A comprehensive comparison of phylogenomics and orthology methods applied to the Fungi

Bas E. Dutilh, Vera van Noort, René T. J. M. van der Heijden, Martijn A. Huynen, Berend Snel

CMBI / NCMLS, Radboud University Nijmegen
The Netherlands www.cmbi.ru.nl/~dutilh

Phylogenomics integrates all the phylogenetic information present in complete genome sequences, and is quickly becoming the standard for inferring reliable species trees. While phylogenetic trees (based on single genes) may be unreliable due to e.g. horizontal gene transfer or parallel gene loss, the large scale of phylogenomic trees can average out biases (*Fig. 1*).
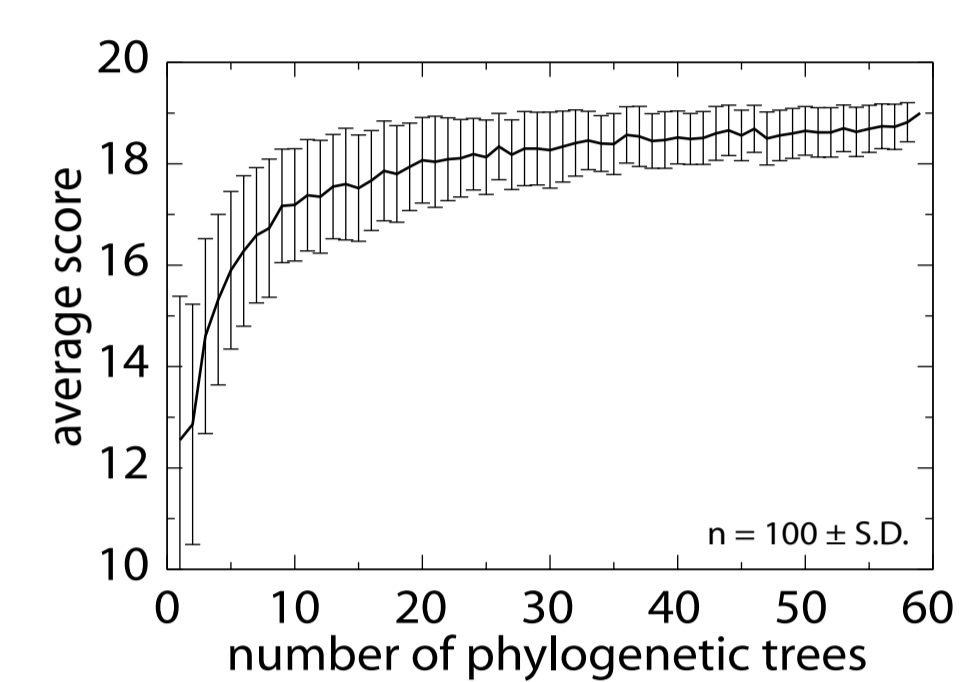

Figure1. More high quality data leads to a better phylogeny. Supertrees (Consense) were composed from an increasing number of phylogenetic distance trees of cluster pan-orthologs.

Several phylogenomic approaches have surfaced that fundamentally differ in the level at which they integrate the phylogenetic information: gene content, superalignment, superdistance and supertree (*Fig. 2*).

We systematically test these phylogenomic methods on the Fungi, the eukaryotic clade with the largest number of sequenced genomes. Many of the taxonomic relationships in the fungal kingdom have been resolved while several details remain unclear (*Fig. 3*)
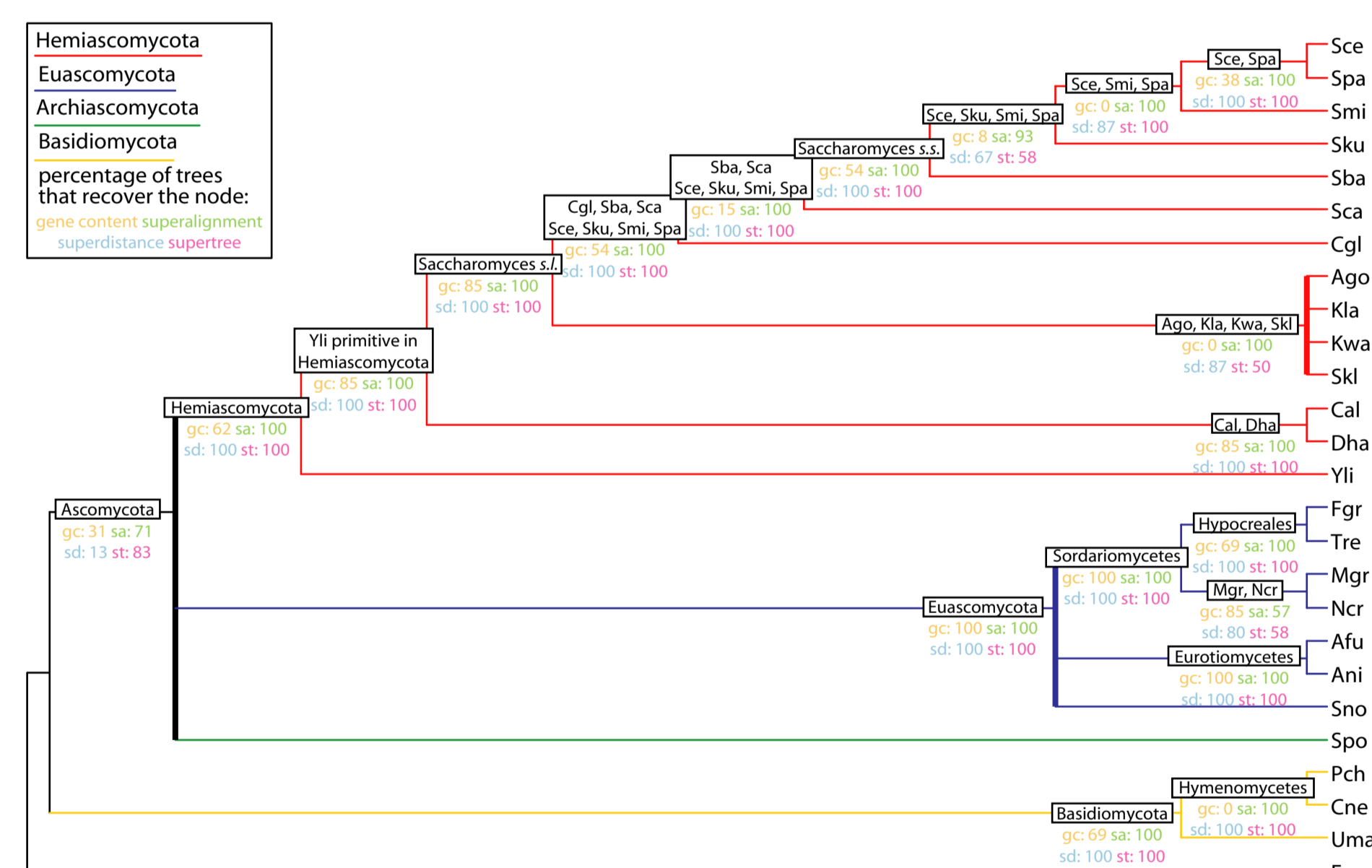

Figure 3. Consensus phylogeny of the Fungi. The labeled nodes are supported by several literature references. We use these nodes to score the reconstructed phylogenomic trees. Unresolved issues are indicated by multifurcating nodes (bold lines). The colored numbers at each node indicate the percentage of the trees in each of the phylogenomic approaches that recovered this node correctly.

In parallel to these four phylogenomic approaches, we also compare the performance of different orthology prediction methods with higher and lower resolution: pairwise orthology (Inparanoid), cluster orthology (a COG-like algorithm) and tree-based orthology.

Several relevant trends in the resulting topologies can be explained from the methods. The gene content trees, for example, cluster the Euascomycota with the Basidiomycota, which can be explained by their filamentous lifestyle, leading to an apparent convergence in gene content (*Fig. 4a*).
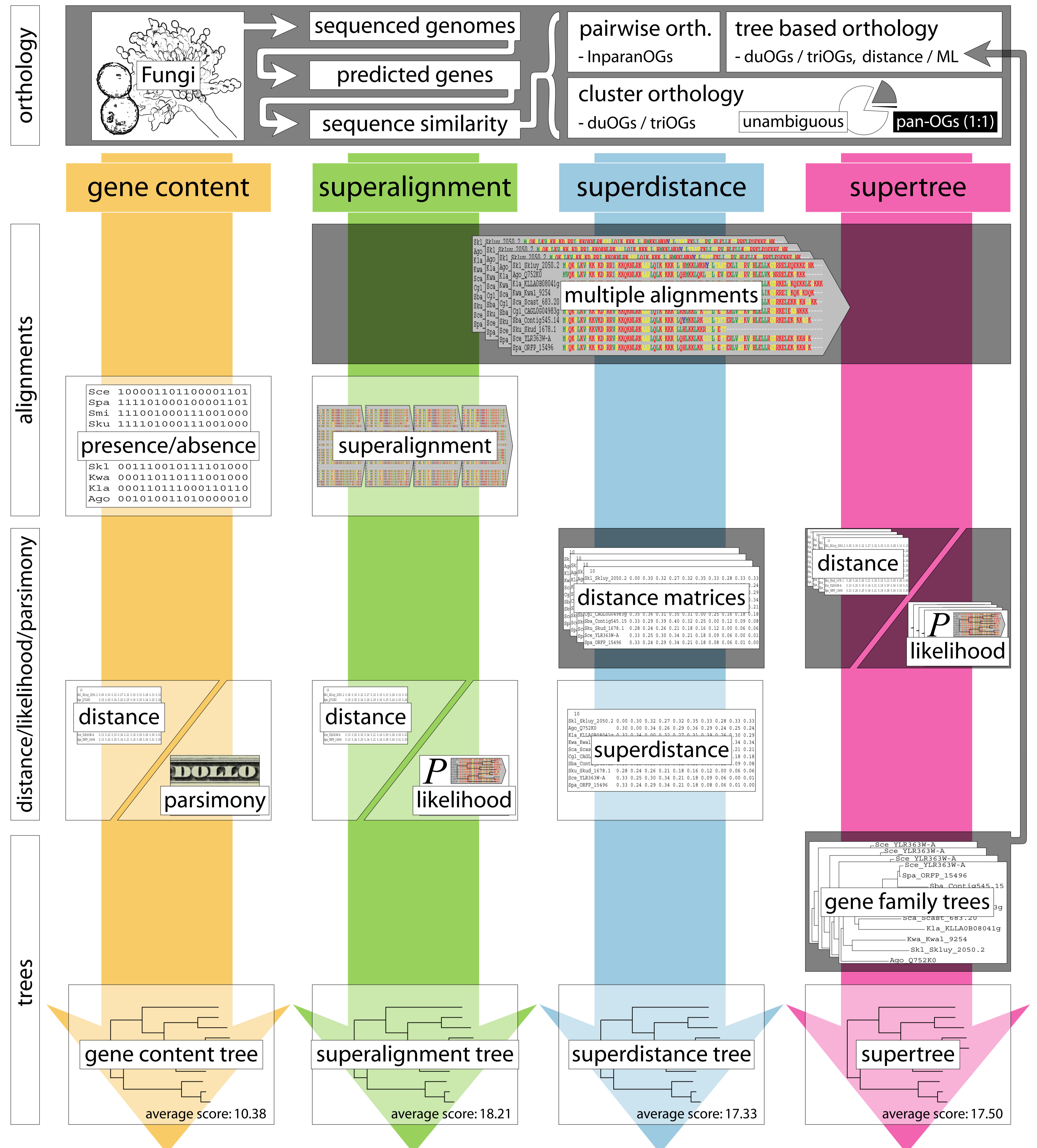

Figure 2. Making phylogenomic trees. Phylogenomics follows the steps of phylogenetics: (1) definition of orthologs, (2) multiple alignment, (3) distance, likelihood or parsimony and (4) tree reconstruction. The parallel phylogenetic information (gray boxes) can be integrated to a phylogenomic scale (white boxes) in each step; this defines the phylogenomic approach.

The phylogenomic methods that are most successful at recovering the nodes of the consensus phylogeny are found among the superalignment and the supertree approaches (*Fig. 4b*).

The orthology definition and the phylogenomic approach have more influence on the final topology than the precise methodological details, such as using a distance or likelihood approach for the reconstruction of the tree.

Apart from the gene content trees, all phylogenomic trees recover many target nodes. They are also quite unanimous about the solutions for the unresolved nodes of the fungal taxonomy (*Table 1*).
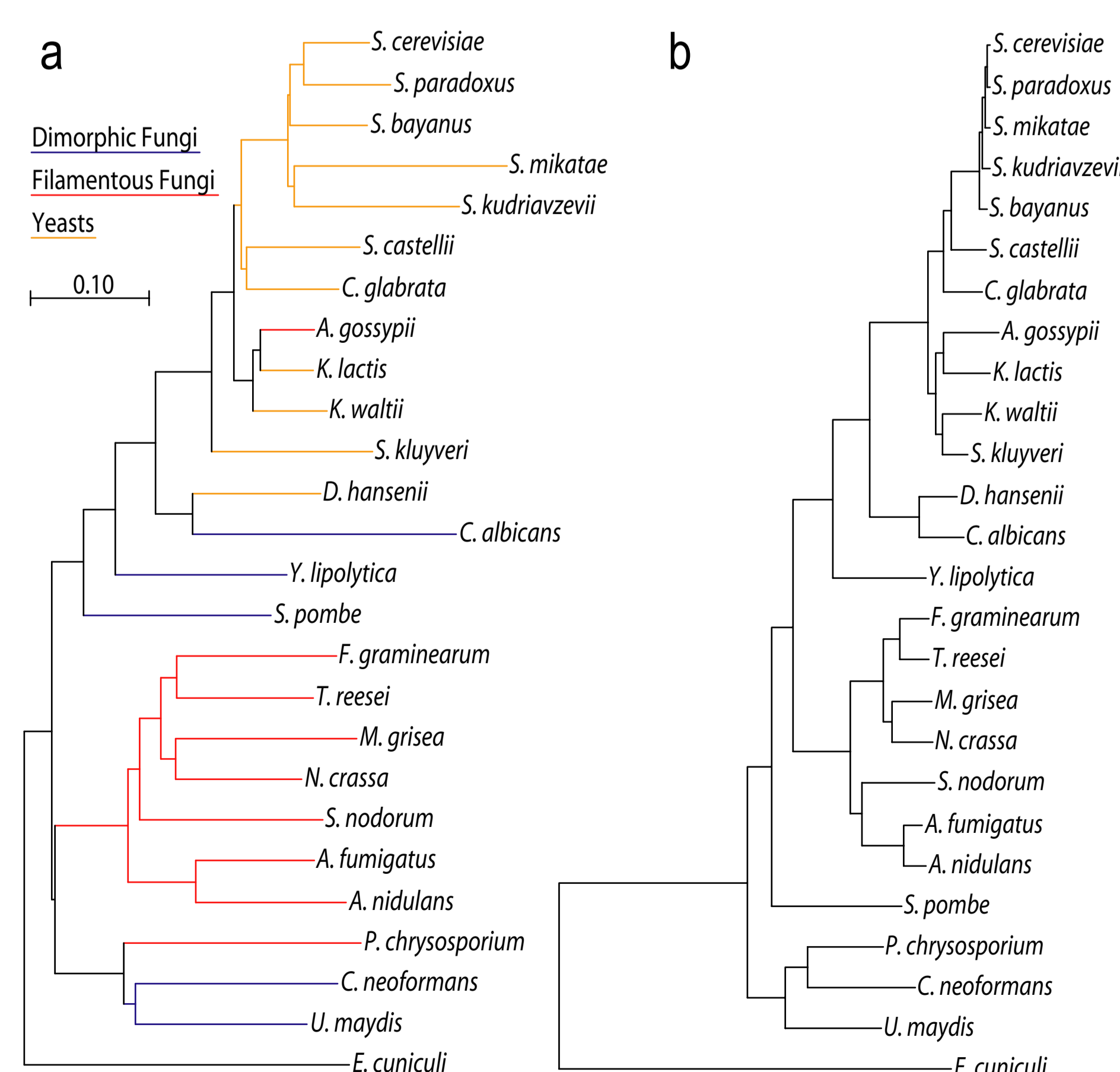

Figure 4. Phylogenomic trees. (a) Gene content tree. BioNJ distance tree based on pairwise orthologs. (b) One of the two highest scoring phylogenomic trees. This topology was recovered by four superalignment trees and one supertree using cluster orthology (unambiguous orthologs or pan-orthologs).

Table 1. Number of trees in each phylogenomic approach that support the possible branchings in the unresolved nodes.

	Ago, Kla	Ago, Kwa	Kla, Skl	Kluyveromyces	Kla, Skl	Kwa, Skl	Ago, Kla, Kwa	Ago, Kla, Skl	Kla, Kwa, Skl	Sordariomycetes prim. in Euascomycota	Eurotiomycetes prim. in Euascomycota	S. nodorum prim. in Euascomycota	Hemiascomycota prim. in Ascomycota	Euascomycota prim. in Ascomycota	Archiascomycota prim. in Ascomycota
gene content (13)	10	0	0	1	0	0	14	0	0	8	0	0	4	5	4	0	11	0
superalignment (14)	14	0	0	0	0	0	14	0	0	0	14	0	0	0	0	0	0	12
superdistance (15)	14	0	0	0	0	0	15	0	0	1	13	0	2	2	0	1
supertree (12)	12	0	0	0	0	0	12	0	0	0	12	0	0	4	4