

Figure S1. Phylogenetic tree of Mcm proteins inferred from maximum likelihood analysis. Numbers on the branches indicate bootstrap support from 1000 replicates. Among-site substitution rate heterogeneity was corrected using two gamma distributed substitution rate categories and LG substitution model for amino acid substitutions (LG+G). Diatom sequences are indicated in bold. The codes before the taxon name are the corresponding sequence IDs (For respective databases refer to materials and methods section).

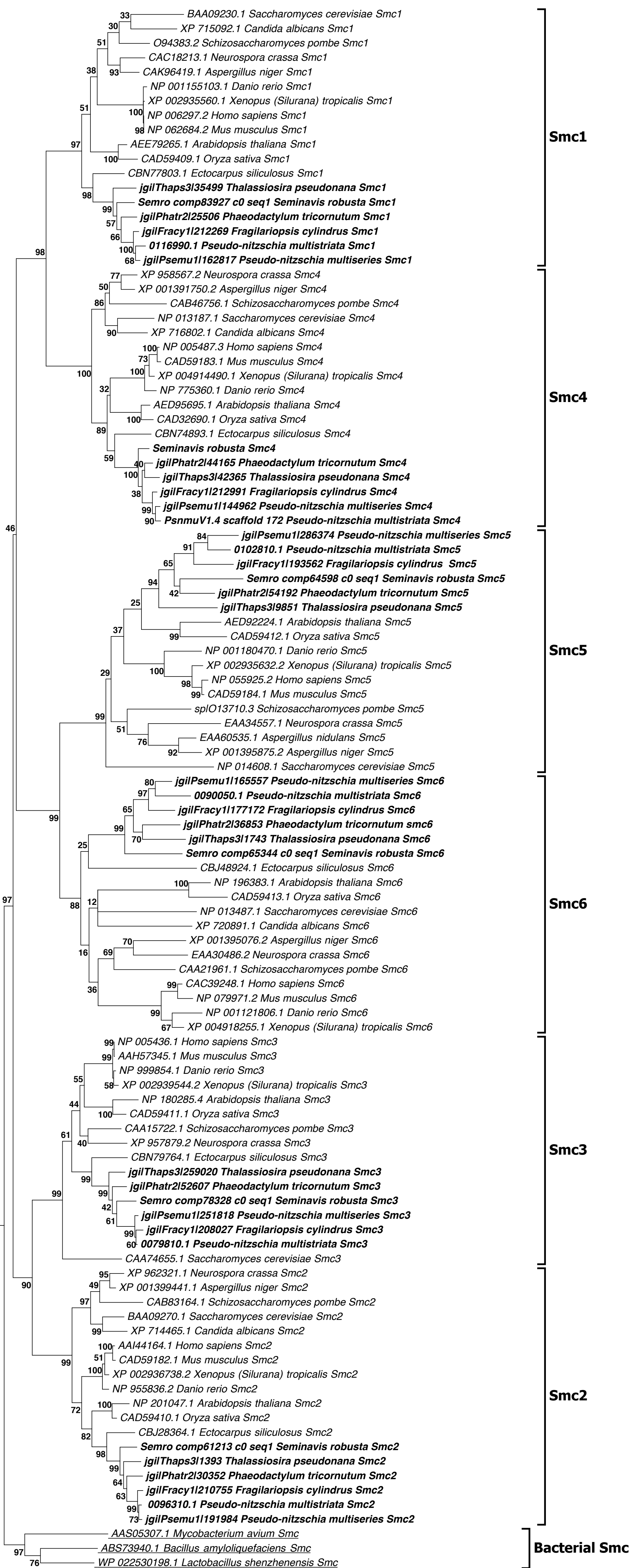
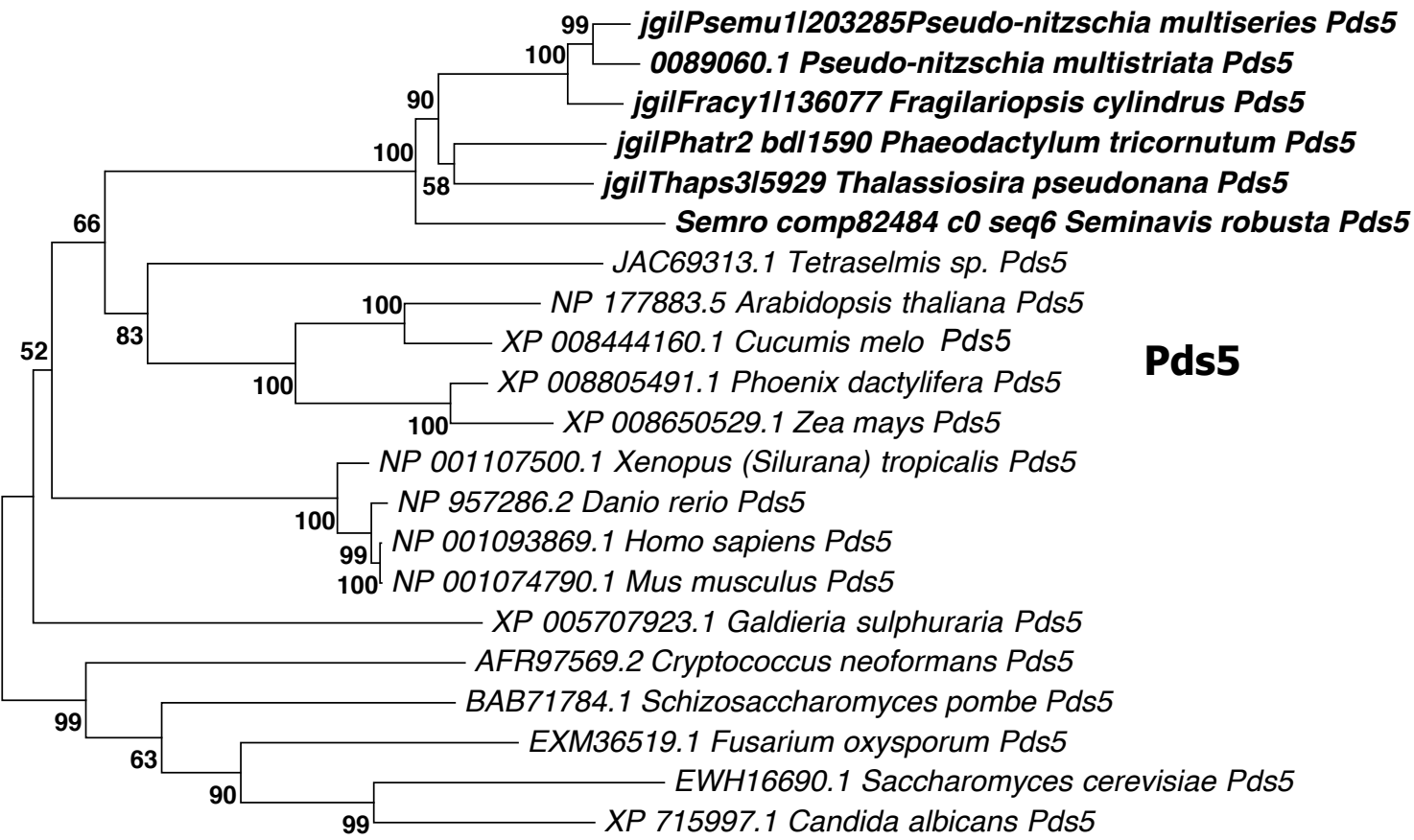


Figure S2. Phylogenetic tree of Smc proteins inferred from maximum likelihood analysis, rooted with bacterial Smc proteins as an outgroup (underlined). Numbers on the branches indicate bootstrap support from 1000 replicates. Among-site substitution rate heterogeneity was corrected using two gamma distributed substitution rate categories and LG with frequencies substitution model for amino acid substitutions (LG+G+F). Diatom sequences are indicated in bold. The codes before the taxon name are the corresponding sequence IDs (For respective databases refer to materials and methods section).



0.5

Figure S3. Phylogenetic tree of Pds5 proteins inferred from maximum likelihood analysis. Numbers on the branches indicate bootstrap support from 1000 replicates. Among-site substitution rate heterogeneity was corrected using three gamma distributed substitution rate categories and LG substitution model for amino acid substitutions (LG+G). Diatom sequences are indicated in bold. The codes before the taxon name are the corresponding sequence IDs (For respective databases refer to materials and methods section).

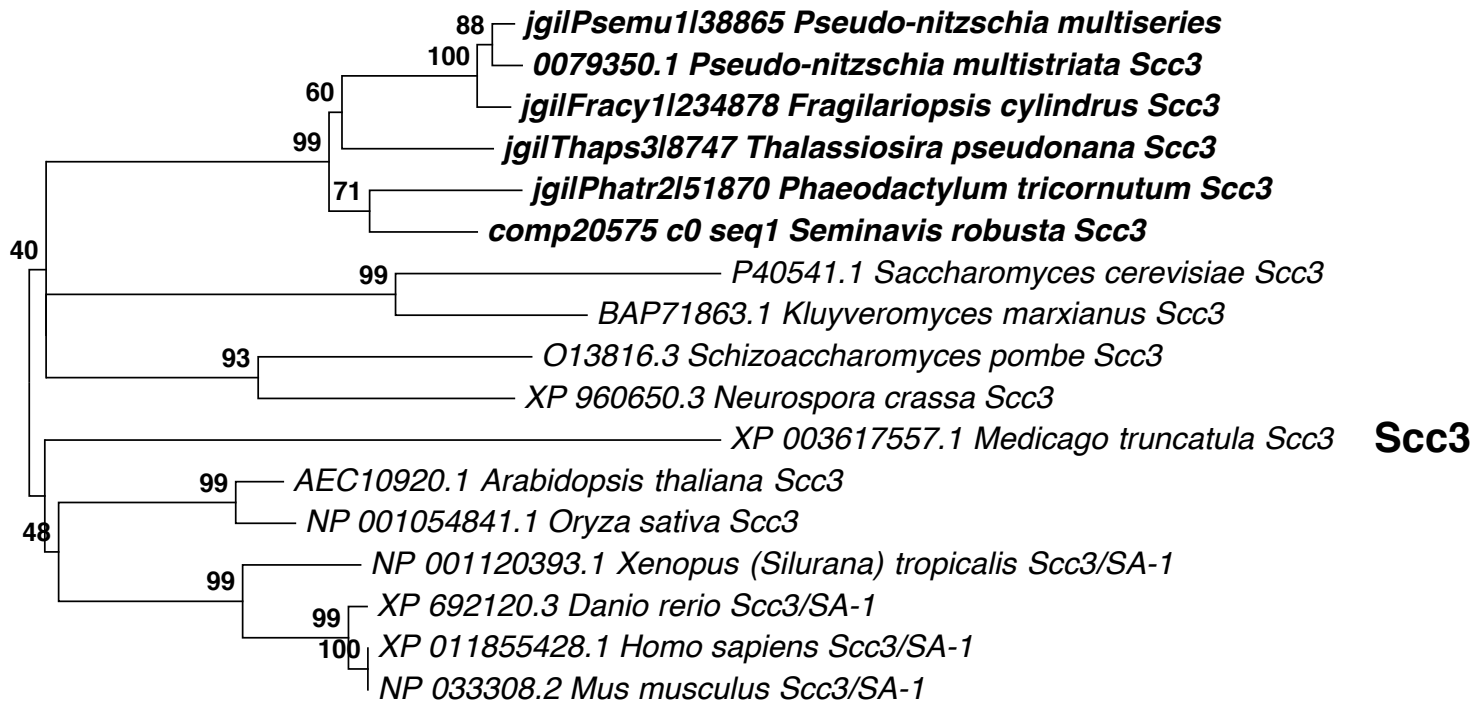
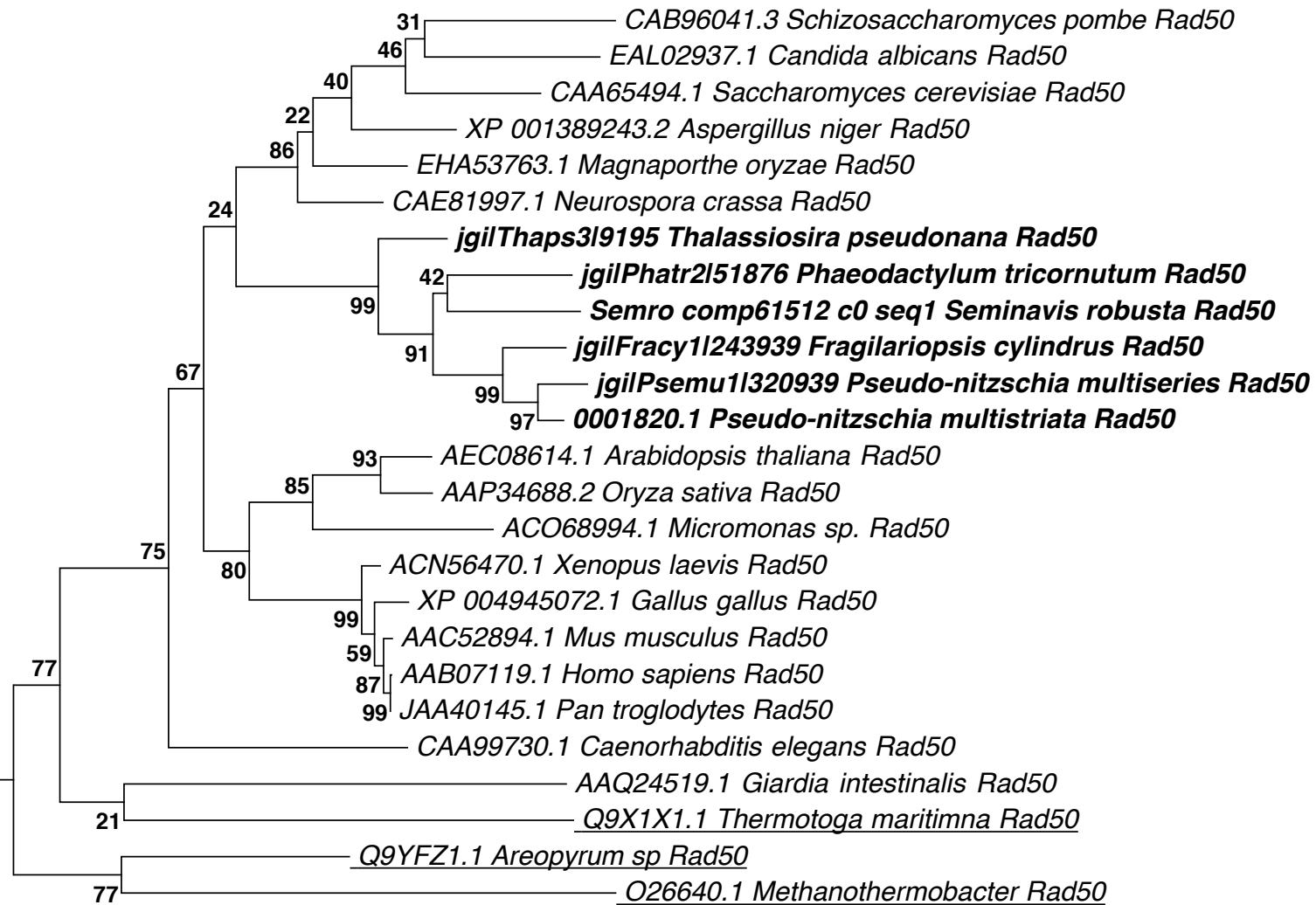


Figure S4. Phylogenetic tree of Scc3 proteins inferred from maximum likelihood analysis. Numbers on the branches indicate bootstrap support from 1000 replicates. Among-site substitution rate heterogeneity was corrected using an invariable and three gamma distributed substitution rate categories and LG substitution model for amino acid substitutions (LG+G+I). Diatom sequences are indicated in bold. The codes before the taxon name are the corresponding sequence IDs (For respective databases refer to materials and methods section).

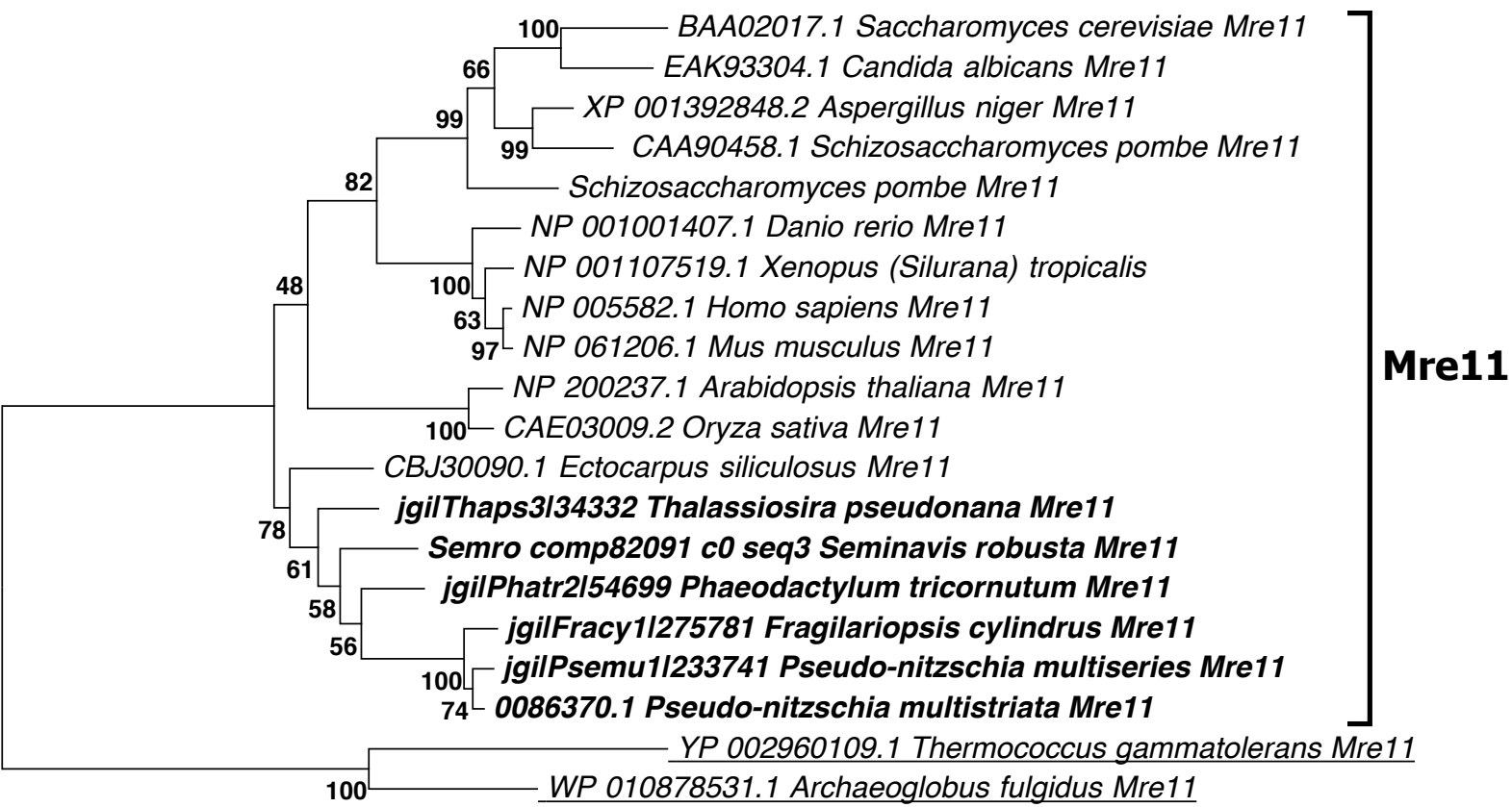




Rad50

0.5

Figure S5. Phylogenetic tree of Rad50 proteins inferred from maximum likelihood analysis, rooted with bacterial Rad50 proteins as an outgroup (underlined). Numbers on the branches indicate bootstrap support from 1000 replicates. Among-site substitution rate heterogeneity was corrected using two gamma distributed substitution rate categories and LG with frequencies substitution model for amino acid substitutions (LG+G+F). Diatom sequences are indicated in bold. The codes before the taxon name are the corresponding sequence IDs (For respective databases refer to materials and methods section).



0.5

Figure S6. Phylogenetic tree of Mre11 proteins inferred from maximum likelihood analysis, rooted with archaeal Mre11 proteins as an outgroup (underlined). Numbers on the branches indicate bootstrap support from 1000 replicates. Among-site substitution rate heterogeneity was corrected using two gamma distributed substitution rate categories and LG substitution model for amino acid substitutions (LG+G). Diatom sequences are indicated in bold. The codes before the taxon name are the corresponding sequence IDs (For respective databases refer to materials and methods section).

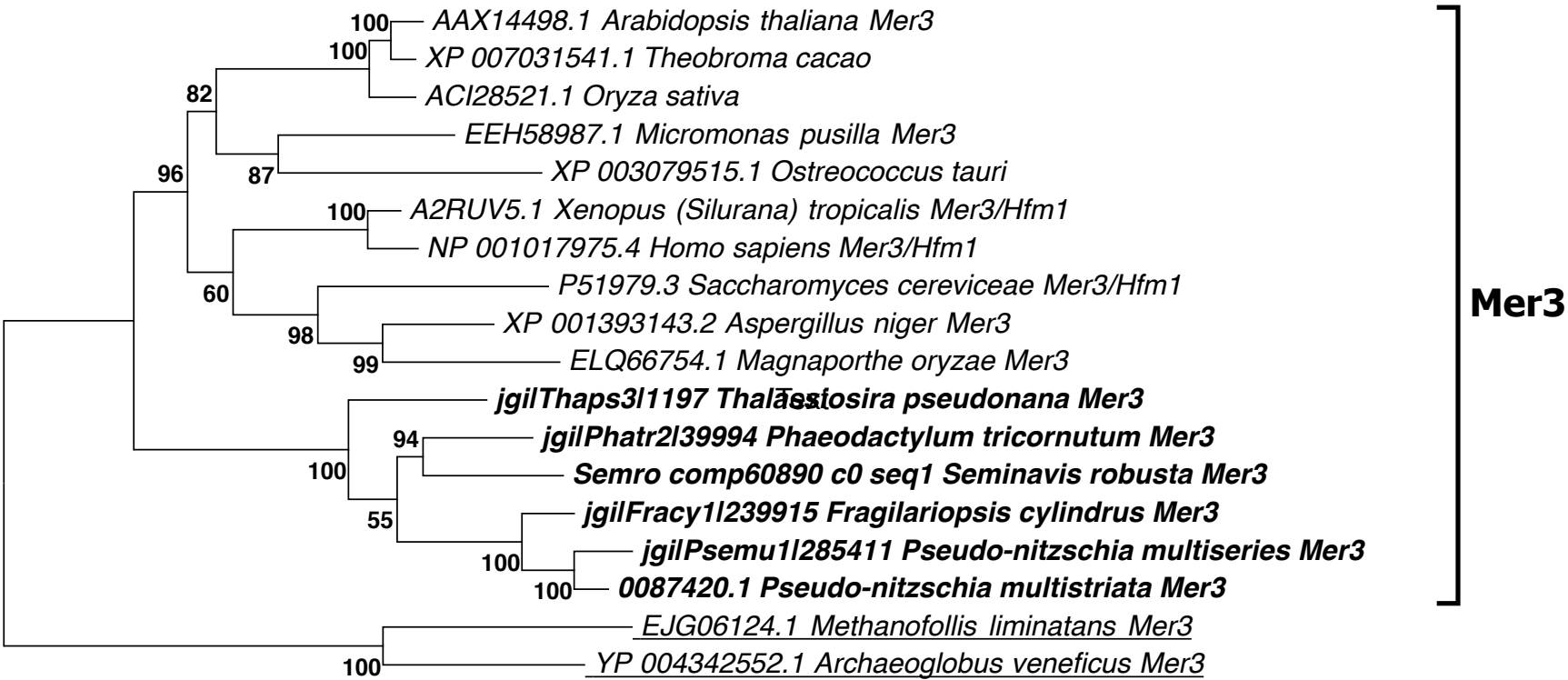
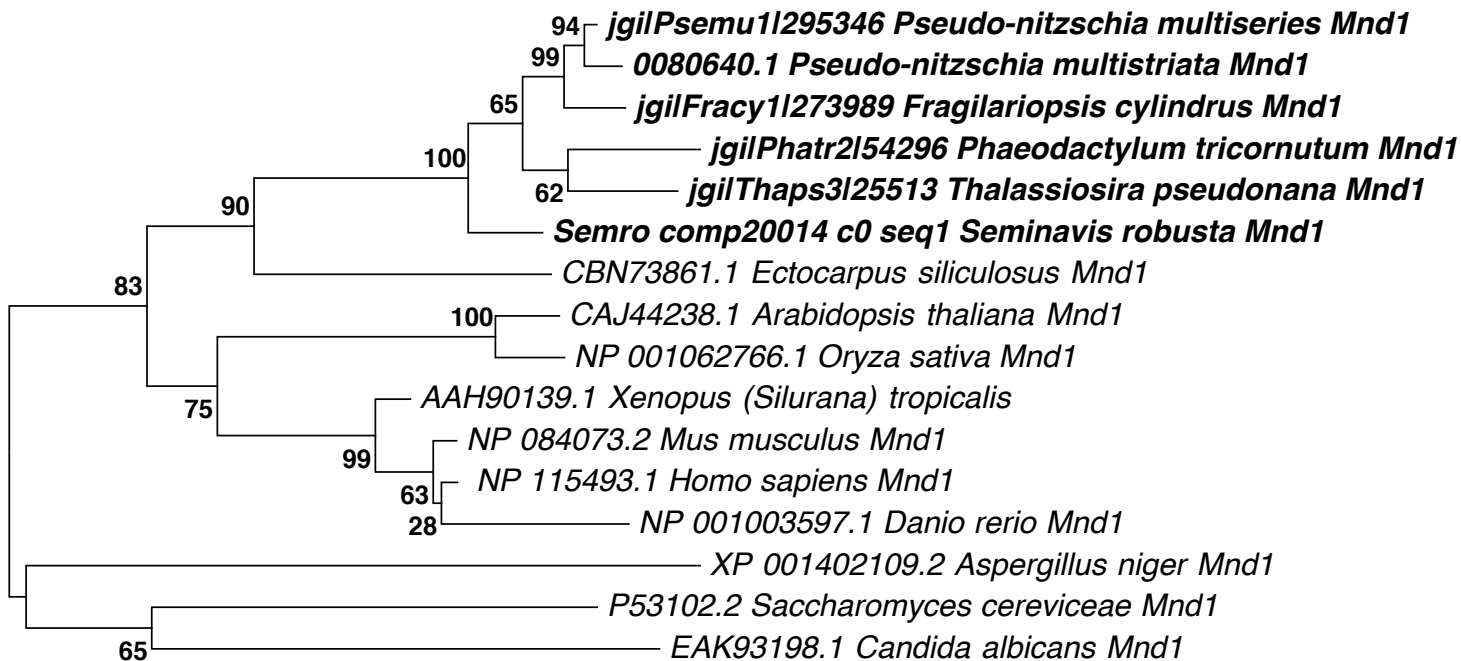


Figure S7. Phylogenetic tree of Mer3 proteins inferred from maximum likelihood analysis, rooted with archaeal Mer3 proteins as an outgroup (underlined). Numbers on the branches indicate bootstrap support from 1000 replicates. Among-site substitution rate heterogeneity was corrected using an invariable and two gamma distributed substitution rate categories and LG substitution model for amino acid substitutions (LG+G+I). Diatom sequences are indicated in bold. The codes before the taxon name are the corresponding sequence IDs (For respective databases refer to materials and methods section).

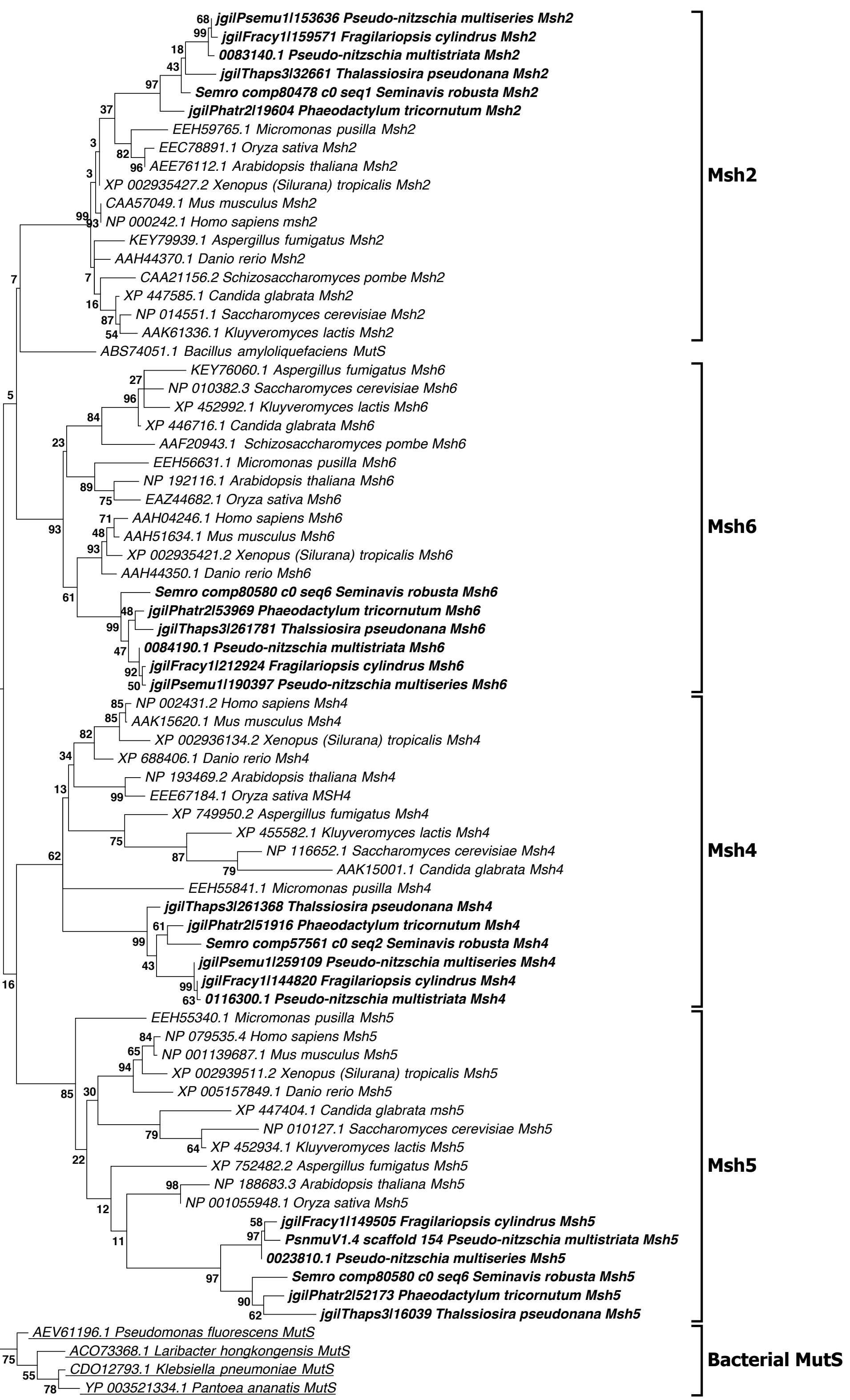
# Mnd1



0.2

Figure S8. Phylogenetic tree of Mnd1 proteins inferred from maximum likelihood analysis. Numbers on the branches indicate bootstrap support from 1000 replicates. Among-site substitution rate heterogeneity was corrected using two gamma distributed substitution rate categories and LG substitution model for amino acid substitutions (LG+G). Diatom sequences are indicated in bold. The codes before the taxon name are the corresponding sequence IDs (For respective databases refer to materials and methods section).

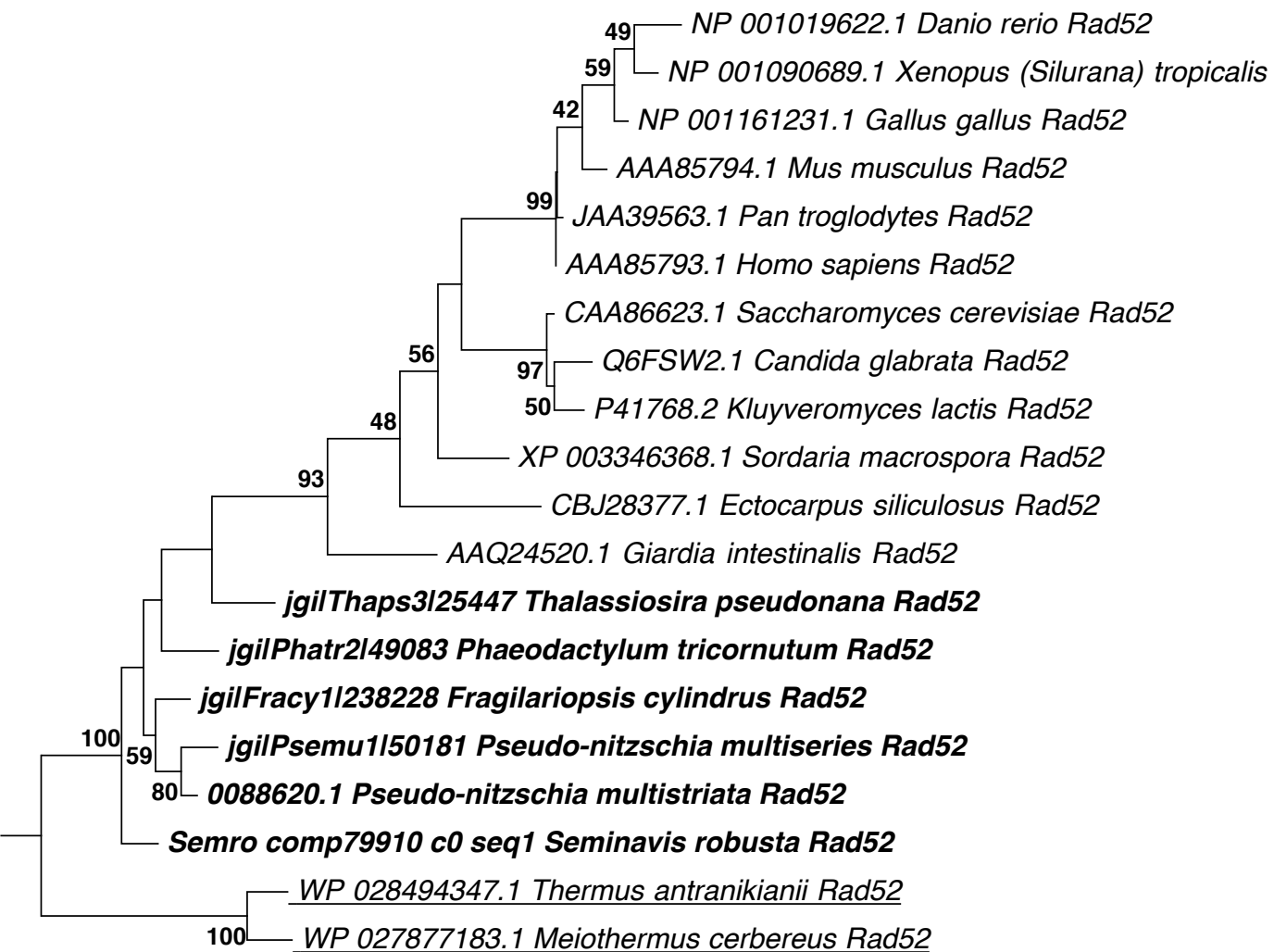




0.1

Figure S9. Phylogenetic tree of Msh proteins inferred from maximum likelihood analysis, rooted with bacterial MutS proteins as an outgroup (underlined). Numbers on the branches indicate bootstrap support from 1000 replicates. Among-site substitution rate heterogeneity was corrected using two gamma distributed substitution rate categories and LG substitution model for amino acid substitutions (LG+G). Diatom sequences are indicated in bold. The codes before the taxon name are the corresponding sequence IDs (For respective databases refer to materials and methods section).

# Rad52



0.5

Figure S10. Phylogenetic tree of Rad52 proteins inferred from maximum likelihood analysis, rooted with archaeal Rad52 proteins as an outgroup (underlined). Numbers on the branches indicate bootstrap support from 1000 replicates. Among-site substitution rate heterogeneity was corrected using two gamma distributed substitution rate categories and JTT substitution model for amino acid substitutions (JTT+G). Diatom sequences are indicated in bold. The codes before the taxon name are the corresponding sequence IDs (For respective databases refer to materials and methods section).

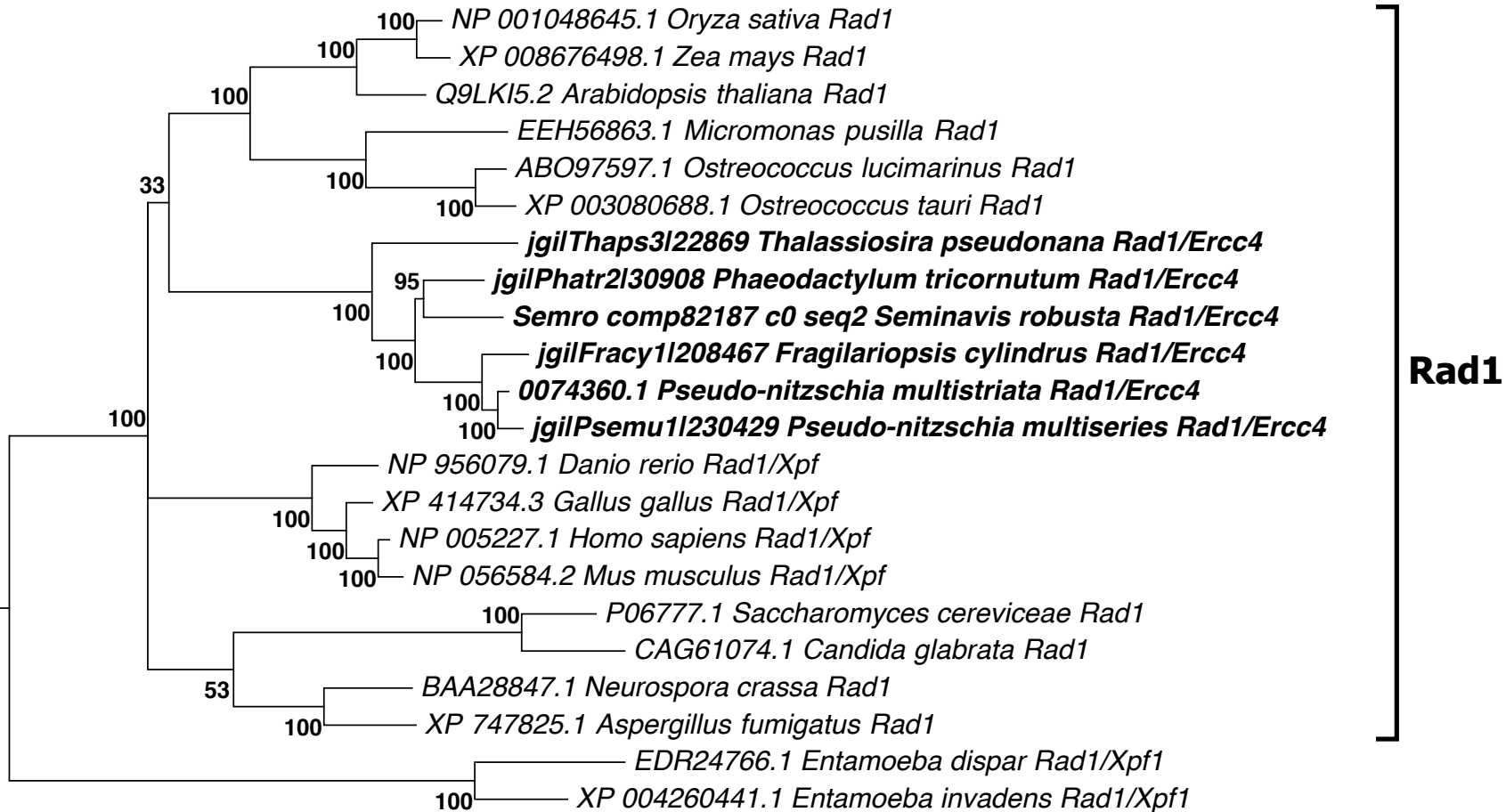


Figure S11. Phylogenetic tree of Rad1 proteins inferred from maximum likelihood analysis. Numbers on the branches indicate bootstrap support from 1000 replicates. Among-site substitution rate heterogeneity was corrected using an invariable two gamma distributed substitution rate categories and LG substitution model for amino acid substitutions (LG+G+I). Diatom sequences are indicated in bold. The codes before the taxon name are the corresponding sequence IDs (For respective databases refer to materials and methods section).

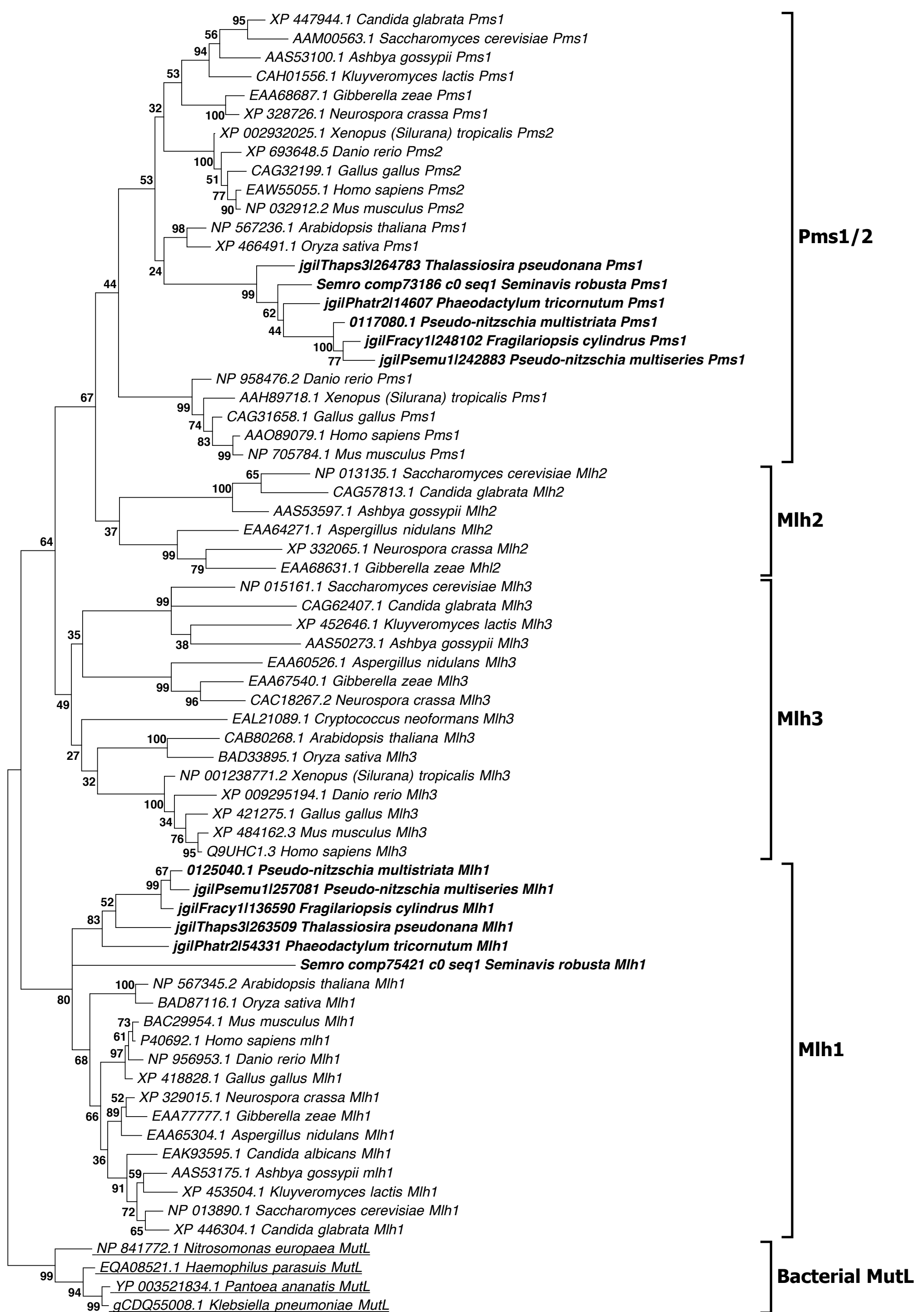
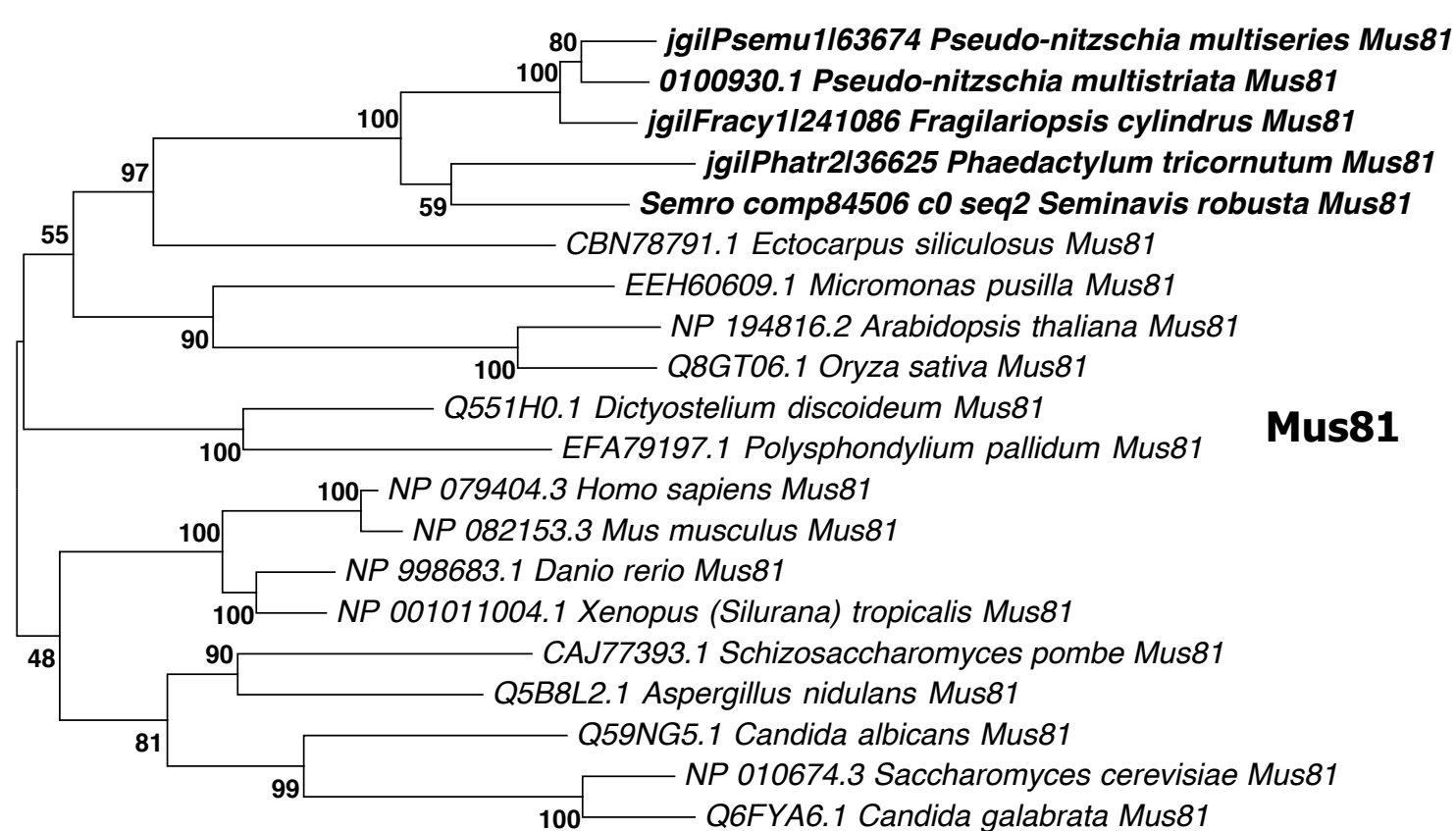


Figure S12. Phylogenetic tree of Mlh proteins inferred from maximum likelihood analysis, rooted with bacterial MutL proteins (underlined). Numbers on the branches indicate bootstrap support from 1000 replicates. Among-site substitution rate heterogeneity was corrected using three gamma distributed substitution rate categories and LG substitution model for amino acid substitutions (LG+G). Diatom sequences are indicated in bold. The codes before the taxon name are the corresponding sequence IDs (For respective databases refer to materials and methods section).

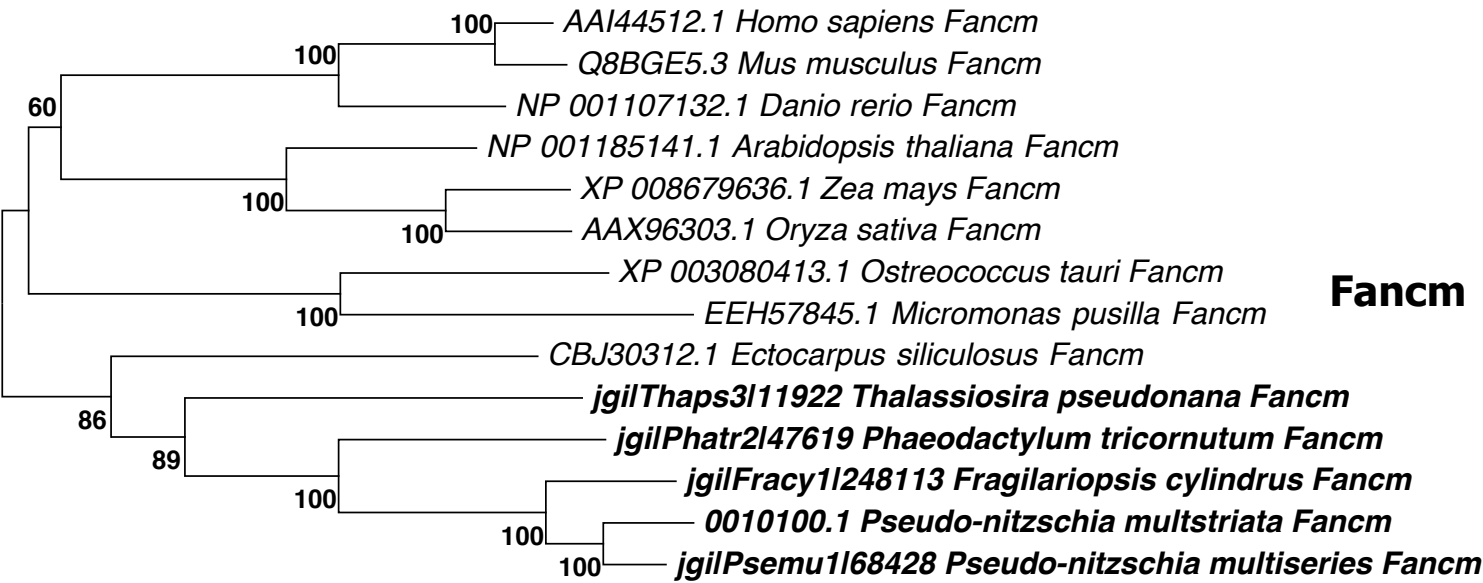




**Mus81**

0.2

Figure S13. Phylogenetic tree of Mus81 proteins inferred from maximum likelihood analysis. Numbers on the branches indicate bootstrap support from 1000 replicates. Among-site substitution rate heterogeneity was corrected using an invariable and two gamma distributed substitution rate categories and LG substitution model for amino acid substitutions (LG+G+I). Diatom sequences are indicated in bold. The codes before the taxon name are the corresponding sequence IDs (For respective databases refer to materials and methods section).



0.5

Figure S14. Phylogenetic tree of Fancm proteins inferred from maximum likelihood analysis. Numbers on the branches indicate bootstrap support from 1000 replicates. Among-site substitution rate heterogeneity was corrected using an invariable and two gamma distributed substitution rate categories and WAG substitution model with frequencies for amino acid substitutions (WAG+G+I+F). Diatom sequences are indicated in bold. The codes before the taxon name are the corresponding sequence IDs (For respective databases refer to materials and methods section).

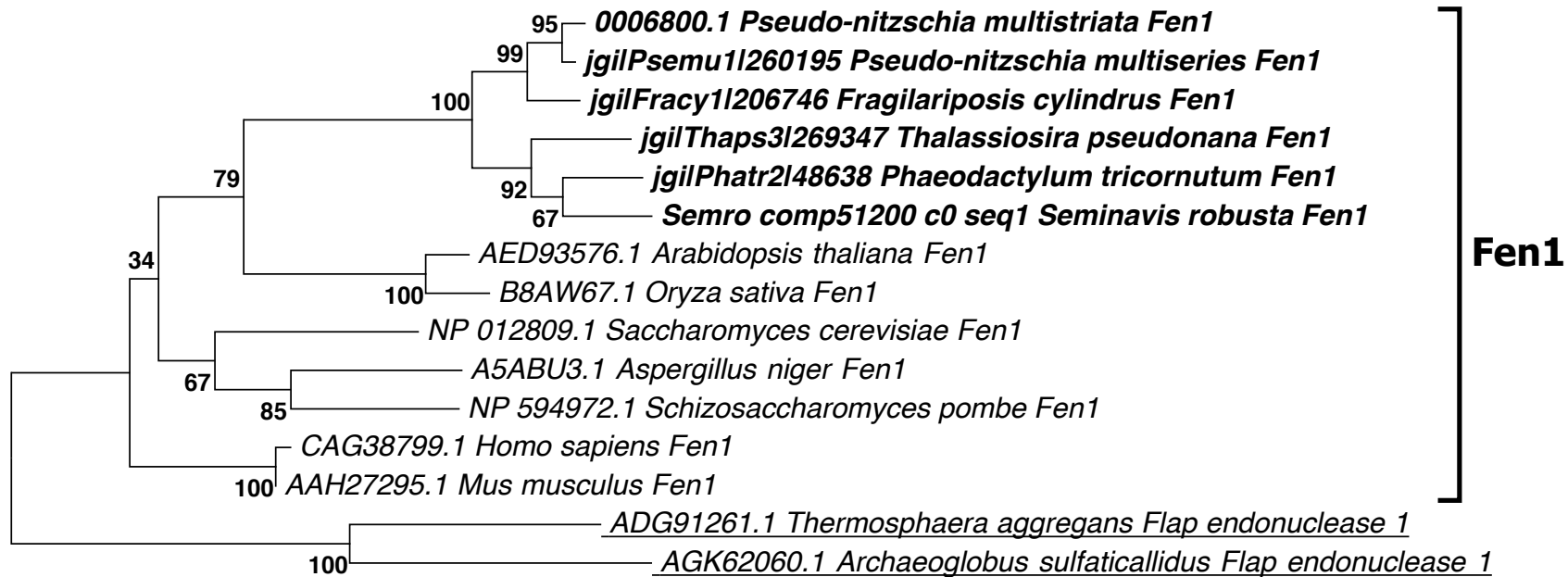
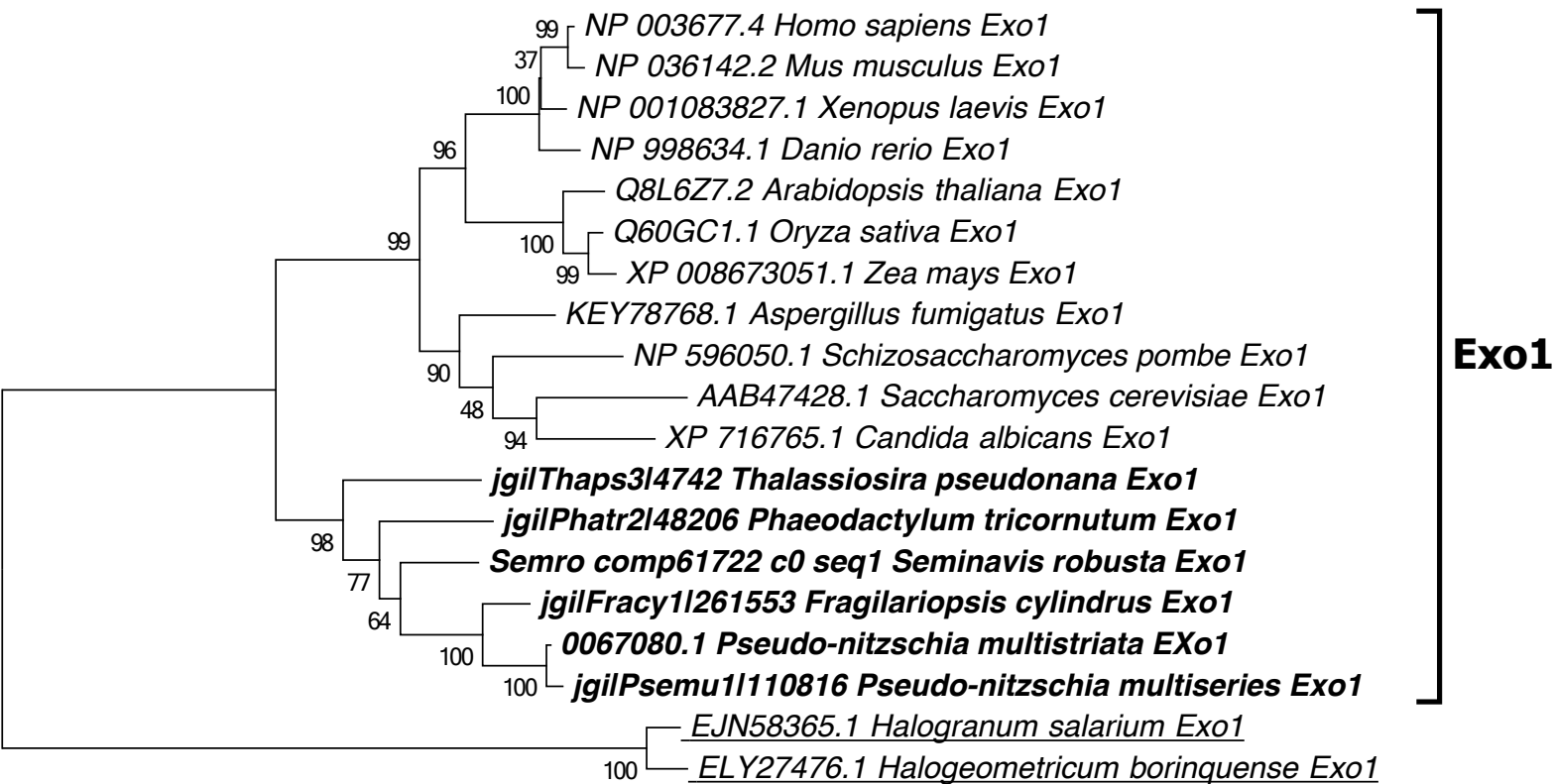


Figure S15. Phylogenetic tree of Fen1 proteins inferred from maximum likelihood analysis, rooted with archaeal Fen1 proteins as an outgroup (underlined). Numbers on the branches indicate bootstrap support from 1000 replicates. Among-site substitution rate heterogeneity was corrected using two gamma distributed substitution rate categories and LG substitution model for amino acid substitutions (LG+G). Diatom sequences are indicated in bold. The codes before the taxon name are the corresponding sequence IDs (For respective databases refer to materials and methods section).

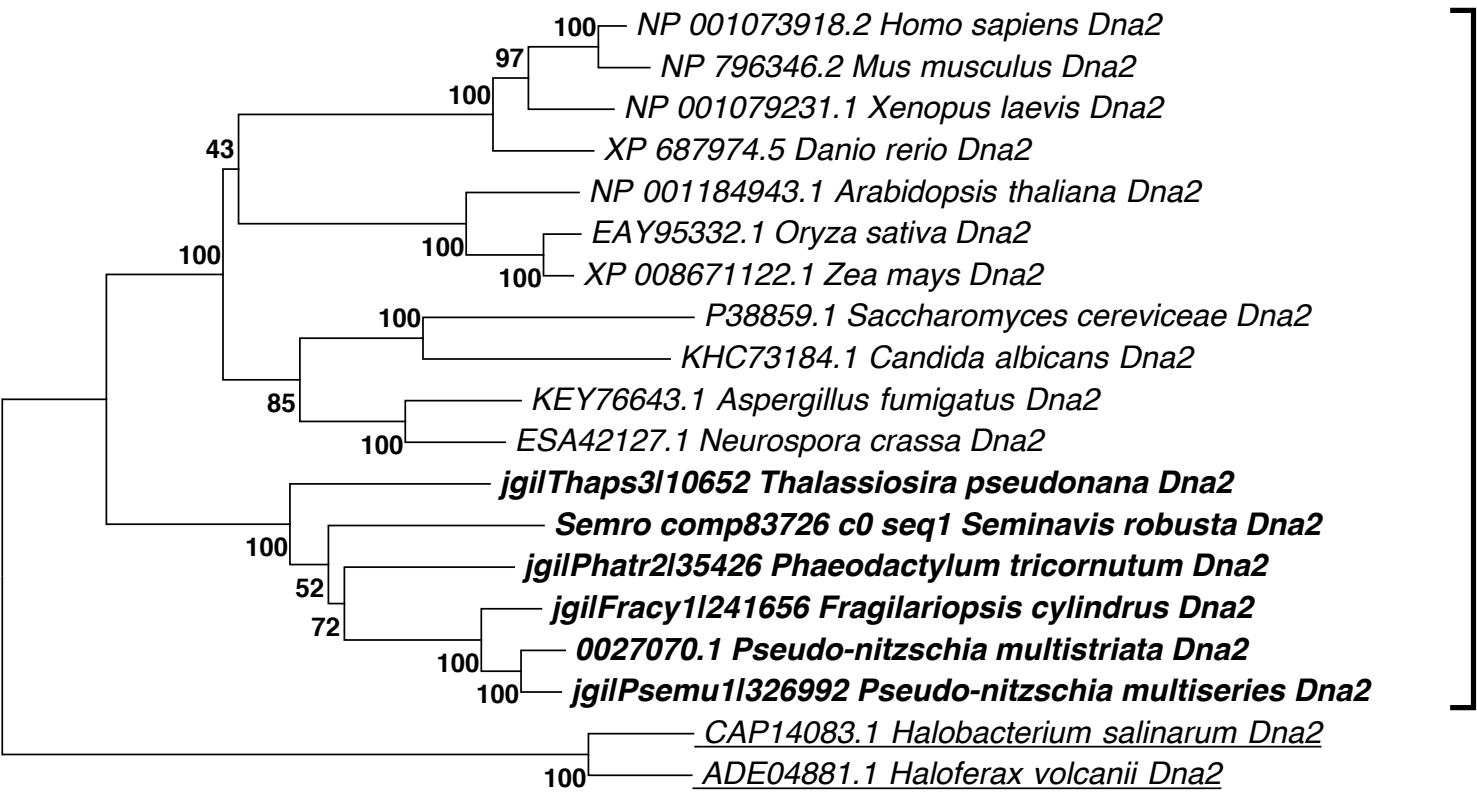


0.5

Figure S16. Phylogenetic tree of Exo1 proteins inferred from maximum likelihood analysis, rooted with archaeal Exo1 proteins as an outgroup (underlined). Numbers on the branches indicate bootstrap support from 1000 replicates. Among-site substitution rate heterogeneity was corrected using an invariable and two gamma distributed substitution rate categories and LG substitution model for amino acid substitutions (LG+G+I). Diatom sequences are indicated in bold. The codes before the taxon name are the corresponding sequence IDs (For respective databases refer to materials and methods section).

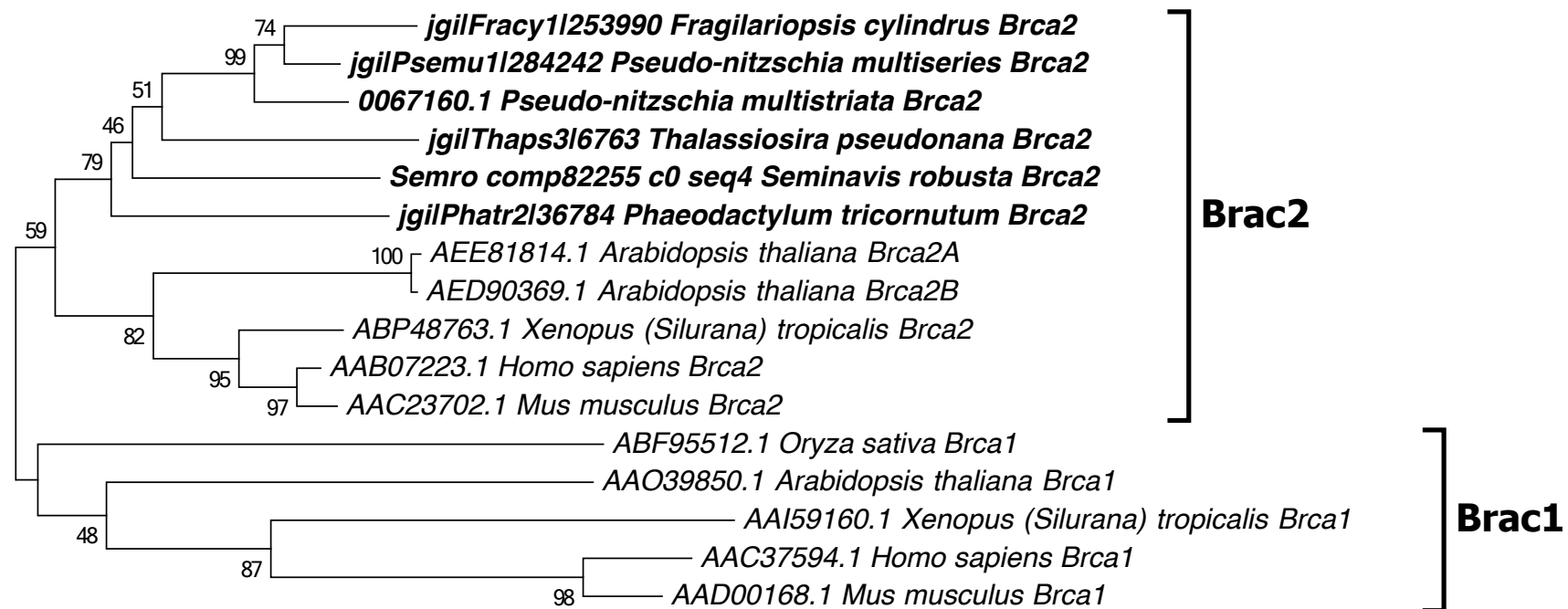


# Dna2



0.5

Figure S17. Phylogenetic tree of Dna2 proteins inferred from maximum likelihood analysis, rooted with archaeal Dna2 proteins (underlined). Numbers on the branches indicate bootstrap support from 1000 replicates. Among-site substitution rate heterogeneity was corrected using an invariable and two gamma distributed substitution rate categories and LG substitution model for amino acid substitutions (LG+G+I). Diatom sequences are indicated in bold. The codes before the taxon name are the corresponding sequence IDs (For respective databases refer to materials and methods section).



0.5

Figure S18. Phylogenetic tree of Brca proteins inferred from maximum likelihood analysis rooted. Numbers on the branches indicate bootstrap support from 1000 replicates. Among-site substitution rate heterogeneity was corrected using three gamma distributed substitution rate categories and WAG substitution model with frequencies for amino acid substitutions (WAG+G+F). Diatom sequences are indicated in bold. The codes before the taxon name are the corresponding sequence IDs (For respective databases refer to materials and methods section).