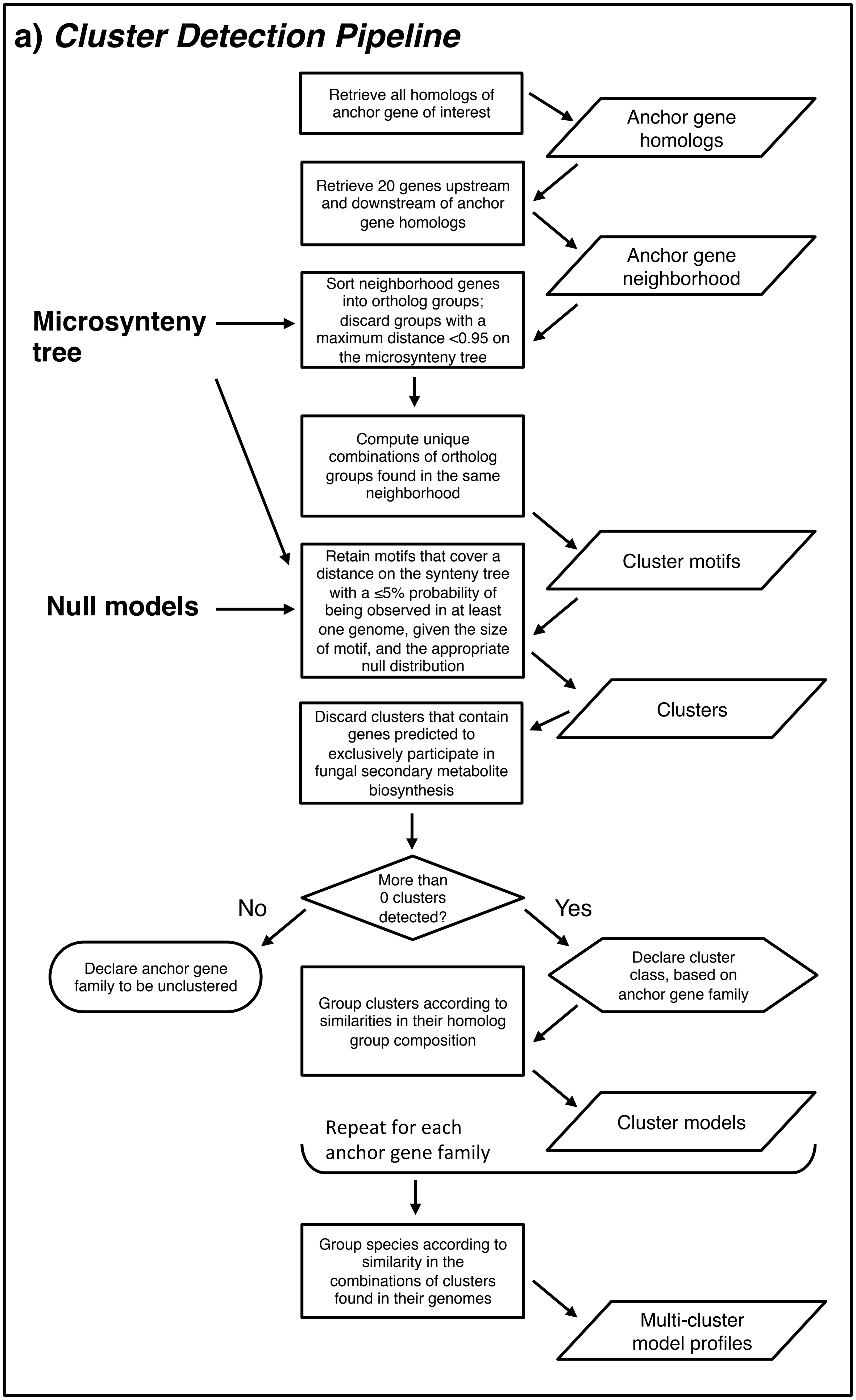
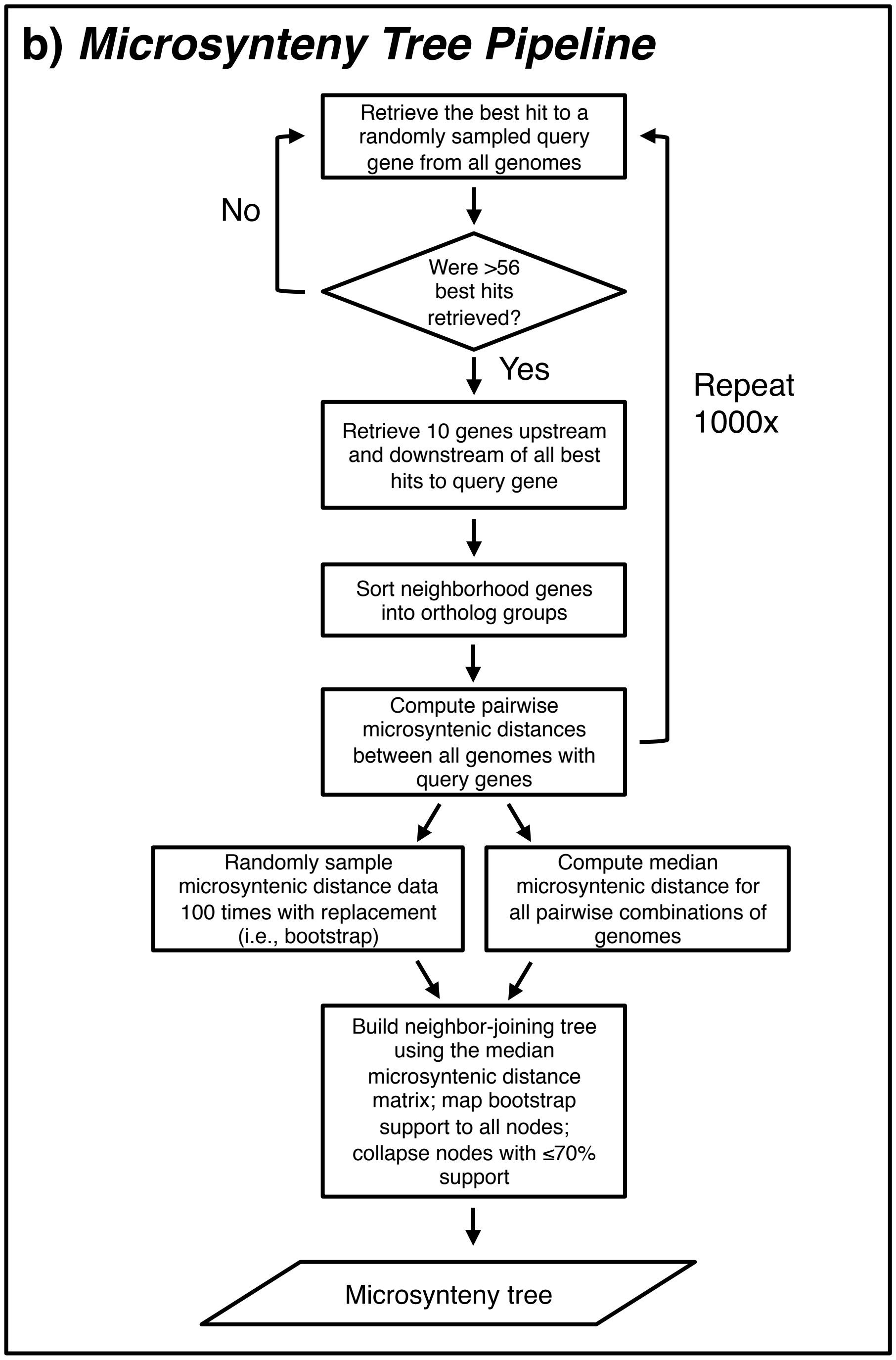
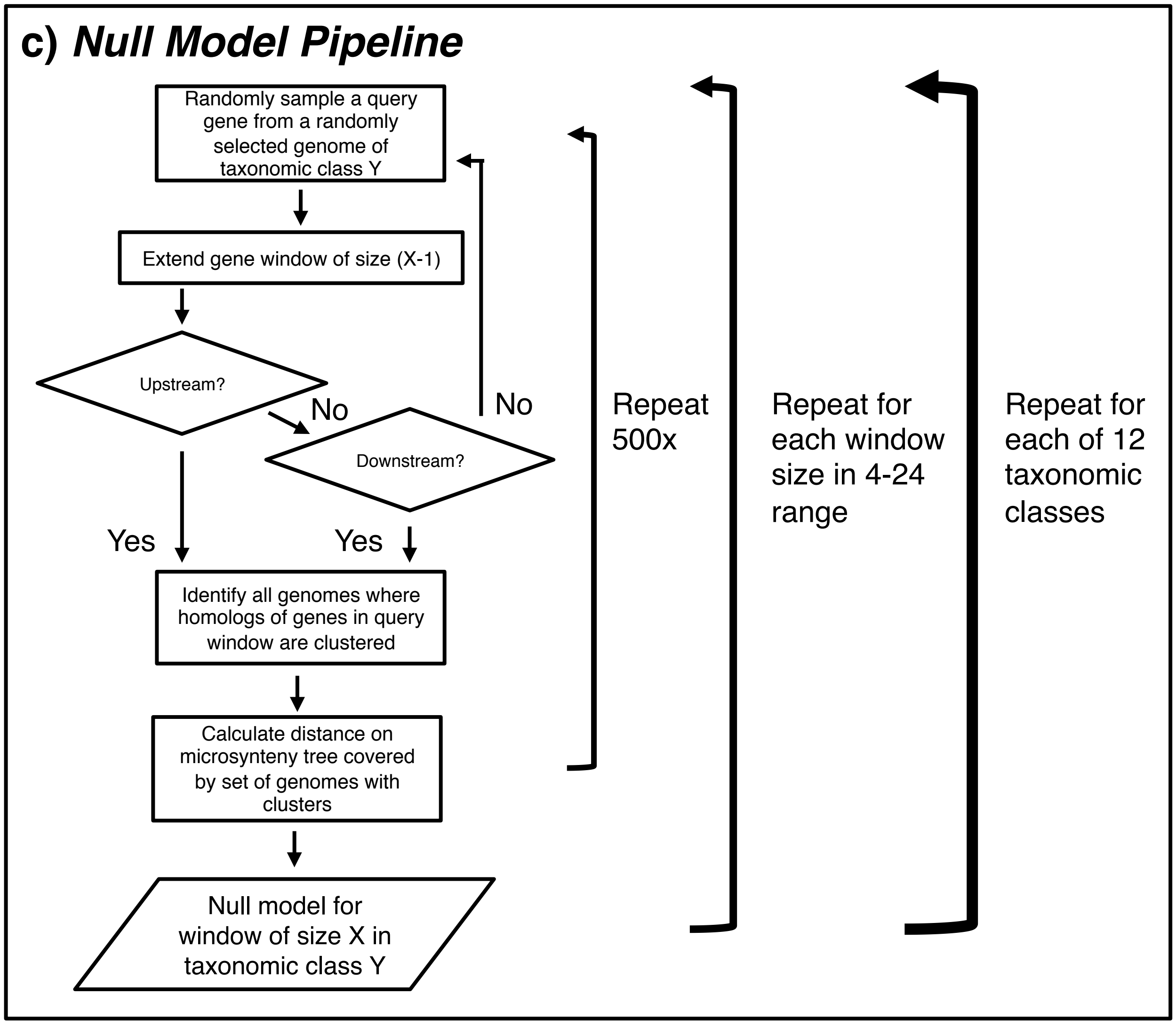

a) Cluster Detection Pipeline
Retrieve all homologs of anchor gene of interest
Anchor gene homologs
Retrieve 20 genes upstream and downstream of anchor gene homologs
Anchor gene neighborhood
Microsynteny tree
Sort neighborhood genes into ortholog groups; discard groups with a maximum distance <0.95 on the microsynteny tree
Compute unique combinations of ortholog groups found in the same neighborhood
Cluster motifs
Null models
Retain motifs that cover a distance on the synteny tree with a ≤5% probability of being observed in at least one genome, given the size of motif, and the appropriate null distribution
Clusters
Discard clusters that contain genes predicted to exclusively participate in fungal secondary metabolite biosynthesis
More than 0 clusters detected?
No
Yes
Declare anchor gene family to be unclustered
Declare cluster class, based on anchor gene family
Group clusters according to similarities in their homolog group composition
Cluster models
Repeat for each anchor gene family
Group species according to similarity in the combinations of clusters found in their genomes
Multi-cluster model profiles



b) *Microsynteny Tree Pipeline*

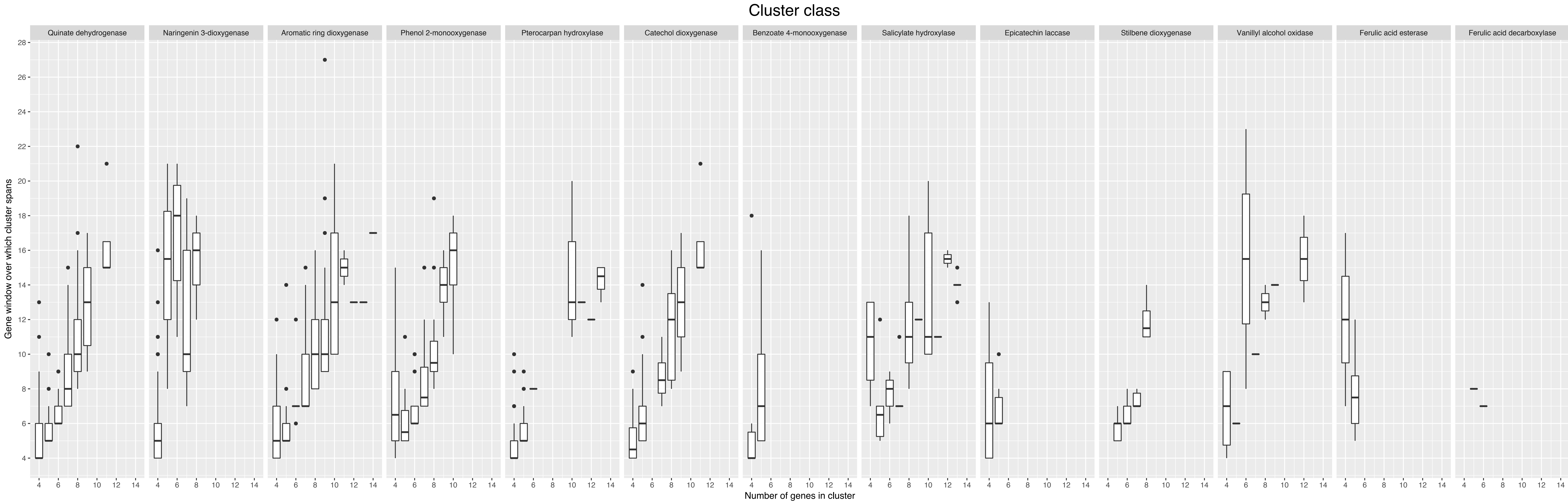

c) Null Model Pipeline
Randomly sample a query gene from a randomly selected genome of taxonomic class Y
Extend gene window of size (X-1)
Upstream?
No
Downstream?
No
Yes
Yes
Identify all genomes where homologs of genes in query window are clustered
Calculate distance on microsynteny tree covered by set of genomes with clusters
Null model for window of size X in taxonomic class Y
Repeat 500x
Repeat for each window size in 4-24 range
Repeat for each of 12 taxonomic classes

d) Generalized contingency table for one tail Fisher's exact tests of enrichment

	Number of (genomes/ species) associated WITH ecological lifestyle	Number of (genomes/ species) NOT associated with ecological lifestyle
Number of (genomes/ species) WITH (cluster/ cluster model/multi-cluster model profile)	n11	n12
Number of (genomes/ species) WITHOUT (cluster/cluster model/ multi-cluster model profile)	n21	n22

Supplementary Figure 1: *Schematics of all computational pipelines and enrichment tests*. a) Description of the main cluster detection algorithm. b) Description of microsynteny tree pipeline. c) Description of null model pipeline. d) Generalized representation of the contingency tables used for all enrichment analyses.

Supplementary Figure 2: *The distribution of candidate phenylpropanoid-degrading gene clusters in fungi*. The data shown in this figure are identical to those presented in Figure 1, except that the full species names of published fungal genomes are displayed on the microsynteny tree, along with their taxonomic order and class. Branch length distances (in black) and bootstrap support values (in red) are additionally displayed above their corresponding branches.



Supplementary Figure 3: *Distributions of gene window size for all gene cluster sizes.* The distributions of gene window size (i.e., the number of genes in the genomic region over which the gene cluster spans) for each gene cluster size in each cluster class are depicted as Tukey-style box-and-whisker plots, where the 25th, 50th, and 75th percentiles are represented by the lower, middle and uppoer hinges, respectively. Lower whiskers extend to the 25th percentile - 1.5 * IQR, and upper whiskers extend to the 75th percentile + 1.5 * IQR, where IQR is the interquartile range (defined as the distance between the 75th and 25th percentiles). Outliers (as determined by Tukey's method) are represented as black points that lie beyond the whisker boundaries. Not all gene cluster sizes are present in all cluster classes.

Supplementary Figure 4: *Associations between cluster model presence and fungal lifestyle.* Odds ratios representing the strength of the association between cluster model presence and ecological lifestyle are shown for all cluster models in all cluster classes, for each of 6 fungal ecological lifestyles, using data from Pezizomycotina species. Dark grey bars indicate enrichment below a significance level of 0.05, while black outlines indicate enrichment below a significance level of 0.01. Error bars indicate the 95% confidence interval (CI) for each odds ratio measurement. CIs of 0 are not shown.

Supplementary Figure 5: *The distribution of candidate phenylpropanoid-degrading gene cluster models in fungi.* Displayed on the left is the microsynteny species tree containing all 529 genomes used in this study. Taxonomic order and class names are indicated below species names. Branch length distances (in black) and bootstrap support values (in red) are displayed above their corresponding branches. To the right are heatmaps displaying the ecological lifestyles associated with each genome (color coded by lifestyle) and the presence/absence of all 56 unique cluster types (cluster models) detected in each genome. The 6 largest taxonomic classes are additionally indicated to the right of the heatmaps.