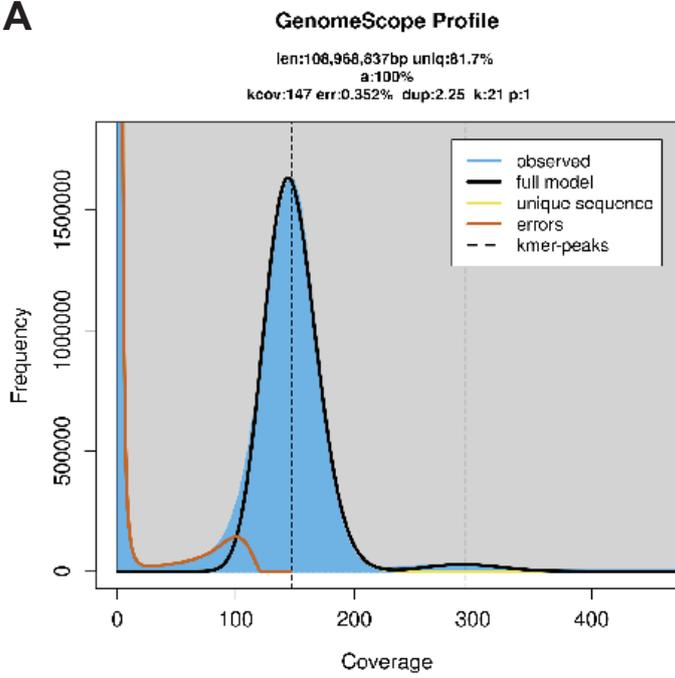
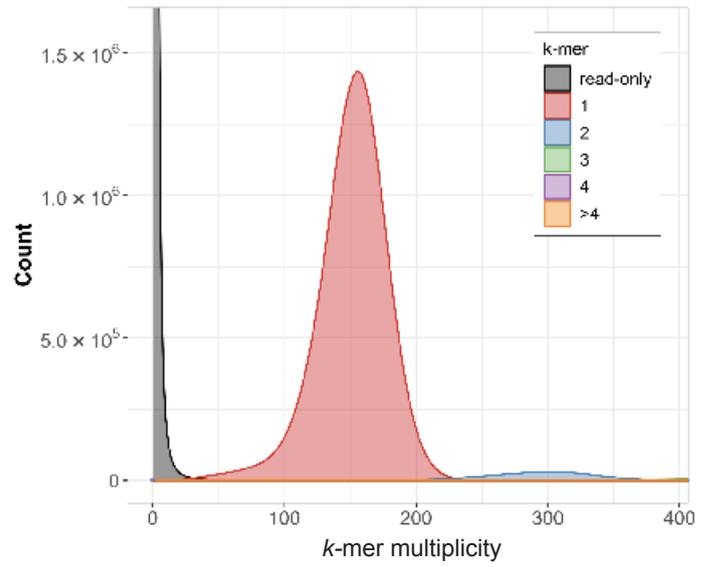
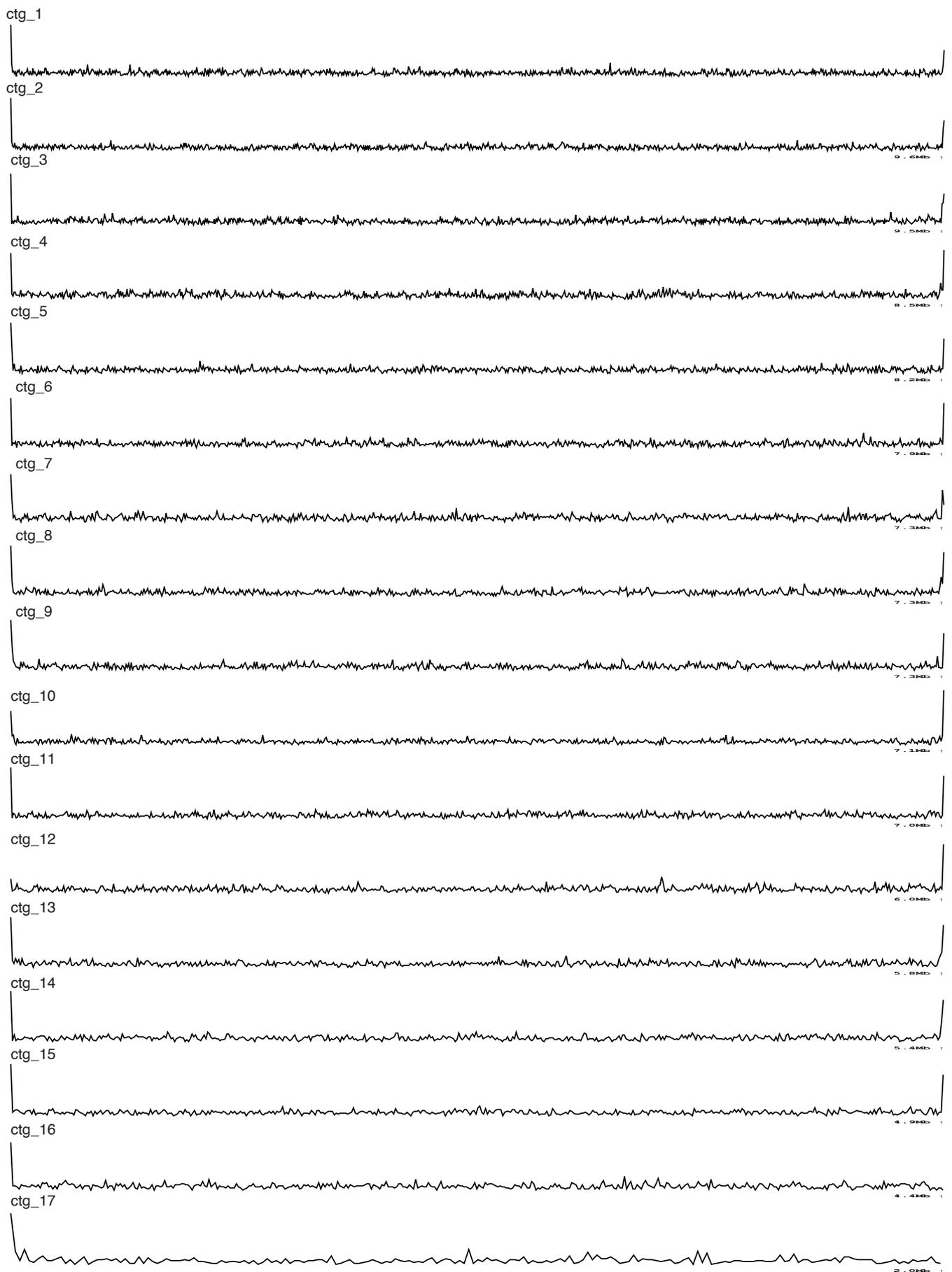


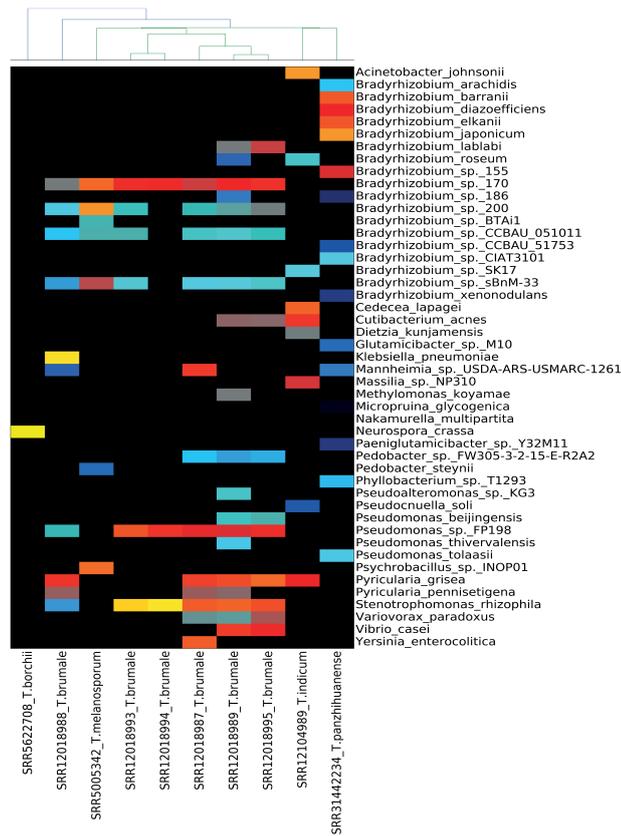
A**B**

Supplemental Figure S1: (A) Short-reads *k*-mer based genome survey using genomescope with a *k*-mer size of 21. (B) Copy number *k*-mer spectrum obtained with Merqury and a *k*-mer size of 21.



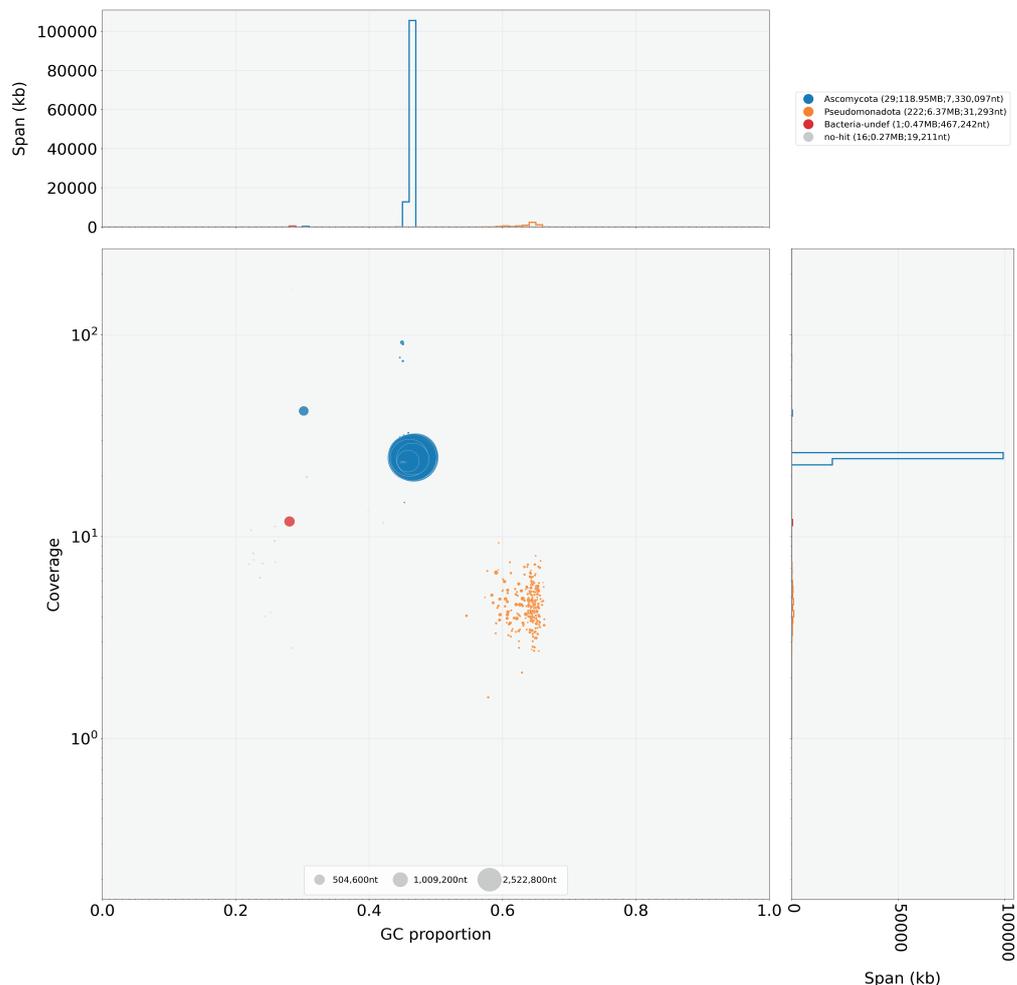
Supplemental Figure S2: Count of TTAGGG telomeric repeats over 10 kb genomic windows along the 17 main nuclear contigs.

A

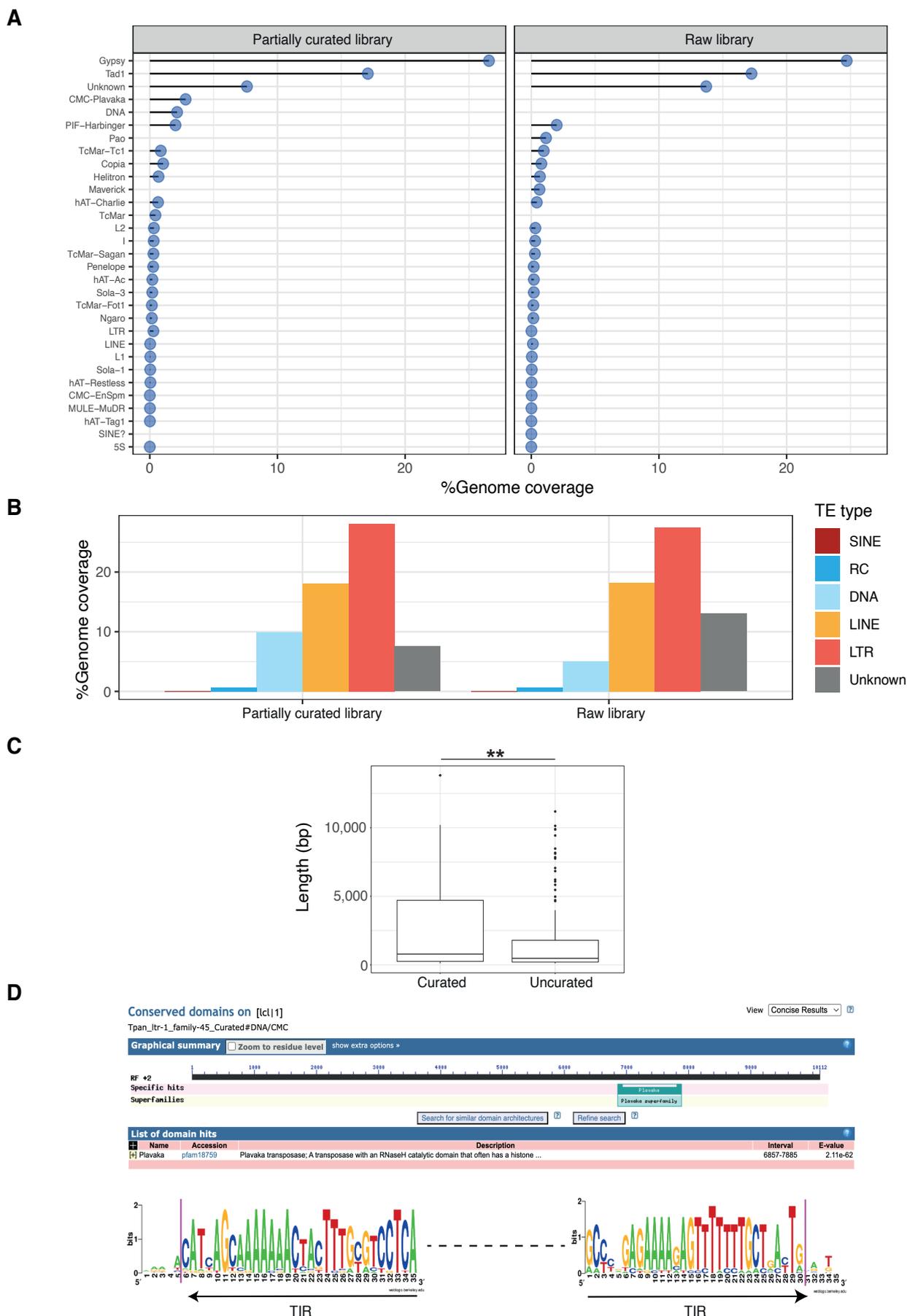


B

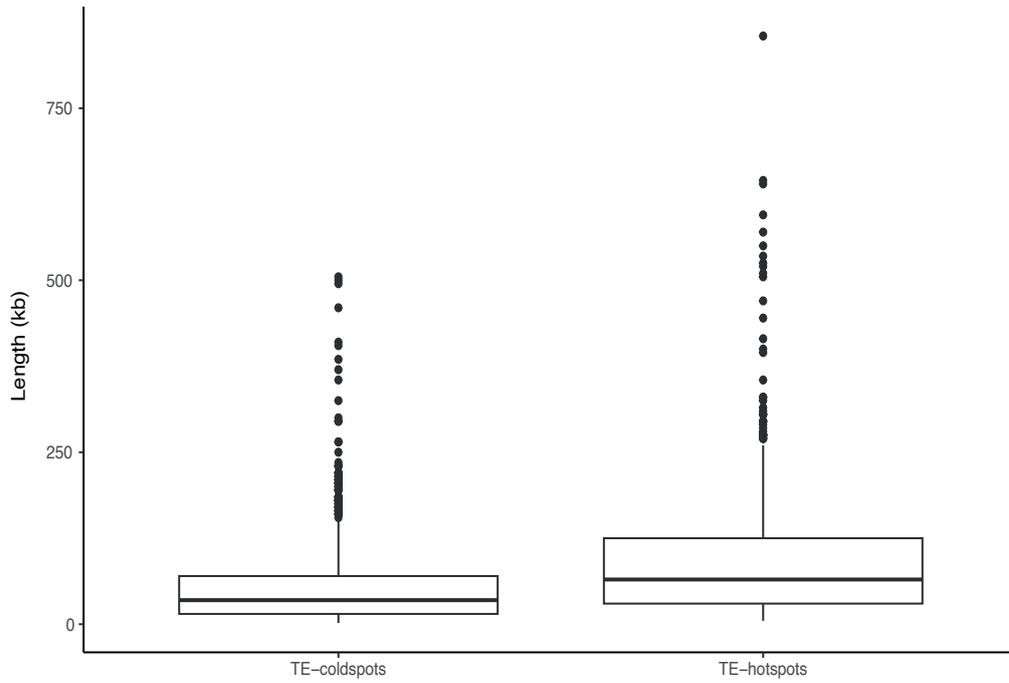
tubl_blobtools1_plot_out.tubl_createblob1_output.blobDB.json.bestsum.phylum.p8.span.100.blobplot.bam0



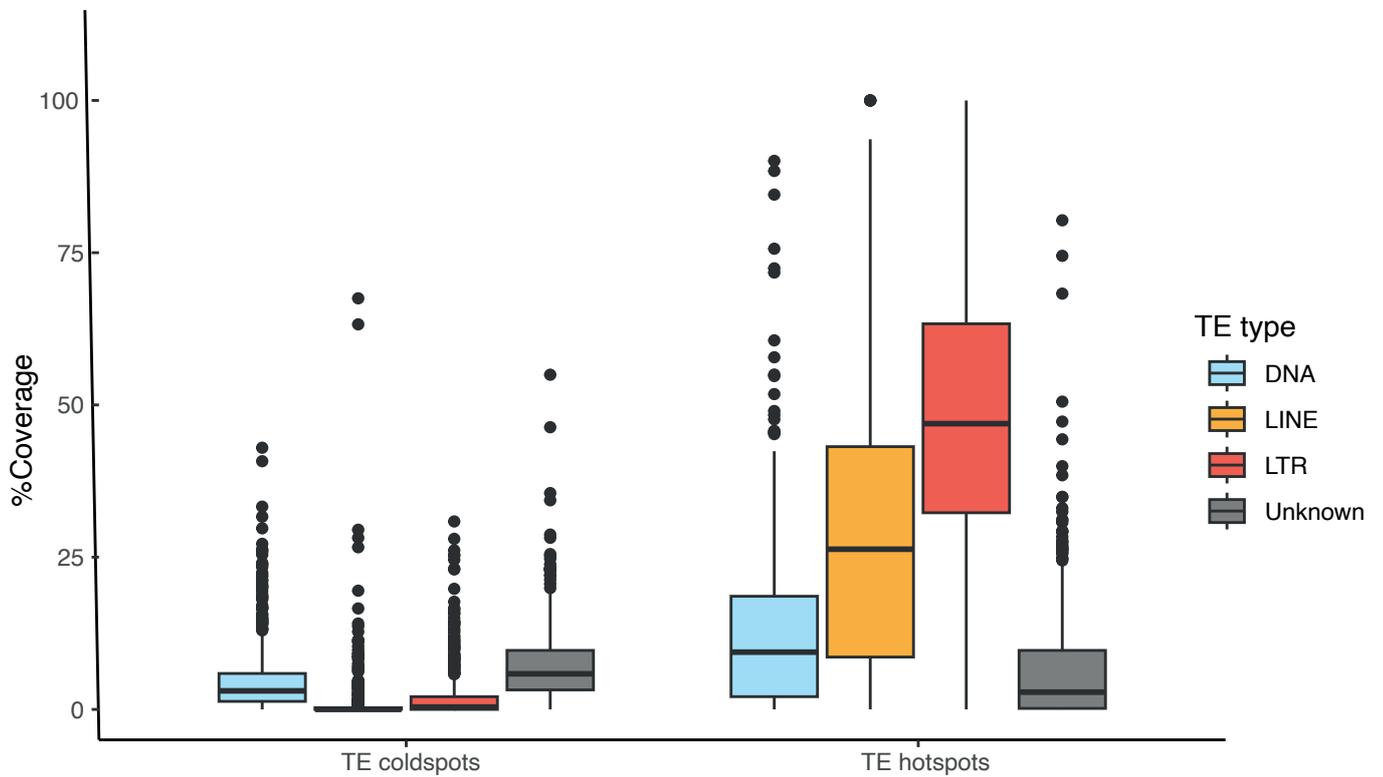
Supplemental Figure S3: (A) Results of metagenomic profiling of the gleba of different truffle samples. SRR codes refer to NCBI SRA archive. **(B)** Blobtools plot of the first, unfiltered version of the assembly. Reads mapping to bacterial contigs were removed and the genome reassembled to obtain the final version of the assembly.



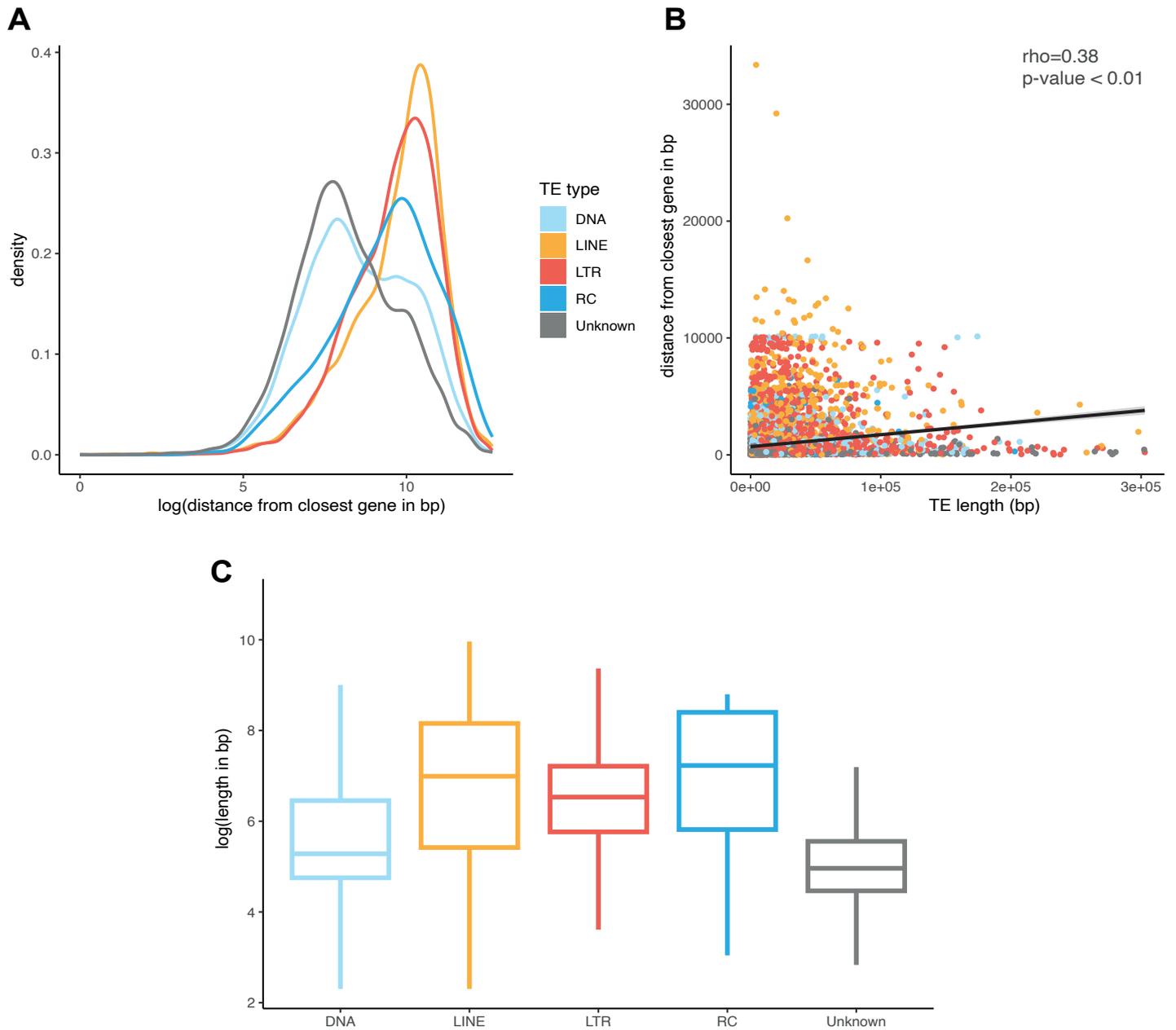
Supplemental Figure S4: Percentage of genome covered by **(A)** different transposon groups and **(B)** different TE class using the raw and the partially curated repeat libraries on the *T. panzihuanense* genome. **(C)** Comparison of the length of TE consensus sequences selected for manual curation before (Uncurated) and after (Curated) the curation process. **(D)** Manual classification of a Plavaka DNA transposon (CMC superfamily) previously classified as LTR/Pao in the raw repeat library. The transposon is characterized by a characteristic transposase domain (identified through CD-Search against NCBI's Conserved Domain Database) and the presence of characteristic terminal inverted repeats (TIRs) with CATCA motif. The vertical purple lines denote the boundaries of the element. The sequence logo was obtained based on the alignment of the top 25 hits of the consensus sequences against the genome.



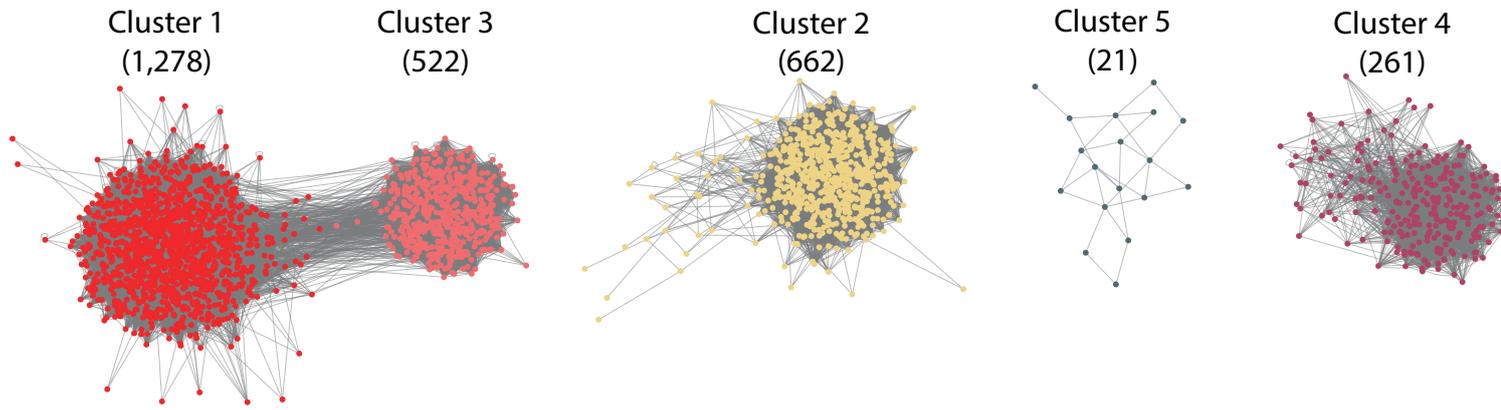
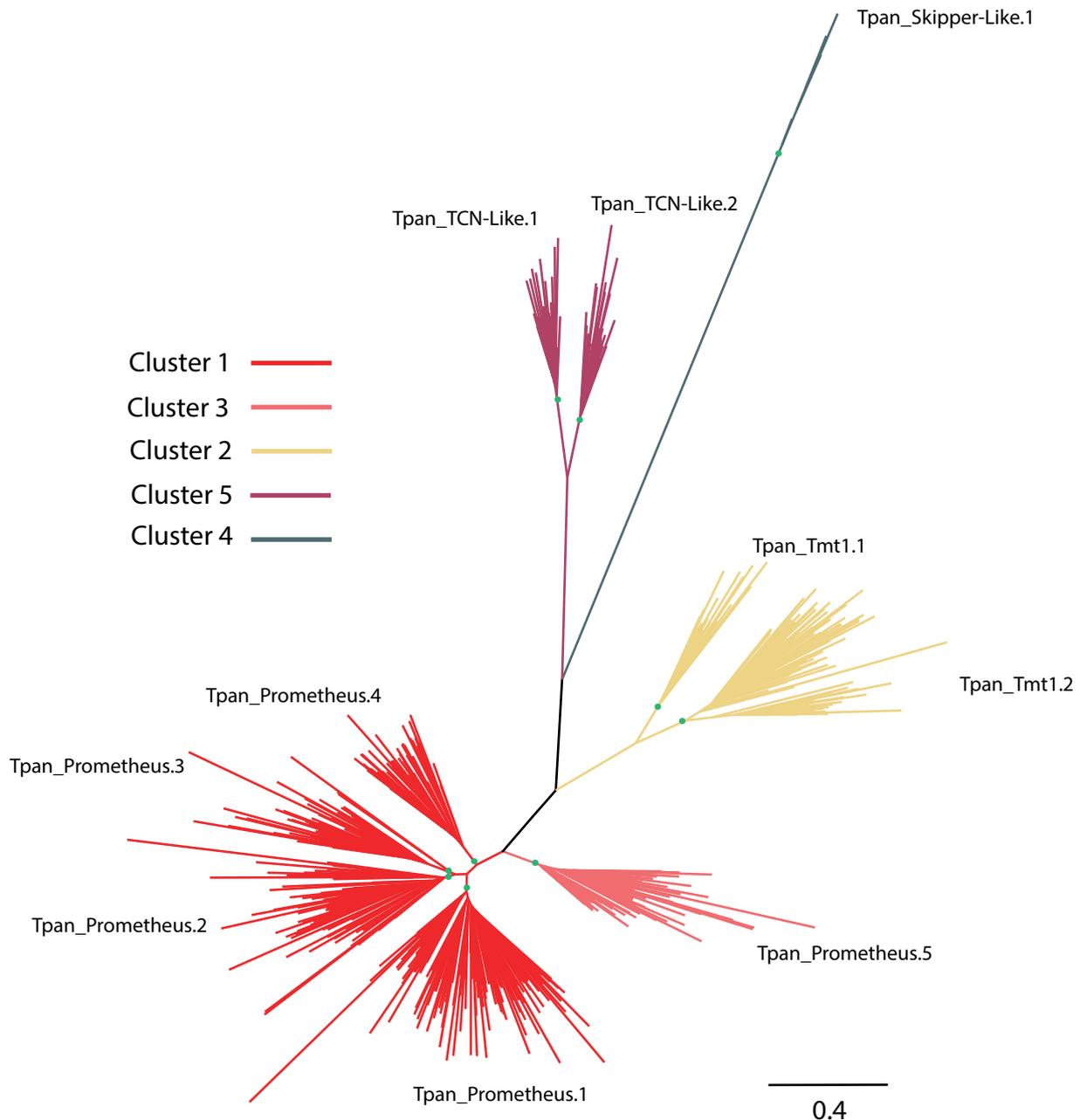
Supplemental Figure S5: Length in kilobases of significantly enriched (TE hotspots) and significantly depleted (TE coldspot) genomic regions in the *T. panzihuanense* genome.



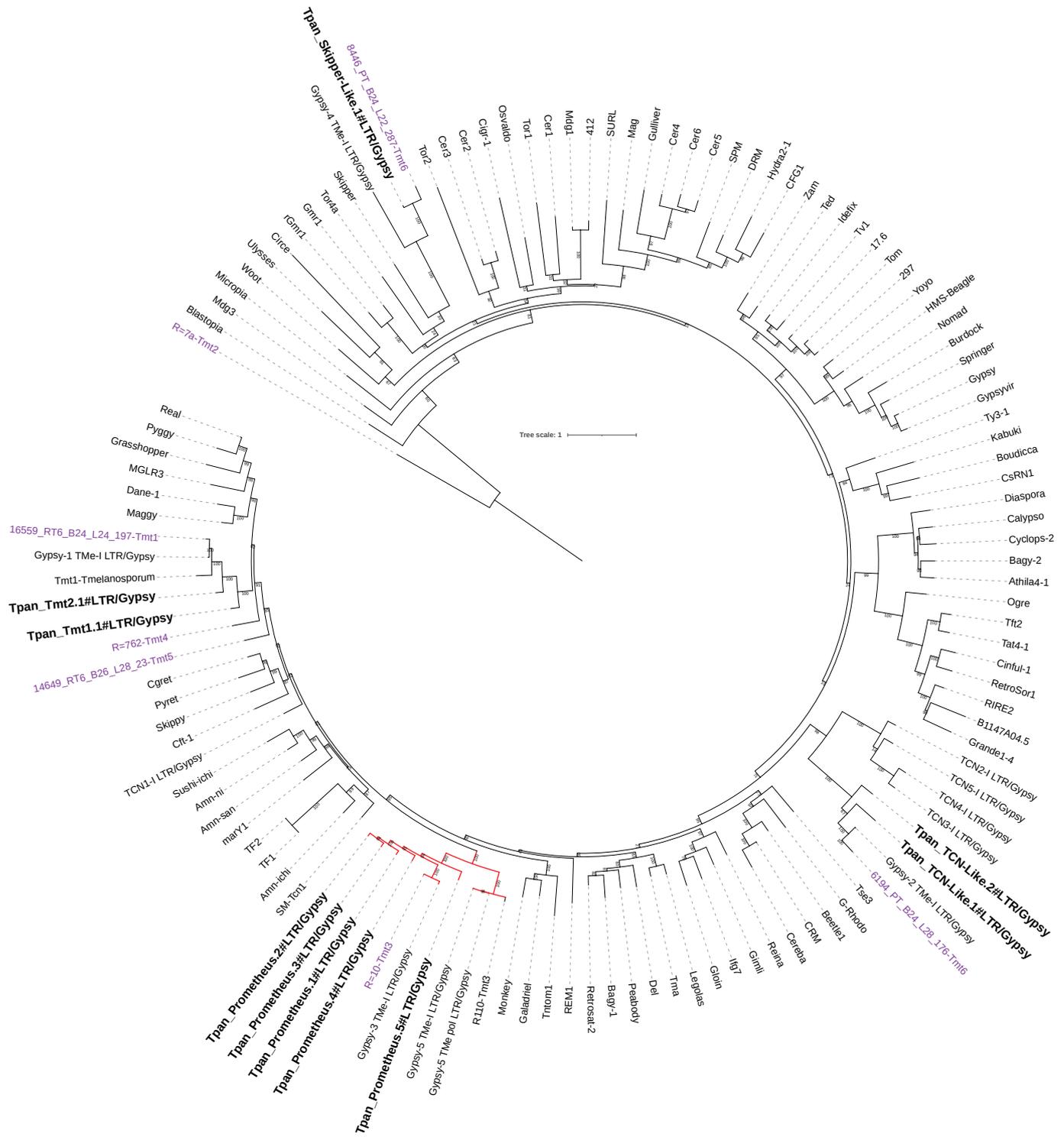
Supplemental Figure S6: Percentage of genomic coverage of the four most abundant transposon groups (DNA, LINEs, LTRs and Unknown) across TE coldspots and TE hotspots.



Supplemental Figure S7: (A) Distance distribution from the closest gene for the 5 main transposon groups identified in the *T. panzihuanense* genome. **(B)** Correlation between transposon length (y axis) and distance from the closest gene (x axis). **(C)** Length distribution of the 5 main transposon groups. In all plots Short Interspersed Nuclear Elements (SINEs) were excluded due to the low number of identified insertions.

A**B**

Supplemental Figure S8: Supplementary Figure S6: Identification of the main long terminal repeats (LTR) Gypsy lineages in the *T. panzihuanense* genome. (A) Communities identified through network analyses (using the greedy_modularity_communities algorithm in NetworkX) based on homologous relationships between reverse transcriptase (RT) nucleotide fragments mined from the *T. panzihuanense* genome. Each cluster is labelled with its name, and the number of sequences in the respective cluster is shown in parentheses. (B) Phylogenetic relationships between representative RT protein segments. Colours correspond to those in panel A and represent clusters identified through network analyses. Families were identified based on cluster results, with some of them further split to highlight highly diverging, well-supported branches, represented by green circles (ultrafast bootstrap value ≥ 75).



Lorem ipsum

Supplemental Figure S9: Phylogenetic classification of *T. panzhihuanense* Gypsy families, highlighted in bold, based on reverse transcriptase domains and reference Gypsy sequences. Red branches represent the Prometheus clade. Sequences highlighted in violet (16559_RT6_B24_L24_197-Tmt1; R=762-Tmt4; 14649_RT6_B26_L28_23-Tmt5; R=10-Tmt3; 6194_P-T_B24_L28_176-Tmt6; 8446_PT_B24_L22_287-Tmt6) denote *T. melanospurum* Gypsy clades identified in Payen et al., (2016). Note that *T. melanospurum* Tmt3 clade corresponds to Prometheus elements, Tmt6 was found to be paraphyletic representing both TCN-Like and Skipper-like clades whereas Tmt4, Tmt5 and Tmt2 appear to be absent from *T. panzhihuanense*. Numbers at each node represent ultra-fast bootstrap values.

Tpan_Prometheus.1



Tpan_Prometheus.2



Tpan_Prometheus.3



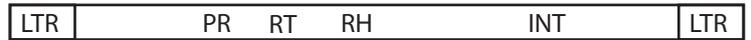
Tpan_Prometheus.4



Tpan_Prometheus.5



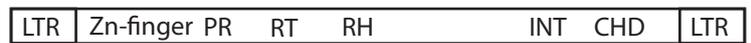
Tpan_Skipper-Like.1



Tpan_TCN-Like.1



Tpan_TCN-Like.2



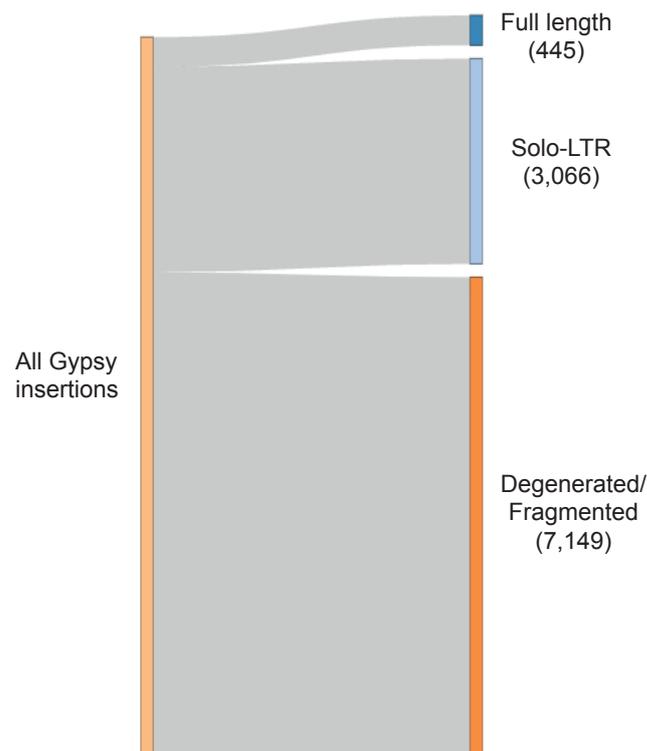
Tpan_Tmt1.1



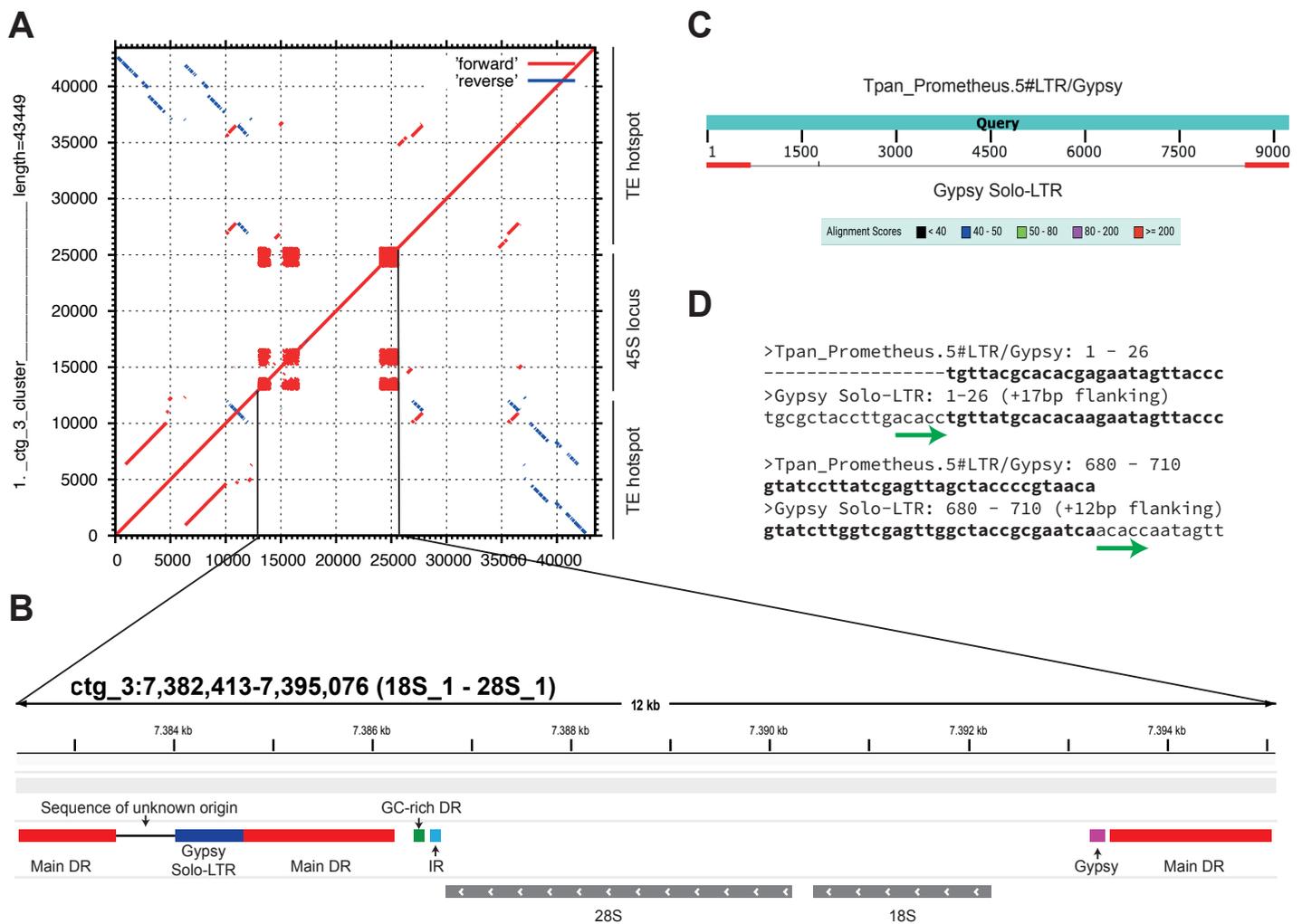
Tpan_Tmt1.2



Supplemental Figure S10: Domain structure of reconstructed Gypsy families. For Tpan_Tmt1.1 we could not reconstruct a full length copy which included long terminal repeats segments. Zn-finger = RING zinc finger; zf-CCHC = zinc knuckle binding motif; PR = proteinase; RT = Reverse transcriptase; RH = Ribonuclease H; dUTPase = Trimeric dUTP diphosphatases; INT = Integrase; CHD = Chromodomain; LTR = Long terminal repeats.



Supplemental Figure S11: Number of full length, solo-LTR and degenerated/fragmented Gypsy insertions identified in the *T. panzhihuanense* genome. Full length copies were defined as insertion covering the 90% of their parental consensus sequence. solo-LTR when both the insertion and on the two LTR segments of the parental consensus reciprocally aligned for at least 90%. All other instances were considered as degenerated/fragmented.



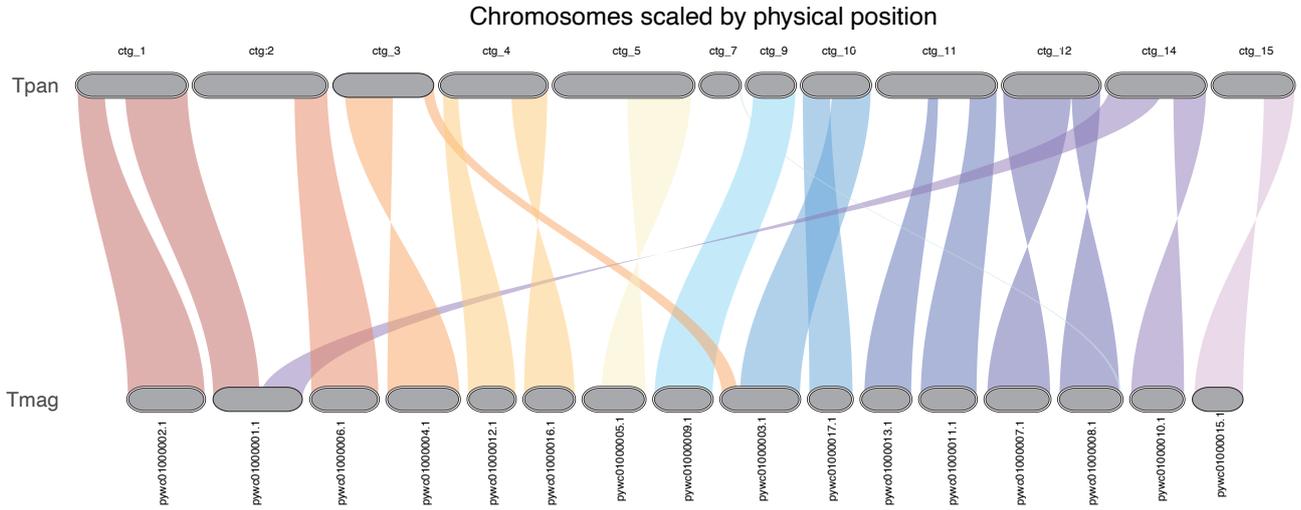
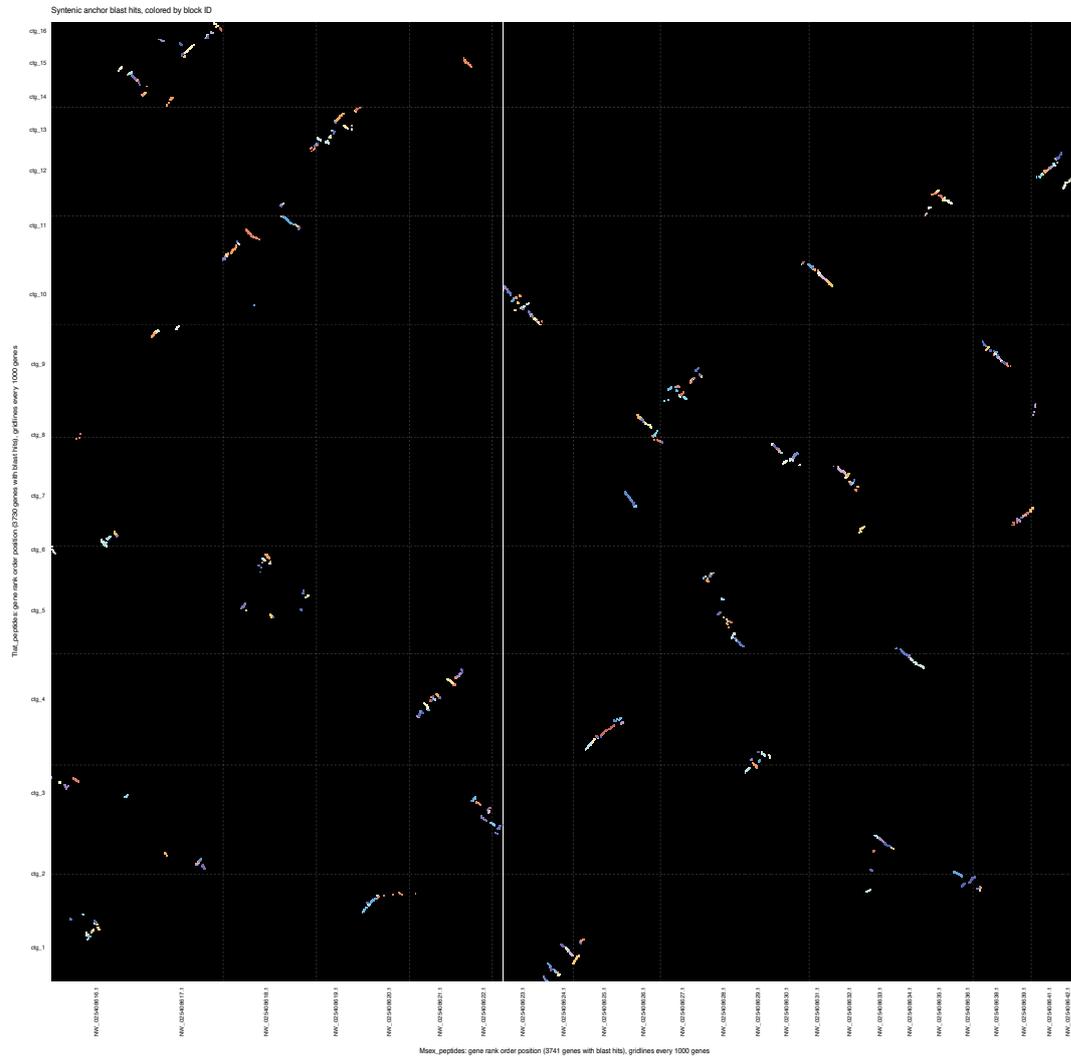
Supplemental Figure S12: Structure of the isolated 45S rDNA locus identified on ctg_3 and included in a complex genomic region corresponding to a TE-hotspot. **(A)** Self-alignment dot plot of the genomic region that includes the rDNA locus. Red lines represent alignments in the same orientation, while blue lines indicate alignments in the opposite orientation. **(B)** Detailed view of the structure and components of the rDNA locus. DR = direct repeats; IR = inverted repeats. **(C)** BLASTn alignment of the LTR fragment identified in the longer IGS region with its parental consensus sequence. **(D)** A more detailed view of the alignment, including the flanking regions of the solo-LTR insertion. Green arrows highlight a 5 bp target site duplication.

Tree scale: 1

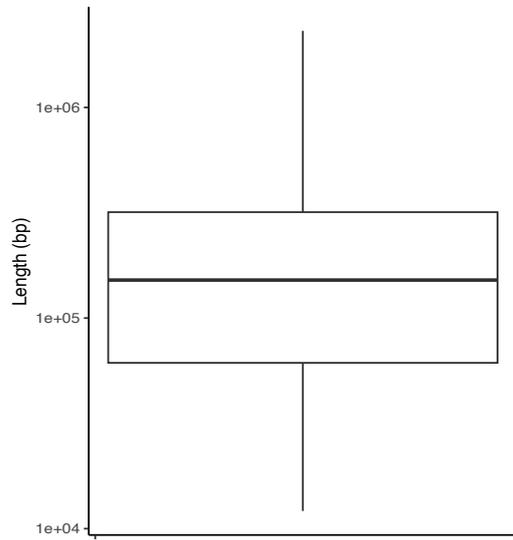
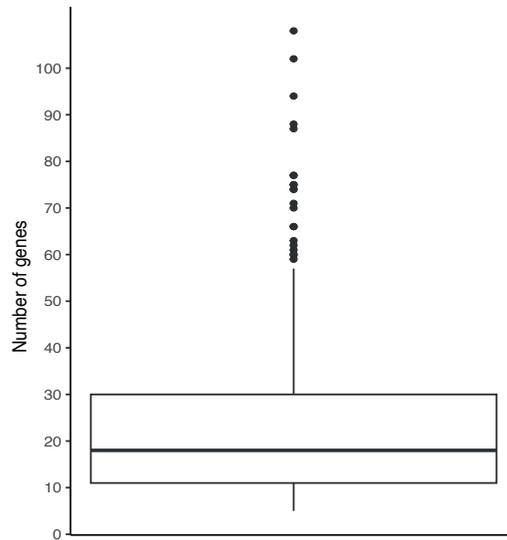
- Skipper-like
- TCN-like
- Prometheus
- Tmt1



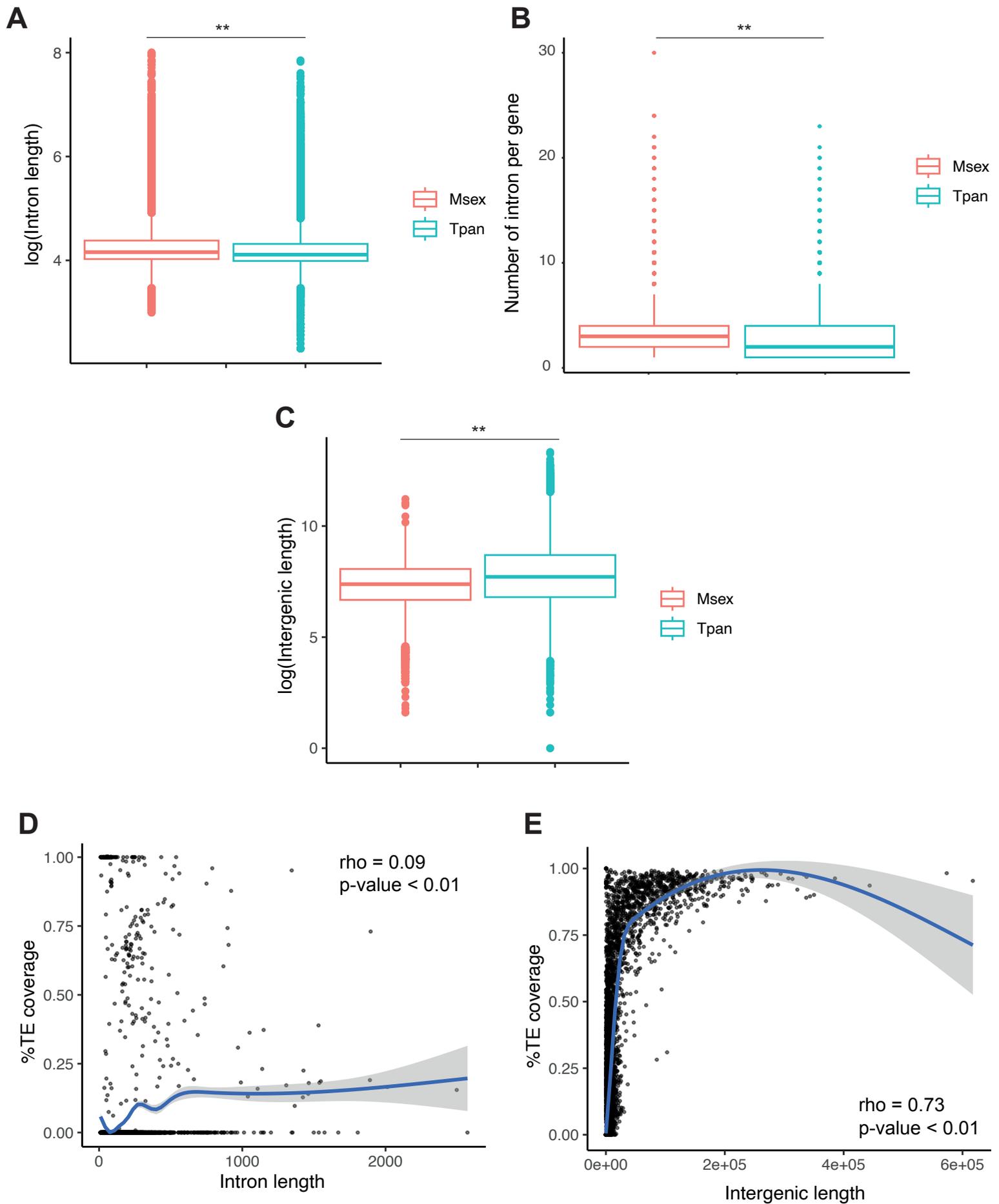
Supplemental Figure S13: Classification of Pezizales LTR Gypsy elements (in bold) based on phylogenetic relationships with known elements estimated on their reverse transcriptase protein domains. Different colours highlight the four clades identified also in *T. panzhihuanense*. Numbers at each node represent ultrafast bootstrap values. Each Pezizales element has a prefix reporting the source genome: Pcon = *Pyronema confluens*; Gesc = *Gyromitra esculenta*; Mimp = *Morchella importuna*; Msex = *Morchella sextelata*; Vcon = *Verpa coninca*; Cven = *Choiromyces venosus*; Tpan = *Tuber panzhihuanense*; Tbor = *Tuber borchii*; Tmag = *Tuber magnatum*; Taes = *Tuber aestivum*; Tbru = *Tuber brumale*; Tmela = *Tuber melanosporum*; Tind = *Tuber indicum*.

A**B**

Supplemental Figure S14: (A) Collinearity between *T. panzhihuanense* and the sixteen longest scaffolds of *T. magnatum* and **(B)** between *T. panzhihuanense* (y axis) and *M. sextelata* (x axis) genomes.

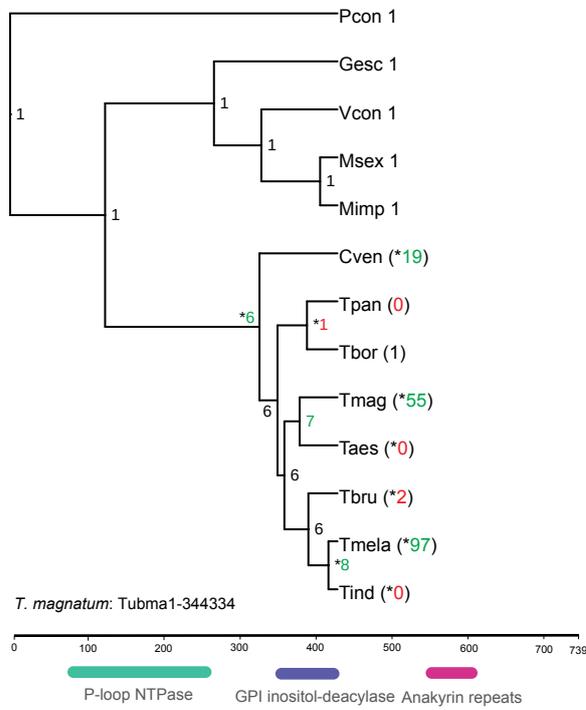
A**B**

Supplemental Figure S15: (A) Length of syntenic blocks with *M. sextelata* in the *T. panzhihua-nense* genome and **(B)** number of genes contained within these regions.

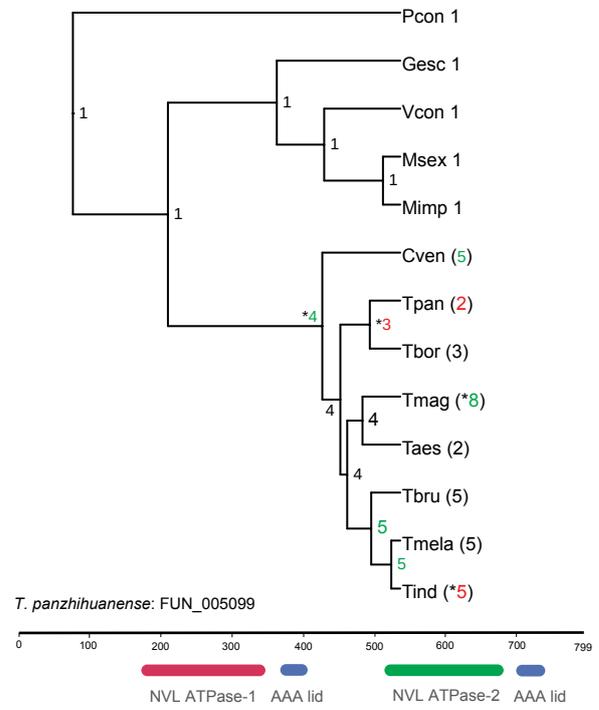


Supplemental Figure S16: Comparison between **(A)** intron length, **(B)** mean number of introns per gene and **(C)** intergenic genomic regions length between *T. panzhihuanense* (Tpan) and *M. sextelata* (Msex). **(D)** and **(E)** correlations between percentage of repeat coverage and intron and intergenic genomic regions length in *T. panzhihuanense*, respectively. ρ = Spearman's rank correlation coefficient.

OG0000009: Ankyrin repeat-containing



OG0000138: AAA ATPase



Supplemental Figure S17: CAFE results for the ECM-induced gene families OG0000009 and OG0000138 encoding for Ankyrin repeat containing proteins and AAA ATPase, respectively. These families together with those represented in Fig. 6 were found significantly expanded in the stem branch of Tuberaceae. For each node, the inferred and observed gene family counts are reported for internal and terminal branches, respectively. Green colours highlight expansion events, red colours indicate contractions, and significant changes are marked with an asterisk. The domain annotation was obtained with InterProScan on a representative protein from *T. panzhihuanense* or *T. magnatum* when the gene family was absent from *T. panzhihuanense*. Pcon = *Pyronema confluens*; Gesc = *Gyromitra esculenta*; Mimp = *Morchella importuna*; Msex = *Morchella sextelata*; Vcon = *Verpa coninca*; Cven = *Choiromyces venosus*; Tpan = *Tuber panzhihuanense*; Tbor = *Tuber borchii*; Tmag = *Tuber magnatum*; Taes = *Tuber aestivum*; Tbru = *Tuber brumale*; Tmela = *Tuber melanosporum*; Tind = *Tuber indicum*.