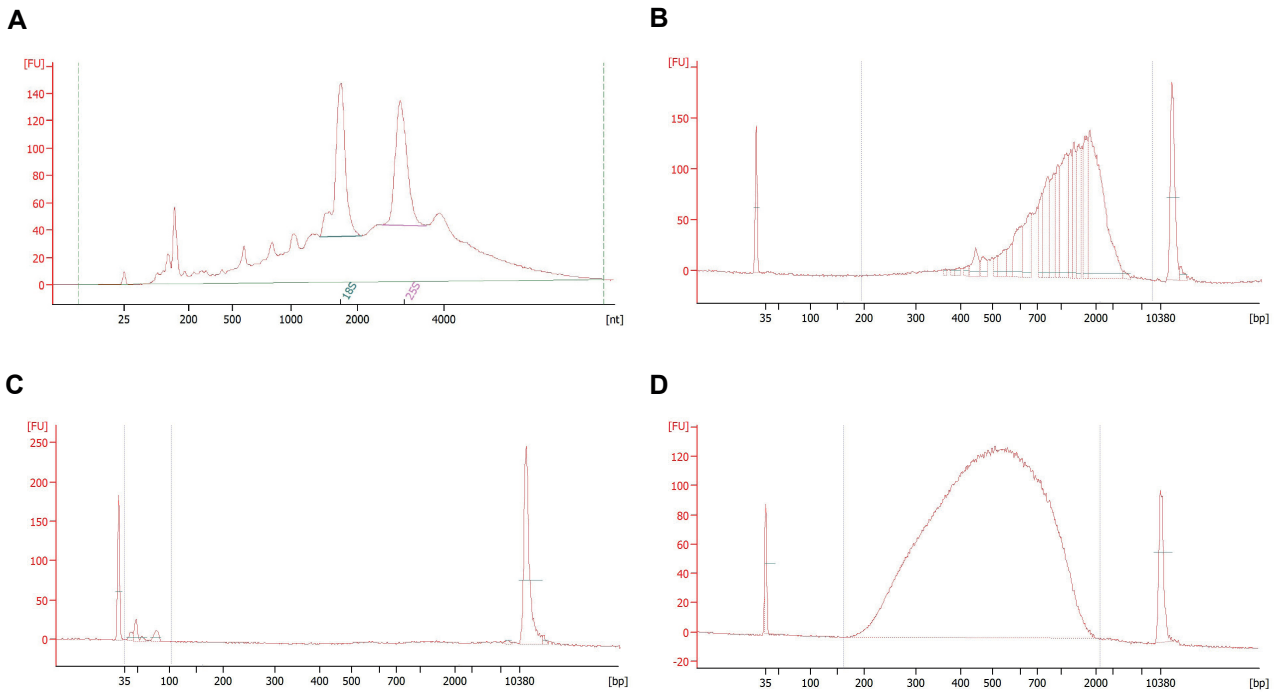

Figure S1. Quality controls of sample preparation for RNA-Seq. Electropherograms obtained with Agilent 2100 Bioanalyzer and Plant RNA Pico Assay (A) and High Sensitivity DNA Assays (B - D). Representative electropherograms are depicted of (A) total RNA extracted from LAM control sample with RNA integrity number (RIN) = 7.0, (B) amplified cDNA from LAM nucelli sample. Length distribution of cDNA fragments ranged from approx. 350 bp up to 3,500 bp roughly reflecting the length distribution of coding sequences in Boechera (Lee et al., 2017), and thus also long or even full-length transcripts were represented. (C) Empty control using water instead of RNA for cDNA amplification to exclude contamination or unspecific amplification and (D) electropherogram of equimolar library pool (30 nmol each) prior to sequencing. In general, all quality controls demonstrated good quality for RNA, cDNA and libraries well suitable for RNA-Seq.

Reference

Lee CR, Wang B, Mojica JP, et al. 2017. Young inversion with multiple linked QTLs under selection in a hybrid zone. Nature Ecology and Evolution **1**, 119.

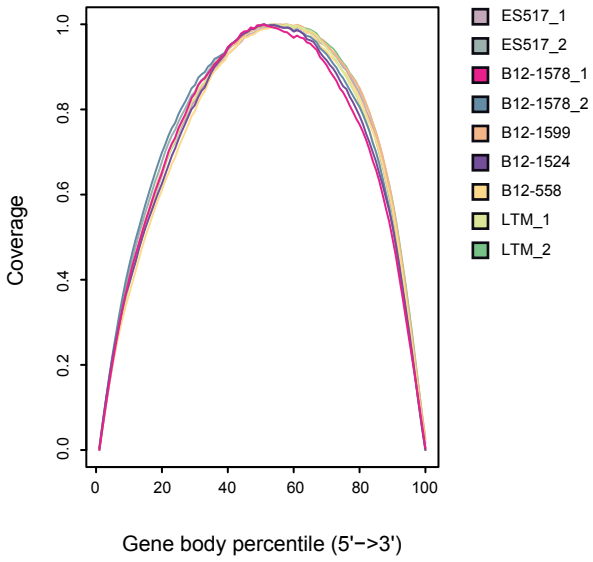

Figure S2. Coverage of gene bodies by mapped reads. Coverage is represented as density plot and was generated using geneBody_coverage.py from the RseQC package version 2.6.4 (Wang et al., 2012).

Reference

Wang L, Wang S, Li W. 2012. RSeQC: quality control of RNA-seq experiments. Bioinformatics **28**, 2184-21.

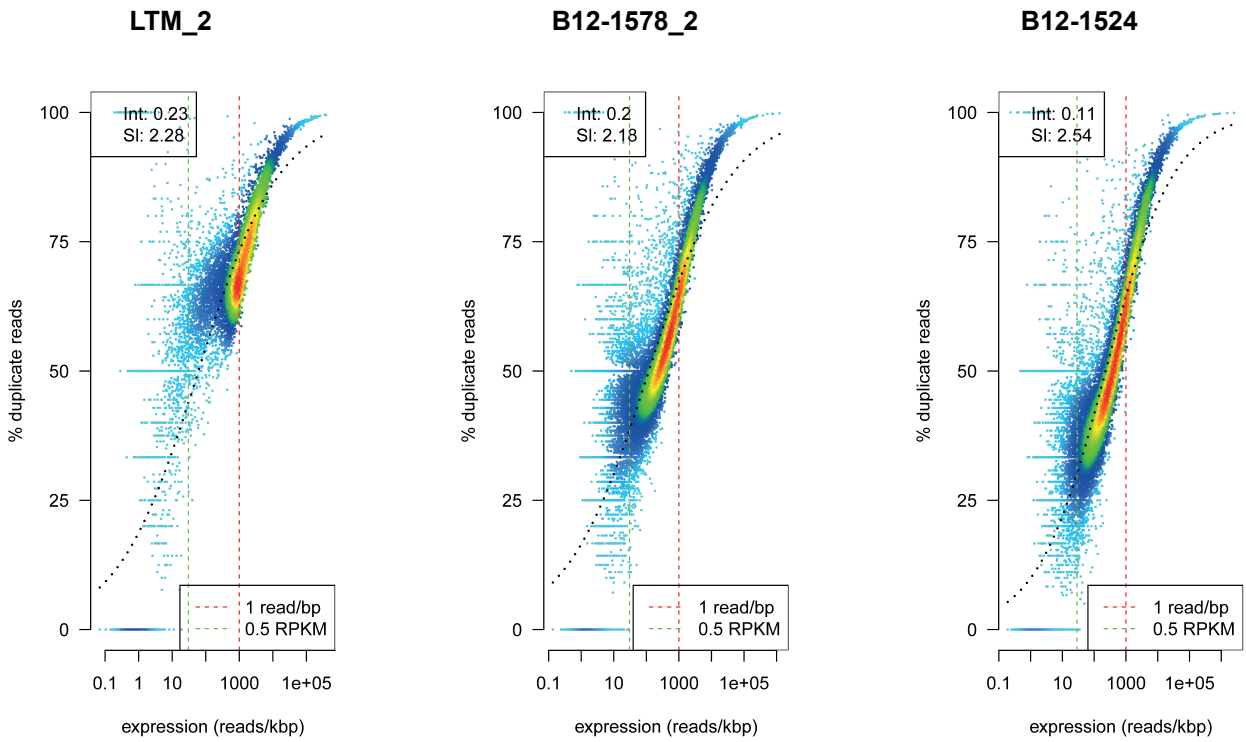

Figure S3. Assessment of read duplication rates. To estimate potential PCR artefacts in the libraries, read duplication of representative libraries was calculated and plotted in relation to estimated gene expression (in reads per kilo base pairs) with dupRadar (Sayols et al., 2016). Selected libraries covered the range of input material from 38 nucelli sections in LTM_2, to 142 in B12-1578_2 and 226 in B12-1524. In libraries B12-1524 and B12-1578_2 the duplication rate is logistically proportional to gene expression, indicating natural read duplication of highly expressed genes. Duplication rates in LTM_2 were higher than 50% also for lowly expressed genes and thus PCR artefacts affecting data quality cannot be excluded in this library (Sayols et al., 2016). Coefficients of the fitted model reflect initial duplication rate at low read counts (Int = intercept) and the progression of the duplication rate along with the progression of the read counts (Sl = Slope).

Reference

Sayols S, Scherzinger D, Klein H. 2016. dupRadar: a Bioconductor package for the assessment of PCR artifacts in RNA-Seq data. BMC Bioinformatics **17**, 428

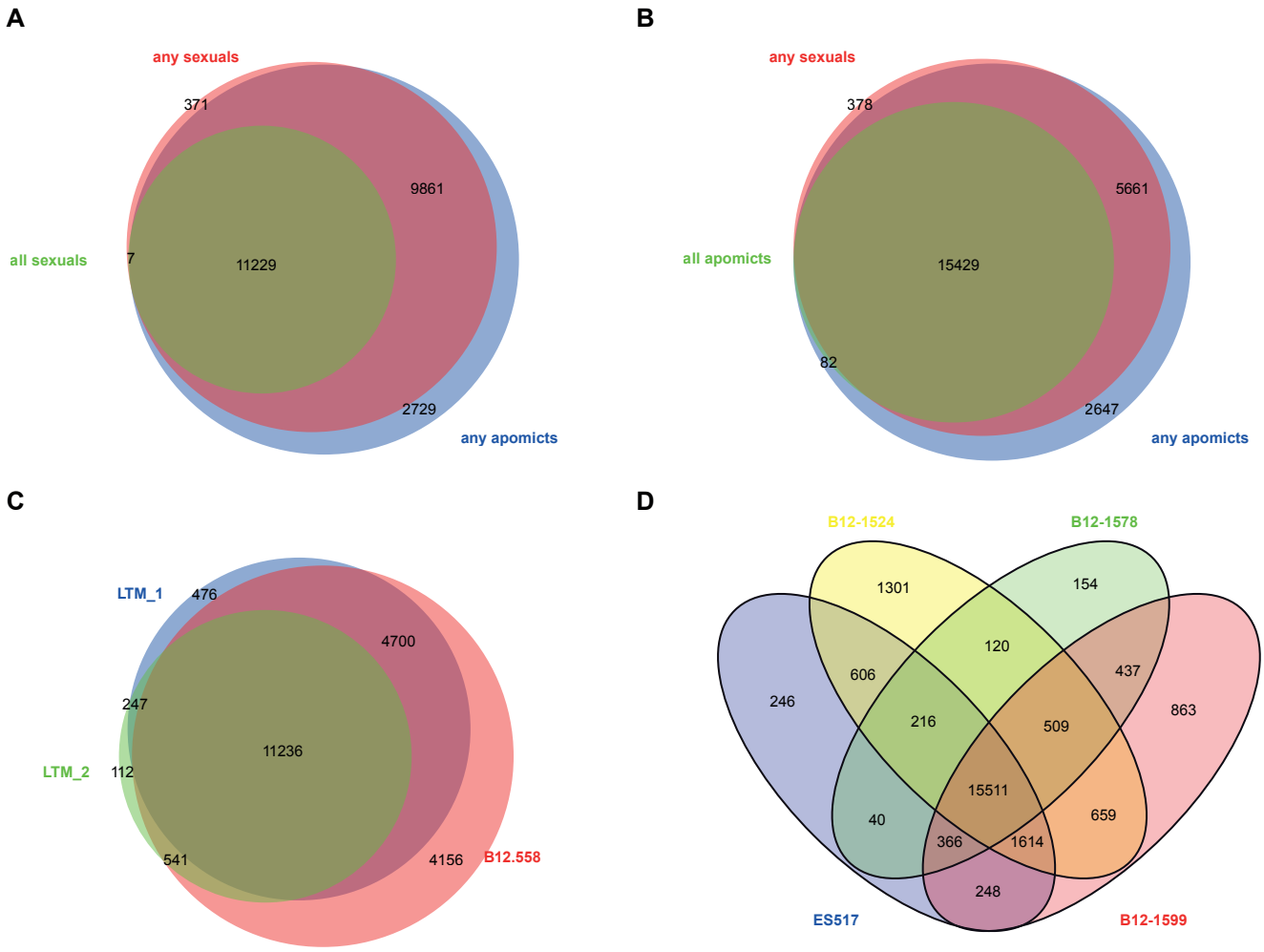

Figure S4. Overlap of gene expression. (A) Overlap of gene expression shared between all sexual samples, or in any sample from apomictic or any sample from sexual accessions (B) Overlap of gene expression shared between all apomictic samples, or in any sample from apomictic or any sample from sexual accessions. Overlap of gene expression in sexual accessions (C) or apomictic accessions (D). To allow presentation in a Venn diagram, intersection of expressed genes was calculated of replicates from ES517 and for B12-1578 each. Genes were considered expressed at ≥ 10 TMM normalized read counts.

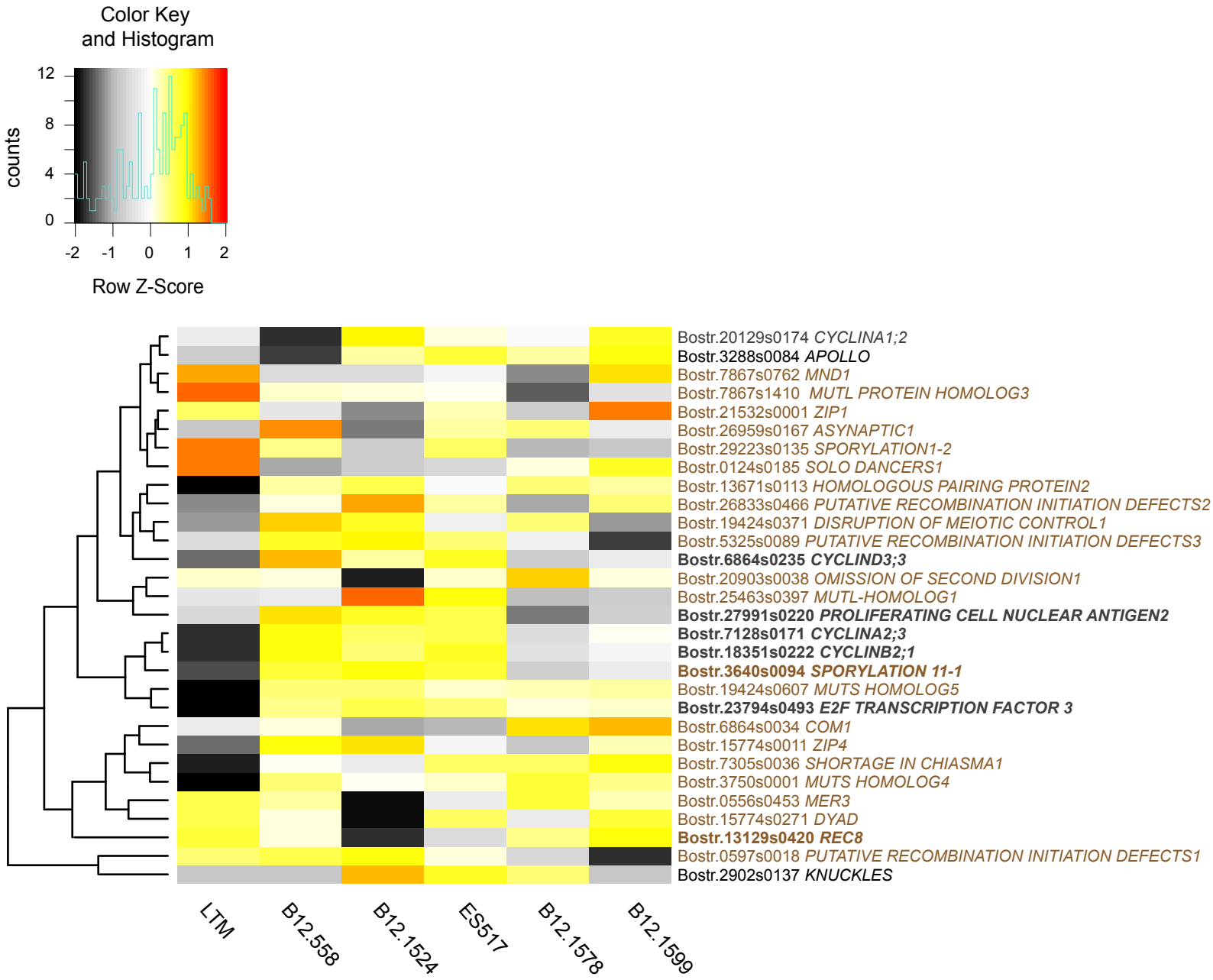

Figure S5. Heatmap of log2 transformed read counts of selected genes. Selection comprises normalized read counts of meiotic genes, core cell cycle regulators, *APOLLO* and *KNUCKLES* as in Table S4. Hierarchical agglomerative clustering of genes was based on Euclidean distance. Genes significantly differentially regulated between any of the accessions are indicated with bold letters. Names of core cell cycle regulators are depicted in dark gray, meiotic genes in brown.

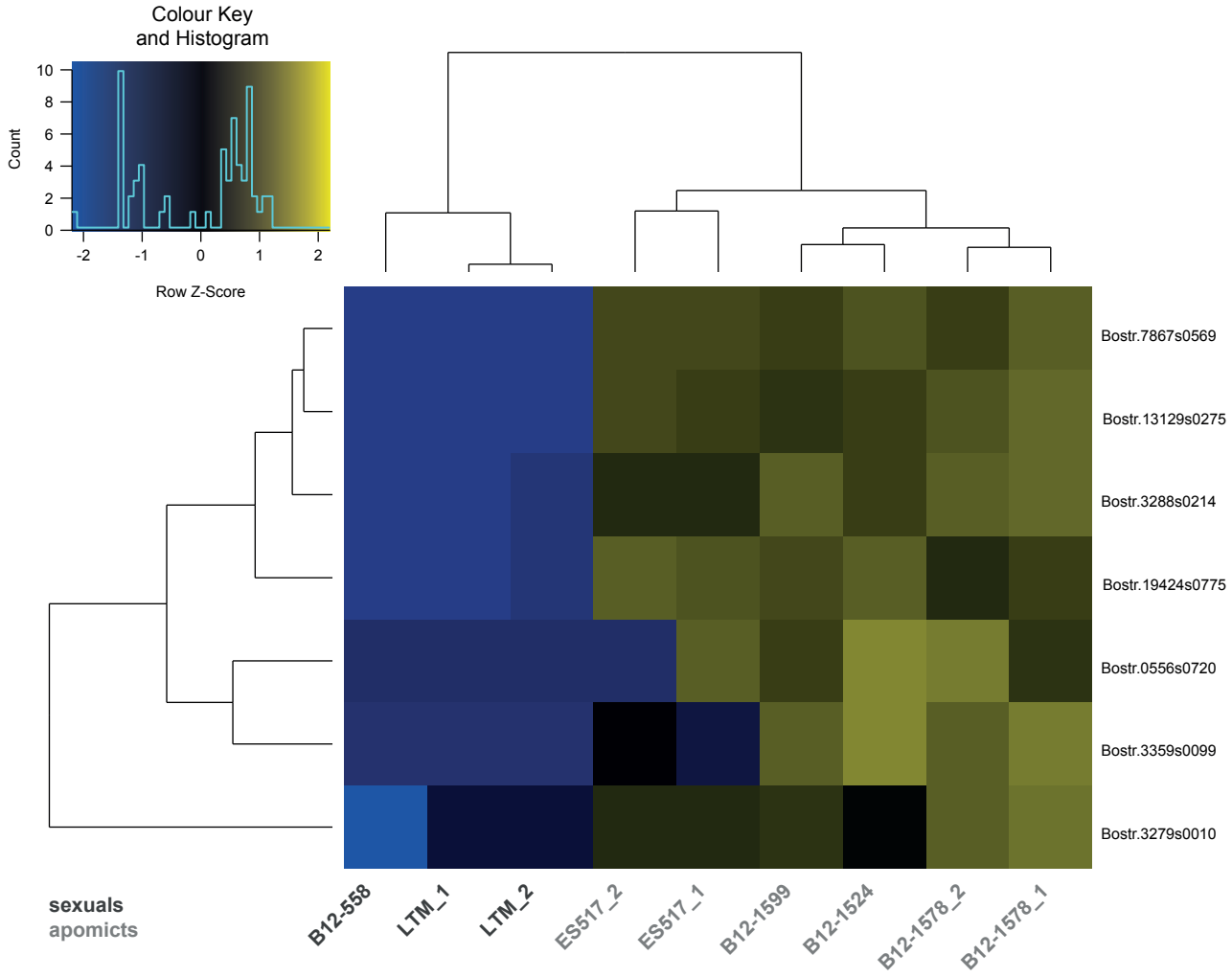

Figure S6. Heatmap of log2 transformed read counts of common, differentially expressed genes. The heatmap is based on normalized read counts of differentially expressed genes identified in all pairwise comparisons between sexual and apomictic accessions. Hierarchical agglomerative clustering and row scaling was applied to the data.