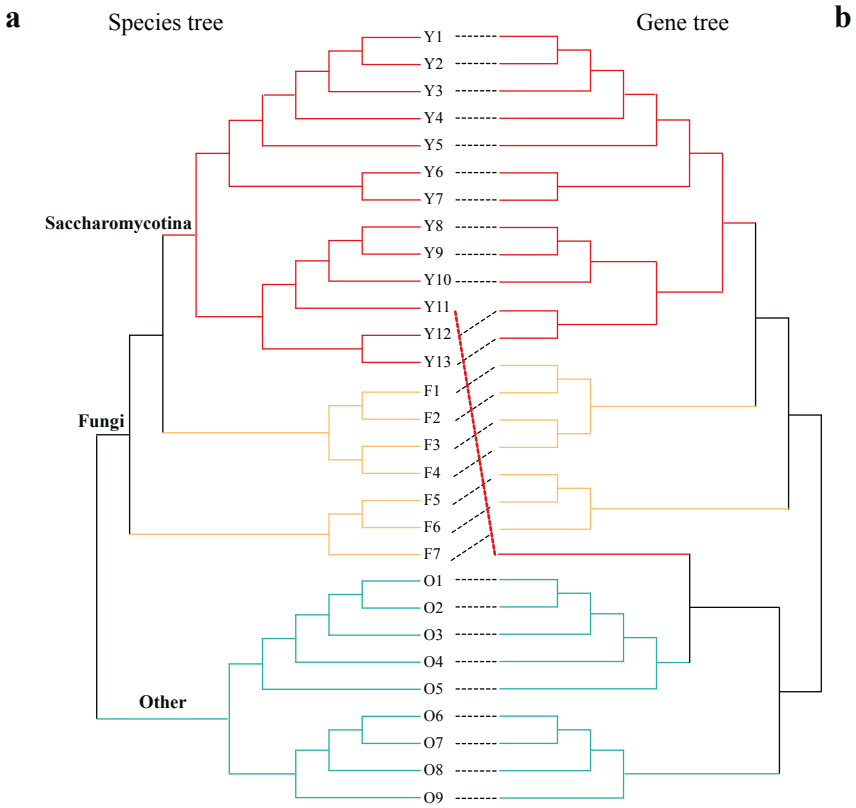
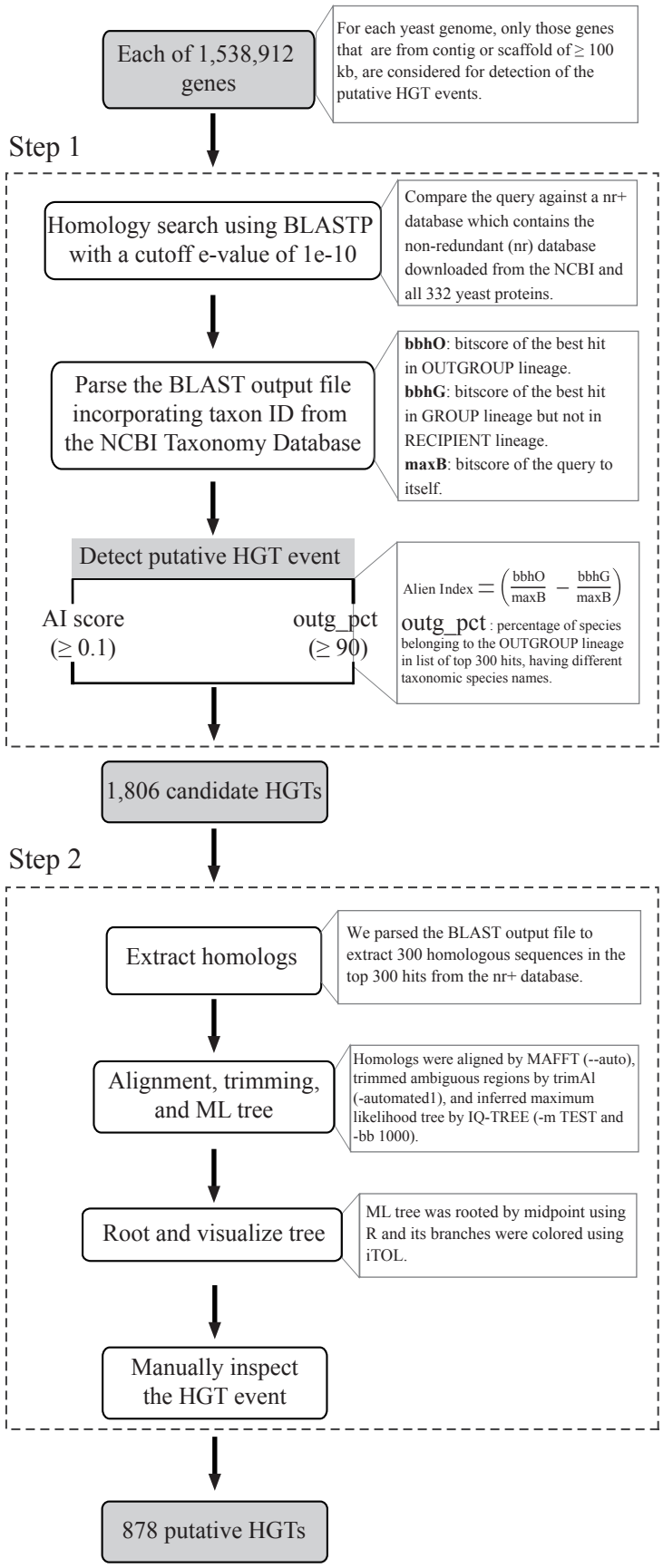
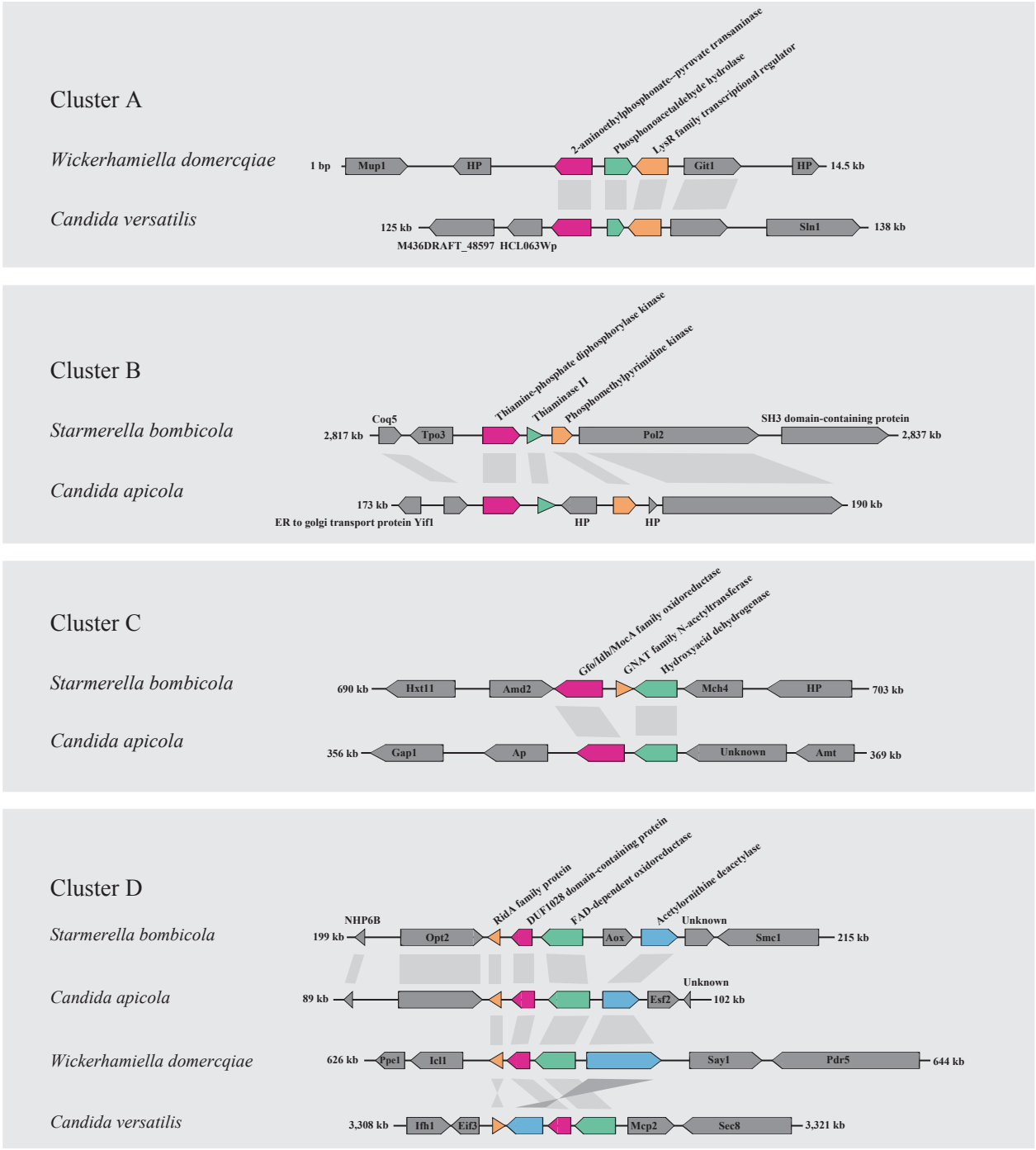


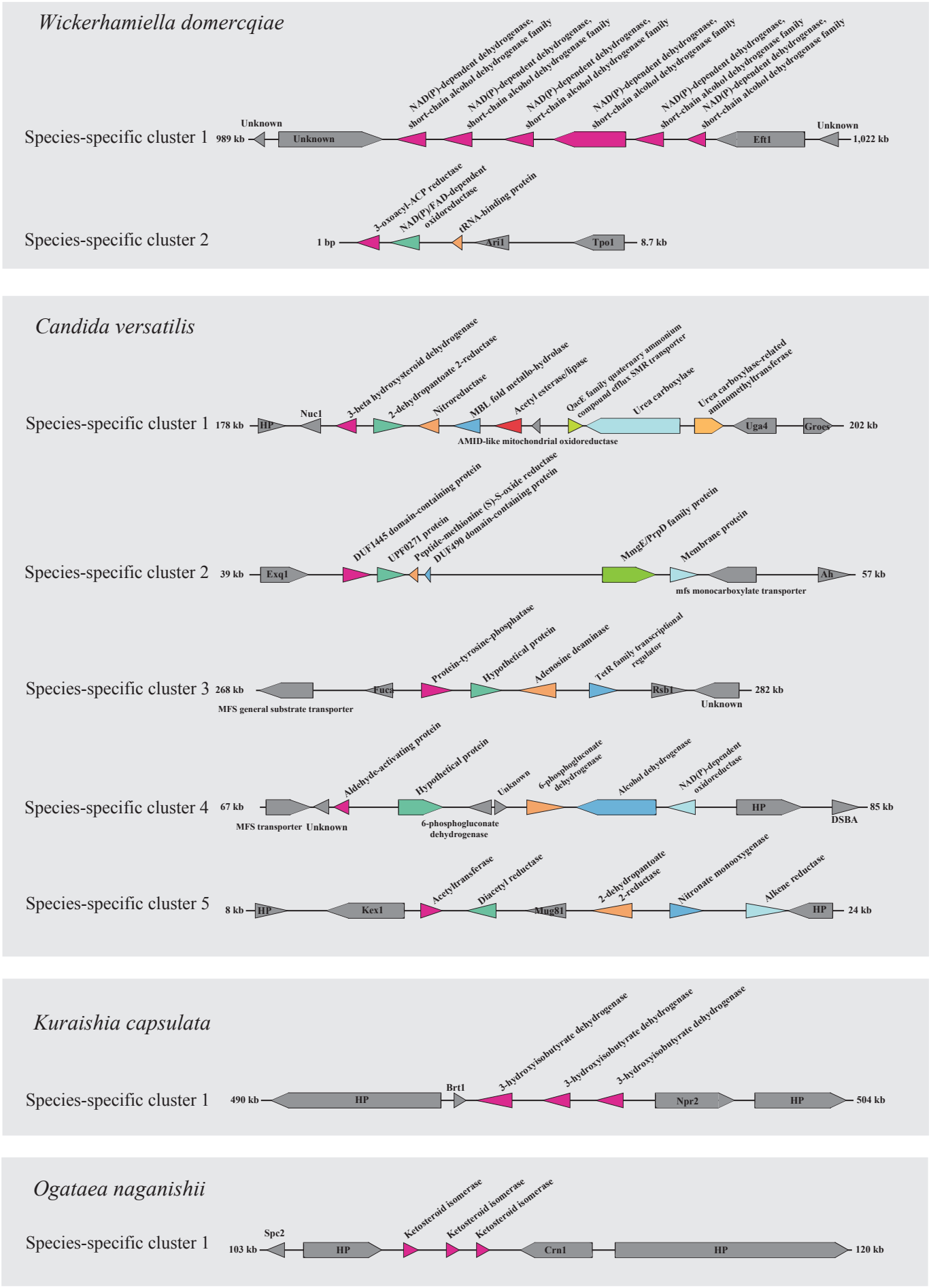
Additional Figure 27. Depiction of a hypothetical horizontal gene transfer (HGT) event between the Saccharomycotina yeasts and a non-fungal species and how gene tree-species tree reconciliation can be used to identify it. (a) Schematic species phylogeny of Saccharomycotina yeasts (red branches and taxon names starting with the letter Y), other fungi (yellow branches and taxon names starting with the letter F), and non-fungal taxa (green branches and taxon names starting with the letter O (other)). (b) Schematic phylogeny for a single gene for the same set of taxa as in panel a. Note that taxon Y11 is nested within a clade comprised of non-fungal taxa and not, as expected on the basis of species relationships, within Saccharomycotina. This incongruence between the species tree and the gene tree can be used to identify Saccharomycotina genes that may have been acquired via HGT from non-fungal organisms.



Additional Figure 28. The workflow used for the identification of genes found in Saccharomycotina yeasts that were likely acquired by horizontal gene transfer (HGT) from non-fungal species. A detailed description of the analyses performed in each step of the workflow is provided in the "Detection of horizontal transfer (HGT)" section of the Supplementary Methods.



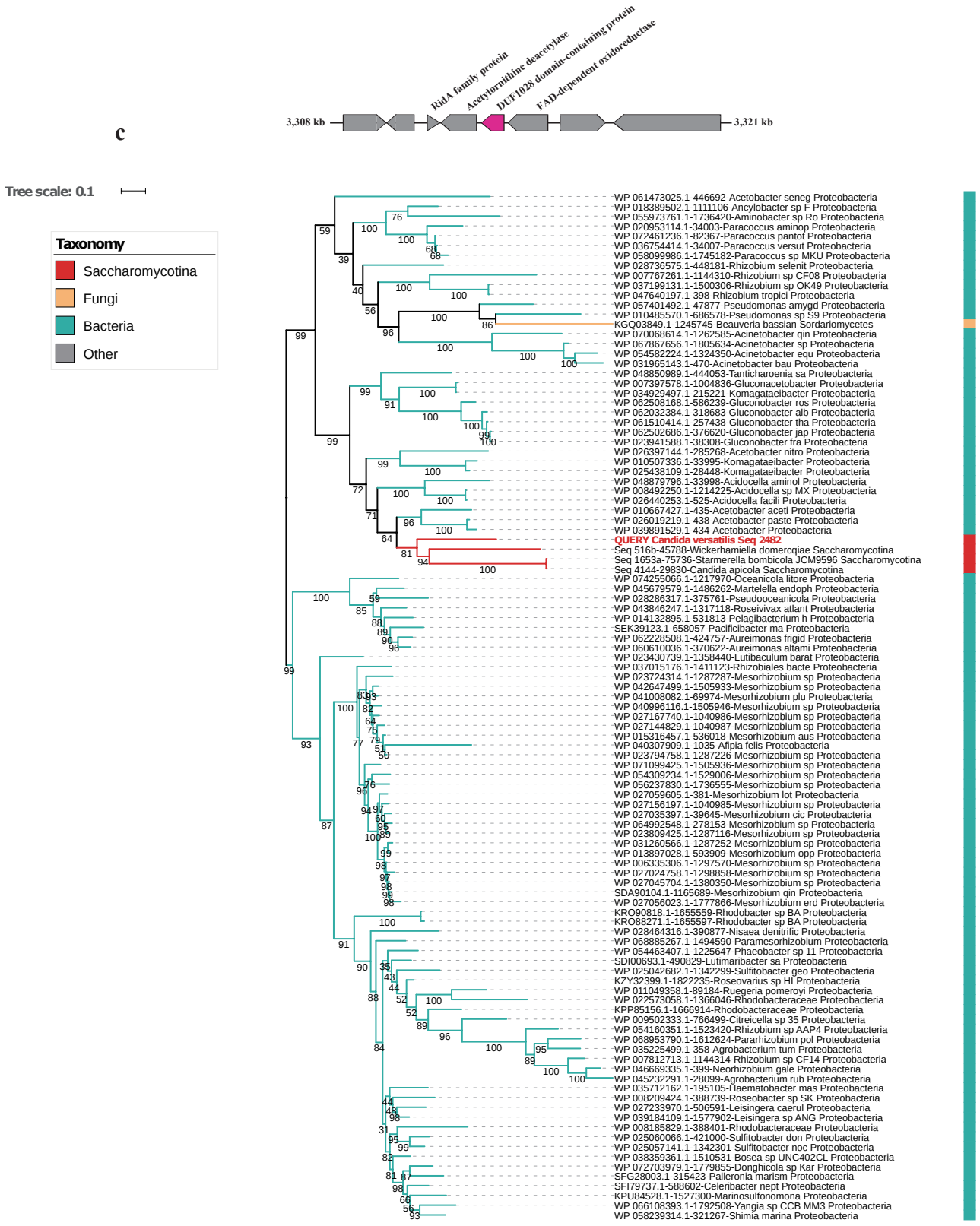
Additional Figure 29. Representative clusters of horizontally transferred genes found in two or more budding yeast species. Native (i.e., acquired via vertical descent) genes are shown as gray polygons and horizontally acquired genes are shown in all other colors. Horizontally acquired genes of the same color correspond to homologs. Detailed information of molecular function and putative donor organisms of horizontally acquired genes found as gene clusters are provided in Table S3, whereas the multiple sequence alignments and gene trees are provided in the Figshare depository. The names of encoded proteins or predicted enzymatic functions are shown for each gene. HP = hypothetical protein.



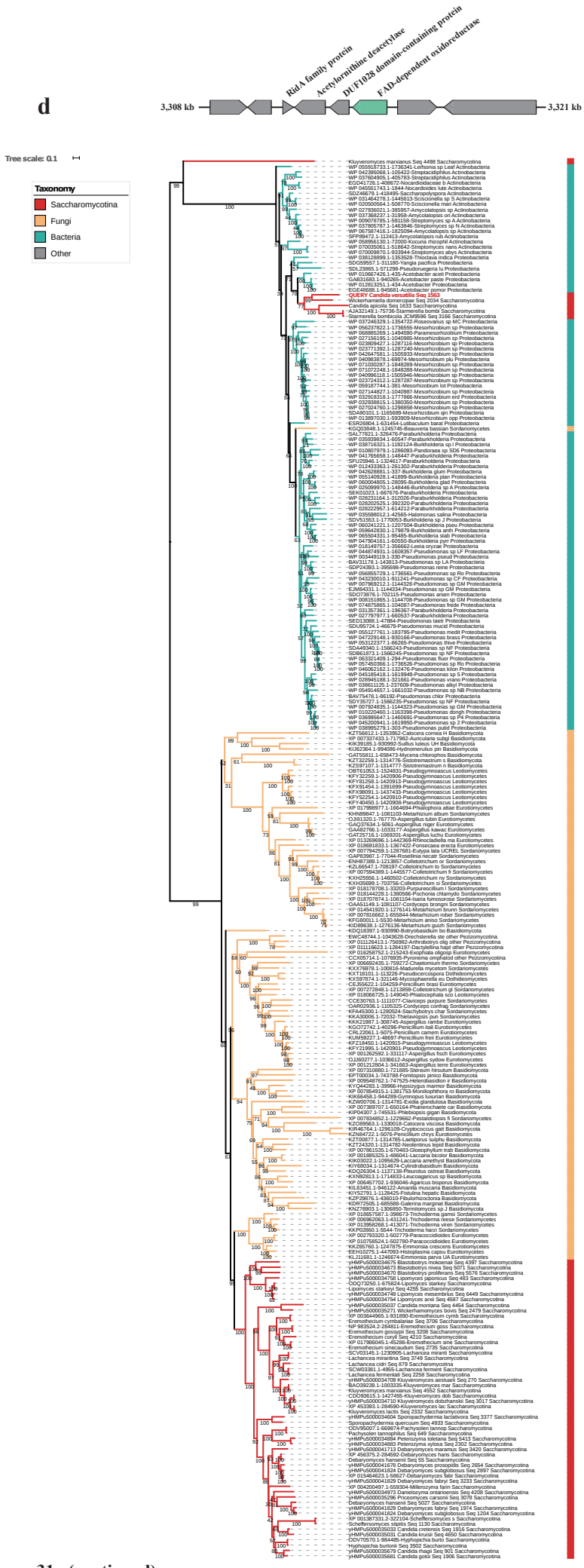
Additional Figure 30. Clusters of horizontally transferred genes found in a single budding yeast species. Native (i.e., acquired via vertical descent) genes are shown as gray polygons, and horizontally acquired genes are shown in all other colors. Horizontally acquired genes of the same color correspond to homologs. Detailed information of molecular function and putative donor organisms of horizontally acquired genes found as gene clusters are provided in Table S3, whereas the multiple sequence alignments and gene trees are provided in the Figshare depository. The names of encoded proteins or predicted enzymatic functions are shown for each gene. HP = hypothetical protein.

Additional Figure 31. Maximum likelihood trees of genes in an exemplar gene cluster that was horizontally acquired by several budding yeasts. The four different panels (a – d) correspond to the gene trees of the 4 genes in the cluster, which encode (a) RidA family protein; (b) Acetylornithine deacetylase; (c) DUF1028 domain-containing protein; and (d) FAD-dependent oxidoreductase. All gene sequences used as queries in the BLAST search are from the *Candida* (*Wickerhamiella*) *versatilis* genome (shown in red font in the gene trees), and their orthologs also can be found in *Candida apicola*, *Wickerhamiella domercqiae*, and *Starmerella bombicola* genomes. The gene used as a query in the BLAST search is shown in red. The tree was midpoint rooted, and the branch supports were evaluated with 1000 ultrafast bootstraps implemented in IQ-TREE. Detailed information of molecular function and putative donor organisms of horizontally acquired genes found as gene clusters are provided in Table S3, whereas the multiple sequence alignments and gene trees are provided in the Figshare depository.

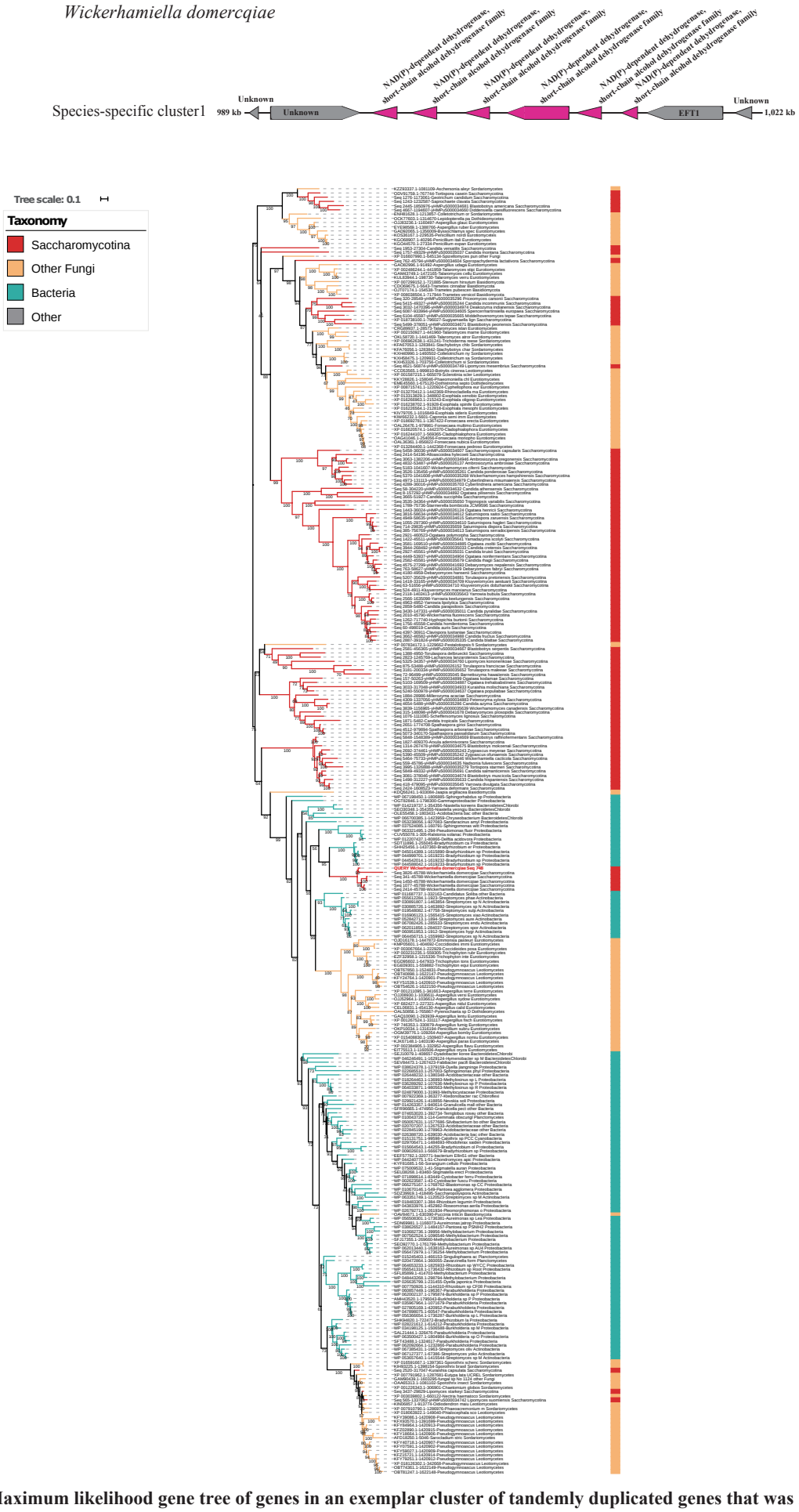
Additional Figure 31. (continued)



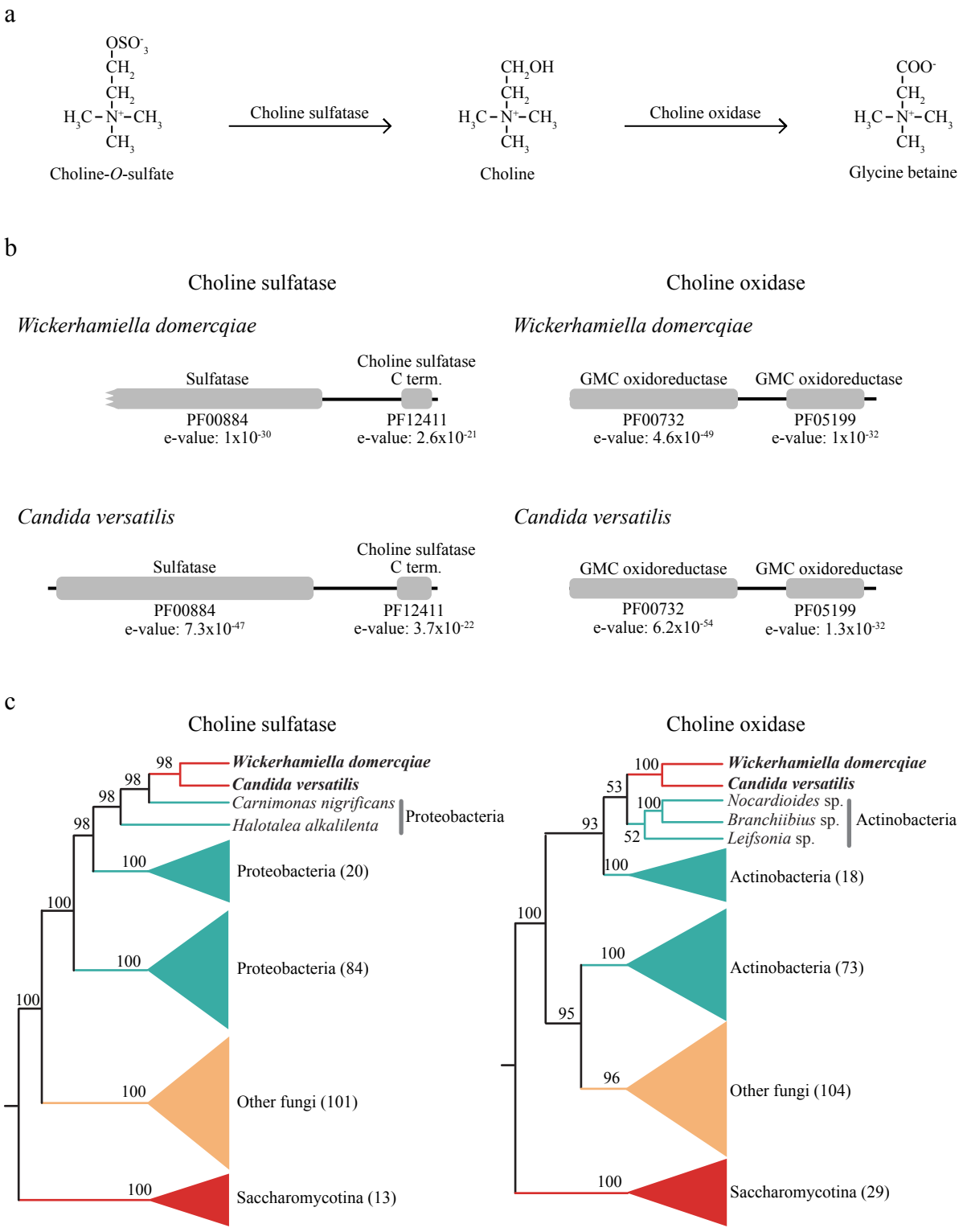
Additional Figure 31. (continued)



Additional Figure 31. (continued)

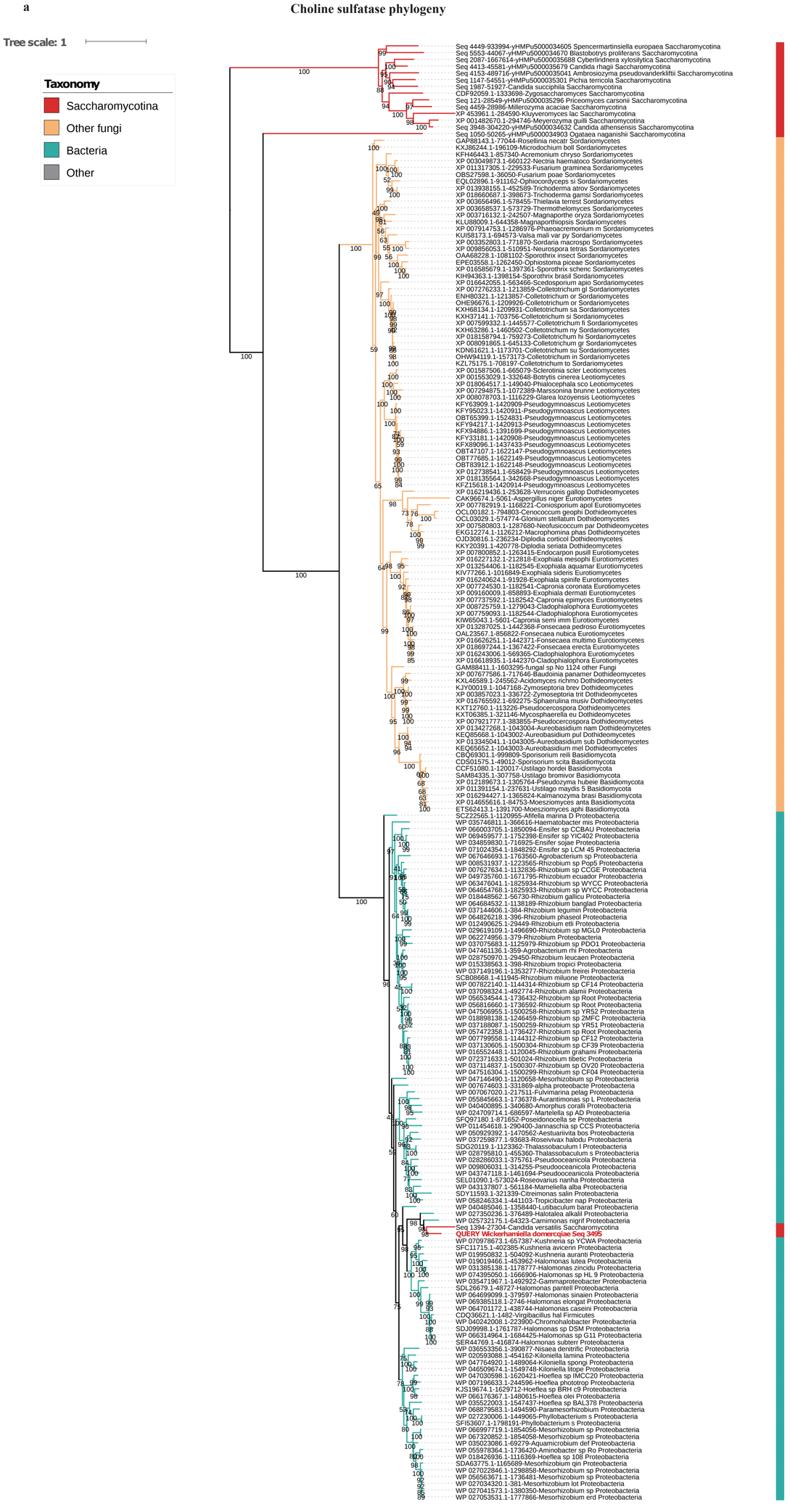


Additional Figure 32. Maximum likelihood gene tree of genes in an exemplar cluster of tandemly duplicated genes that was horizontally acquired by the budding yeast *Wickerhamiella domercqiae.* The gene used as a query in the BLAST search is shown in red. The other five duplicates of the gene form a clade in the gene tree. The tree was midpoint rooted and the branch supports were evaluated with 1000 ultrafast bootstraps implemented in IQ-TREE. Detailed information of molecular function and putative donor organisms of horizontally acquired genes found as gene clusters are provided in Table S3, whereas the multiple sequence alignments and gene trees are provided in the Figshare depository.



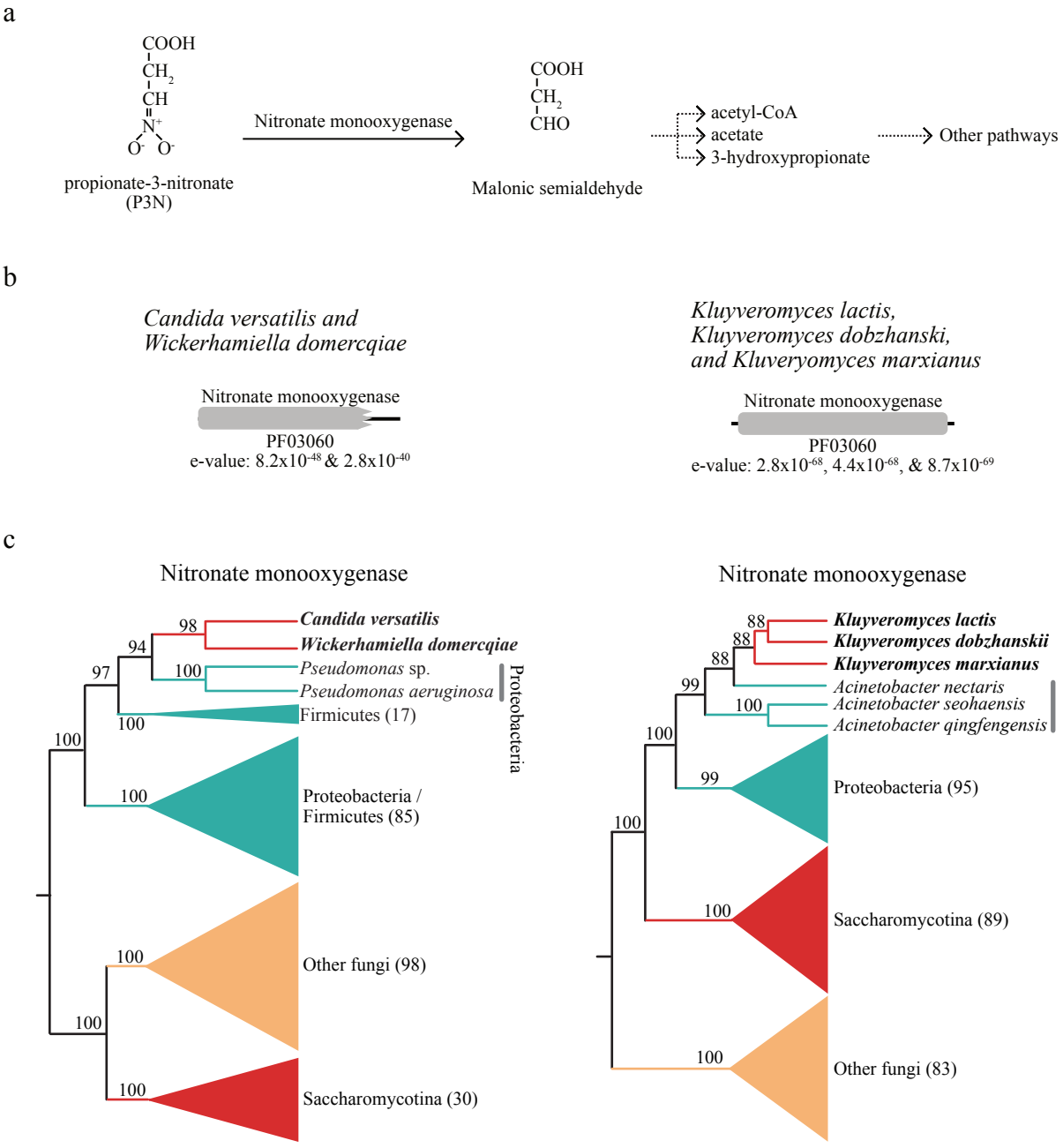
Additional Figure 33. Two independent HGT events provided osmotolerant budding yeasts the genetic machinery to produce the osmoprotectant glycine betaine. (a) Biochemical pathway for the biosynthesis of glycine betaine from choline-*O*-sulfate. Choline-*O*-sulfate is first converted by the action of choline sulfatase into choline; then, choline is converted by the action of choline oxidase into glycine betaine. (b) PFAM protein domain structure and e-values of (left) choline sulfatase and (right) choline oxidase genes present in *Wickerhamiella domercqiae* (top) and *Candida* (*Wickerhamiella*) *versatilis* (bottom). (c) Cartoon gene phylogenies showing support for two independent HGT events of (left) choline sulfatase from the phylum Proteobacteria and (right) choline oxidase from the phylum Actinobacteria. Each gene phylogeny was midpoint rooted, and the values next to collapsed clades (shown by triangles) correspond to the numbers of included sequences. Branch supports were evaluated with 1000 ultrafast bootstraps implemented in IQ-TREE. Detailed information of molecular function and putative donor organisms of horizontally acquired genes are provided in Table S3, whereas the multiple sequence alignments and gene trees are provided in the Figshare depository.

Additional Figure 34. Maximum likelihood trees of two independent HGT events provided osmotolerant budding yeasts the genetic machinery to produce the osmoprotectant glycine betaine. Biochemical pathway for the biosynthesis of glycine betaine from choline-*O*-sulfate. Choline-*O*-sulfate is first converted by the action of choline sulfatase (a) into choline; then, choline is converted by the action of choline oxidase (b) into glycine betaine. The gene used as a query in the BLAST search is shown in red. The other five duplicates of the gene form a clade in the gene tree. Each gene phylogeny was midpoint rooted, and the branch supports were evaluated with 1000 ultrafast bootstraps implemented in IQ-TREE. Detailed information of molecular function and putative donor organisms of horizontally acquired genes found as gene clusters are provided in Table S3, whereas the multiple sequence alignments and gene trees are provided in the Figshare depository.



Additional Figure 34. (continued)

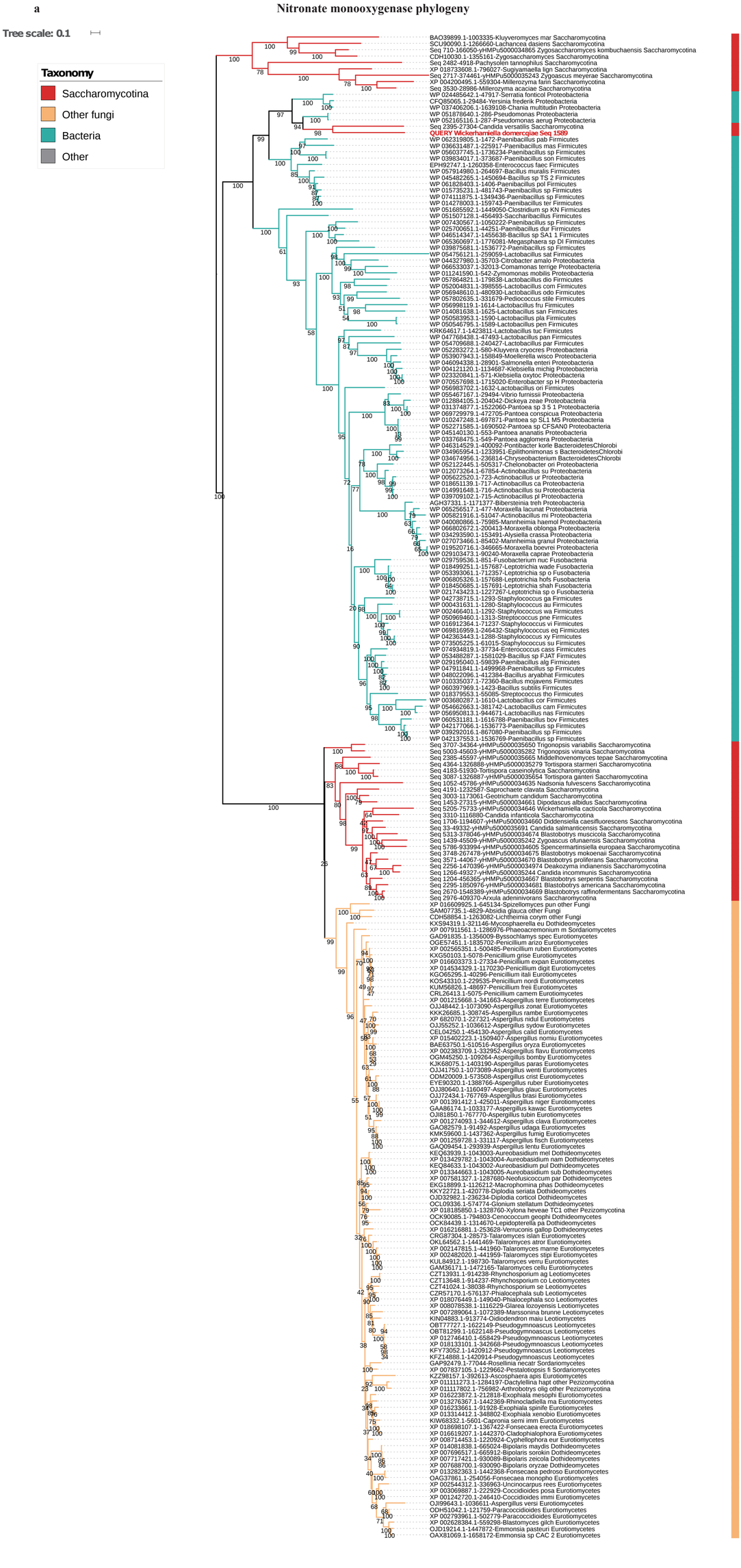
b



Additional Figure 35. Two independent horizontal gene transfer events provide two independent budding yeast lineages the genetic machinery for protection against the mitochondrial toxin P3N. (a) Biochemical pathway for the oxidation of the mitochondrial toxin propionate-3-nitronate (P3N). P3N is converted into malonic semialdehyde via the action of nitronate monooxygenase. Malonic semialdehyde can then be converted into acetyl-CoA, acetate, or 3-hydroxypropionate for downstream use in various pathways. (b) PFAM protein domain structure of nitronate monooxygenase in (left) *Candida* (*Wickerhamiella*) *versatilis* and *Wickerhamiella domercqiae* and in (right) *Kluyveromyces lactis*, *Kluyveromyces dobzhanski*, and *Kluyveromyces marxianus*. (b) Cartoon gene phylogenies showing support for the independent acquisition of nitronate monooxygenase by *C. versatilis* and *W. domercqiae* from *Pseudomonas* donors in the phylum Proteobacteria (left) and by *Kluyveromyces* from *Acinetobacter* donors in the phylum Proteobacteria (right). Each gene phylogeny was midpoint rooted and values next to collapsed clades (shown by triangles) correspond to numbers of included sequences. Branch supports were evaluated with 1000 ultrafast bootstraps implemented in IQ-TREE. Detailed information of molecular function and putative donor organisms of horizontally acquired genes are provided in Table S3, whereas the multiple sequence alignments and gene trees are provided in the Figshare depository.

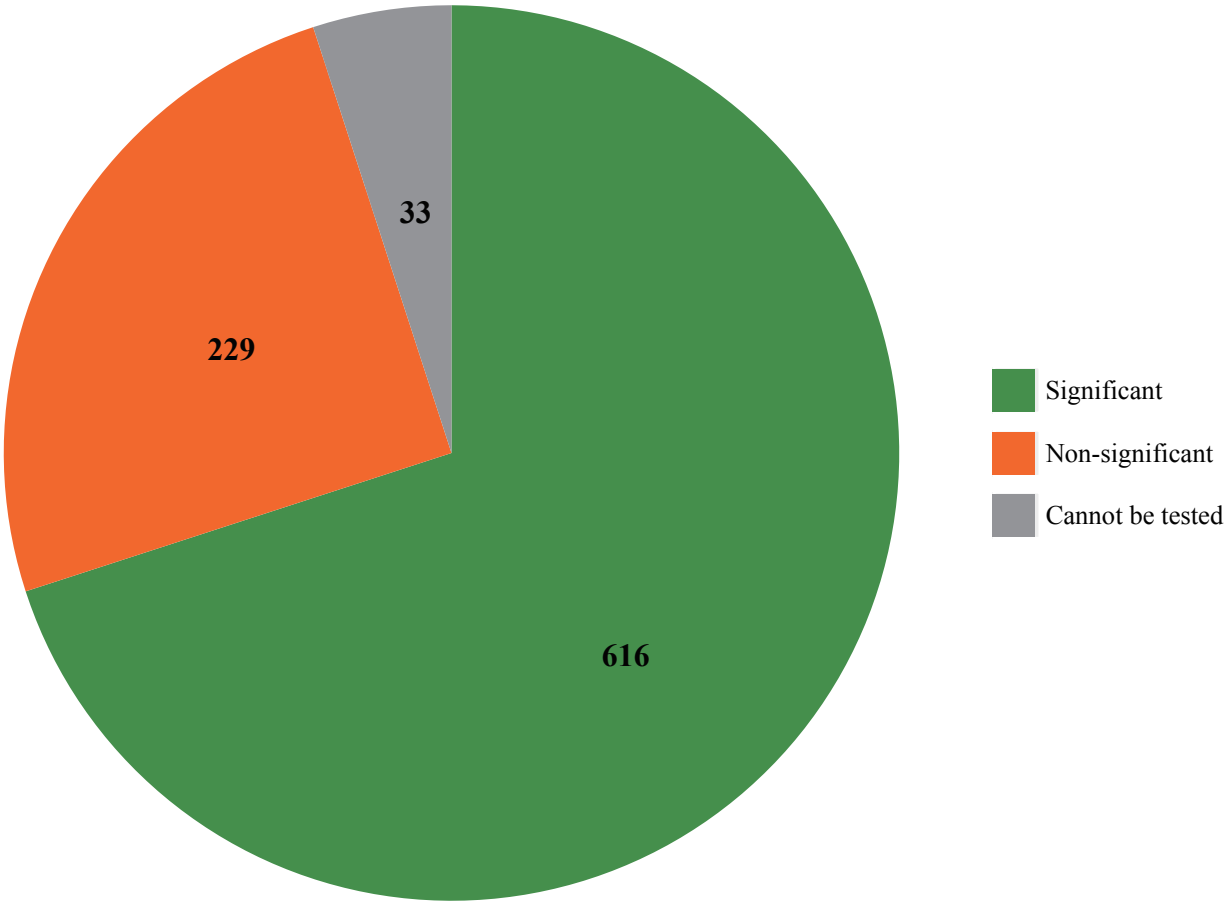
Additional Figure 36. Maximum likelihood trees of two independent horizontal gene transfer events provide two independent budding yeast lineages the genetic machinery for protection against the mitochondrial toxin P3N. Biochemical pathway for the oxidation of the mitochondrial toxin propionate-3-nitronate (P3N). P3N is converted into malonic semialdehyde via the action of nitronate monooxygenase. Gene trees of the nitronate monooxygenase enzymes in (a) *Candida* (*Wickerhamiella*) *versatilis* and *Wickerhamiella domercqiae* and in (b) *Kluyveromyces lactis*, *Kluyveromyces dobzhanski*, and *Kluyveromyces marxianus*, show that both genes were acquired from proteobacteria. The gene used as a query in the BLAST search is shown in red. The other five duplicates of the gene form a clade in the gene tree. Each gene phylogeny was midpoint rooted and the branch supports were evaluated with 1000 ultrafast bootstraps implemented in IQ-TREE. Detailed information of molecular function and putative donor organisms of horizontally acquired genes found as gene clusters are provided in Table S3, whereas the multiple sequence alignments and gene trees are provided in the Figshare depository.

a

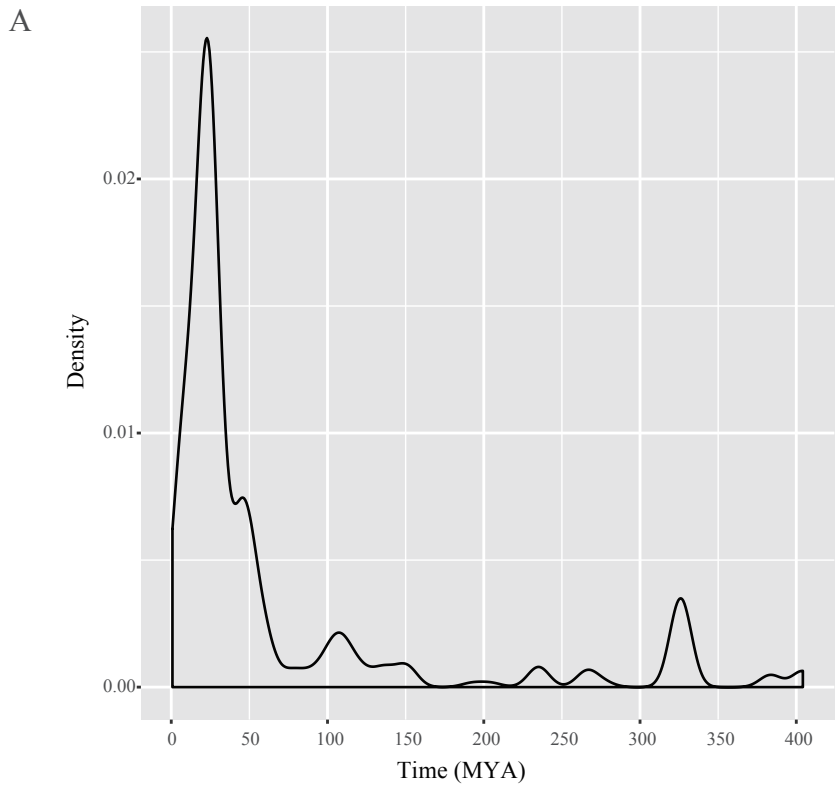


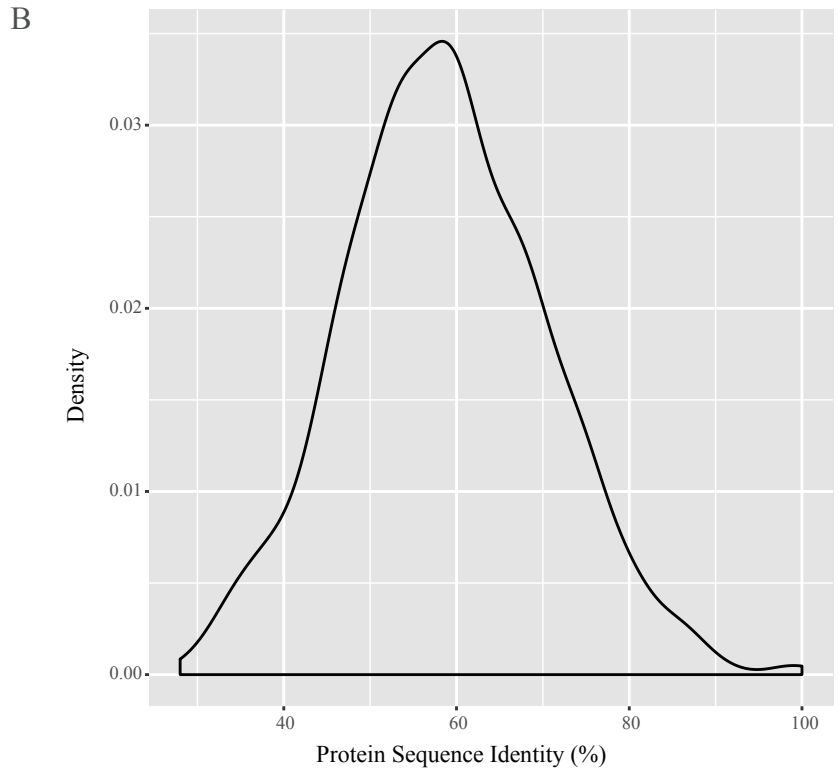
Additional Figure 36. (continued)

b



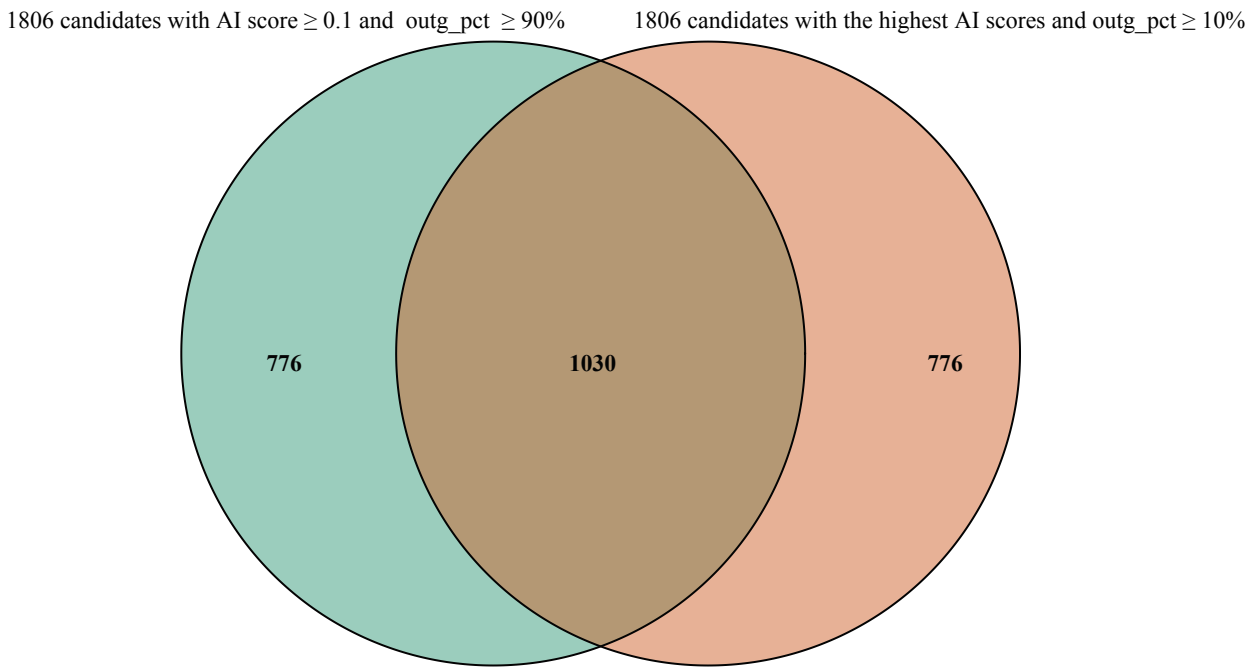
Additional Figure 37. Examination of the 878 horizontally acquired genes using the Approximately Unbiased (AU) test of phylogenetic tree selection. To assess whether the topology inferred by the unconstrained maximum likelihood (ML) gene tree (i.e., the tree with the highest ML score, which shows that the horizontally acquired budding yeast gene was nested within a clade of non-fungal donor genes) was statistically significantly different from a constrained ML tree in which all fungal and yeast genes were forced to be monophyletic, we performed the AU test. The gene trees of 33 of the 878 (4%) horizontally acquired genes did not contain any other fungal sequences and could not be tested. From the remaining 845 horizontally acquired genes, the gene trees of 616 (~73%) were significantly different from the constrained ML trees (p-value <0.05) and 229 (~ 27%) were not. The percentage of HGT-acquired genes supported by the AU test in our study is very similar to the percentage of significantly supported genes identified in previous studies in budding yeasts (15 / 21 or ~71%; Table S3).



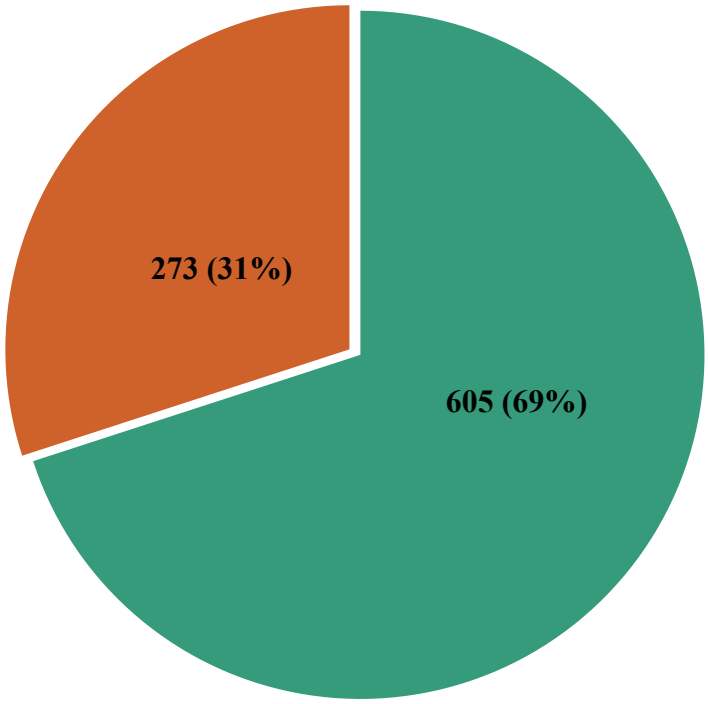


Additional Figure 38. Time of origin and sequence divergence of the 878 genes that were horizontally acquired in budding yeasts. The identification of the 365 events that gave rise to the 878 HGT-acquired genes coupled with the generation of a robust timetree for the budding yeast subphylum enabled us to time the origin of each HGT event in the budding yeast phylogeny. (A) Distribution of 365 HGT events over evolutionary time. (B) Distribution of sequence identities between the protein sequence of each of the 878 HGT-acquired budding yeast protein sequences and the protein sequence of their closest non-fungal donor.

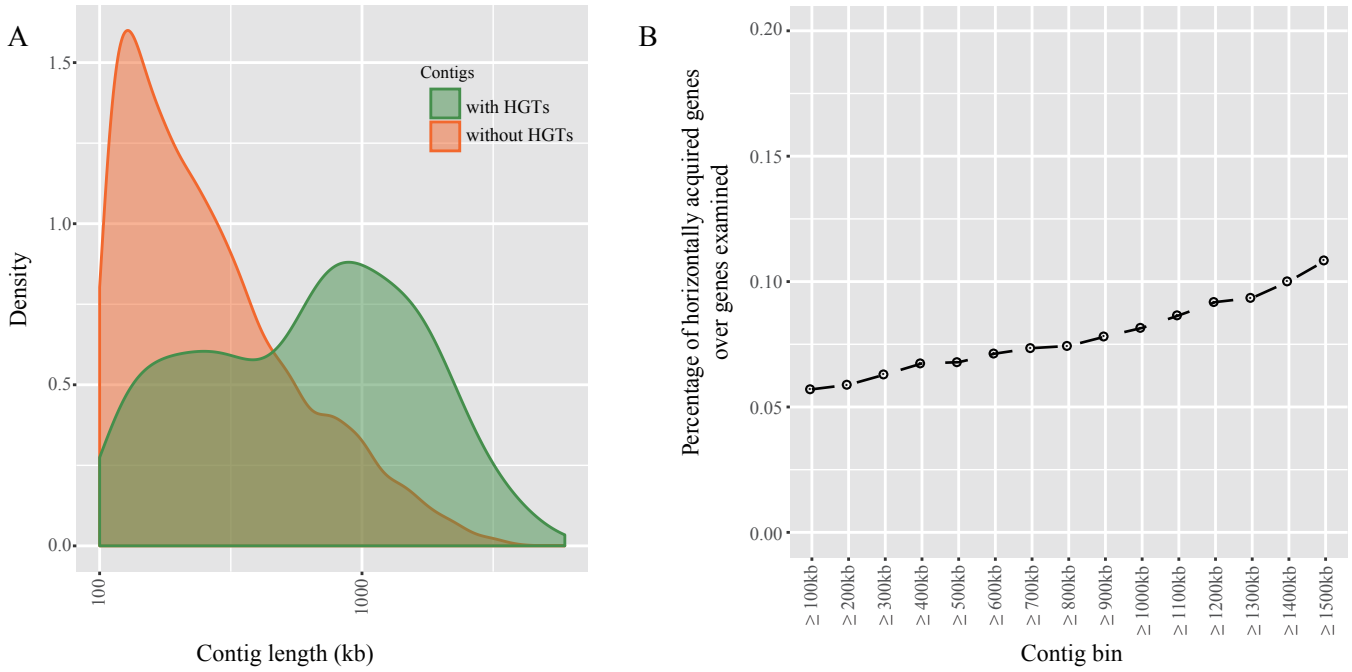
a



b



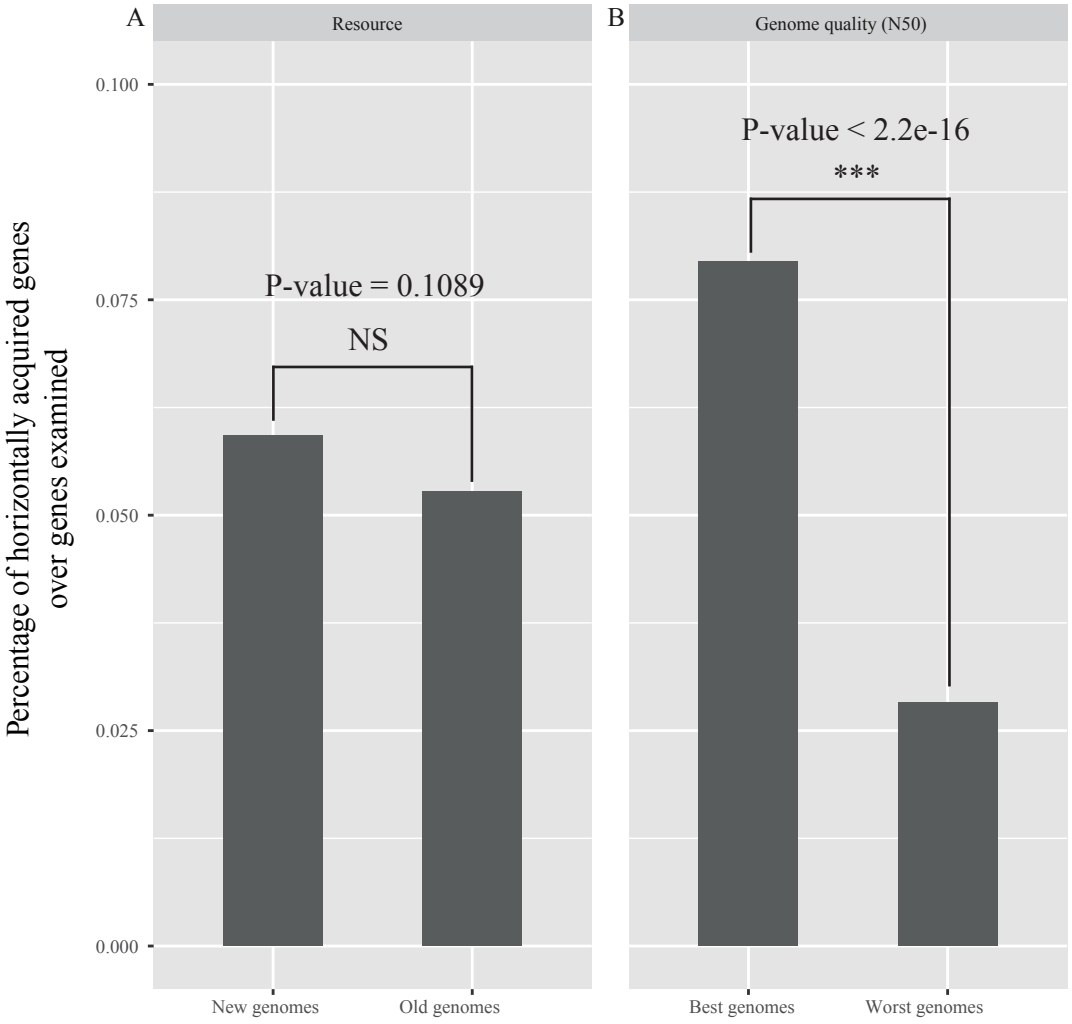
Additional Figure 39. Gene overlap between genes inferred to have been horizontally acquired by budding yeasts when using two different thresholds in the workflow of HGT identification. (a) The overlap between the 1,806 genes identified as HGT candidates using the cutoffs of AI score ≥ 0.1 and outg_pct ≥ 90%, and the 1,806 genes identified as HGT candidates because they had the highest AI scores for the cutoff outg_pct ≥ 10%. The two lists of candidates shared 1,030 / 1,806 genes (or 57%), suggesting that the implementation of a strict outg_pct cutoff (i.e., the percentage of non-fungal species in the list of the top 300 BLAST hits that have different taxonomic species names) did not greatly influence the identification of top HGT candidates. (b) Among the set of 878 strongly supported cases of HGT identified in our study, the gene overlap between the two sets shown in panel (a) was even greater (605 /878 or 69%).



C

Contig bin	# of HGT events	Average age (MYA) of HGT events
≥ 100kb	365	66.3
≥ 200kb	342	64.7
≥ 300kb	333	62.2
≥ 400kb	322	62.7
≥ 500kb	301	62.2
≥ 600kb	293	62.4
≥ 700kb	282	61.5
≥ 800kb	265	60.6
≥ 900kb	264	57.9
≥ 1000kb	257	56.6
≥ 1100kb	253	55.8
≥ 1200kb	248	54.5
≥ 1300kb	239	50.4
≥ 1400kb	236	47.6
≥ 1500kb	233	46.8

Additional Figure 40. Distribution of sequence lengths of genomic contigs used to examine the incidence of 878 horizontally acquired genes in the budding yeast subphylum. We investigated a total of 8,792 contigs with sequence lengths ≥ 100 kb from 329 yeast genomes; of these, 480 contigs contain one or more horizontally acquired genes, whereas the remaining 8,312 contigs do not contain any genes that were acquired via HGT. (A) Distributions of sequence lengths of genomic contigs based on whether they contain HGT-acquired genes. The green distribution is that of sequence lengths of contigs that contain HGT-acquired genes, and the orange distribution is that of sequence lengths of contigs that do not contain HGT-acquired genes. (B) We used different cutoff values of contig sequence length (from ≥ 100 kb to ≥ 1,500 kb with steps of 100 kb) to create 15 different bins of contigs. For each bin, we then calculated the percentage of HGT-acquired genes over total number of genes examined. We find that increasing the cutoff value increases the percentage of horizontally acquired genes. (C) For each contig bin, we calculated the number of HGT events and their average age (MYA: million years ago).



Additional Figure 41. Comparisons of the percentage of horizontally acquired genes found in different subsets of budding yeast genomes. Examination of the 1,538,912 genes present in 8,792 genomic contigs with sequence lengths ≥ 100 kb led to the identification of 878 HGT-acquired genes. (A) Comparison of the number of HGT-acquired genes present in the 217 newly sequenced genomes versus the number present in the 112 publicly available genomes did not identify a significant difference (based on Fisher's exact test). (B) Comparison of the number of HGT-acquired genes present in the 164 genomes with the highest N50 values (i.e., the contig or scaffold value above which 50% of the total length of the sequence assembly can be found) versus the number present in the 165 genomes with the lowest N50 values was significantly different (based on Fisher's exact test), with genomes with higher N50 values containing higher numbers of HGT-acquired genes . Note that each one of the percentage values reported is scaled for the total number of genes examined.