

PIPELINE GENOMIKA PORÓWNAWCZA

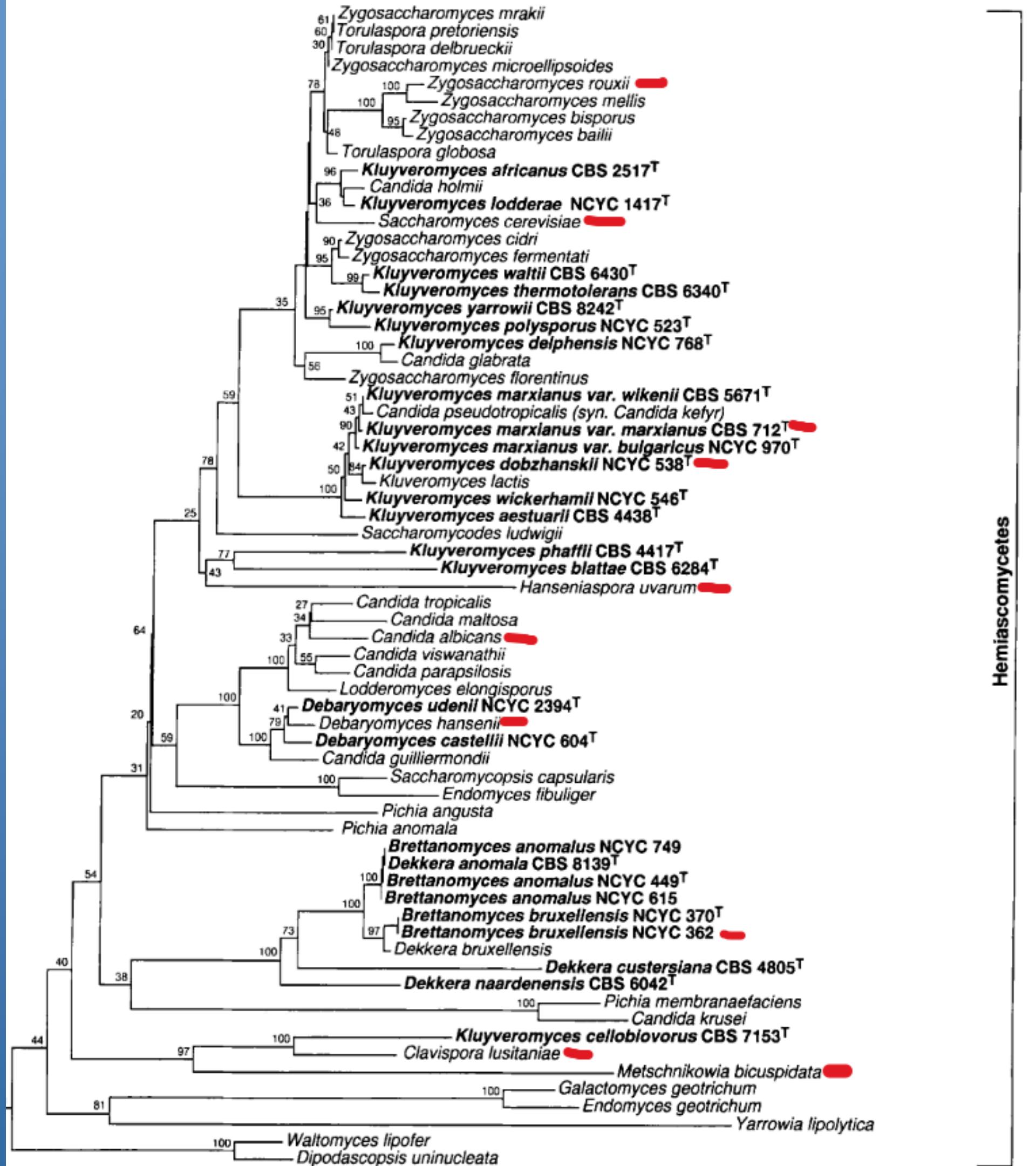
Prezentuje: Zuzanna Surowiec

DROŻDŻE

JUNPENG CAI ET AL. 96

18S rRNA, NJ, Bootstrap.

Wybrałem 20 z 116 taksonów drożdży użytych w pracy.



Hemiascomycetes

Ascomycetes

02

DROŻDŻE

SACCHAROMYCES CEREVISIAE

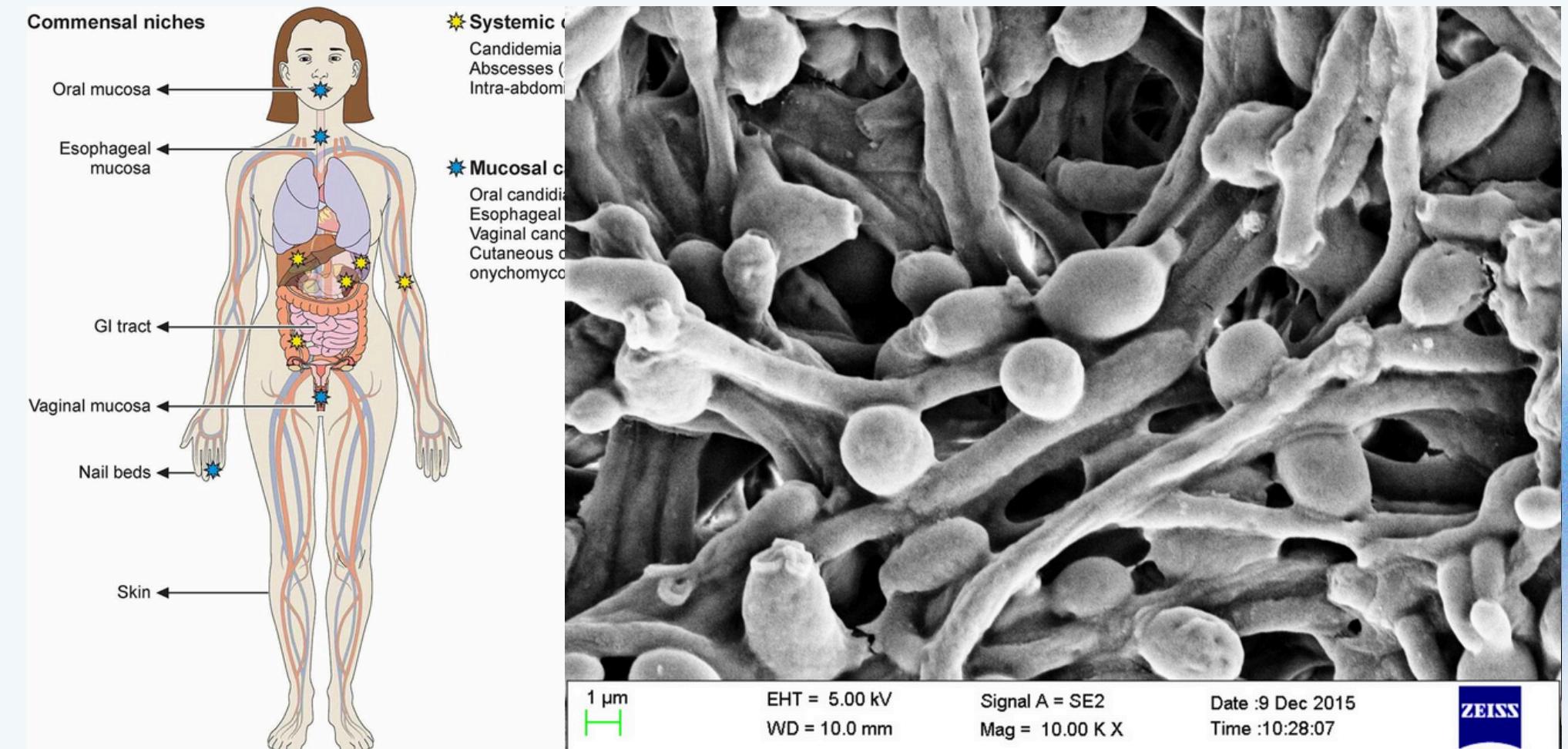
Używane w piwowarstwie, winiarstwie, piekarnictwie i wielu
więcej.

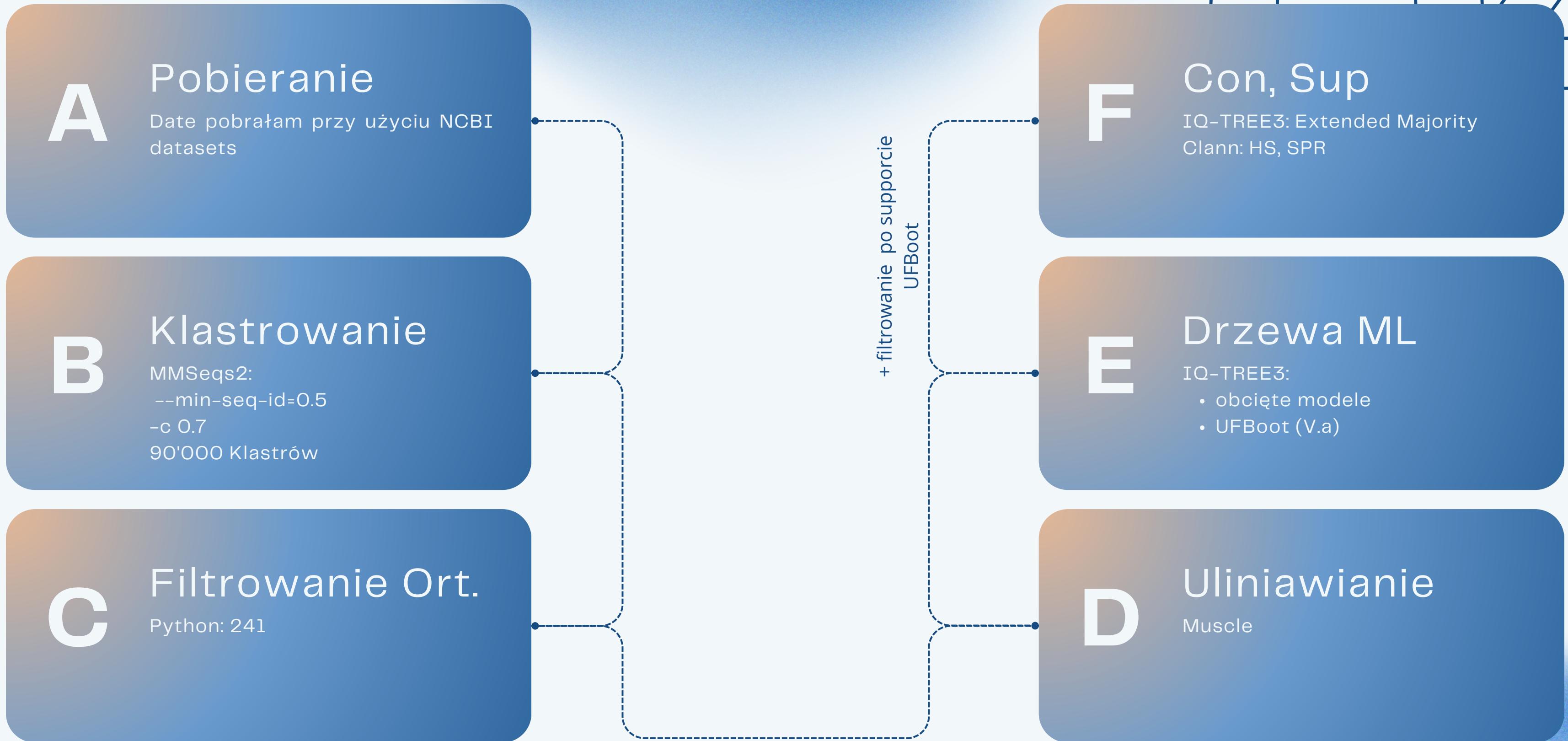


03

DROŻDŻE CANDIDA ALBICANS

Odpowiedzialna za większość infekcji grzybiczych





A

Pobieranie

Dane pobrałem przy użyciu NCBI datasets

B

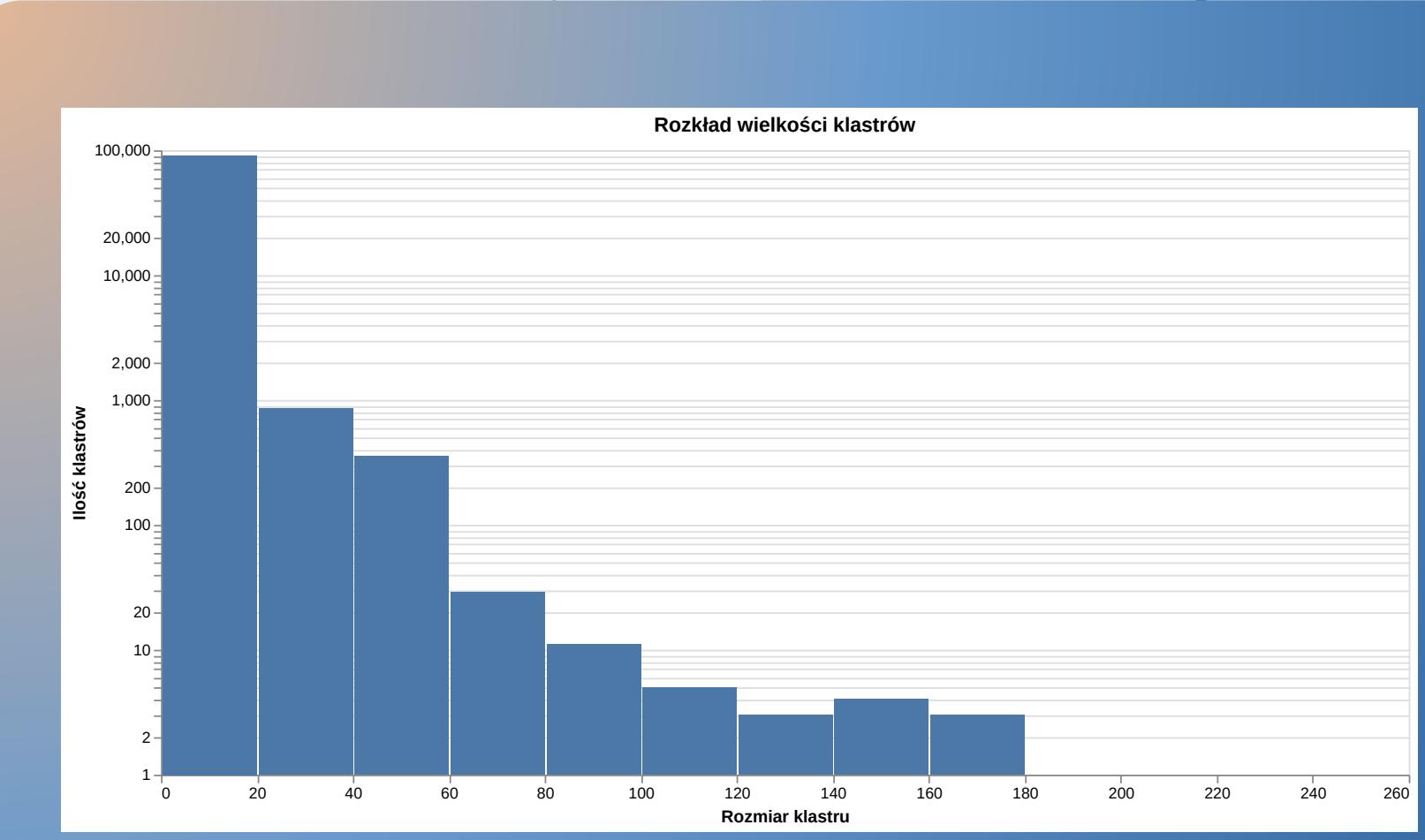
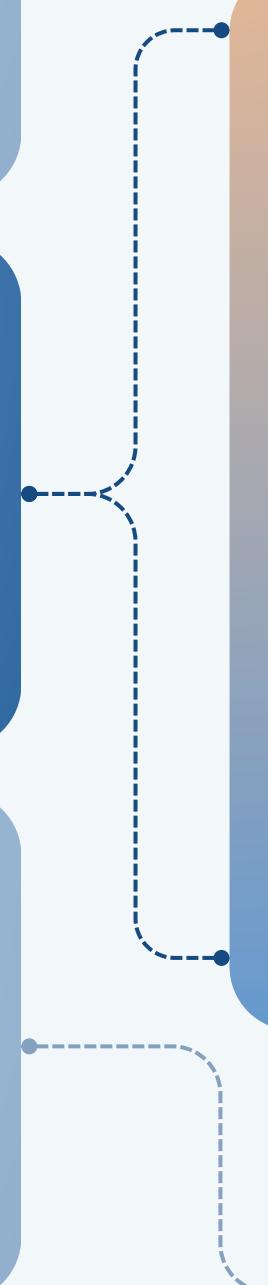
Klastrowanie

MMSeqs2:
--min-seq-id=0.5
-c 0.7
90'000 Klastrów

C

Filtrowanie Ort.

Python: 241



min=2, max=248

D

F

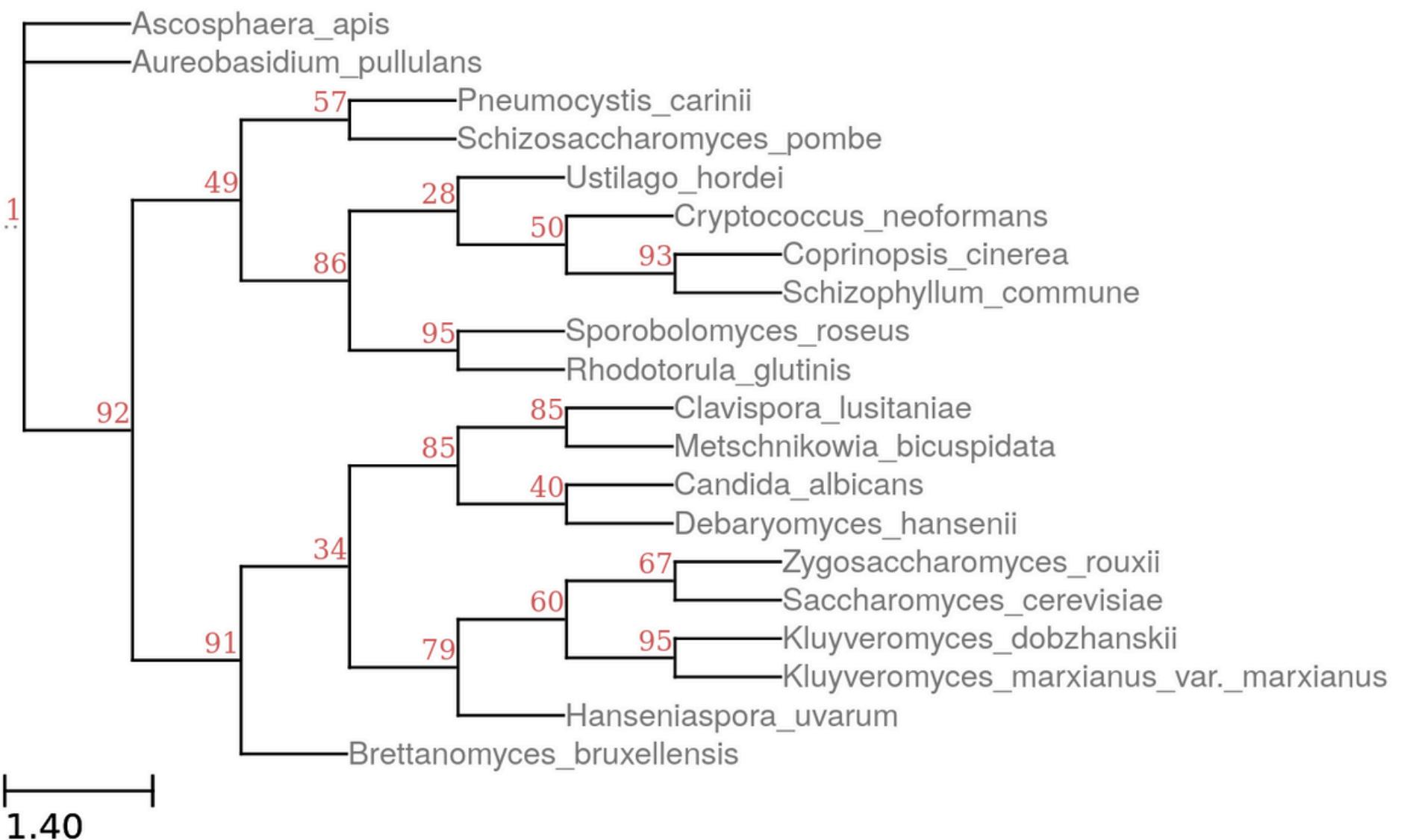
Con, Sup

IQ-TREE3, Clann

Pipeline Overview

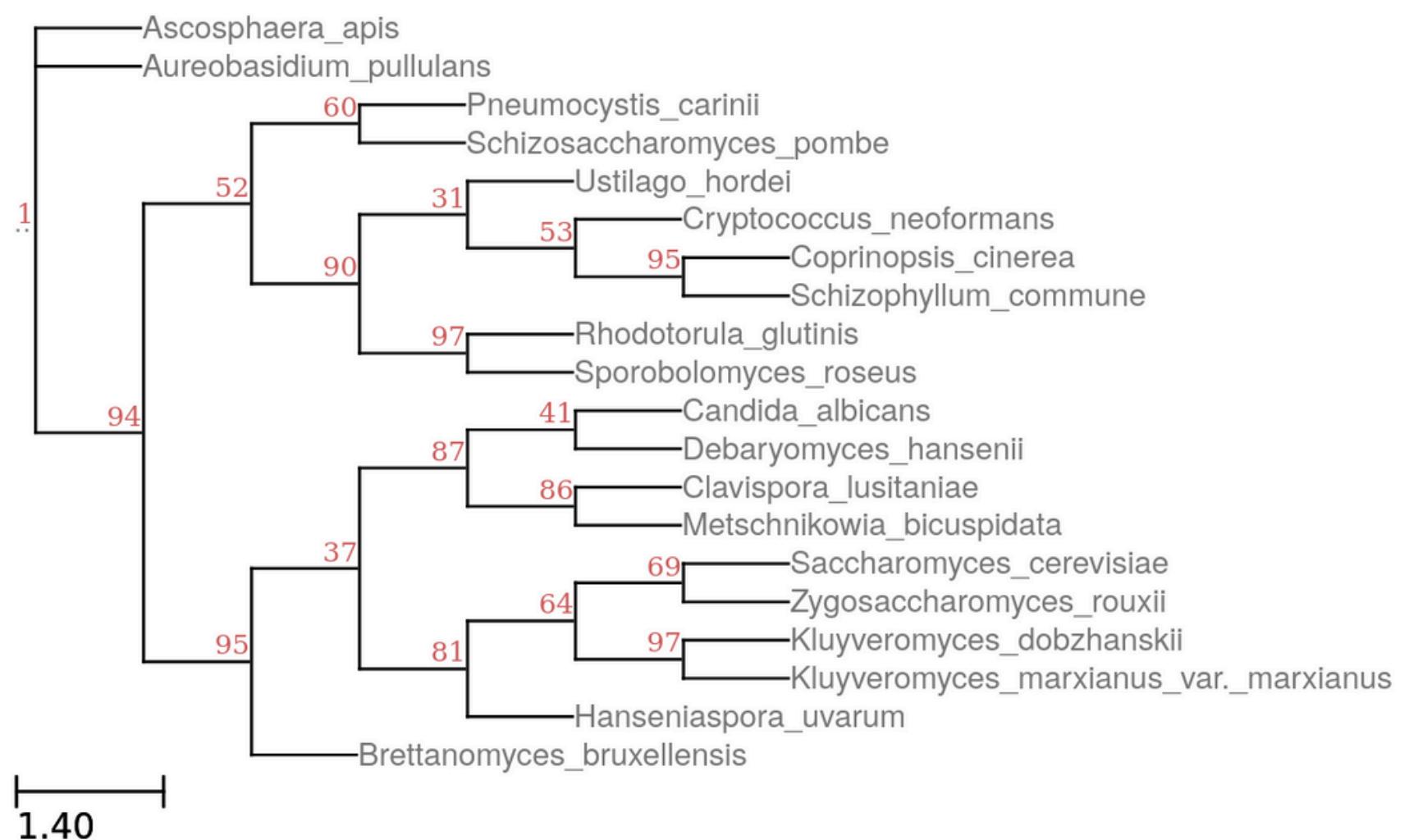
KONSENSUS ORT.

Reference	nRF	nJRF
NCBI	0.333	0.296
Publication	0.235	0.106
Time Tree	0.066	0.040



KONSENSUS ORT. FIL.

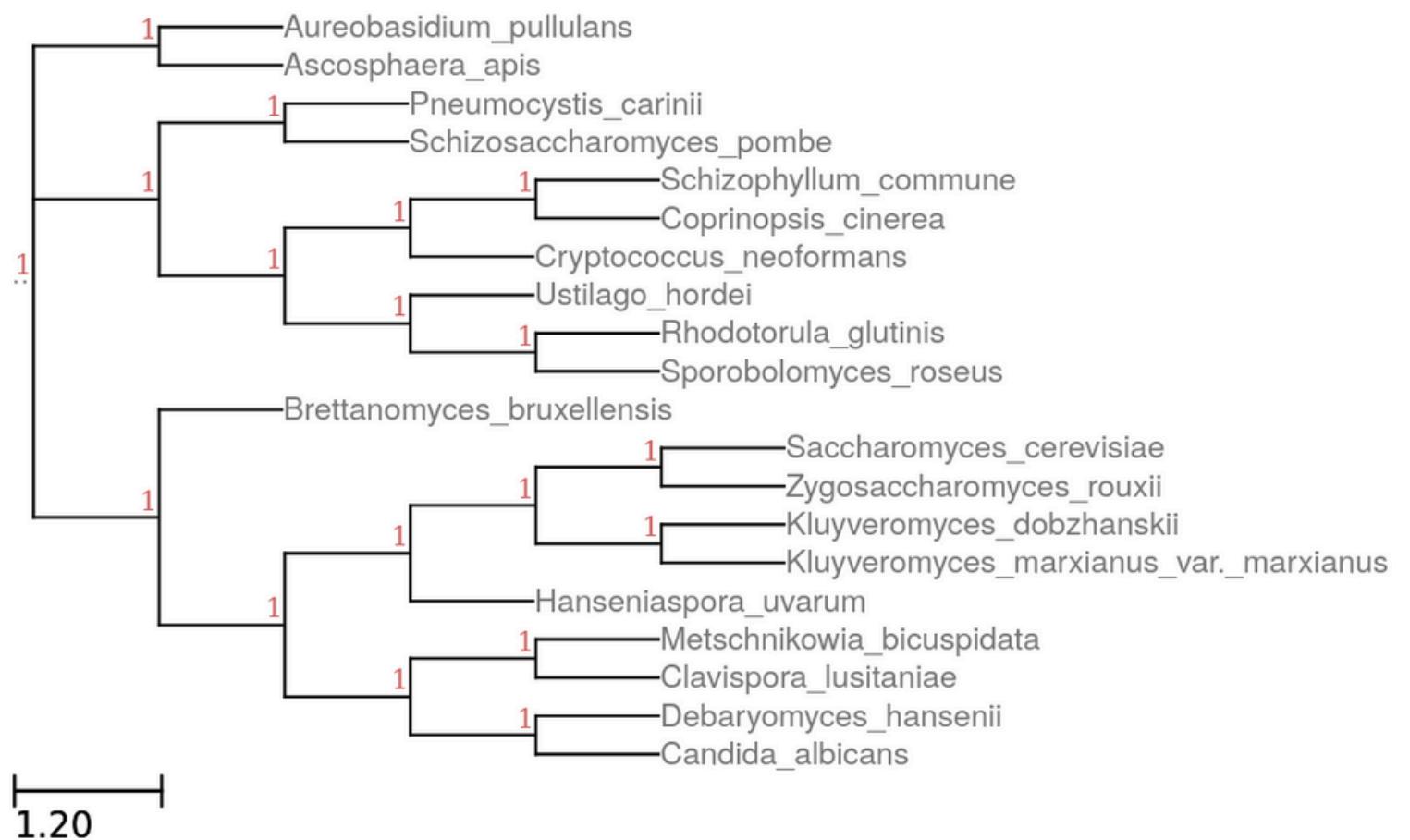
Reference	nRF	nJRF
NCBI	0.333	0.296
Publication	0.235	0.106
Time Tree	0.066	0.040



07

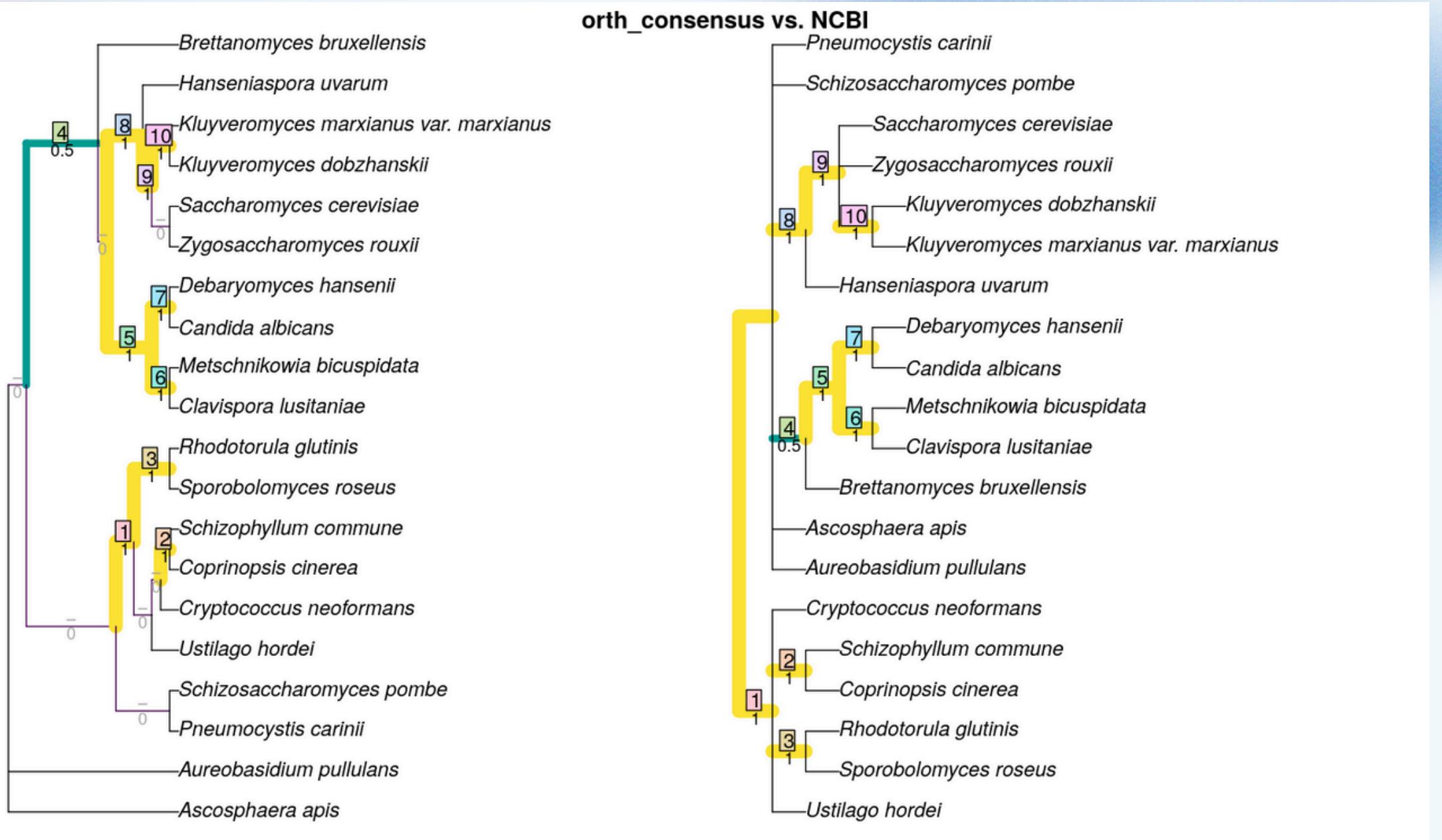
SUPER D. ORT.

Reference	nRF	nJRF
NCBI	0.333	0.296
Publication	0.235	0.116
Time Tree	0.066	0.040



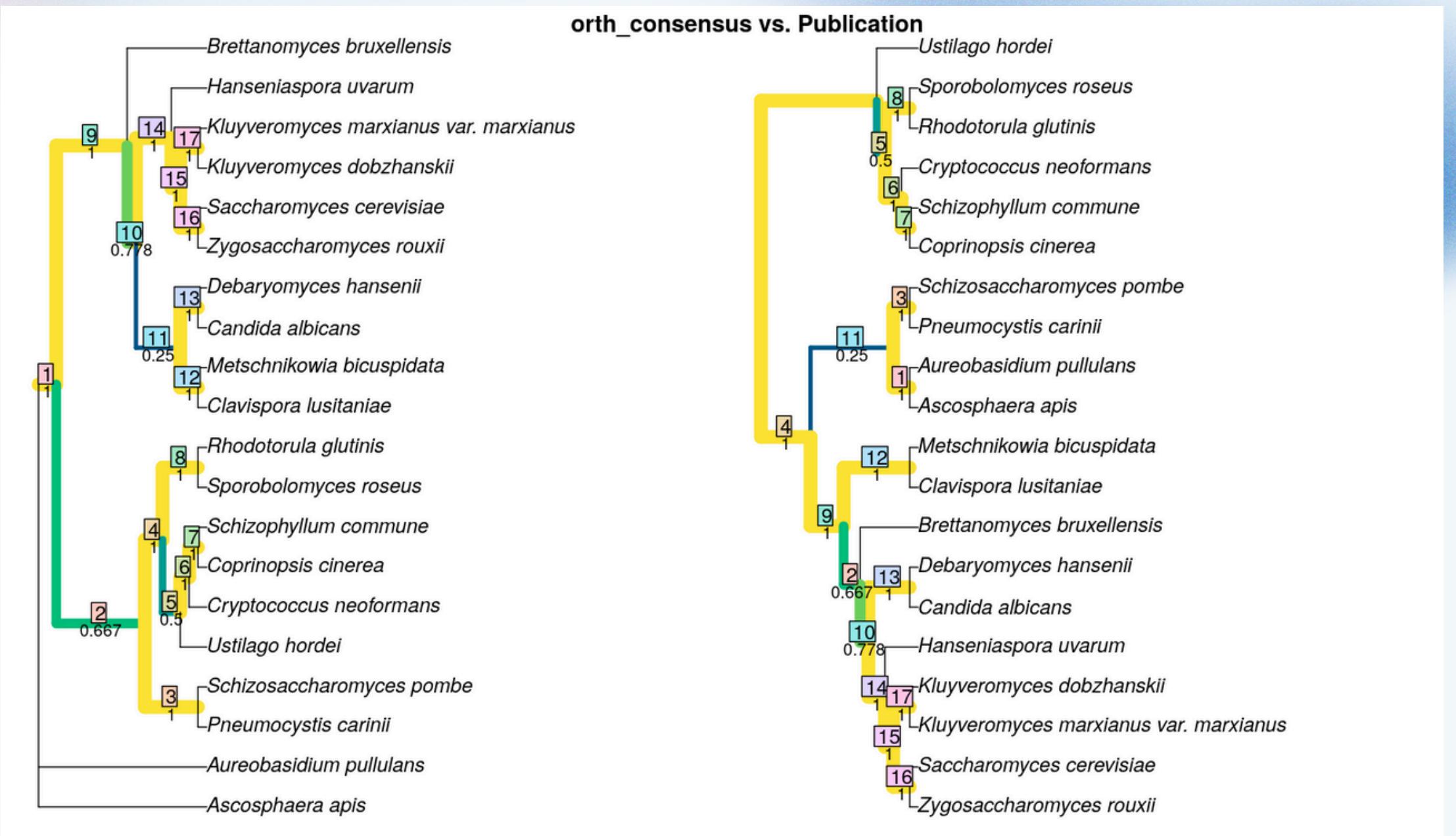
08

KONSENSUS ORT. - CMP



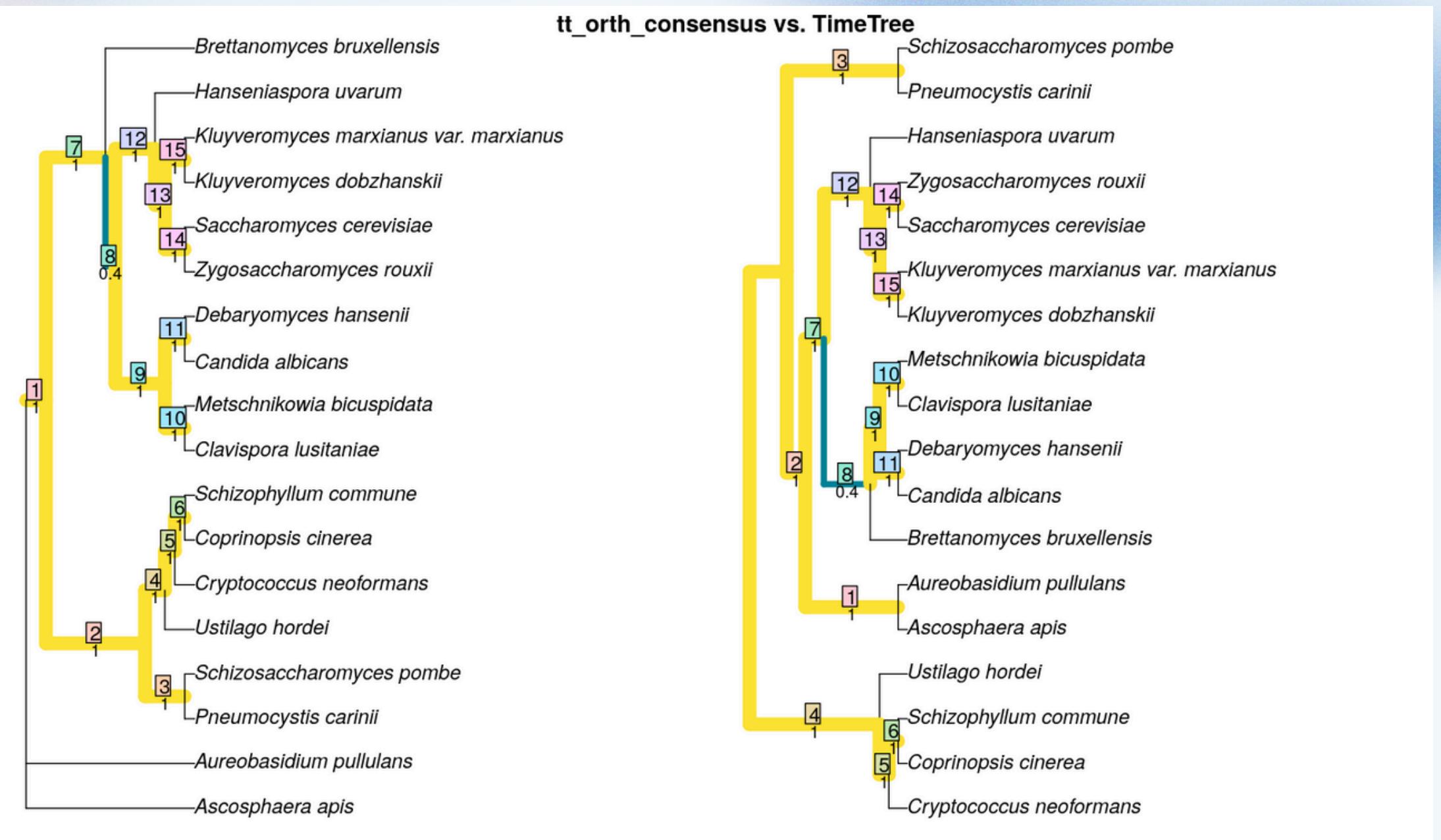
Reference	nRF	nJRF
NCBI	0.333	0.296
Publication	0.235	0.106
Time Tree	0.066	0.040

KONSENSUS ORT. - CMP



Reference	nRF	nJRF
NCBI	0.333	0.296
Publication	0.235	0.106
Time Tree	0.066	0.040

KONSENSUS ORT. - CMP



Reference	nRF	nJRF
NCBI	0.333	0.296
Publication	0.235	0.106
Time Tree	0.066	0.040

BRAKUJĄCE:

- RHODOTORULA GLUTINIS
- SPOROBOLOMYCES ROSEUS



DO ROZWAŻENIA

■ MODELE

Zezwolenie na przetestowanie większej ilości modeli podczas liczenia drzew ML

■ BOOTSTRAP

Użycie nie-param. Bootstrapu zamiast UFBoot do filtrowania — większa skuteczność?

■ RECLUSTERING

Ponowne przeklastrowanie małych klastrów w celu zwiększenia ilości klastrów ortologicznych



**DZIĘKUJĘ
ZA UWAGĘ
ZUZANNA SUROWIEC**

Bibliography

- [1] J. CAI, I. N. ROBERTS, and M. D. COLLINS, "Phylogenetic Relationships among Members of the Ascomycetous Yeast Genera Brettanomyces, Debaryomyces, Dekkera, and Kluyveromyces Deduced by Small Subunit rRNA Gene Sequences," *International Journal of Systematic and Evolutionary Microbiology*, vol. 46, no. 2, pp. 542–549, 1996, doi: https://doi.org/10.1099/00207713_46.2.542.
- [2] C. L. Schoch et al., "NCBI Taxonomy: a comprehensive update on curation, resources and tools," *Database*, vol. 2020, p. baaa62, 2020, doi: 10.1093/database/baaa062.
- [3] S. Kumar et al., "TimeTree 5: An Expanded Resource for Species Divergence Times," *Molecular Biology and Evolution*, vol. 39, no. 8, p. msac174, 2022, doi: 10.1093/molbev/msac174.
- [4] E. W. Sayers, M. Cavanaugh, K. Clark, J. Ostell, K. D. Pruitt, and I. Karsch-Mizrachi, "GenBank," *Nucleic Acids Research*, vol. 47, no. D1, pp. D94–D99, 2018, doi: 10.1093/nar/gky989.
- [5] N. A. O'Leary et al., "Exploring and retrieving sequence and metadata for species across the tree of life with NCBI Datasets," *Scientific Data*, vol. 11, no. 1, p. 732, Jul. 2024, doi: 10.1038/s41597.024.03571.y.
- [6] M. Steinegger and J. Söding, "MMseqs2 enables sensitive protein sequence searching for the analysis of massive data sets," *Nature Biotechnology*, vol. 35, no. 11, pp. 1026–1028, Nov. 2017, doi: 10.1038/nbt.3988.
- [7] C. da Costa-Luis, "tqdm: A Fast, Extensible Progress Meter for Python and CLI," *Journal of Open Source Software*, vol. 4, p. 1277, 2019, doi: 10.21105/joss.01277.
- [8] R. C. Edgar, "Muscle5: High-accuracy alignment ensembles enable unbiased assessments of sequence homology and phylogeny," *Nature Communications*, vol. 13, no. 1, p. 6968, Nov. 2022, doi: 10.1038/s41467.022.34630.w.
- [9] O. Tange, "GNU Parallel 20251222 ('Bondi').," [Online]. Available: <https://doi.org/10.5281/zenodo.18039569>
- [10] H. R. H. B. A. J. R. E. S. C. B. N. D. M. N. G. M. W. H. G. H. R. L. B. Q. M. Thomas K.F. Wong Nhan Ly.Trong, "IQ-TREE 3: Phylogenomic Inference Software using Complex Evolutionary Models ,," 2025, [Online]. Available: <https://doi.org/10.32942/X2P62N>
- [11] C. J. Creevey and J. O. McInerney, "Clann: investigating phylogenetic information through supertree analyses," *Bioinformatics*, vol. 21, no. 3, pp. 390–392, 2004, doi: 10.1093/bioinformatics/bti020.
- [12] C. J. Creevey et al., "Does a tree-like phylogeny only exist at the tips in the prokaryotes?," *Proceedings of the Royal Society B: Biological Sciences*, vol. 271, no. 1557, pp. 2551–2558, 2004, doi: 10.1098/rspb.2004.2864.
- [13] M. R. Smith, "Information theoretic Generalized Robinson-Foulds metrics for comparing phylogenetic trees," *Bioinformatics*, vol. 36, no. 20, pp. 5007–5013, 2020, doi: 10.1093/bioinformatics/btaa614.
- [14] M. R. Smith, "Robust analysis of phylogenetic tree space," *Systematic Biology*, vol. 71, no. 5, pp. 1255–1270, 2022, doi: 10.1093/sysbio/syab100.
- [15] M. R. Smith, "TreeDist: Distances between Phylogenetic Trees. R package version 2.11.1." 2020. doi: 10.5281/zenodo.3528124.