

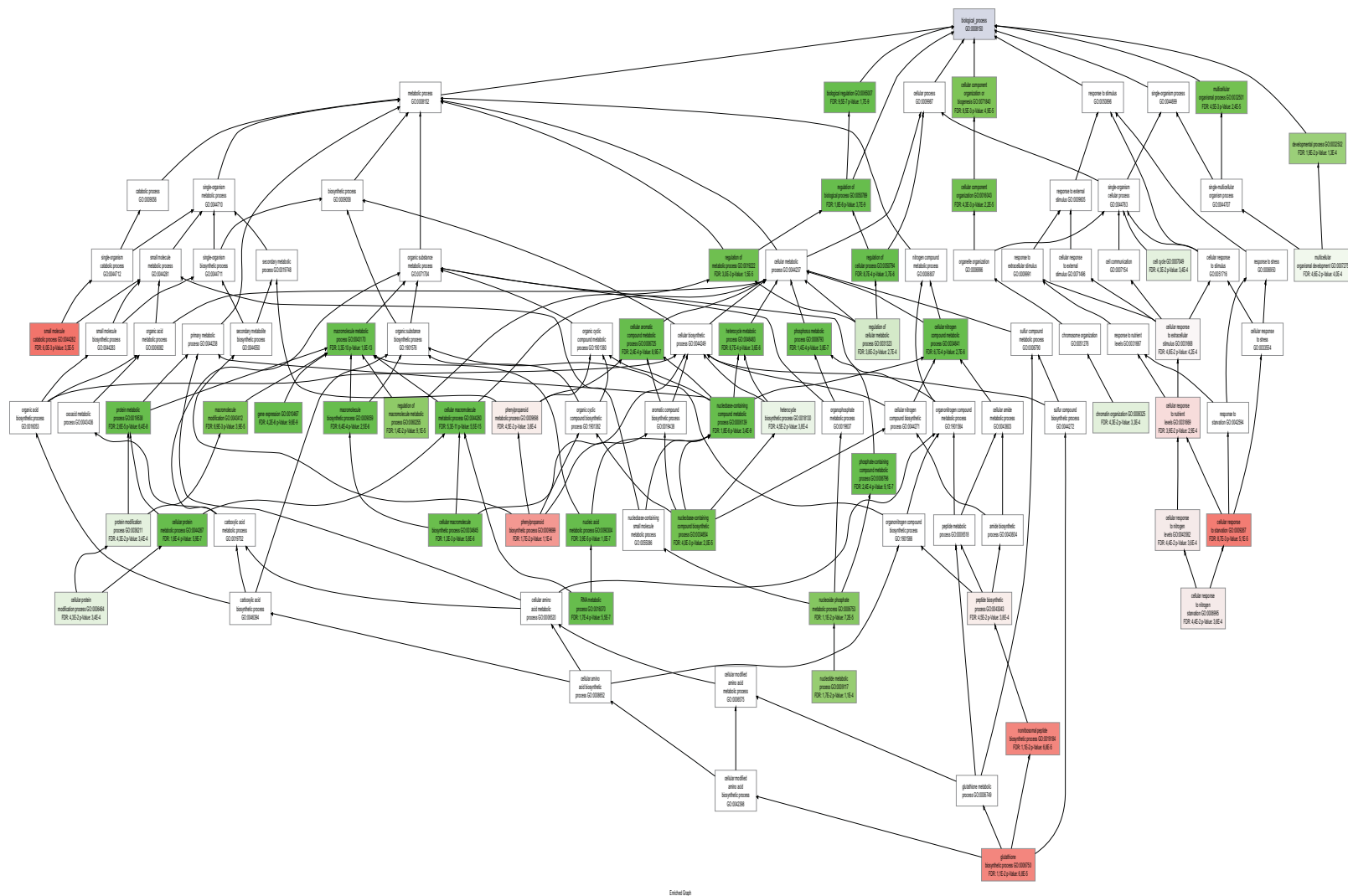
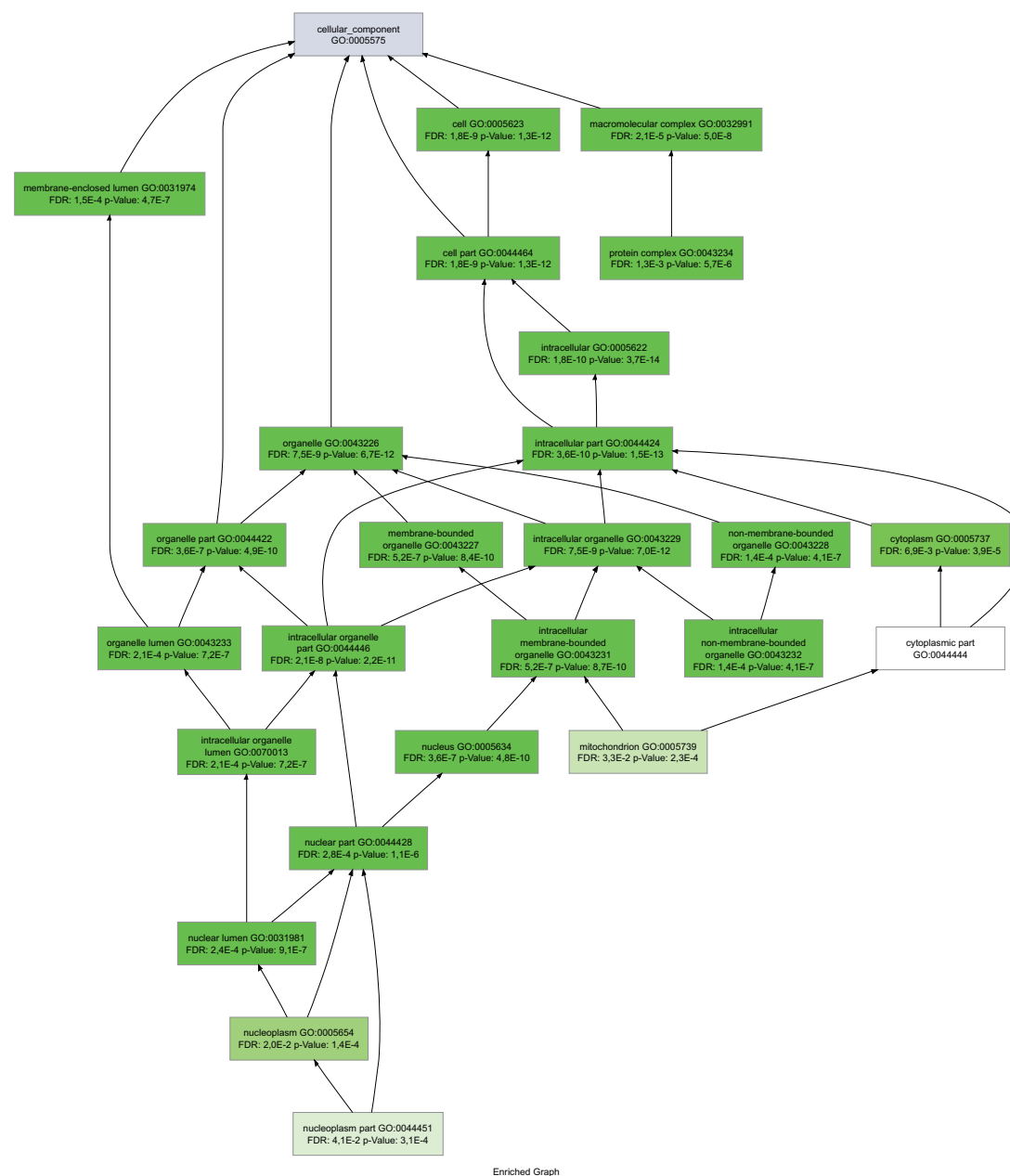


Figure S1. GO enrichment analysis of subtelomeric genes. Biological process.

Directed Acyclic Graph of the enriched GO categories. The 280 -member set of subtelomeric genes (20 genes at the telomere proximal regions of the each of seven chromosomes) was put to gene ontology analysis. The graph shows the significance of overexpressed genes and non-randomly present GO terms. Categories with significantly enriched gene numbers as determined by hypergeometric test are indicated in red, those with depleted - in green.



Enriched Graph

Figure S2. GO enrichment analysis of subtelomeric genes. Cell component.
Graph explanation as on figure S1.

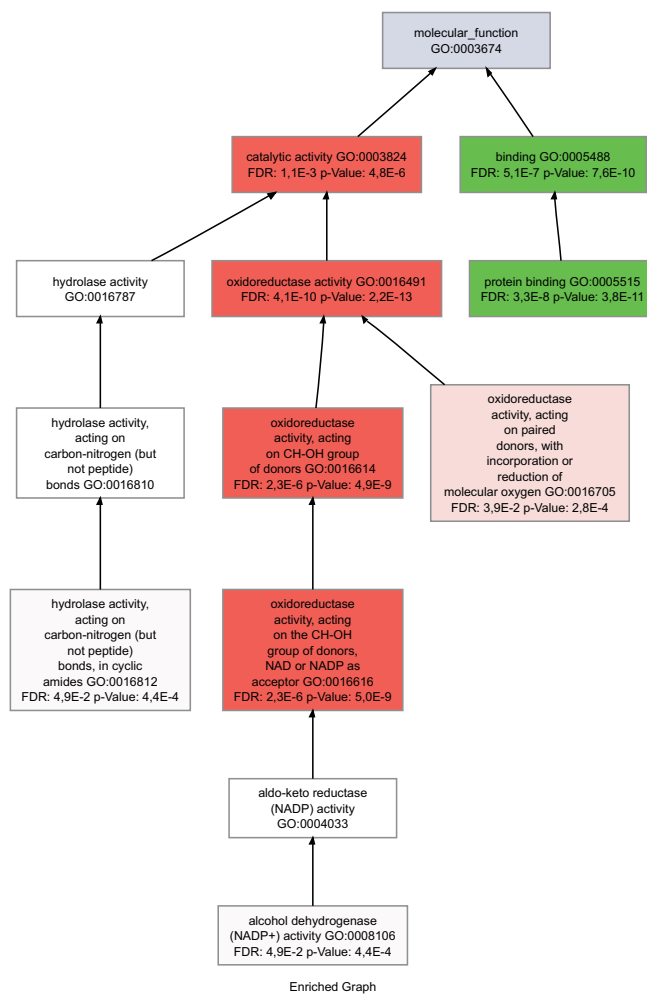


Figure S3. GO enrichment analysis of subtelomeric genes. Molecular function.
Graph explanation as on figure S1.

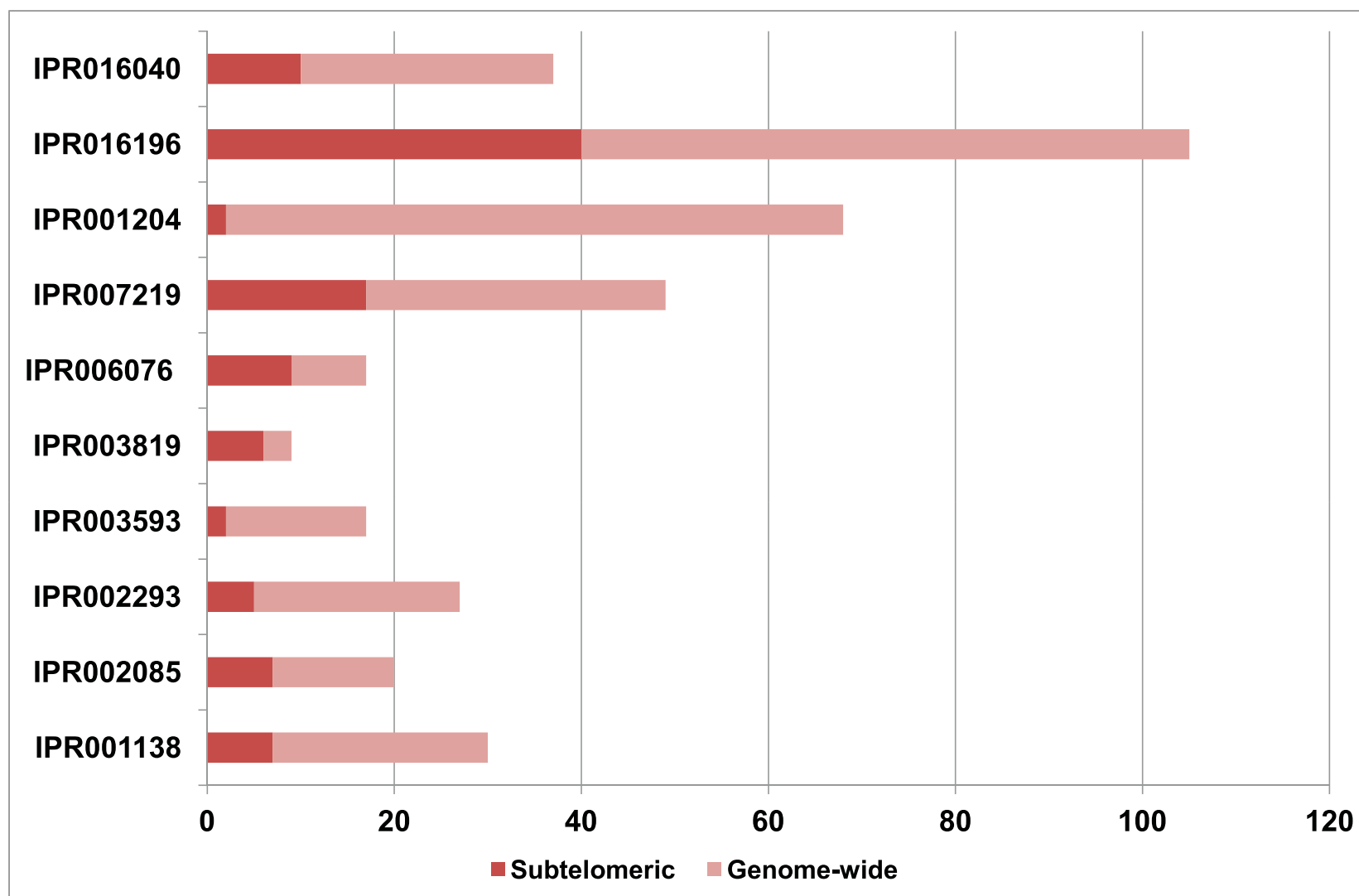
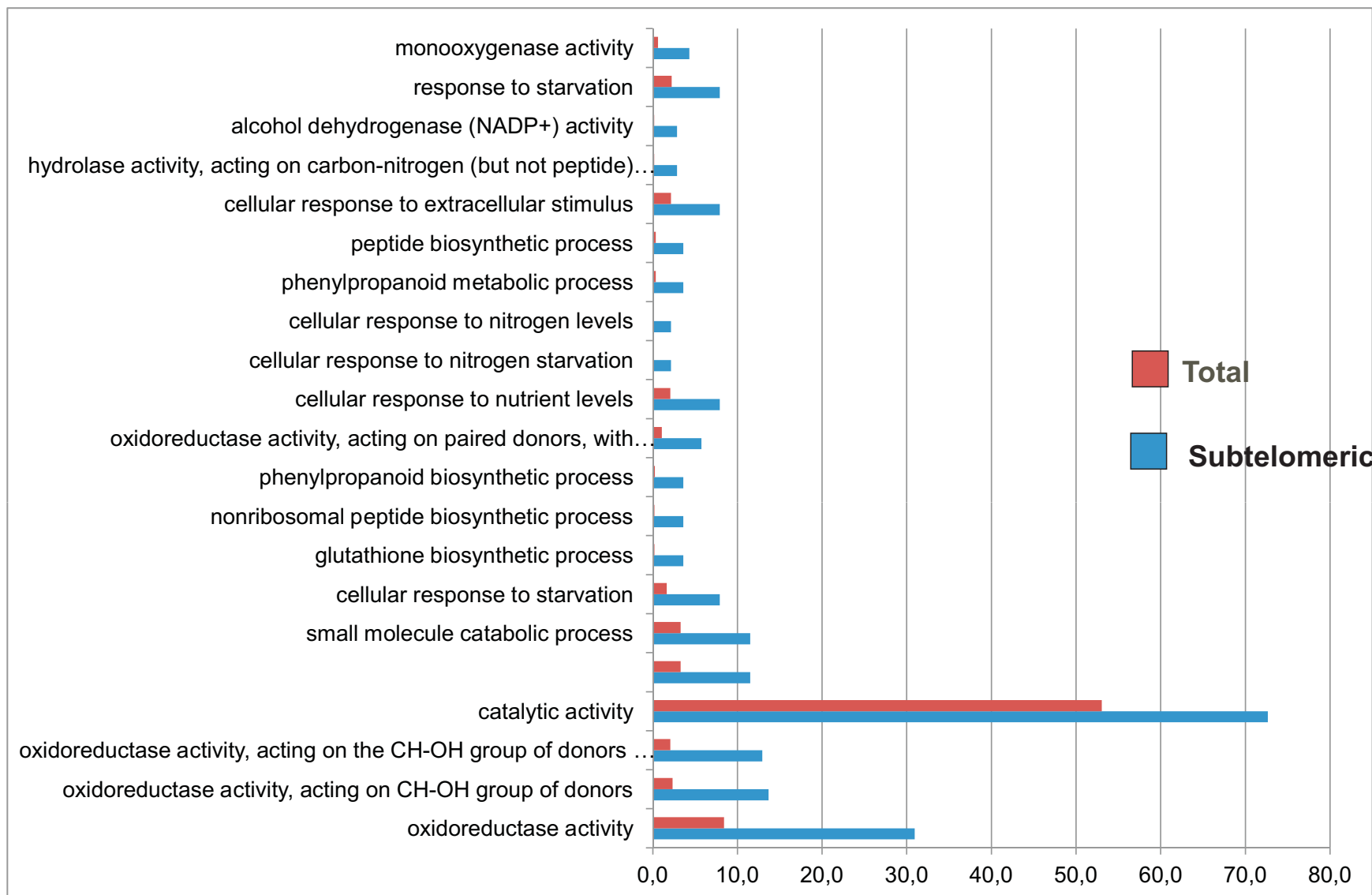


Figure S4. Characteristic subtelomeric and non-subtelomeric genes in *H. polymorpha* genome.

For each protein family the number of subtelomeric (dark purple) and “internal” (light purple) genes is plotted. Classification is based on INTREPRO categories: IPR003819 Taurine catabolism dioxygenase TauD/TfdA, IPR006076 - FAD dependent oxidoreductase, IPR007219 - Transcription factor, fungi, IPR003692- Hydantoinase B/oxoprolinase, IPR016196 - Major facilitator superfamily domain, general substrate transporter, IPR016040 - NAD(P)-binding domain, IPR001138- Zn(2)-C6 fungal-type DNA-binding domain, IPR002085 - Alcohol dehydrogenase superfamily, zinc-type, IPR002293 - Amino acid/polyamine transporter I, IPR003593 - proteins with AAA+ ATPase domain.

GO Terms



% of sequences

Figure S5. Differential GO term distribution. Subtelomeric genes.

Bar diagram of functional category enrichment analysis (Fisher's Exact Test) of *H. polymorpha* subtelomeric genes (Blue) and total set of *H. polymorpha* genes (Red). The Y-axis is the percentage of genes mapped by the term, and represents the abundance of the GO term. The percentage for the subtelomeric genes is calculated by the number of genes mapped to the GO term divided by the number of all genes in the test set. The same calculation was applied to the reference list to generate its percentage. The X-axis is the definition of GO terms.

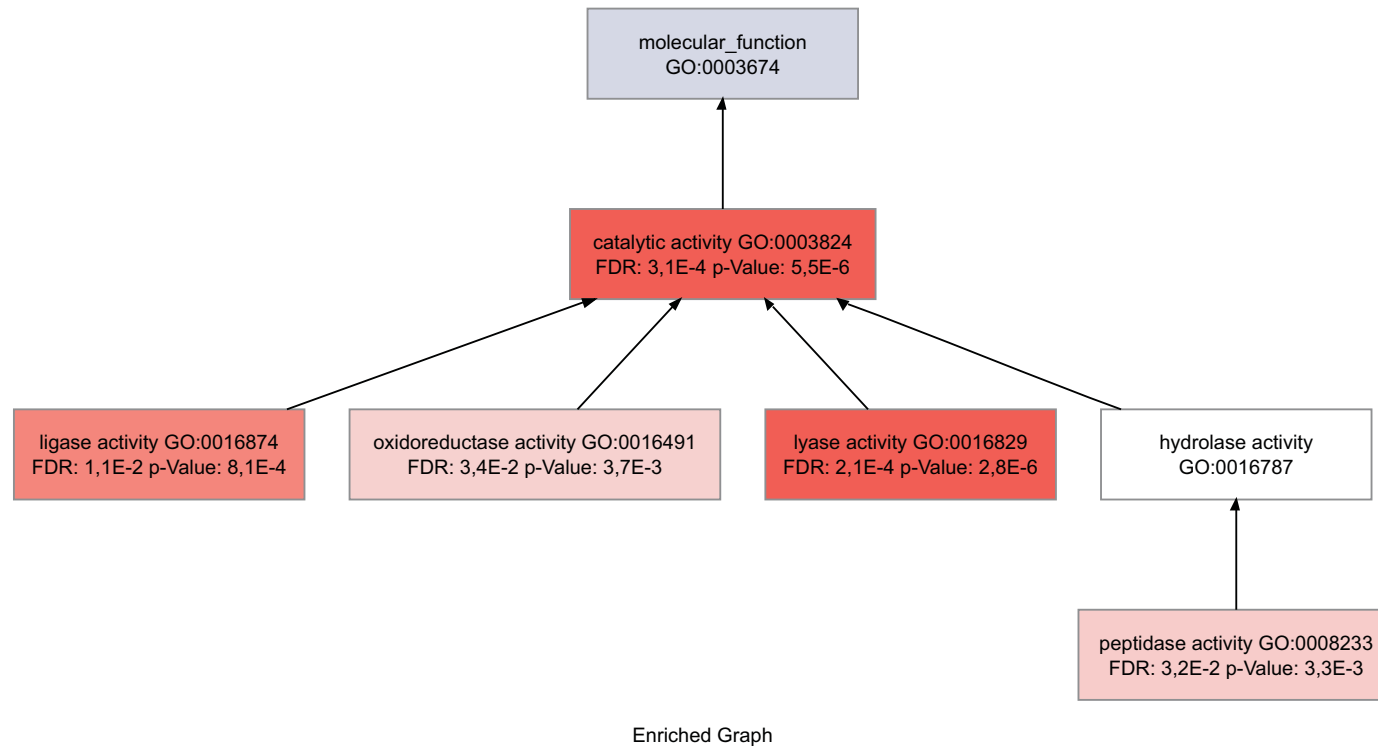
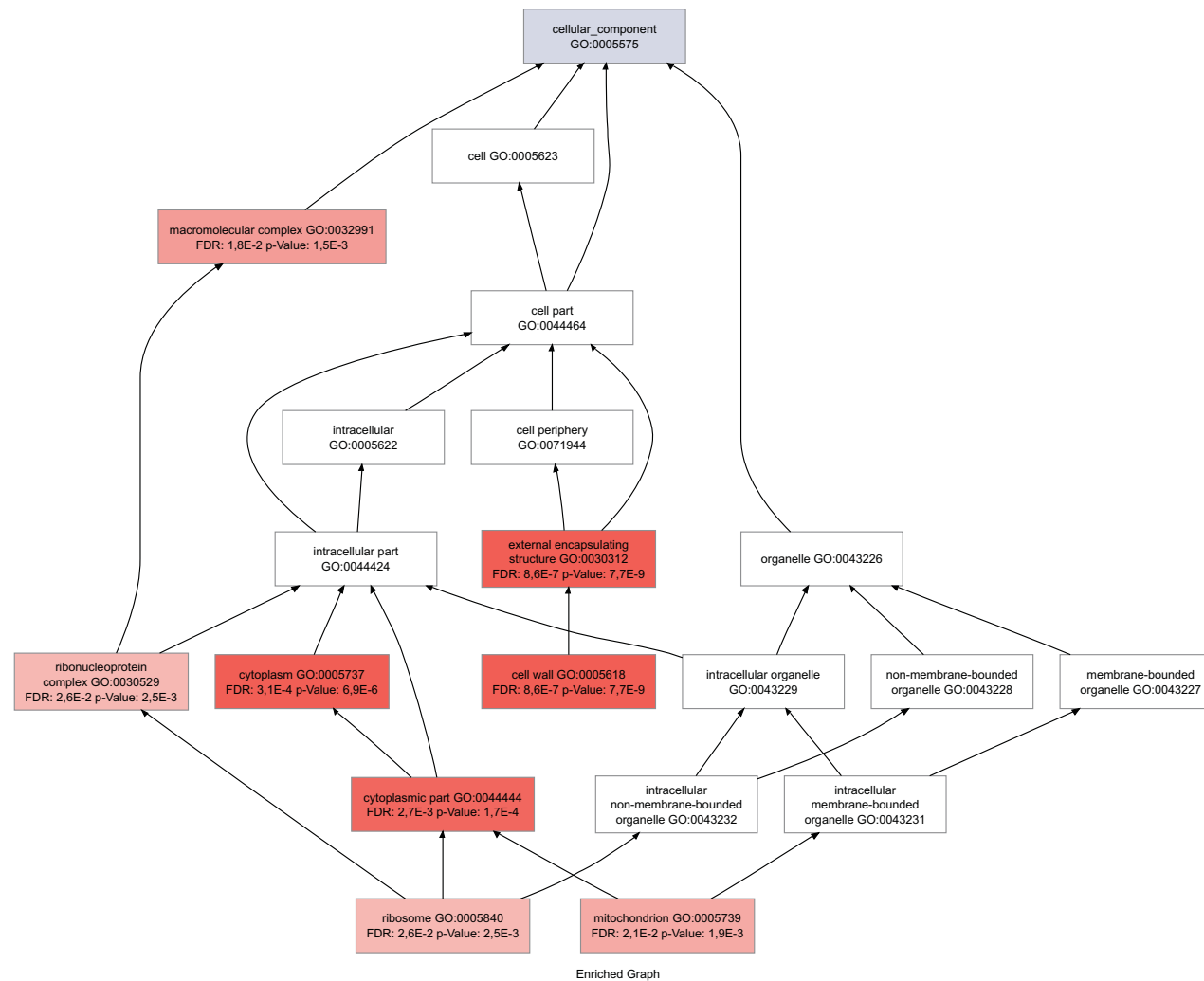


Figure S8. Genes, overexpressed on glucose. Molecular function.
Graph explanation - same as on Figure S7.



Enriched Graph

Figure S9. Genes, overexpressed on glucose. Cell component.
Graph explanation - as on Figure S7.

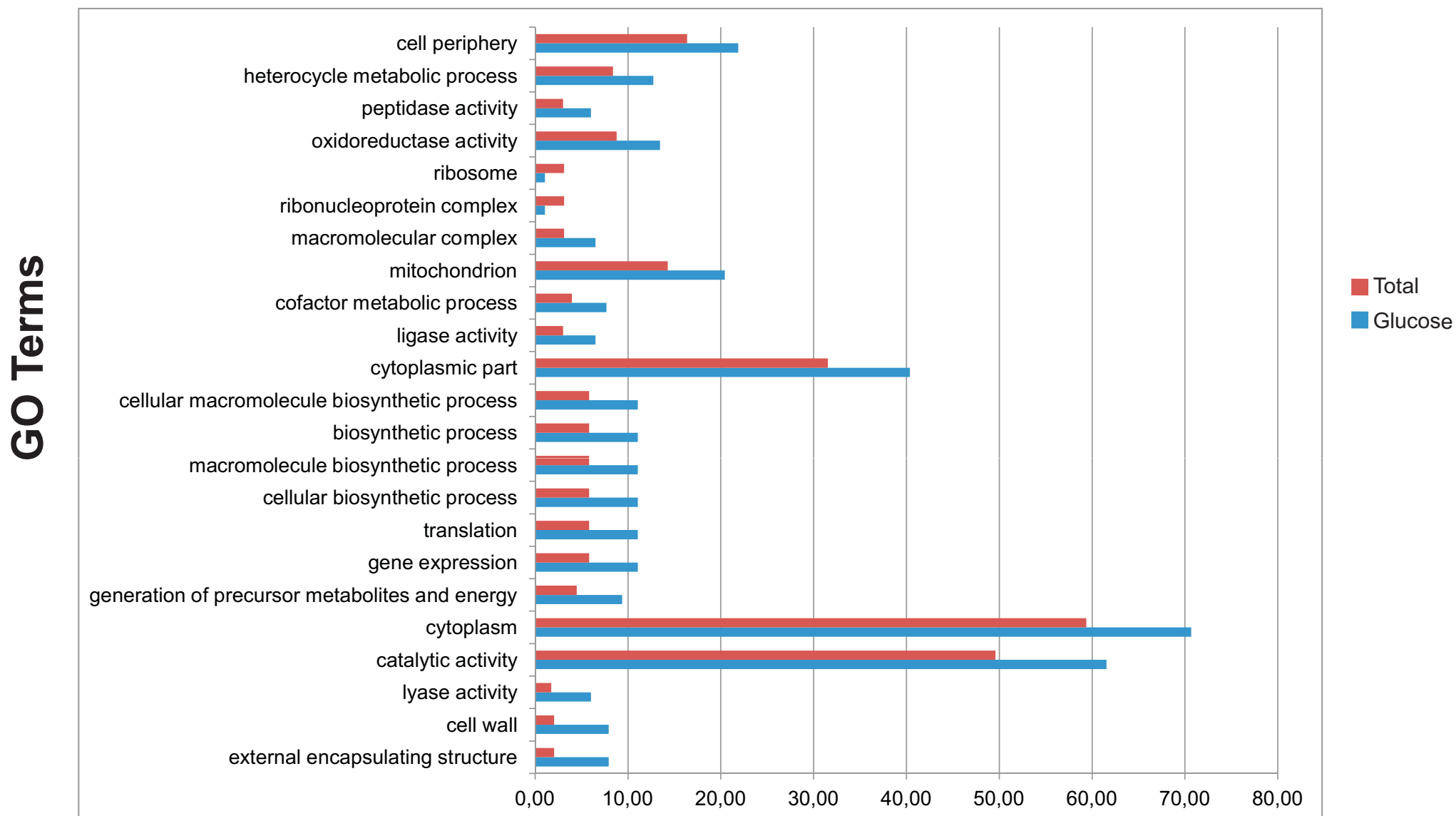


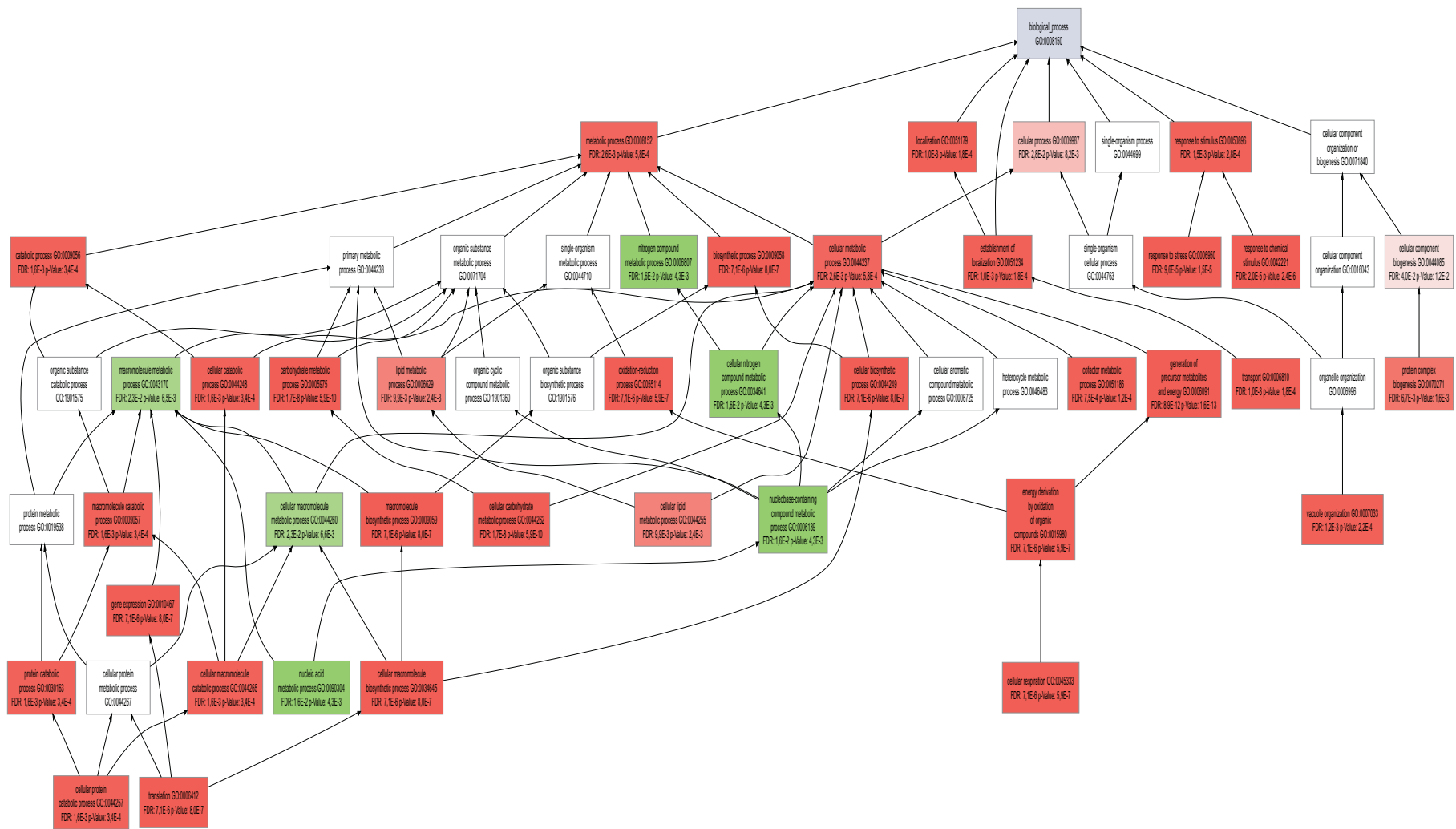
Figure S10. Differential GO term distribution. Genes overexpressed on glucose

Bar diagram of functional category enrichment analysis (Fisher's Exact Test) of *H. polymorpha* glucose overexpressed genes (Test Set) and total set of *H. polymorpha* genes (Reference Set). The Y-axis is the percentage of genes mapped by the term, and represents the abundance of the GO term. The percentage for the (Test Set) is calculated by the number of genes mapped to the GO term divided by the number of all genes in the test set. The same calculation was applied to the reference list to generate its percentage. These two lists are represented using different custom colors. The X-axis is the definition of GO terms..



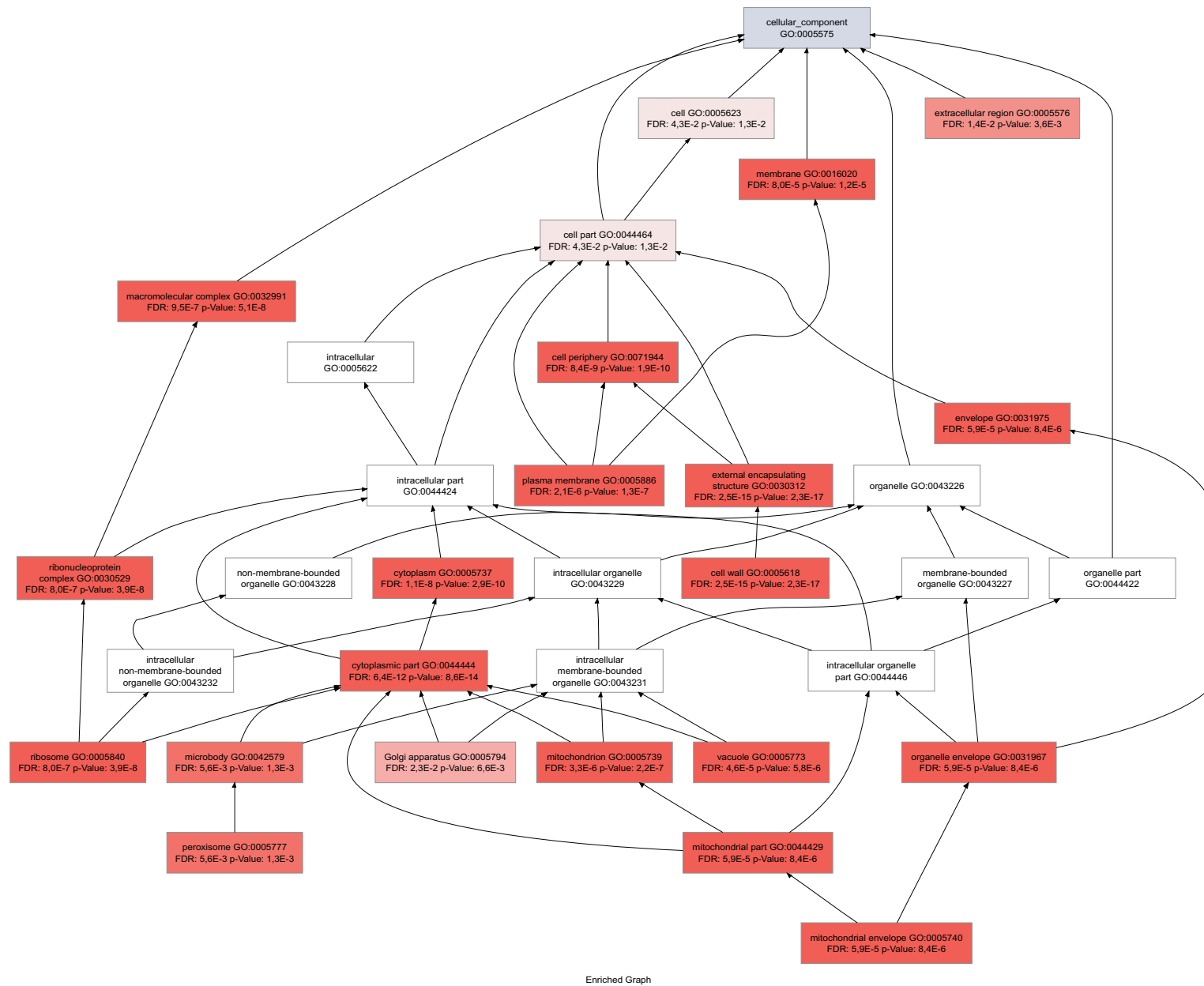
Figure S11. Genes, overexpressed on methanol. Molecular function.

Directed Acyclic Graph of the enriched GO categories. The 500-member set of genes over expressed on methanol ($\log_2$ tag count >8) was put to the gene ontology analysis. The graph shows the significance of overexpressed genes and non-randomly present GO terms. Categories with significantly enriched gene numbers as determined by hypergeometric test are indicated in red.



Enriched Graph

Figure S12. Genes, overexpressed on methanol. Biological process.
Graph explanation - as on Figure S7.



Enriched Graph

Figure S13. Genes, overexpressed on methanol. Cell component.
Graph explanation - same as on Figure S7.

GO Terms

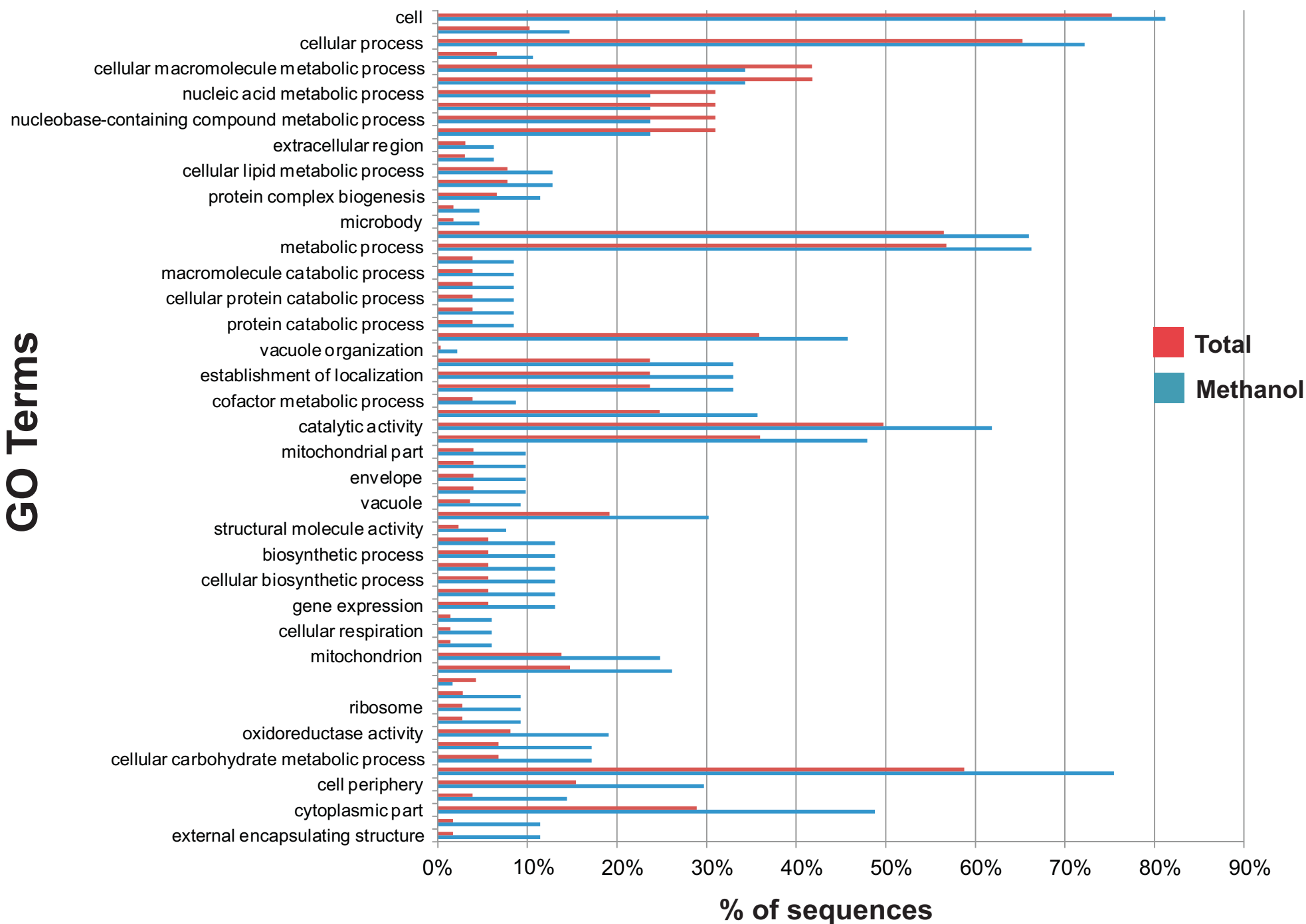


Figure S14. Differential GO term distribution. Genes overexpressed on methanol.
 Chart explanation - as on Figure S10.

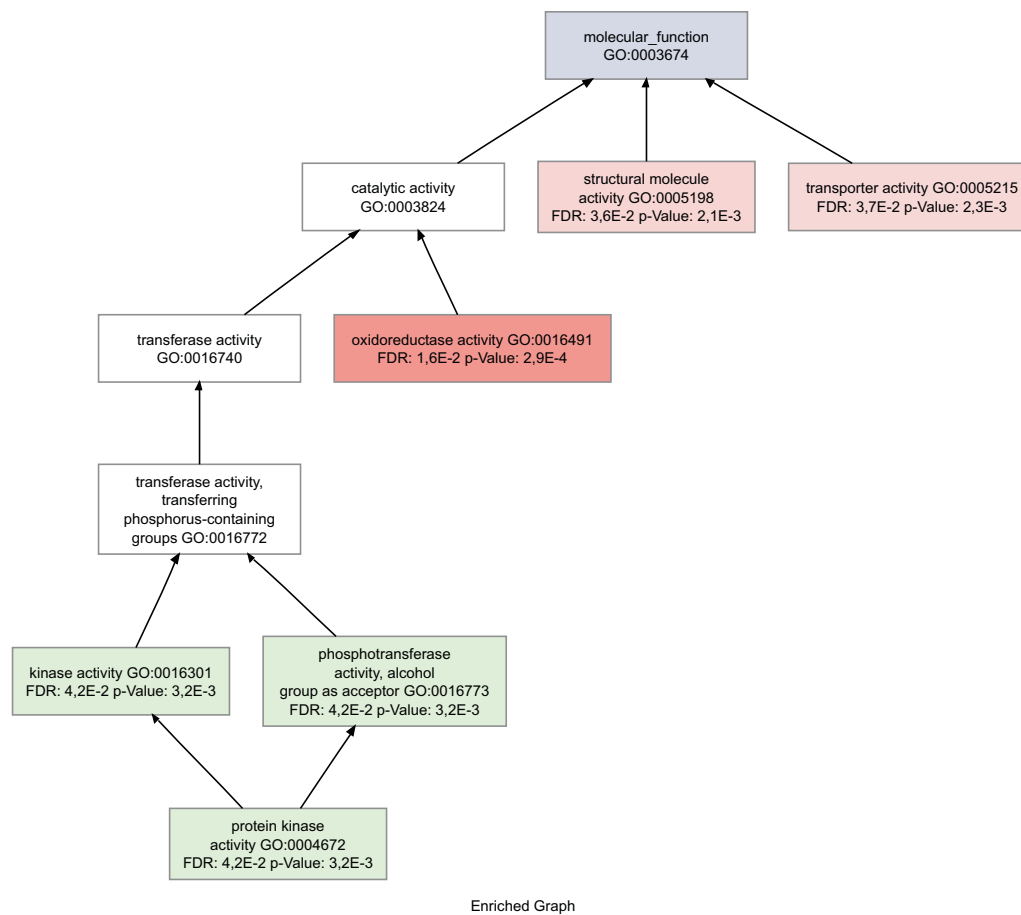


Figure S15. Genes, upregulated on methanol. Molecular function.
Graph explanation - same as on Figure S7.

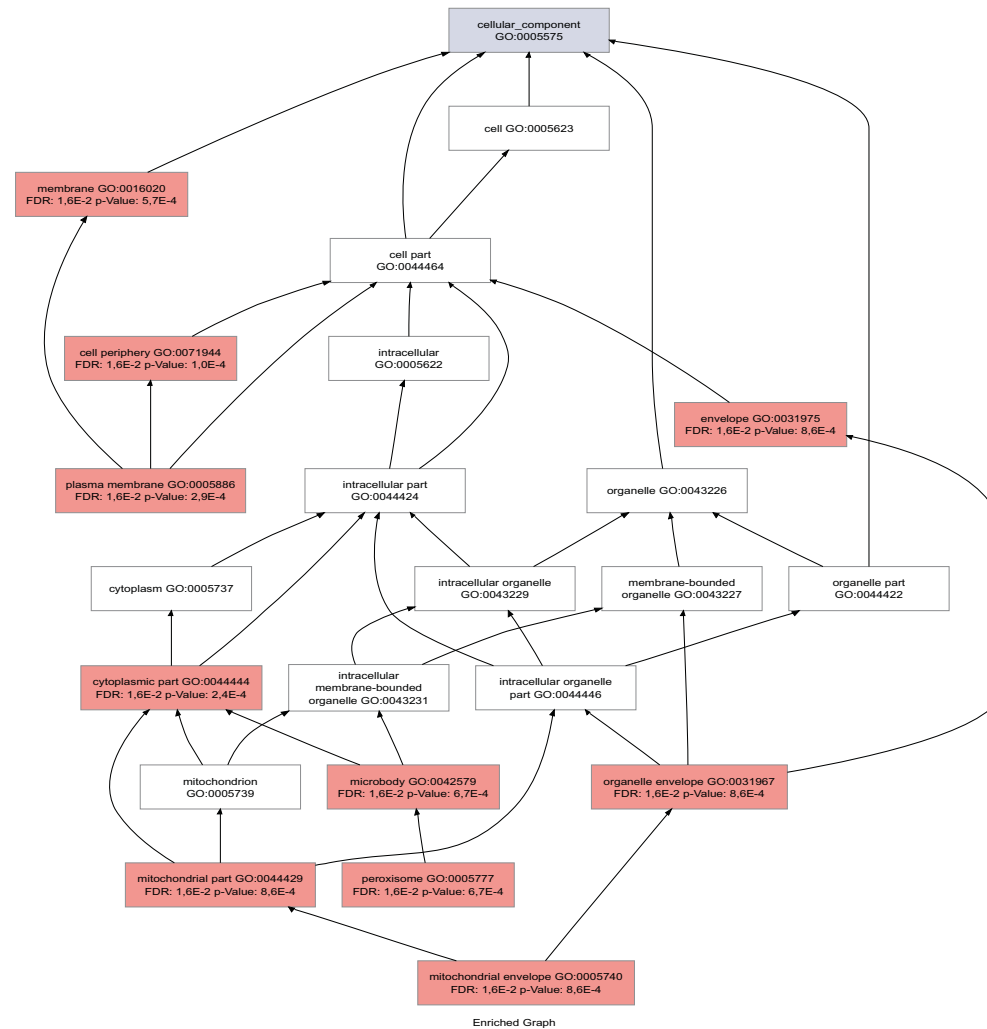


Figure S16. Genes, upregulated on methanol. Cell component.
Graph explanation - same as on Figure S7.

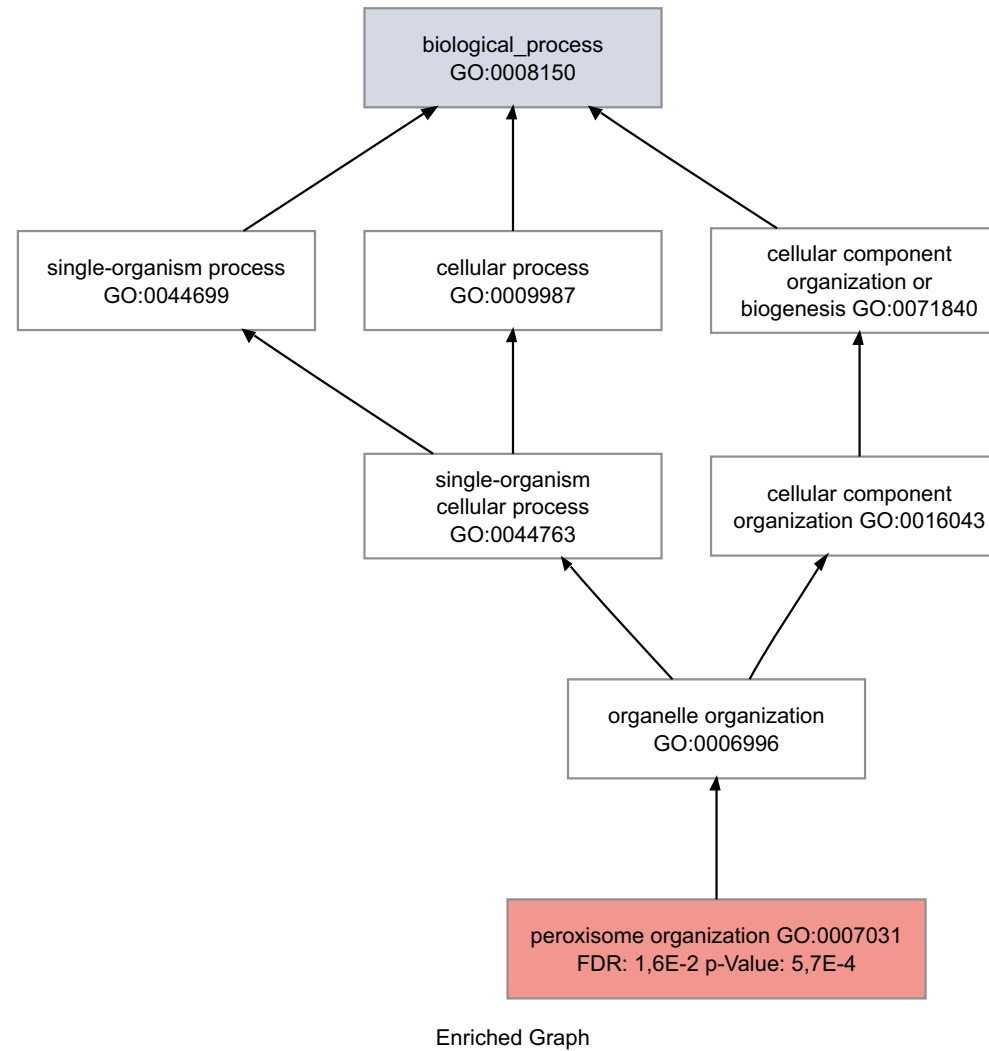


Figure S17. Genes, upregulated on methanol. Biological process.
Graph explanation - as on Figure S7.

GO Terms

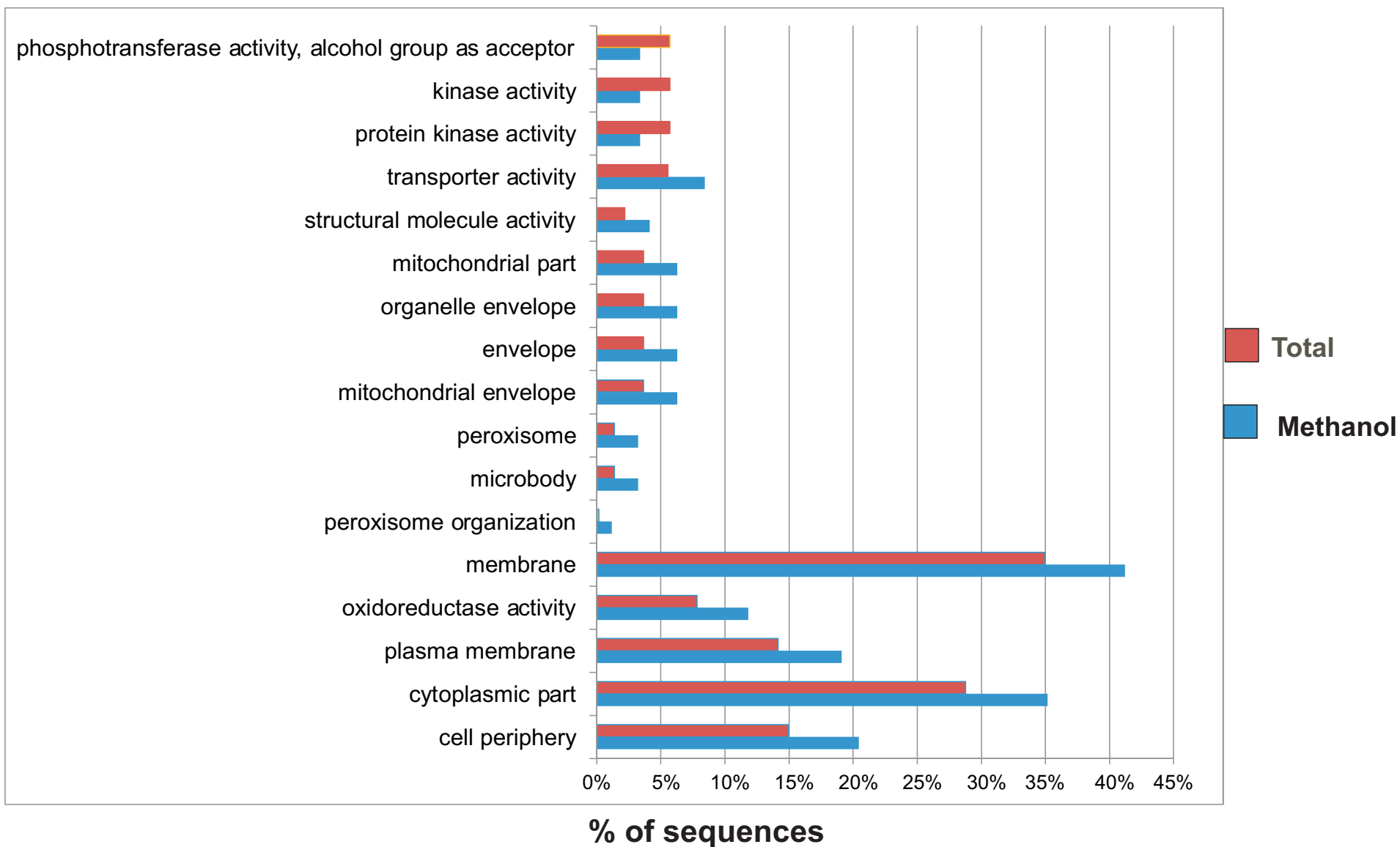
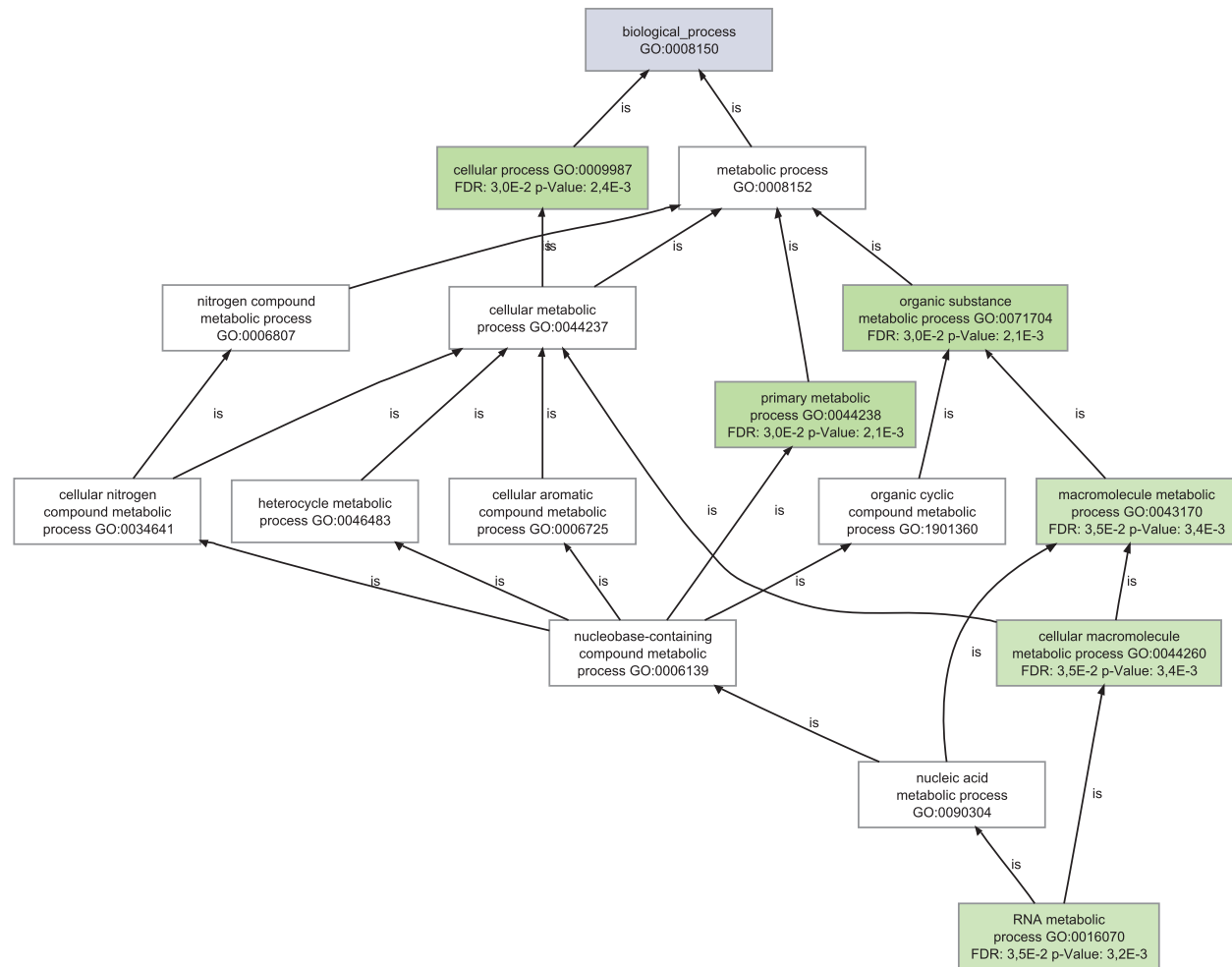


Figure S18 Differential GO term distribution. Genes upregulated on methanol
 Chart explanation - same as on Figure S10.



Enriched Graph

Figure S19. Genes, downregulated on methanol. Biological process
Graph explanation - same as on Figure S7.

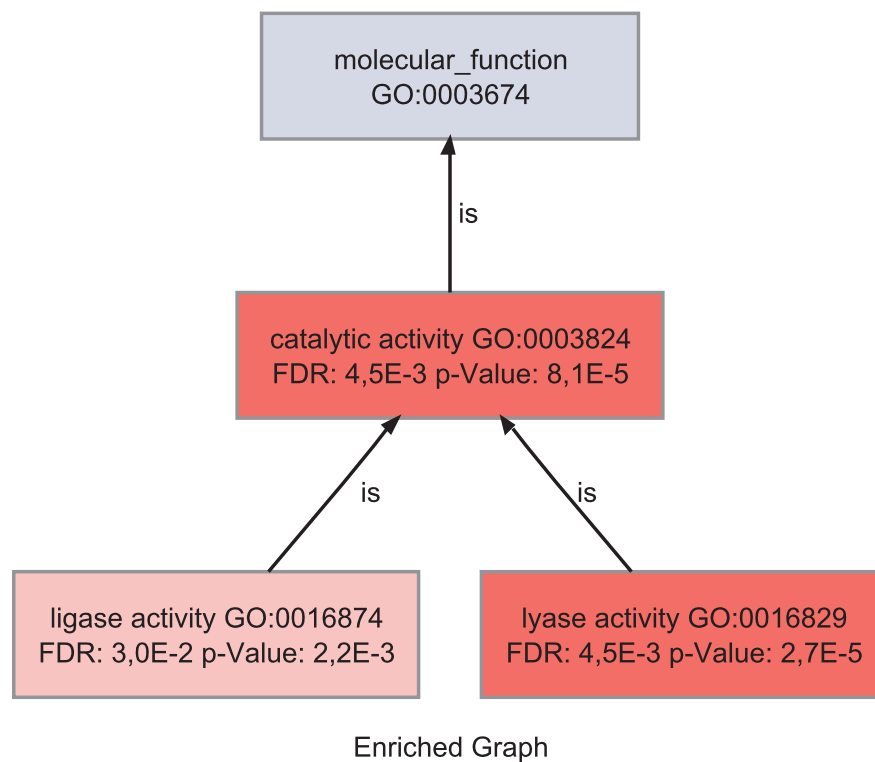


Figure S20. Genes, downregulated on methanol. Molecular function.
Graph explanation - same as on Figure S7.

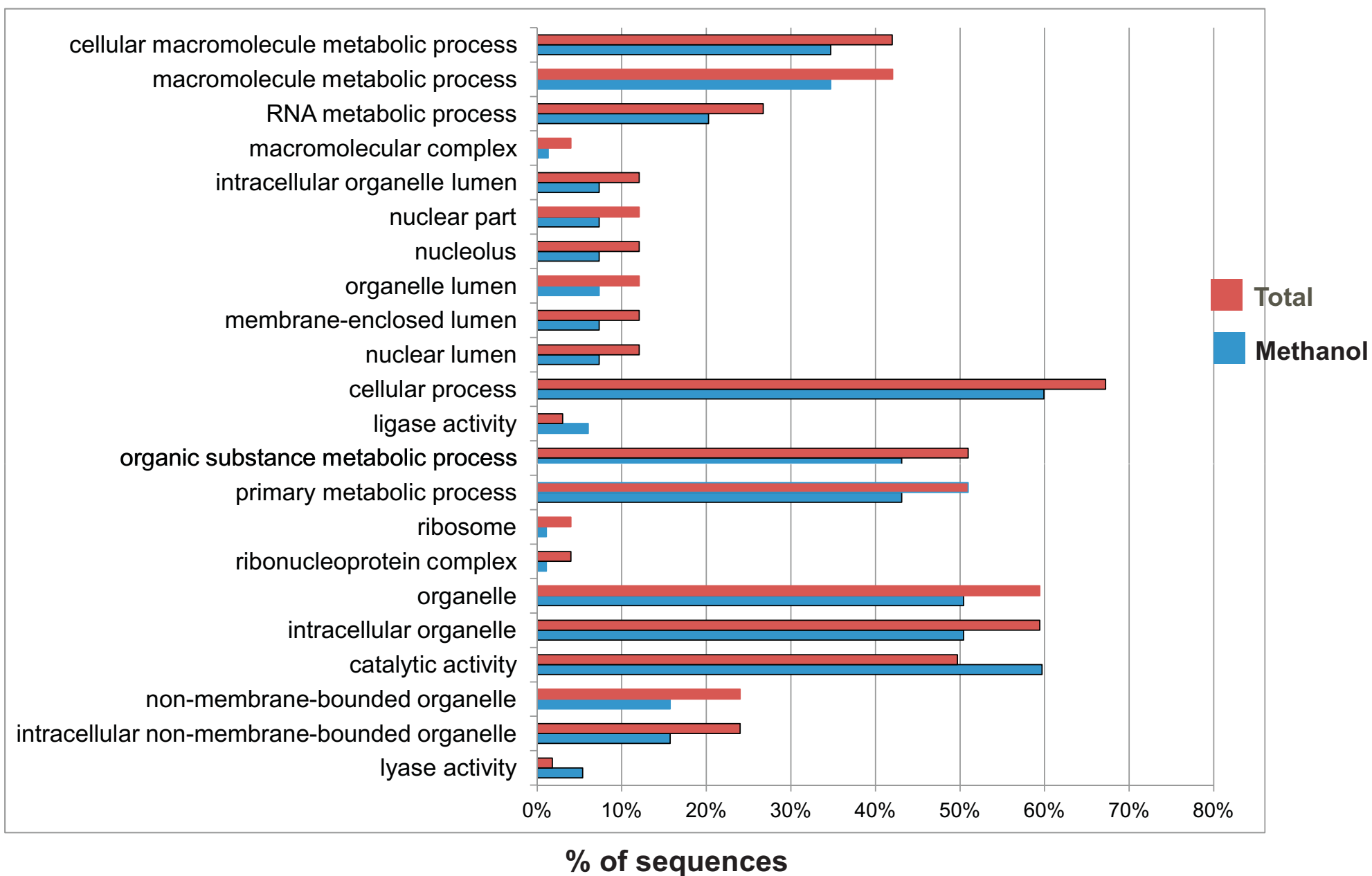


Figure S22.Differential GO term distribution.Genes downregulated on methanol.
 Chart explanation - as on Figure S10.

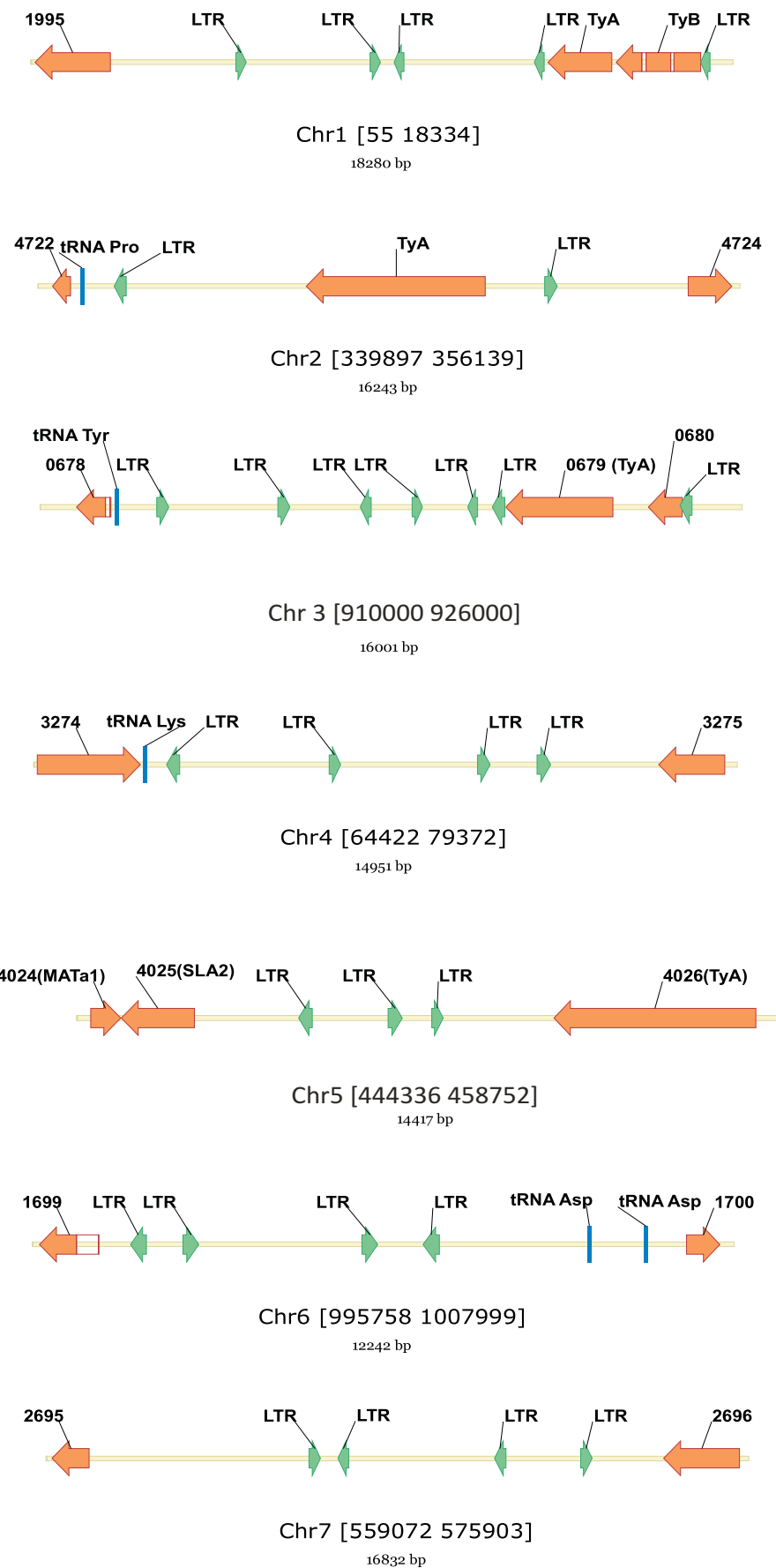


Figure S24: Clusters of LTR elements at proposed *H. polymorpha* DL-1 centromeric regions. LTRs - green arrows, numbers below maps correspond to chromosomal positions of corresponding loci, protein coding genes - orange arrows, numbers correspond to IDs of annotated genes.

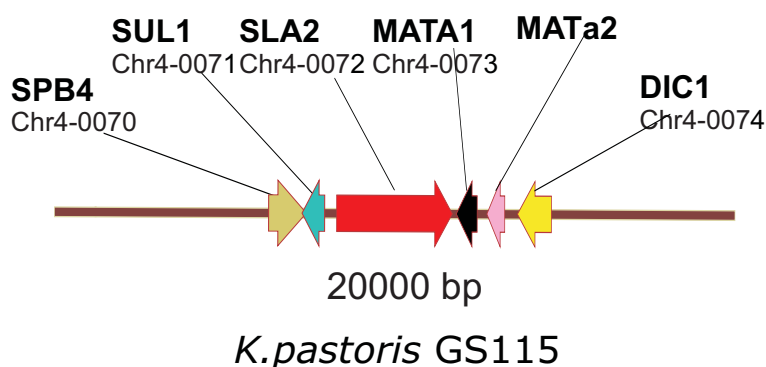
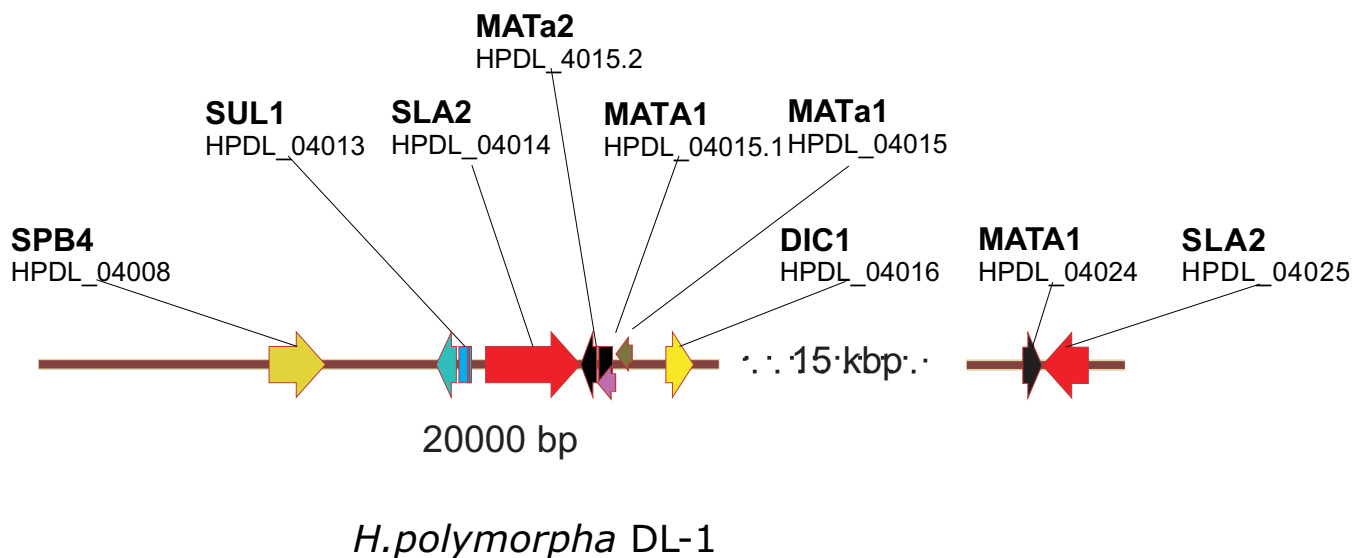
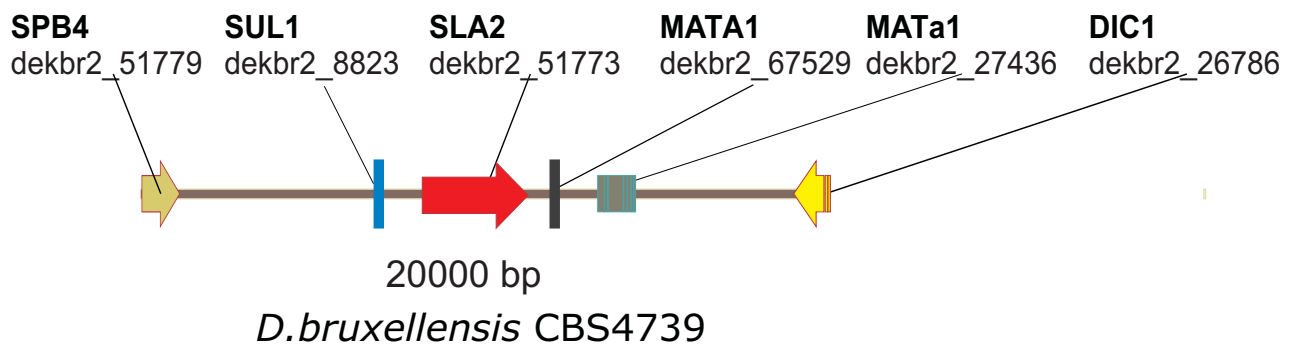


Figure S25. Comparative organization of MAT loci in the three yeast species. Genes are shown as arrowed boxes, identical coloring indicates orthology, gene names correspond to established nomenclature of MAT loci associated genes (Butler et al., 2004). “DIC” gene encodes mitochondrial oxoglutarate/malate carrier protein, “SUL1”- eukaryotic translation initiation factor eIF-1, “SLA2”- Actin-binding protein SLA2/Huntingtin-interacting protein Hip1, “MATA1”- MATa1 transcription factor, “MATa1” - MATalpha1 transcription factor, “MATa2” - MATalpha2 transcription factor, “SPB4” - ATP-dependent RNA helicase.

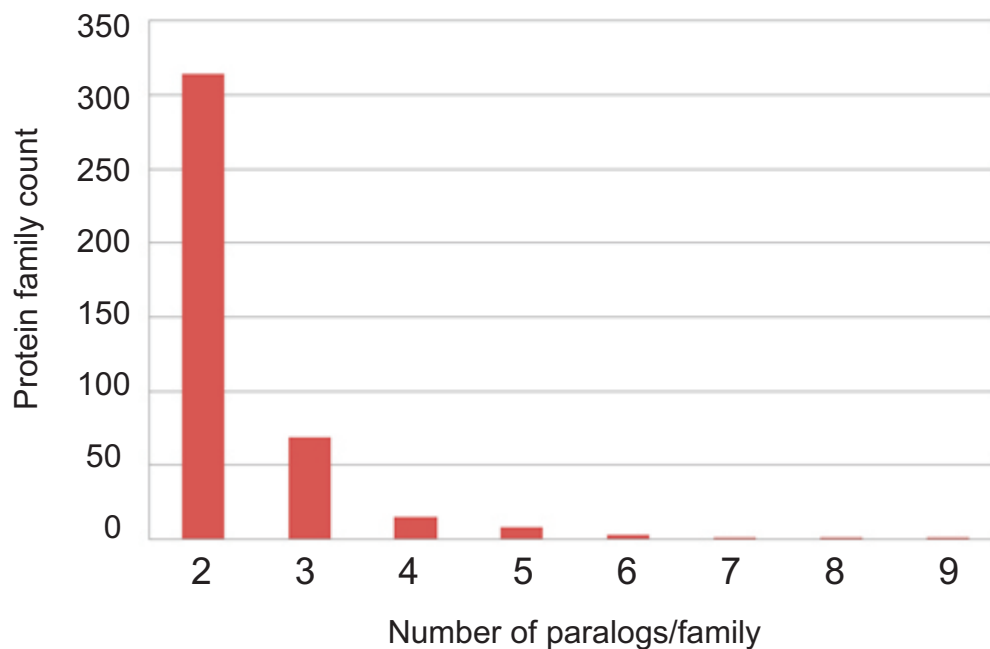
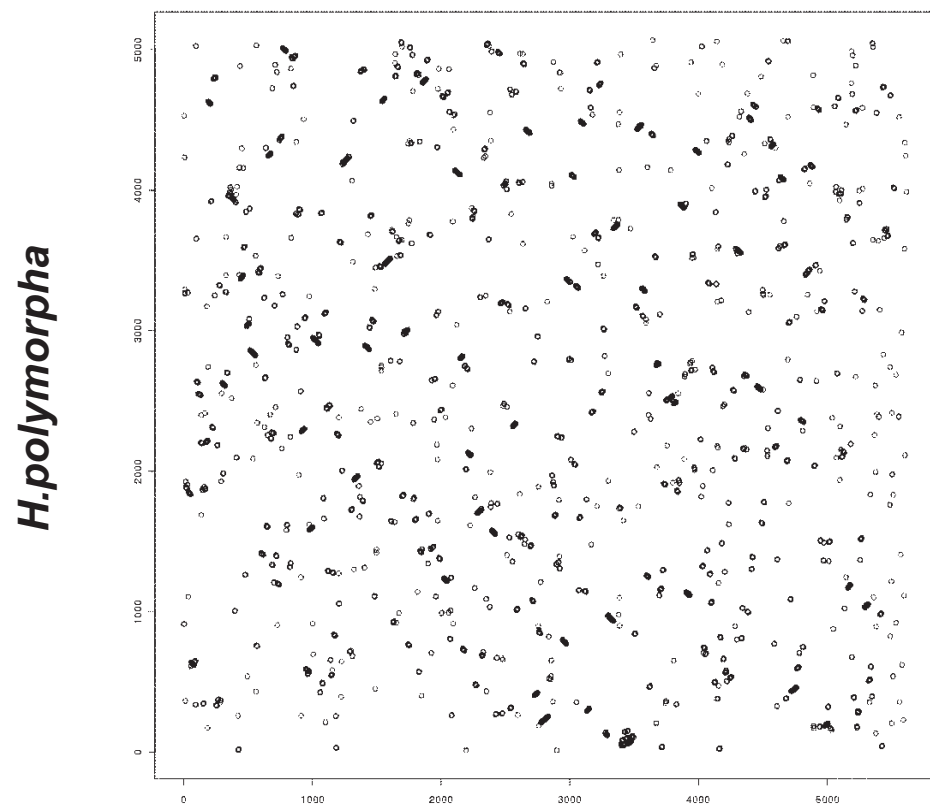
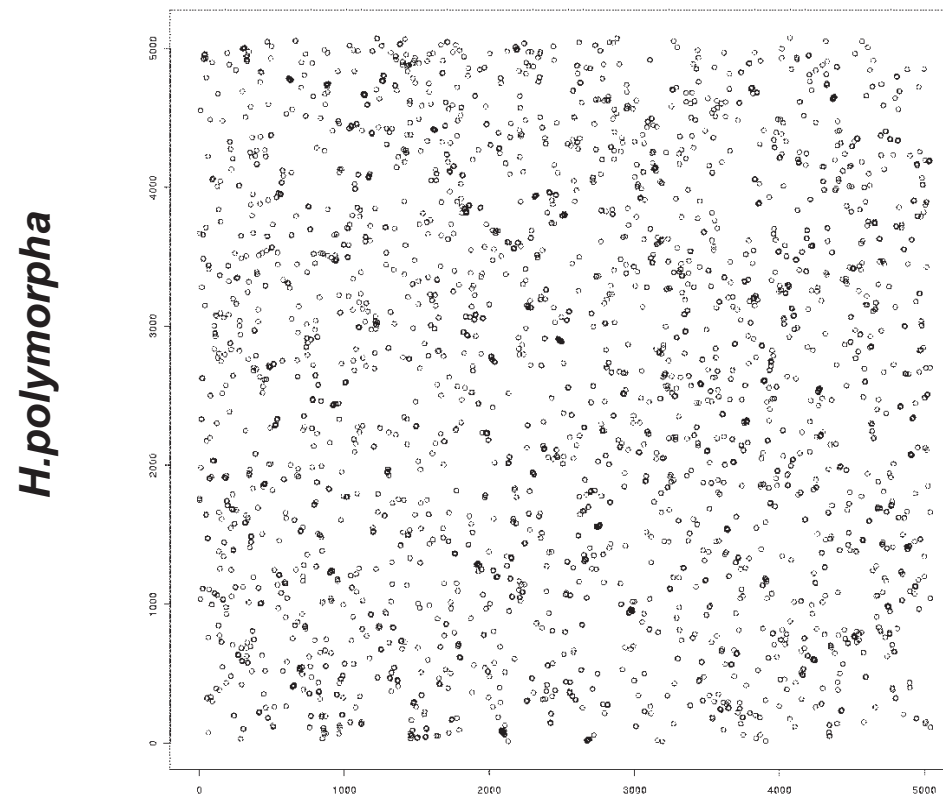


Figure S26. Evaluation of degree of paralogy in *H. polymorpha* genome. Distribution of *H. polymorpha* protein families identified using OrthoMCL cluster analysis according to the number of paralogs per family (X axis).



D.bruxellensis



P.pastoris

Figure S27. Genomic dot-plot for estimation of degree of syntenic conservation in three yeast genomes.

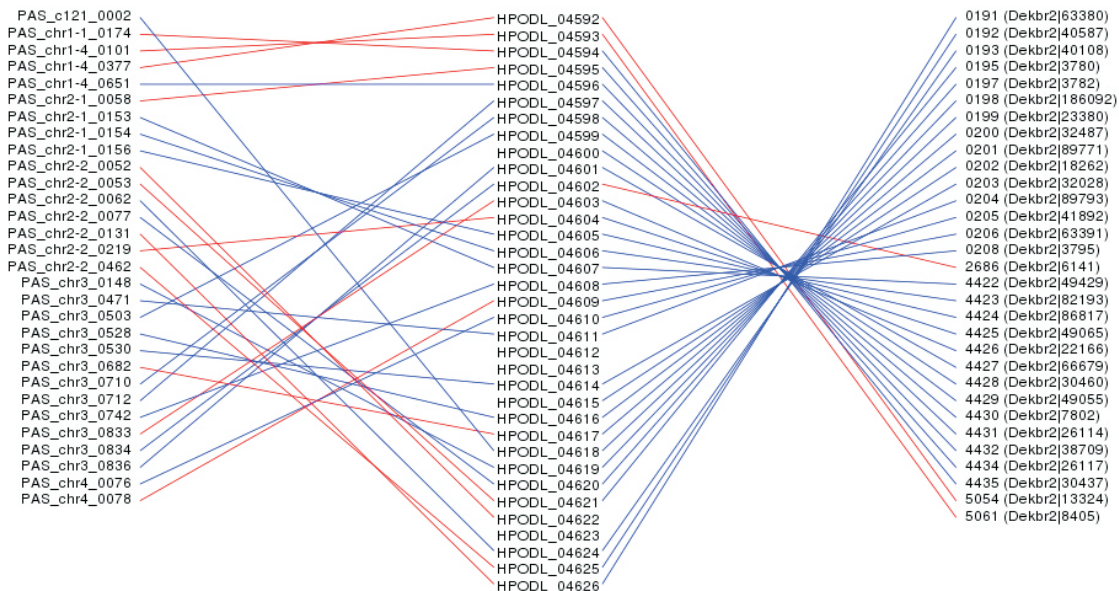
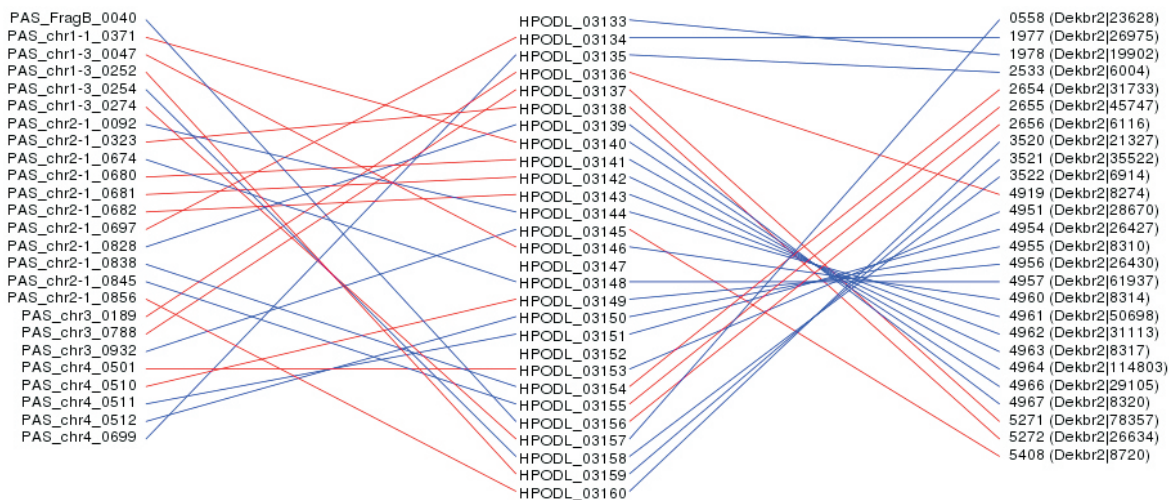
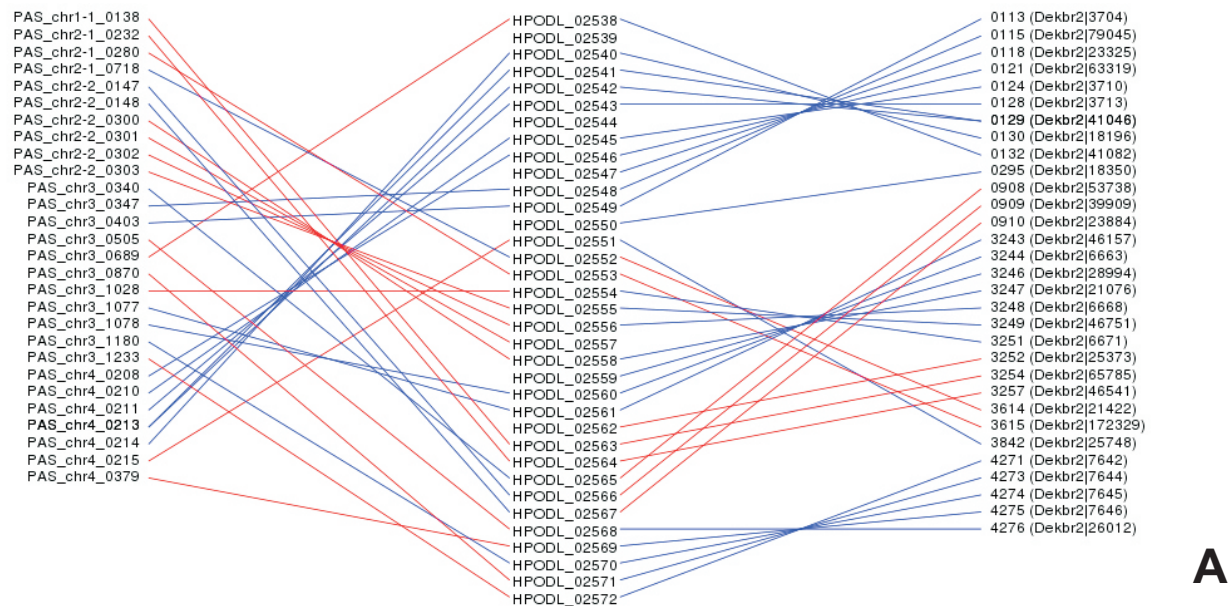


Figure S28. Synteny maps of *H.polymorpha* “MUT” loci (central column) and orthologous loci in *D.bruxellensis* (right column) and *P.pastoris* (left column) genomes

A- FLD locus B - FDH locus C- DAS .

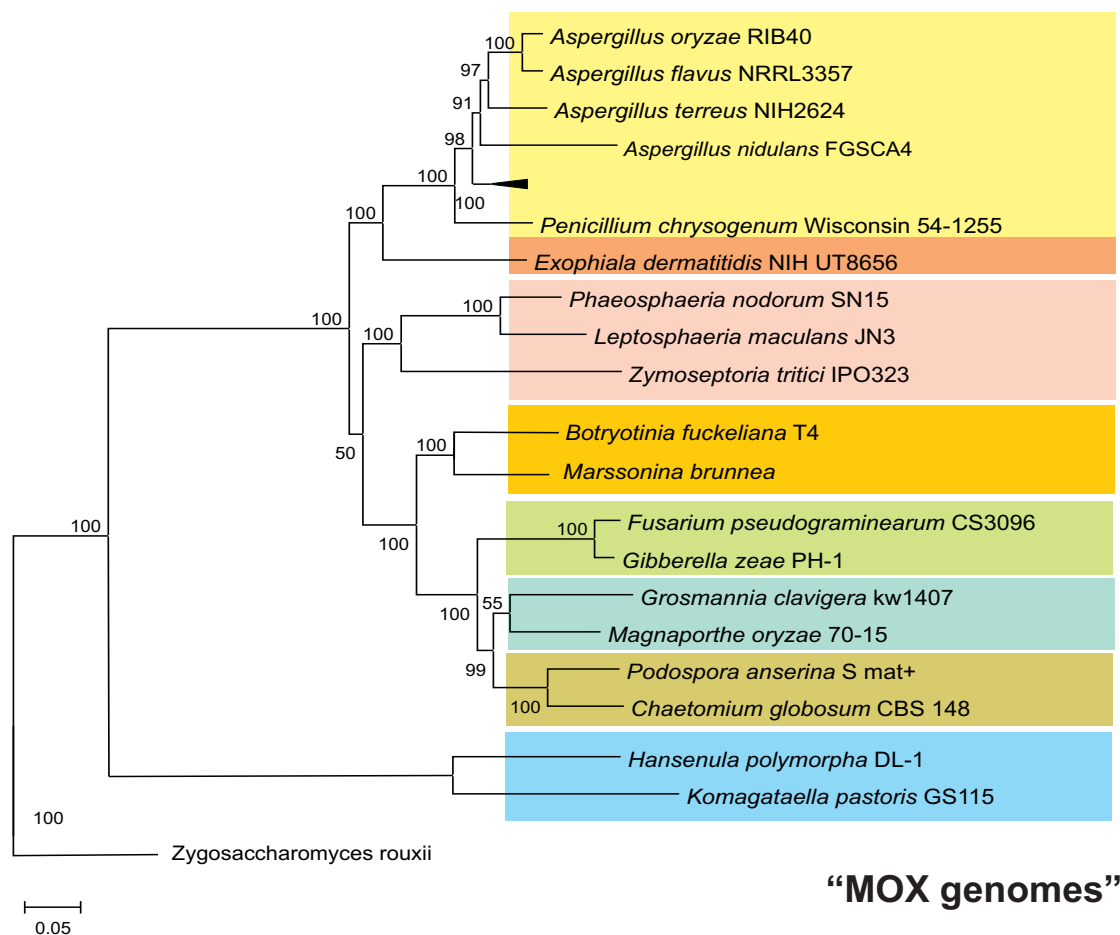
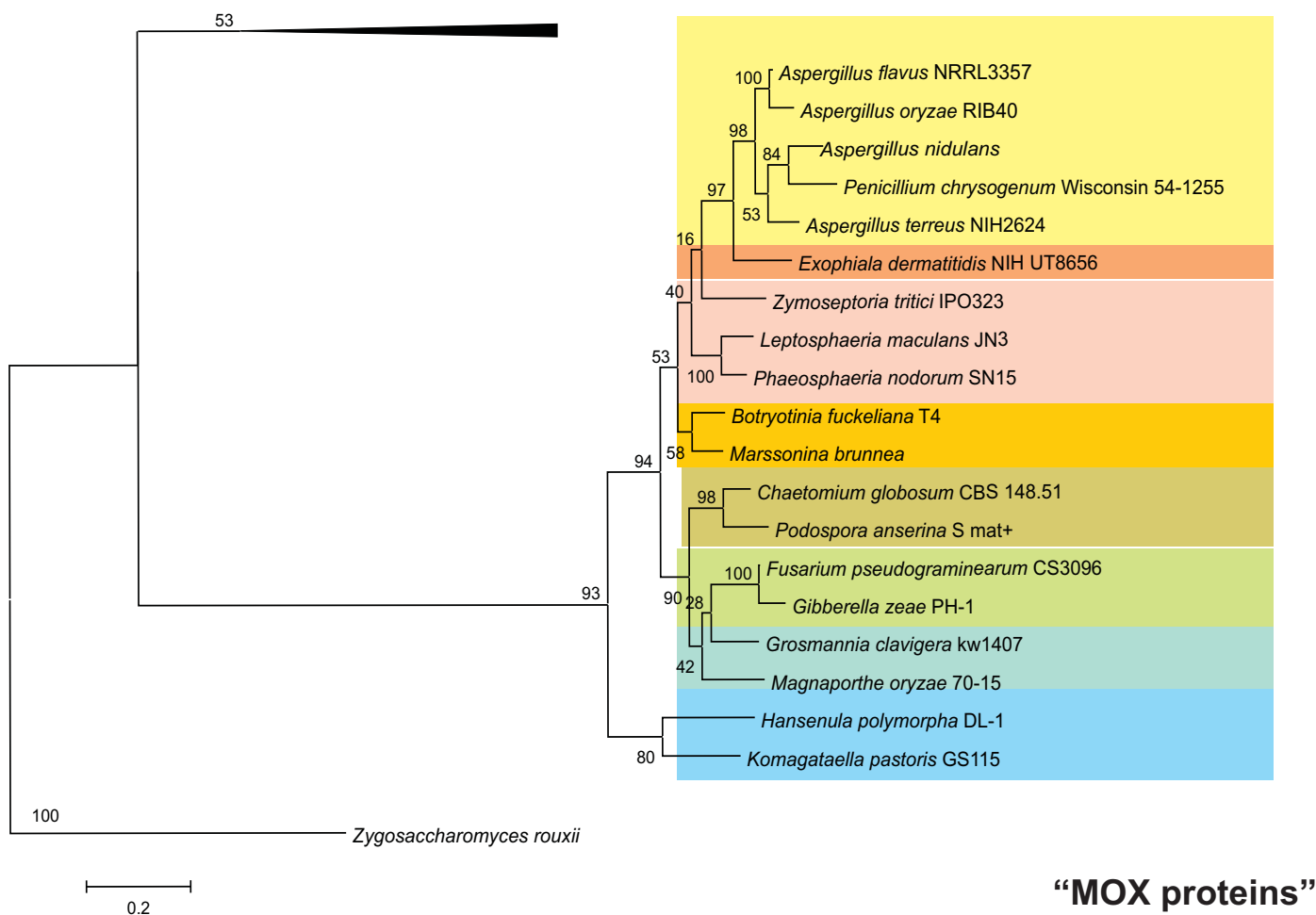


Figure S29. Comparison of phylogenetic trees of MOX proteins and phylogenomic trees of MOX genomes in Ascomycetes.

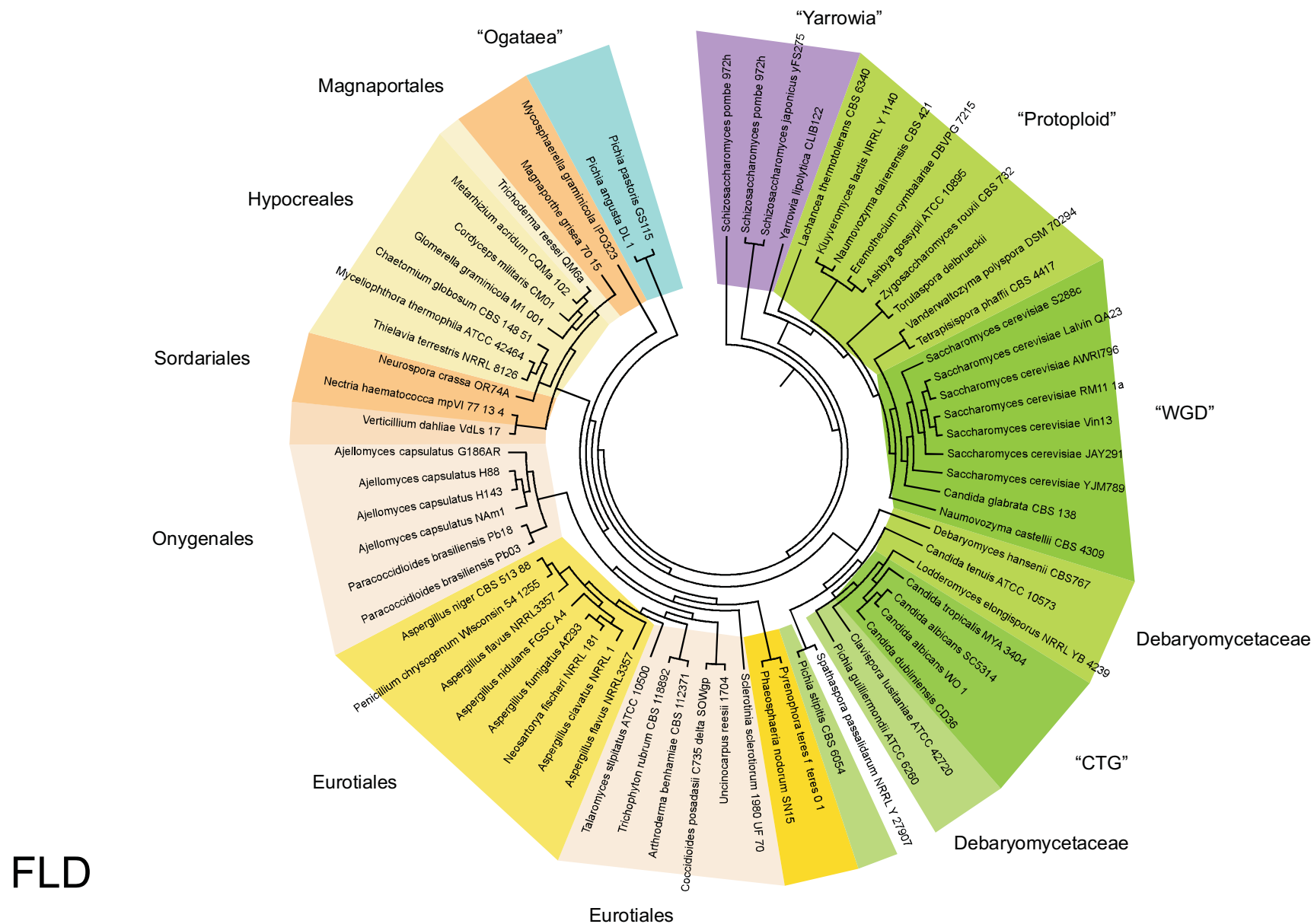


Figure S31. Phylogenetic tree of FLD orthologs in Ascomycetes. Tree construction procedure and color code as on Fig.30

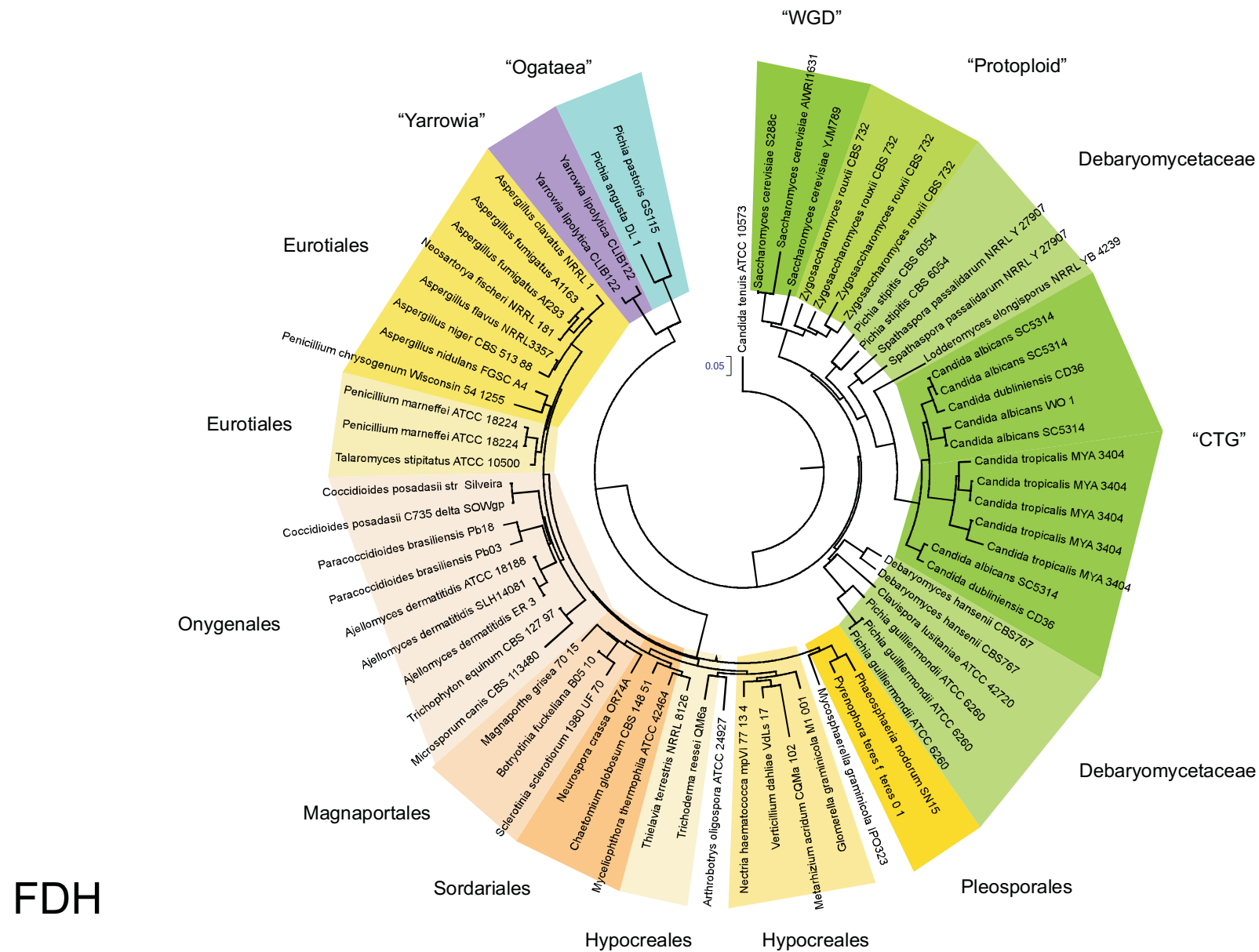


Figure S32. Phylogenetic tree of FDH orthologs in Ascomycetes. Tree construction procedure and color code as on Fig.30