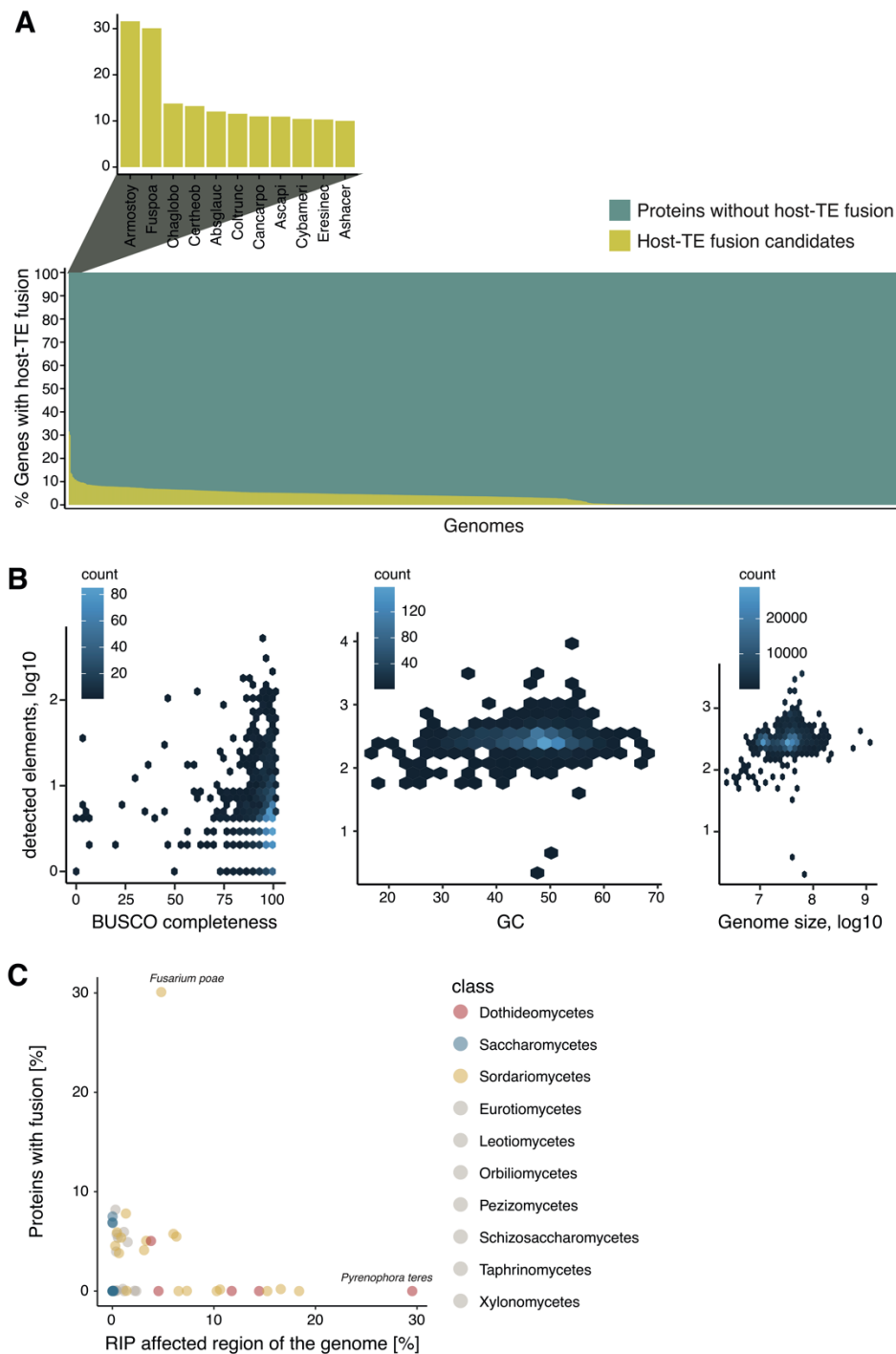
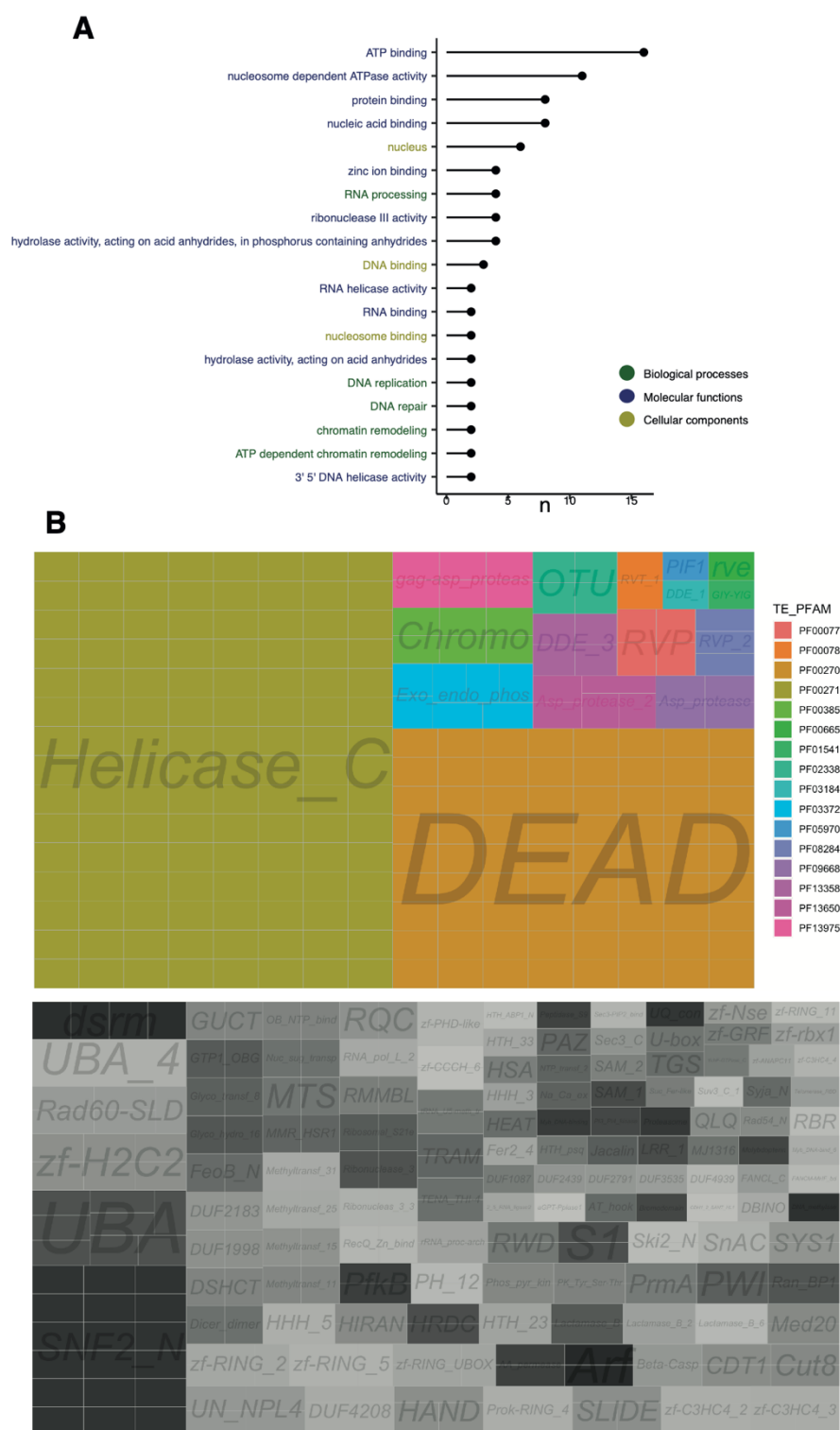


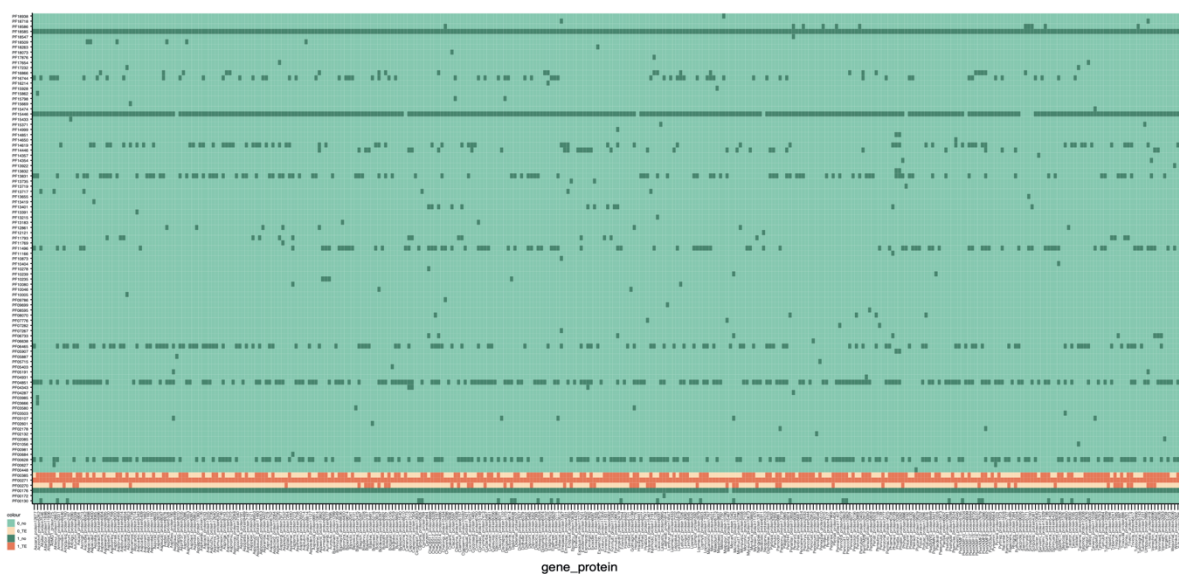
# Supplementary Figures



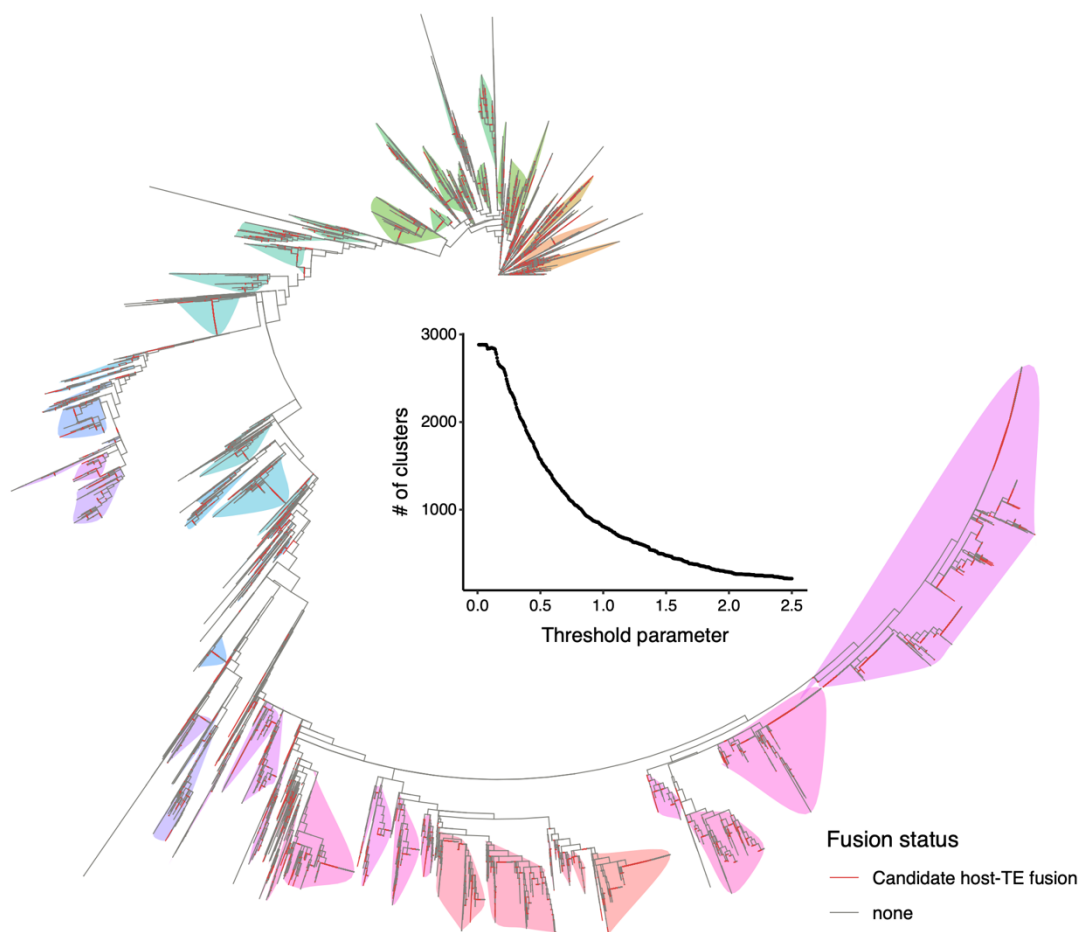
**Supplementary Figure S1: Number of host-TE detection and genome characteristics:** (A) Percentage of proteins assessed to be host-TE fusion candidates per genome. Genomes where more than 10% of all proteins are host-TE fusion candidates are represented in the subplot. (B) BUSCO completeness score, genome-wide GC content, genome size. (C) RIP strength and the detection of host-TE fusions in a subset of genomes (RIP strength is the % of the genome covered by RIP affected regions; data from van Wyk et al, 2021).



**Supplementary Figure S2: Distribution of domains in the filtered set of candidates:** (A) Counts of non-TE derived domain gene ontology terms. Gene ontology with a single occurrence can be found in Supplementary Table S5. (B) After filtering for candidates that are present in at least 20 species, with at least 5 species being closely related. Distribution for the TE-derived domain and the non-TE derived domain.



**Supplementary Figure S3: Additional domains in the host-TE fusion candidate PF00271\_Can16:** Dark orange indicates presence of a TE domain in a specific gene, bright orange indicates absence of a TE domain. Dark green indicates presence of a non-TE derived domain, bright green indicates its absence.



**Supplementary Figure S4: Clustering of Helicase\_C domains (PF00271) with and without host-TE fusion candidates:** Phylogenetic tree of Helicase\_C domains identified across genomes. Clades were retained with a threshold of 2 (see subplot). Background colors indicate clusters containing more than 30 copies. Red tips indicate occurrences of PF00271 as a putative host-TE fusion.

## Supplementary File

**Supplementary File F1: Phylogenetic tree:** phylogenetic tree of the fungal kingdom, based on 100 genes and in nexus format. The tree contains the following metadata: label (a short species ID), organism, species taxonomy ID, assembly name, assembly accession number, taxonomy ID, link to the genome in GenBank, phylum, class, order, family, genus, protein file, genome file, cds file, yeast (marked with 1 for yeast-like growing species), sequences counts, genome size (number of bases), average length, median length, maximum length, minimum length, N50, L50, BUSCO completeness score [%], BUSCO single copy genes [%], BUSCO fragmented genes [%], BUSCO missing genes [%], number of genes in the BUSCO code, BUSCO code, number of proteins detected, predicted lifestyle with CATASrophy, phylum3 includes the differentiation between Saccharomycotina and the other classes of the phylum Ascomycota