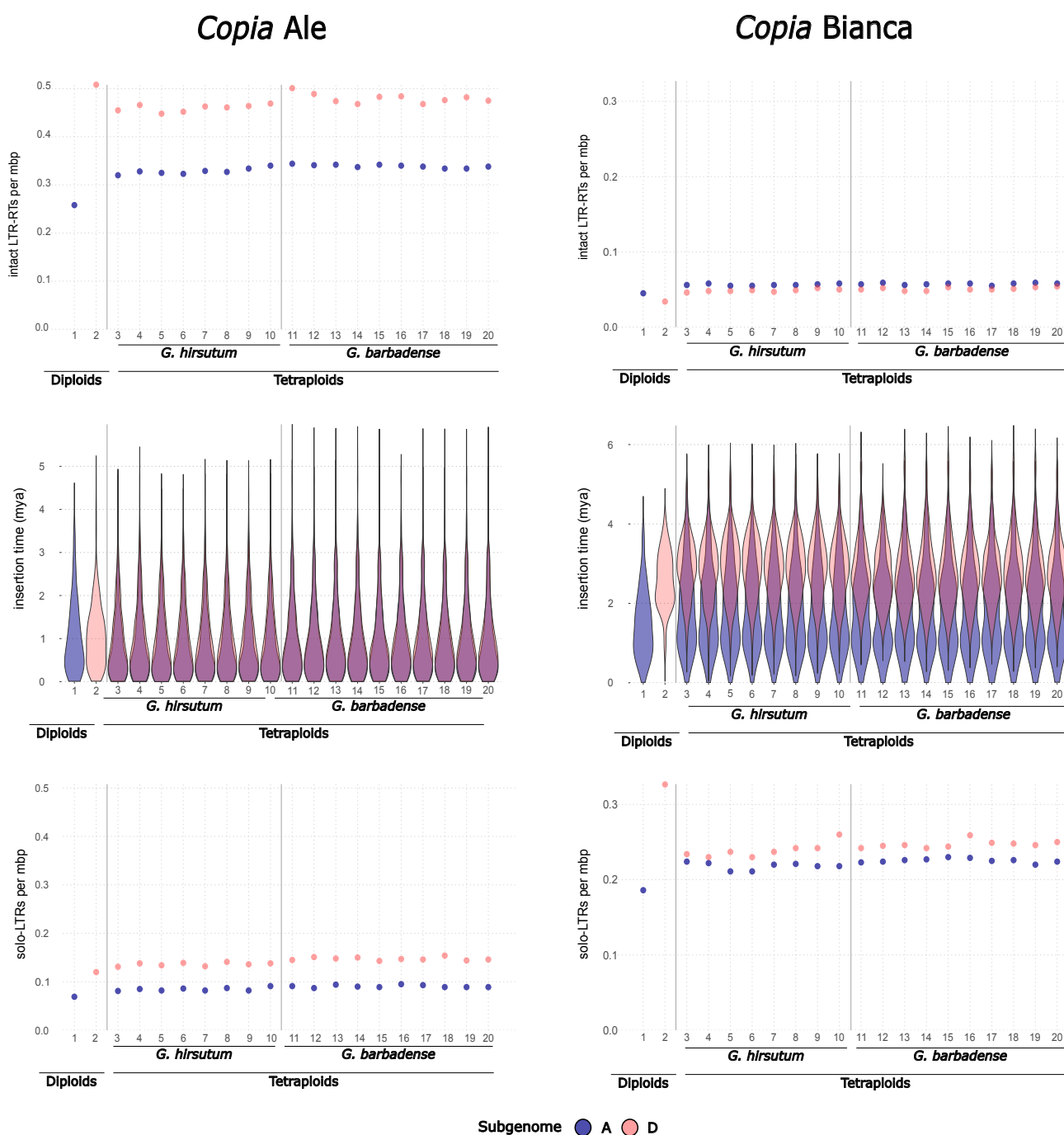
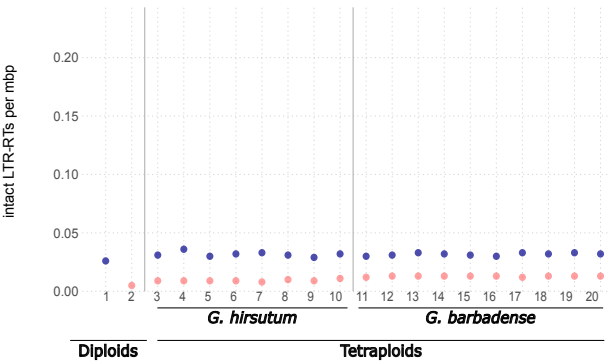


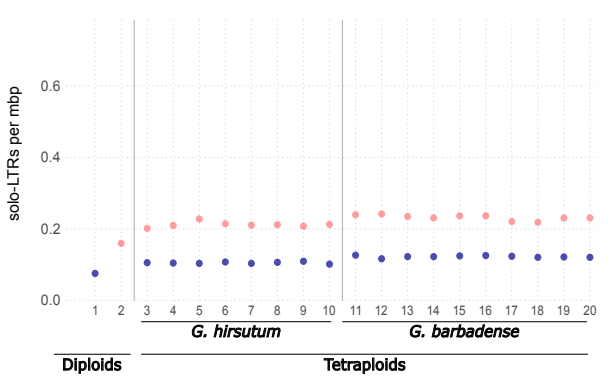
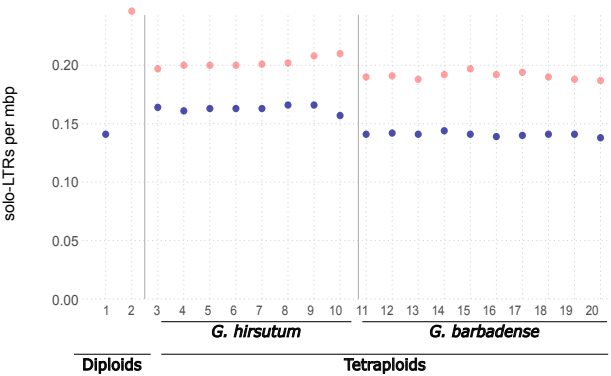
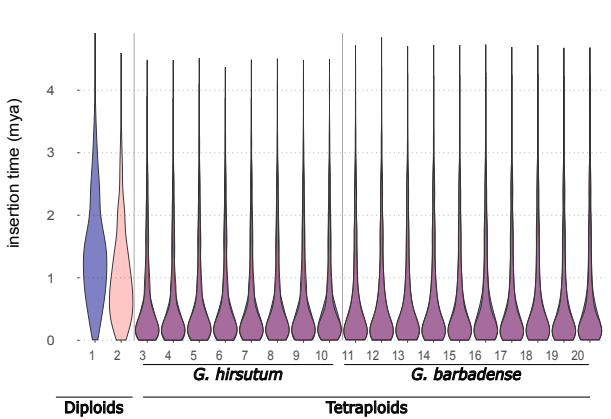
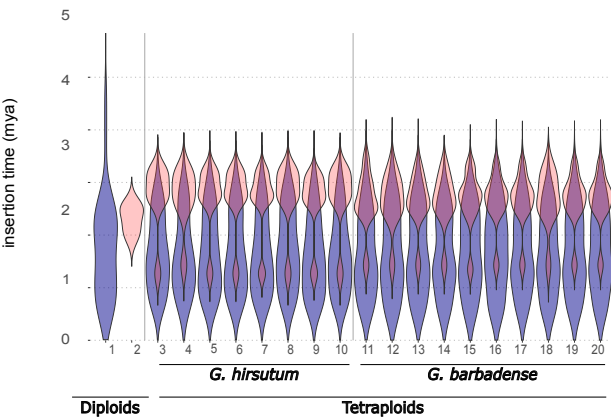
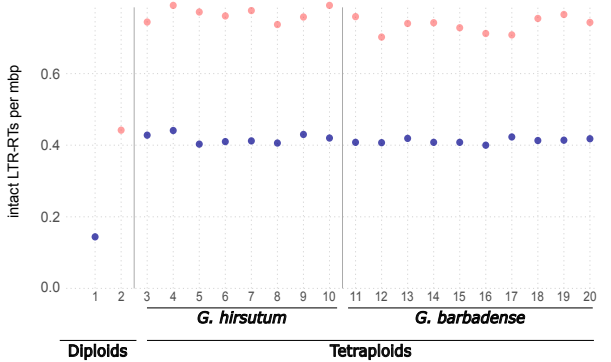
Figure S1. Intact LTR-RT and solo LTR content in diploid and tetraploid cotton by lineage and subgenome. Number of intact LTR-RT per megabase pairs per subgenome in each of the cotton species and varieties analyzed (top panels), violin plots representing the insertion times these intact LTR-RTs per subgenome in each of the cotton species and varieties analyzed (middle panels), and number of solo LTRs per megabase pairs per subgenome in each of the cotton species and varieties analyzed (bottom panels).



Copia Ikeros

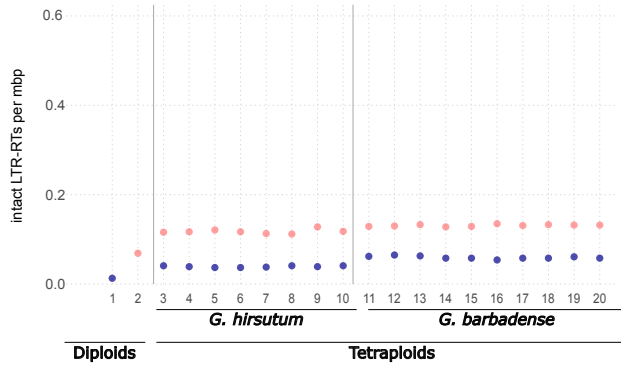


Copia Ivana

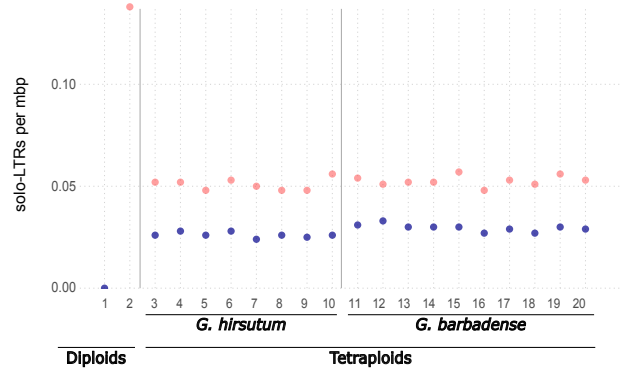
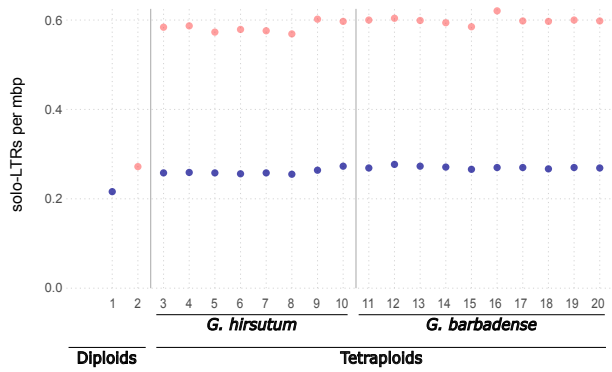
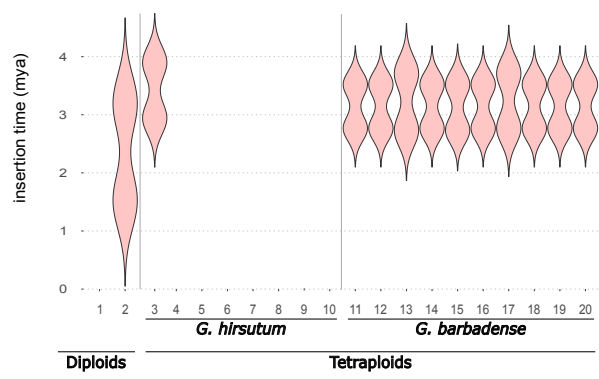
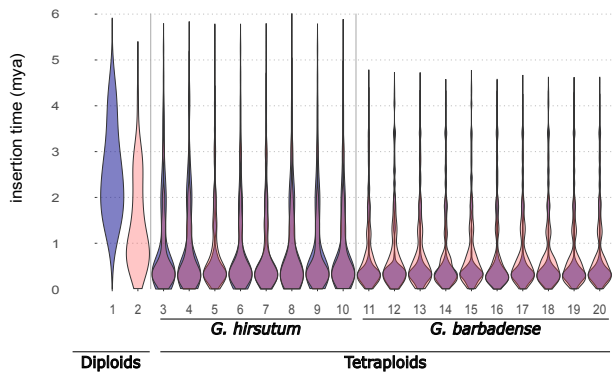
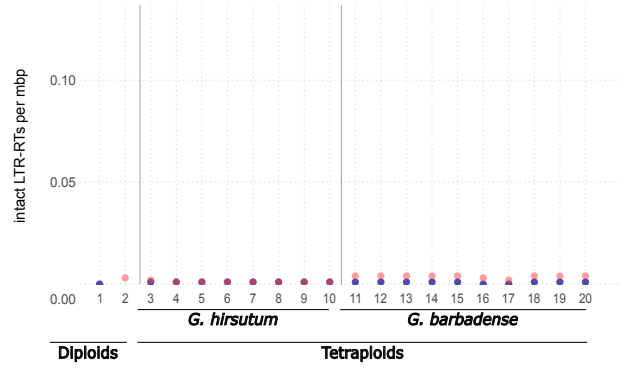


Subgenome ● A ● D

Copia SIRE

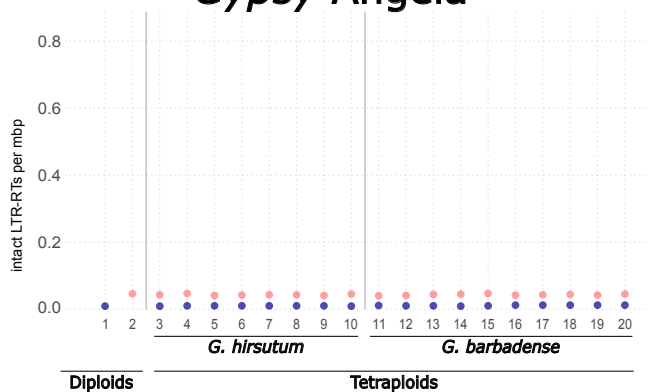


Copia TAR

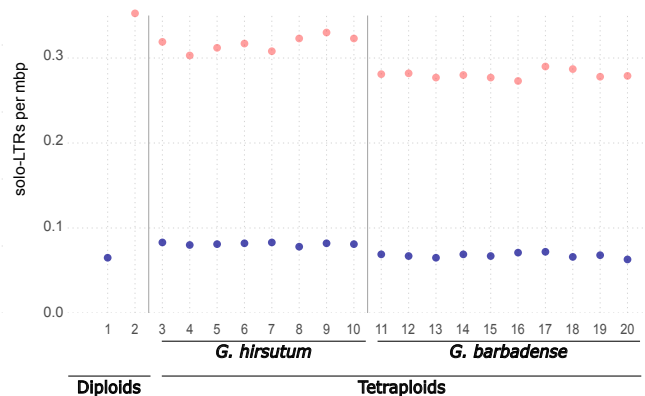
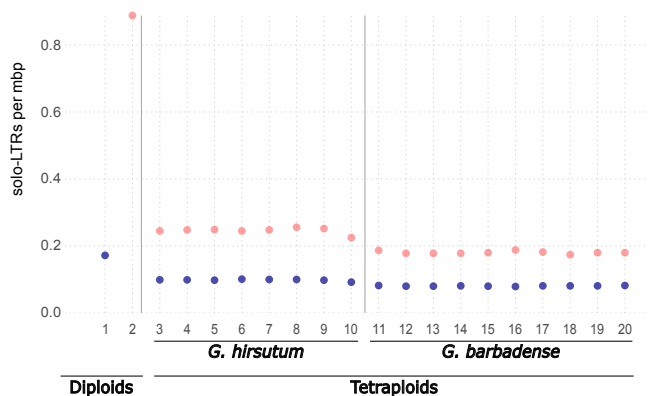
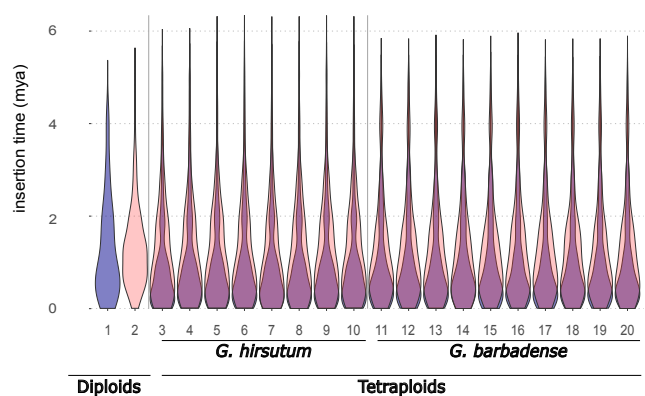
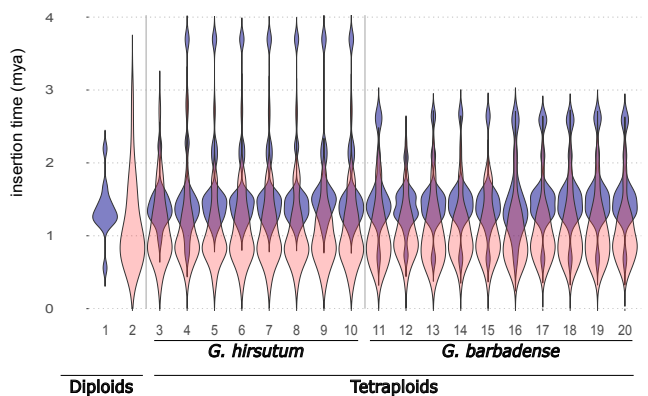
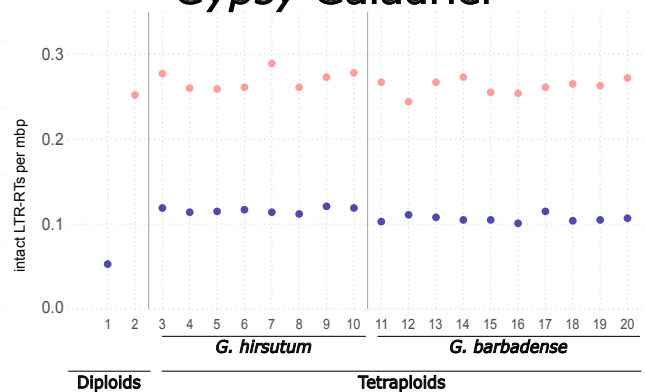


Subgenome ● A ● D

Gypsy Angela

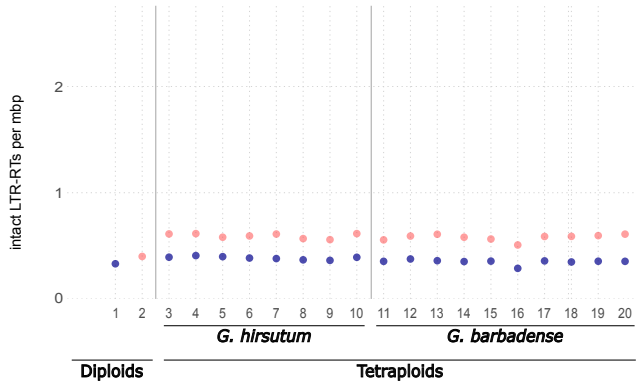


Gypsy Galadriel

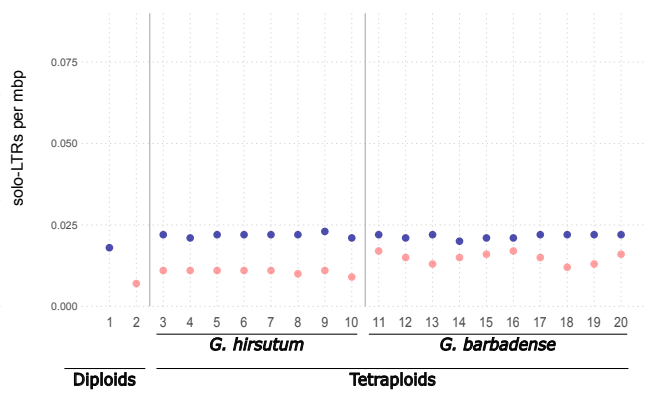
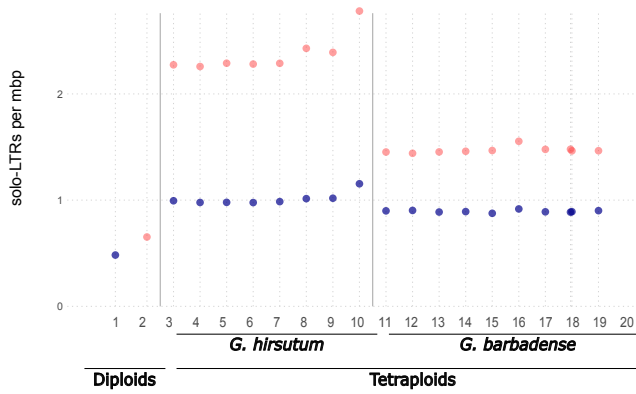
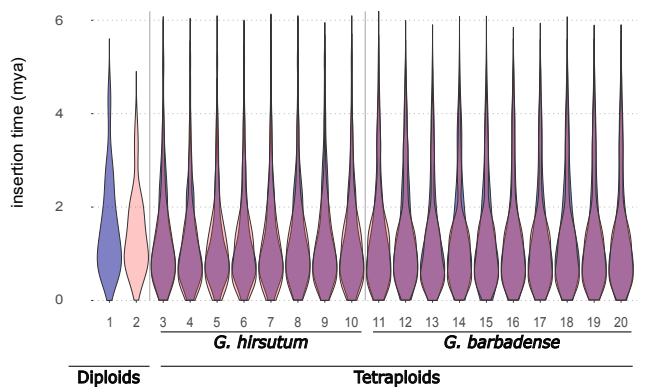
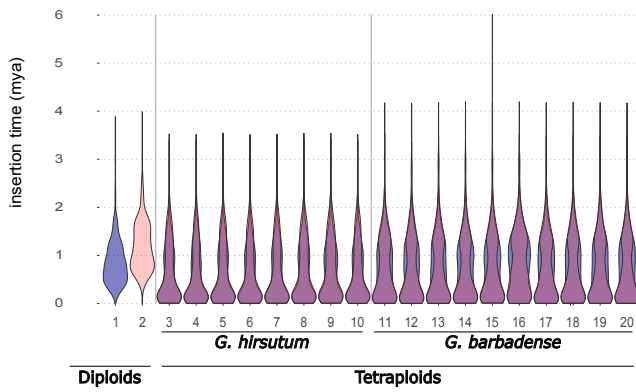
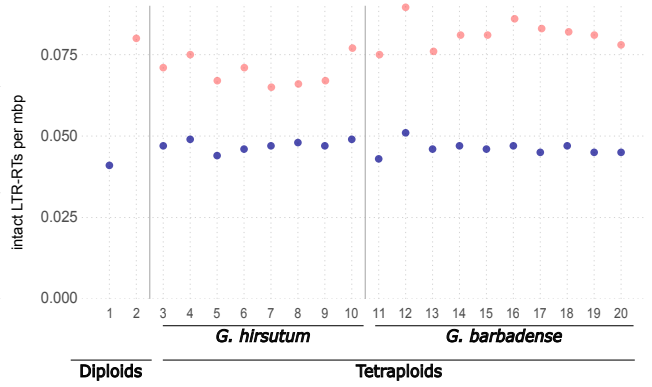


Subgenome ● A ● D

Gypsy Ogre



Gypsy Reina



Subgenome ● A ● D

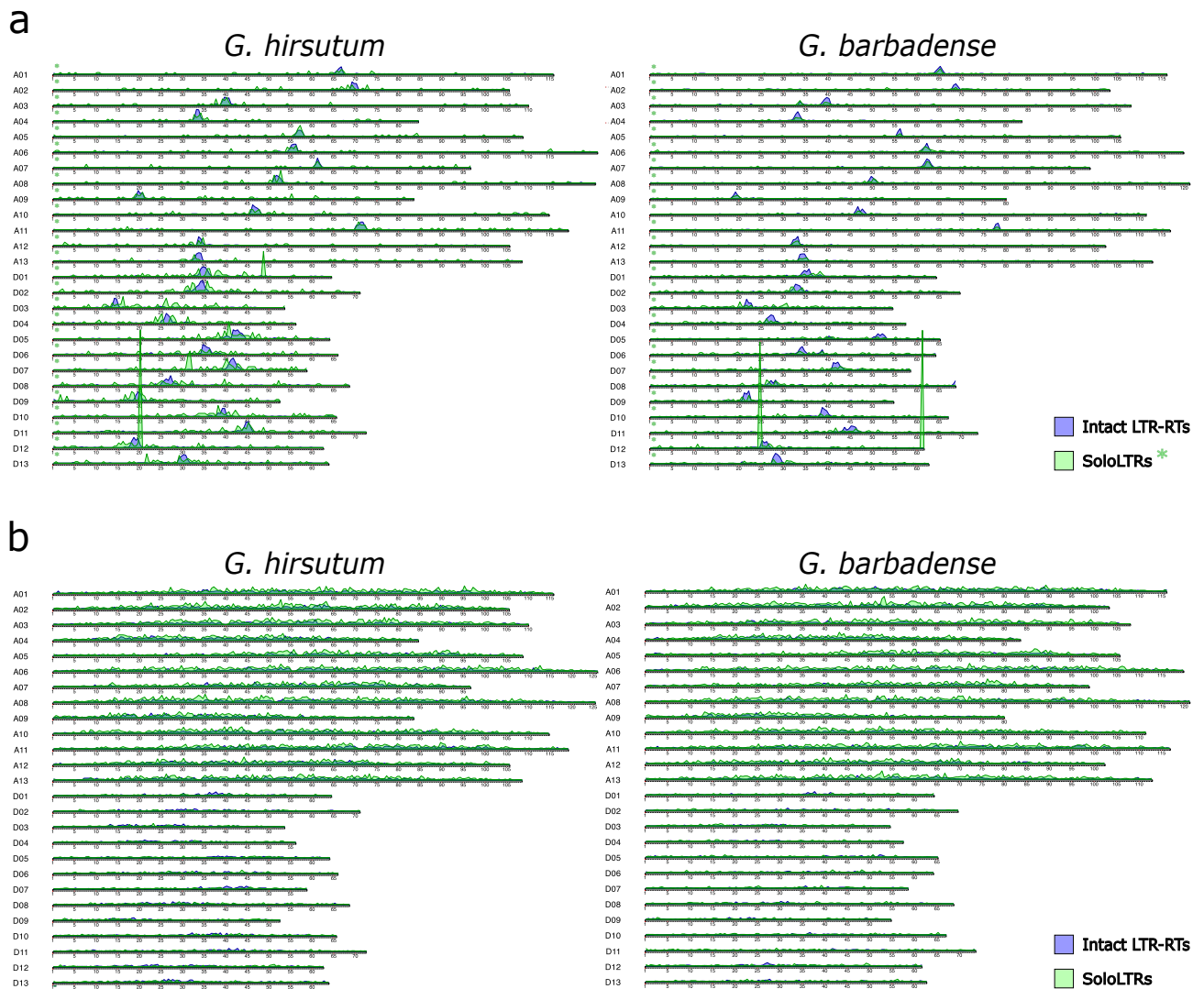


Figure S2. Chromosomal distribution of CRM and Tekay elements along the tetraploids *G. hirsutum* and *G. barbadense*. a) Density plots of CRM intact LTR-RTs and solo-LTRs along all chromosomes of *G. hirsutum* and *G. barbadense*. A green asterisk indicates where the density of solo-LTRs represented has been modified to make their distribution visible along chromosomes (scale of solo-LTRs is 10x the scale of intact). b) Density plots of Tekay intact LTR-RTs and solo-LTRs along all chromosomes of *G. hirsutum* and *G. barbadense*.

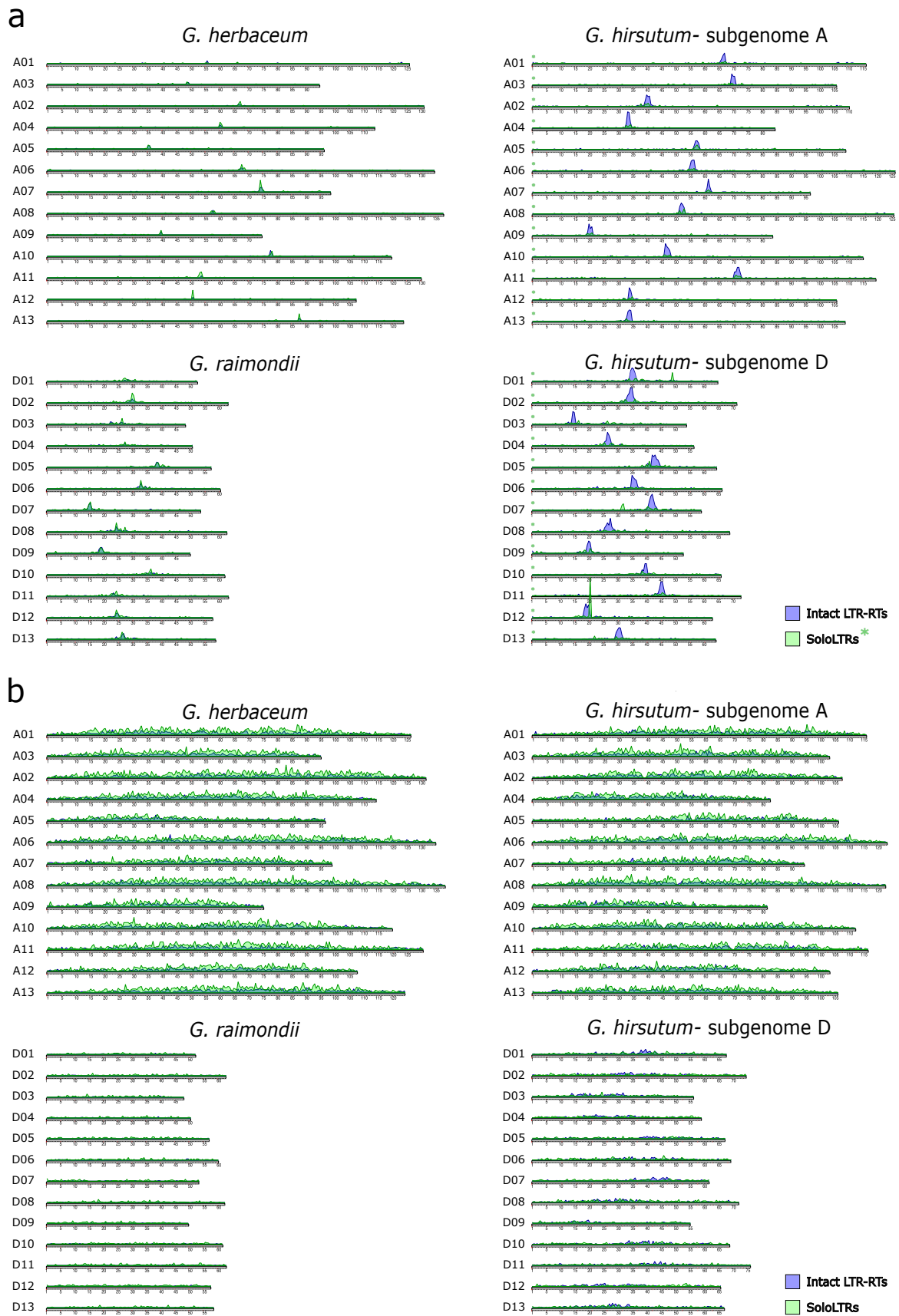
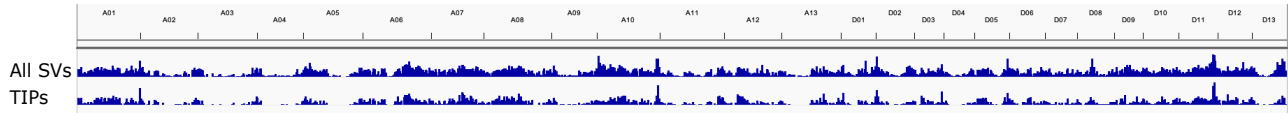


Figure S3. Chromosomal distribution of CRM and Tekay elements along the parental diploids *G. herbaceum* and *G. raimondii*, and the tetraploid *G. hirsutum*. a) Density plots of CRM intact LTR-RTs and solo-LTRs along all chromosomes of *G. herbaceum*, *G. raimondii* and *G. hirsutum*. A green asterisk indicates where the density of solo-LTRs represented has been modified to make their distribution visible along chromosomes (scale of solo-LTRs is 2.5x the scale of intact). b) Density plots of Tekay intact LTR-RTs and solo-LTRs along all chromosomes of *G. herbaceum*, *G. raimondii* and *G. hirsutum*.

a



b

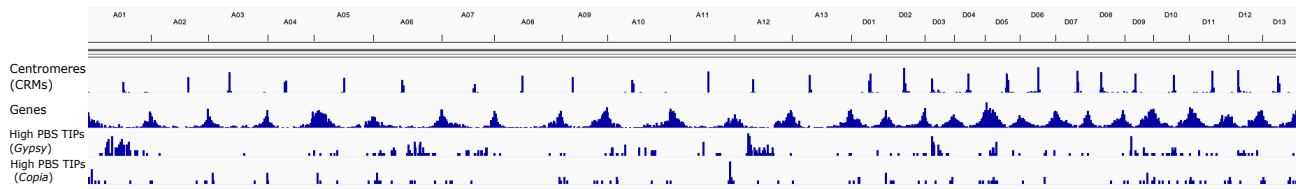
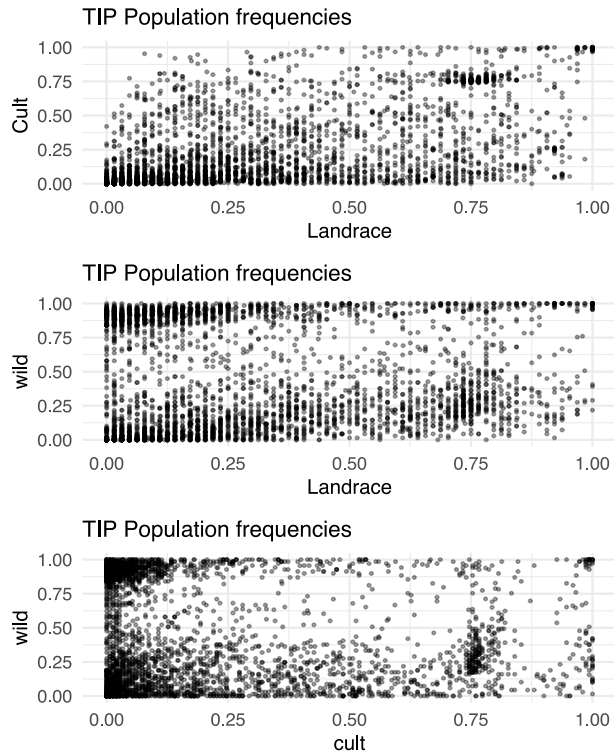


Figure S4. Distribution of SVs and TIPS along the *G. hirsutum* chromosomes.
a) Distribution of SVs (top) and all TIPS (bottom) along the *G. hirsutum* chromosomes. b) Genome distribution of high PBS TIPS (split into *Gypsy* and *Copia* LTR superfamilies) along the *G. hirsutum* chromosomes (two bottom tracks). Centromere (intact CRM elements) and gene distribution in the two top tracks for reference.

a



b

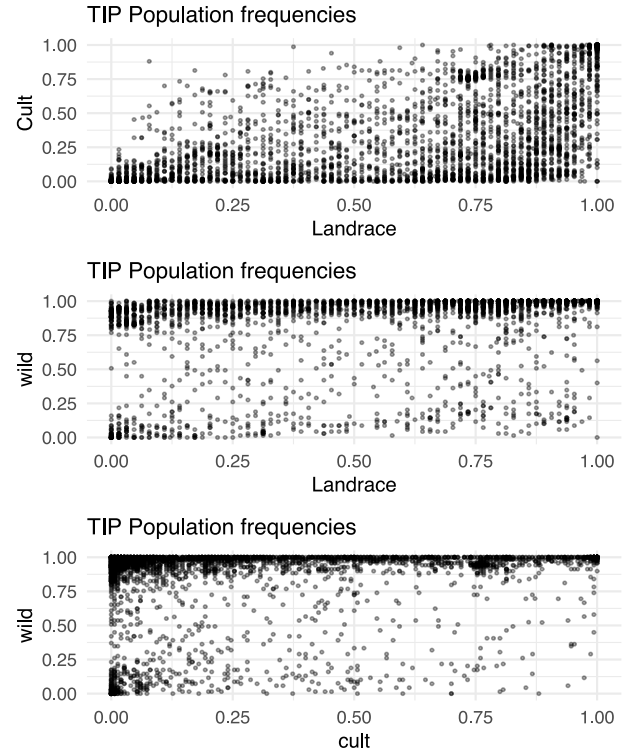


Figure S5. TIP population frequencies in *G. hirsutum*. a) Scatterplots representing the frequency of TIP insertion variants in the three *G. hirsutum* populations. b) Scatterplots representing the frequency of TIP deletion variants in the three *G. hirsutum* populations. Cult = Cultivated accessions.

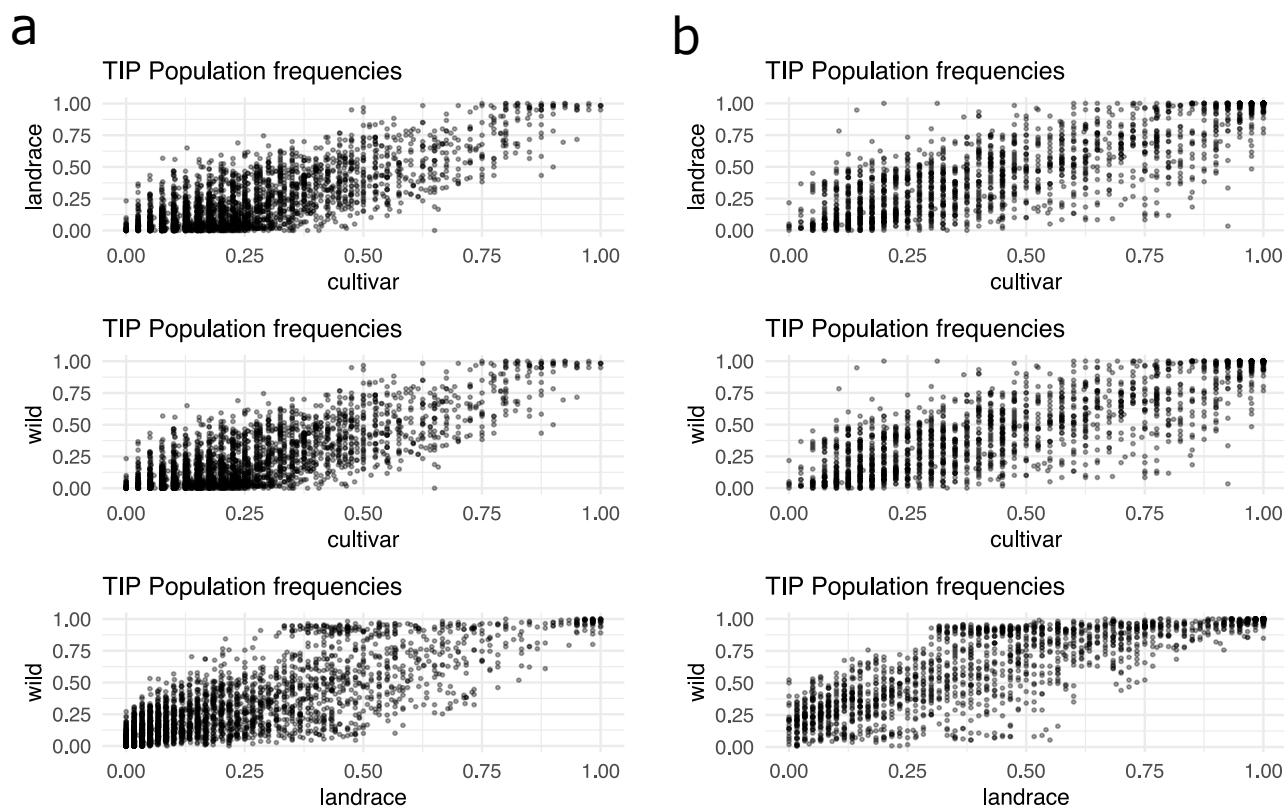


Figure S6. TIP population frequencies in *G. barbadense*. a) Scatterplots representing the frequency of TIP insertion variants in the three *G. barbadense* populations. b) Scatterplots representing the frequency of TIP deletion variants in the three *G. barbadense* populations. Cult = Cultivated accessions.

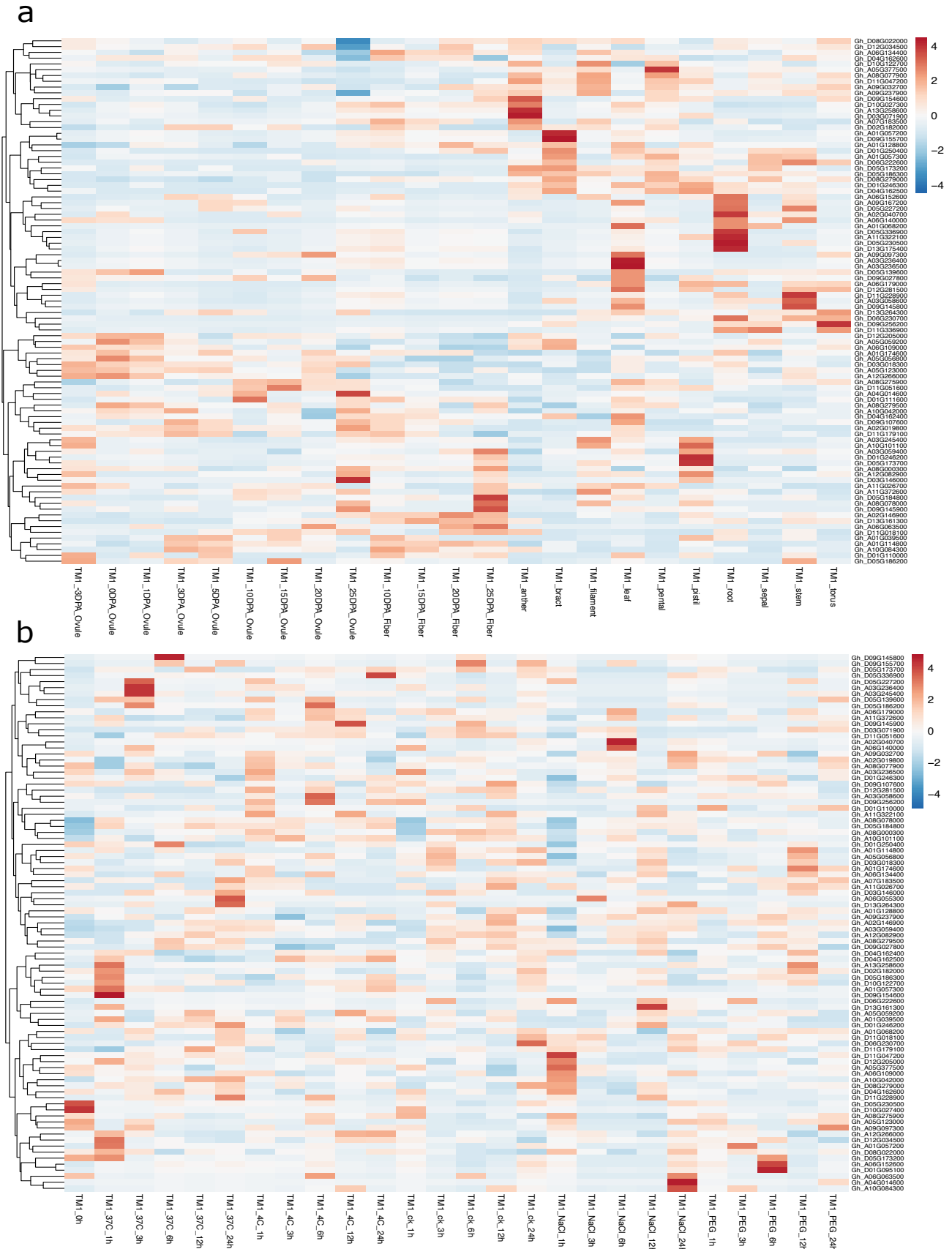


Figure S7. Expression heatmaps of genes near high PBS TIPs in *G. hirsutum*. a) Heatmap representing the relative expression (Rlog values of raw counts) of the genes close to (max 2kbp) high PBS TIPs in different tissues and developmental stages of *G. hirsutum*. Each value is the average of three replicates per sample. b) Heatmap representing the relative expression (Rlog values of raw counts) of the genes close to (max 2kbp) high PBS TIPs in different stress conditions for *G. hirsutum*. Each value is the average of three replicates per sample.