Current Biology, Volume *30*

Supplemental Information

New Lineage of Microbial Predators Adds Complexity to Reconstructing the Evolutionary Origin of Animals

Denis V. Tikhonenkov, Kirill V. Mikhailov, Elisabeth Hehenberger, Sergei A. Karpov, Kristina I. Prokina, Anton S. Esaulov, Olga I. Belyakova, Yuri A. Mazei, Alexander P. Mylnikov, Vladimir V. Aleoshin, and Patrick J. Keeling

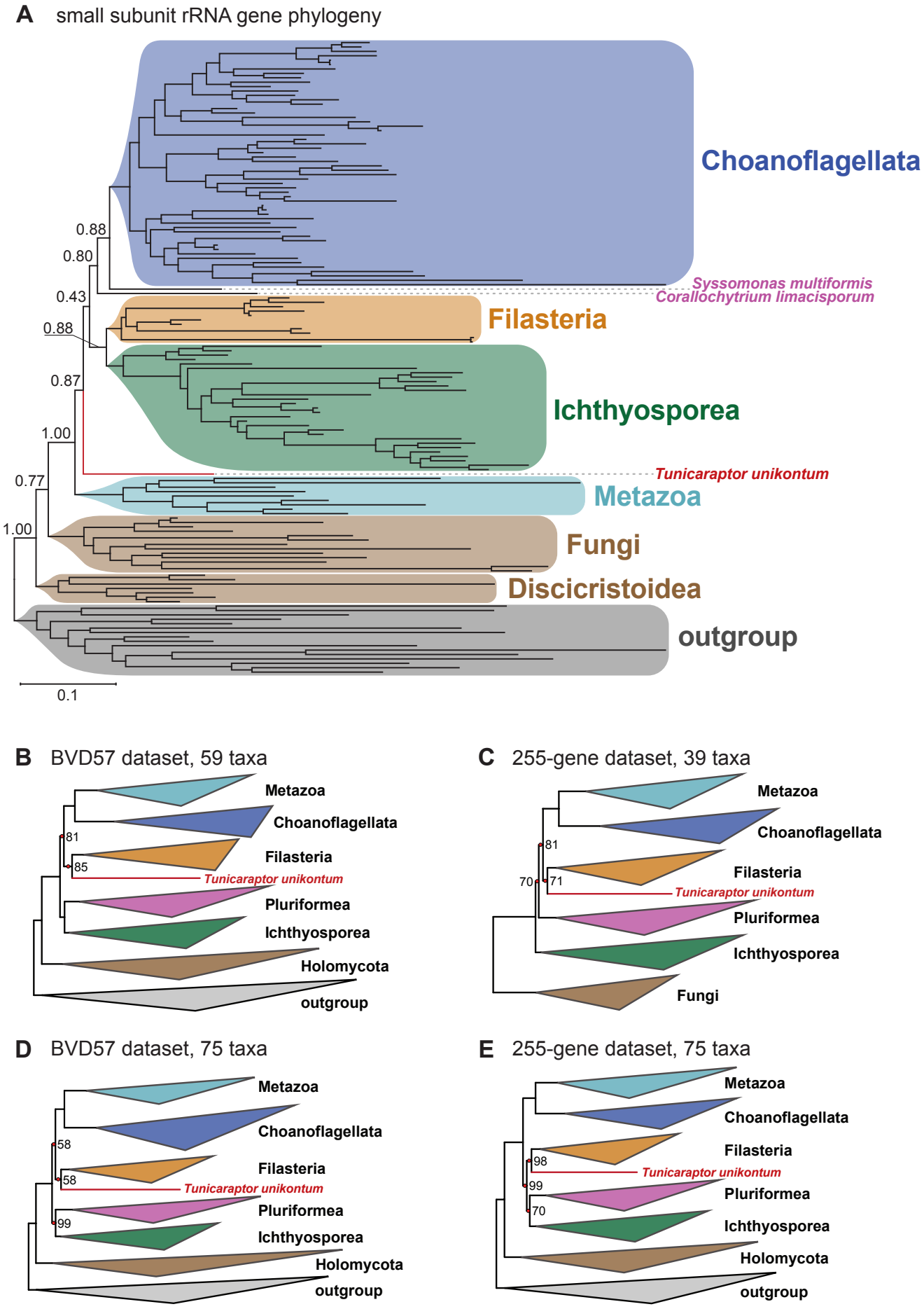

Figure S1. Phylogenetic reconstructions with alignments based on the previously published datasets, Related to Figure 2. (A) Bayesian inference with small subunit rRNA gene sequences including environmental lineages, based on the alignment of rRNA genes of Hehenberger et al. [S1]; support for key inner nodes of the tree is indicated with posterior probability values. **(B)** IQ-TREE inference with the BVD57 dataset [S2] expanded with sequences of *T. unikontum*; reconstructions were performed using the LG+C60+F+G4 model, and the support values estimated with 1000 replicates of ultrafast bootstrap; the established monophyletic groups are abridged, showing only the details of relationships between key holozoan taxa. **(C)** IQ-TREE inference with the 255-gene dataset [S1] expanded with sequences of *T. unikontum*. **(D)** IQ-TREE inference using the BVD57 gene set with the taxonomic sampling adjusted to conform to the sampling used for the 200-gene dataset (see Figure 2). **(E)** IQ-TREE inference with the 255-gene dataset, taxonomic sampling was expanded to conform to the sampling used for the 200-gene dataset.

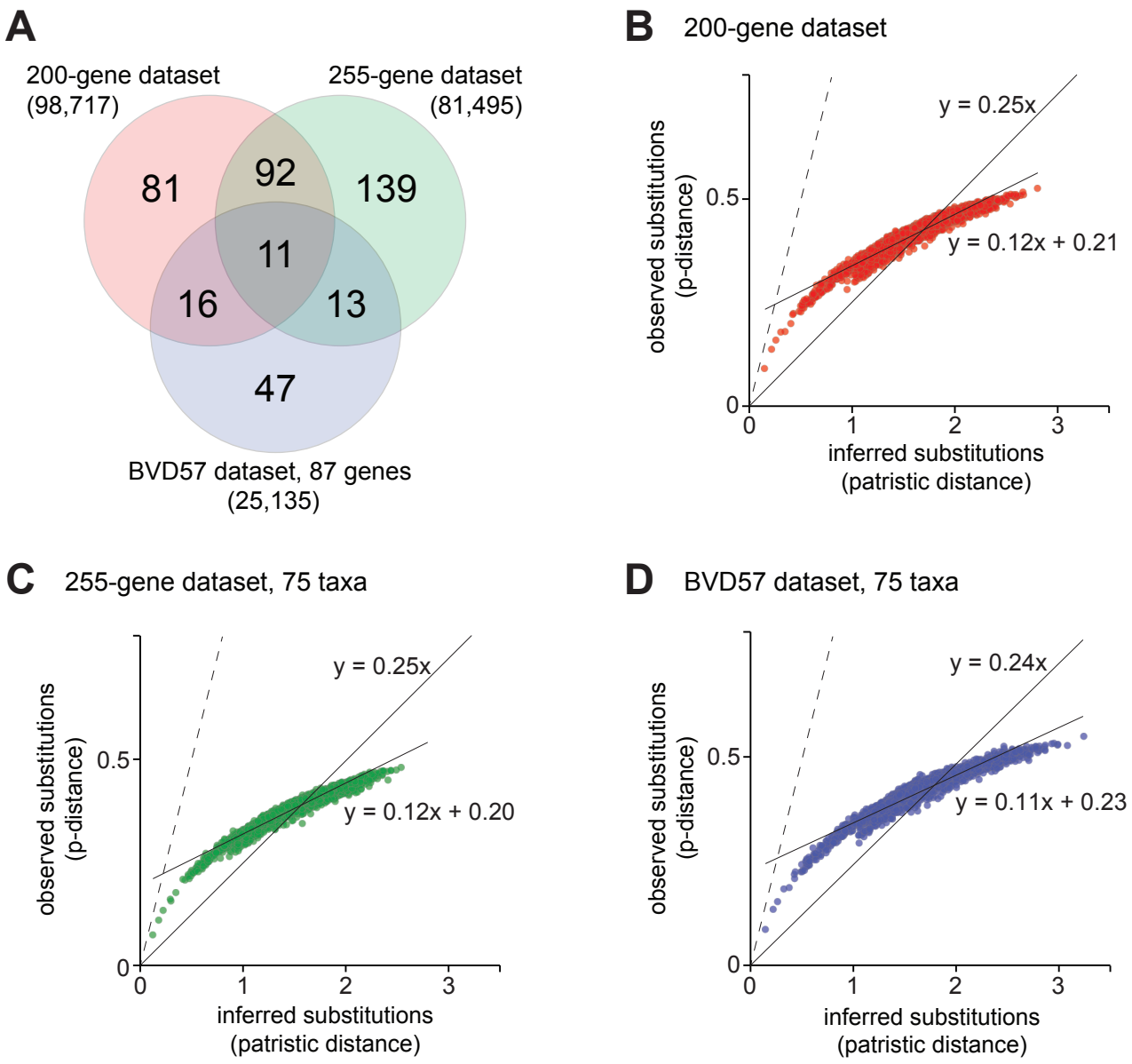

Figure S2. Characteristics of the three datasets employed in the phylogenetic analyses, Related to Figure 2. (A) Venn diagram of shared orthologs for the three datasets; the 200-gene dataset is the newly assembled alignment analyzed in this work (99,000 amino acid sites, average proportion of missing data is 6%, with *T. unikontum* containing 21% missing data), the 255-gene dataset was used in the study by Hehenberger et al. [S1] and derived from eukaryote-wide phylogenomic studies [S3], the BVD57 dataset corresponds to the 87-gene matrix constructed from the set of single-copy protein domains [S2]; the total alignment length (aa sites) is listed for each dataset. The novel 200-gene dataset overlaps with the 255-gene alignment by 103 genes, which constitute approximately 40% of the 99,000-site alignment length, and by 27 genes with the BVD57 alignment (15% of the 99,000-site alignment). **(B)** estimation of mutational saturation in the 200-gene dataset, given by the relation of patristic distances (x-axis) in the ML tree, and the p-distance values (y-axis) calculated from the alignment; fitted linear regression models are provided for each dataset. **(C)** mutational saturation plot for the 255-gene dataset that was expanded to conform to the taxonomic sampling of the 200-gene dataset (75 OTU). **(D)** mutational saturation plot for the expanded BVD57 dataset (75 OTU).

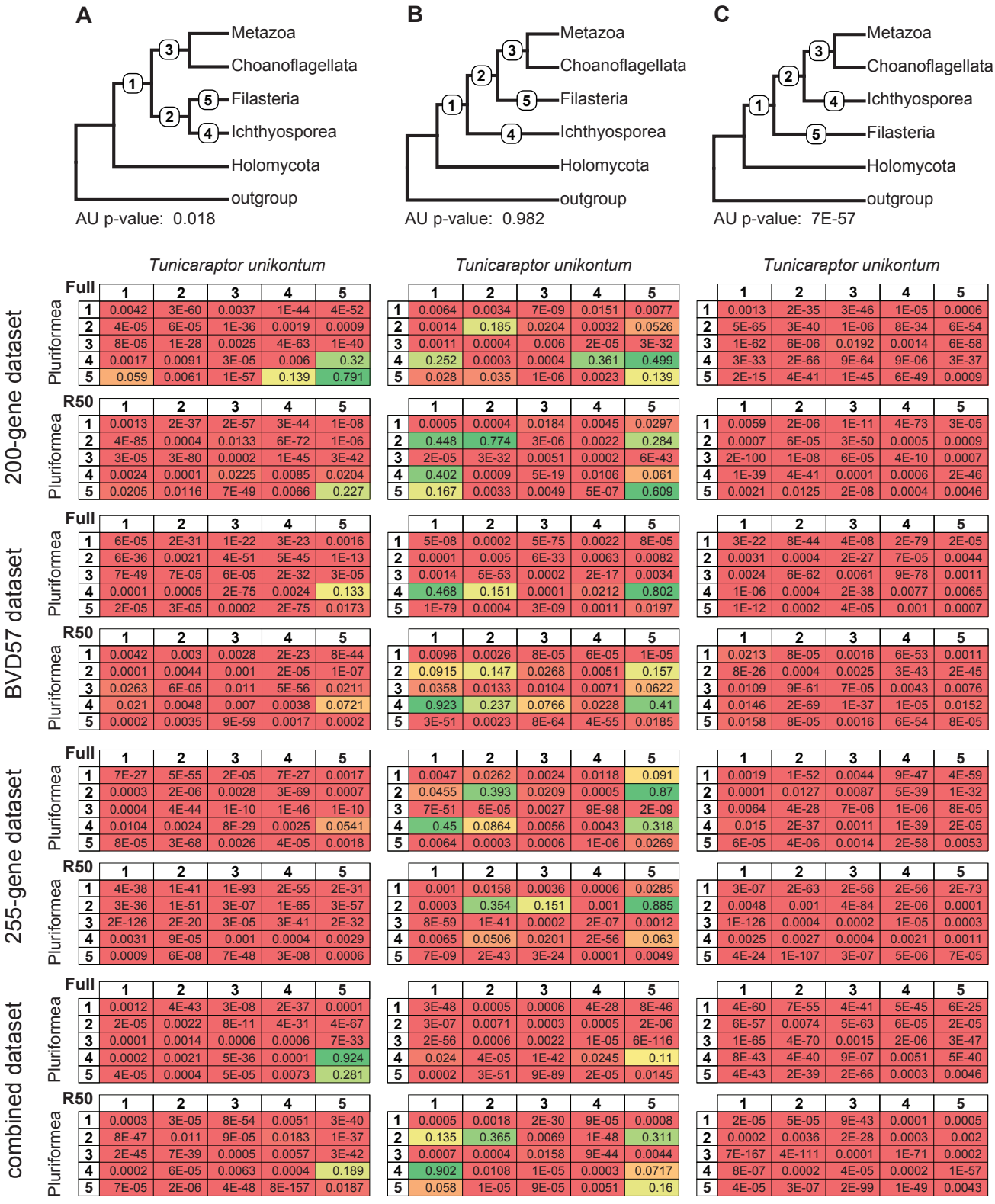

Figure S3. Phylogenetic hypothesis testing, Related to Figure 2. AU test p-values for topologies representing all possible interrelations of Filasteria, Ichthyosporea, Pluriformea, and *T. unikontum* lineages; the test was performed with IQ-TREE under the LG+C60+F+G4 model. Three topologies **(A, B, C)** corresponding to the possible relationships of filasterian and ichthyosporean lineages were tested with the combined dataset (AU test p-values are written under the respective trees) and used as basis for exploring possible positions of Pluriformea and *T. unikontum*. The number at the row or column of each table corresponds to the index on the branch, which marks the position of lineage in the tree (Pluriformea – rows, *T. unikontum* – columns); for topologies where Pluriformea and *T. unikontum* are on the same tree node, three possible arrangements of these taxa were tested, and the highest p-value is written in the table. Four datasets were analyzed using the AU test: the 200-gene dataset, the 255-gene dataset [S1], the BVD57 dataset [S2], and the combined 395-gene alignment. For each dataset two variants of the test are presented: the result obtained with the full alignment (Full), and the result obtained with an alignment where 50% of variable sites are removed starting with the fastest-evolving category (R50) (for full tables see Data S1).

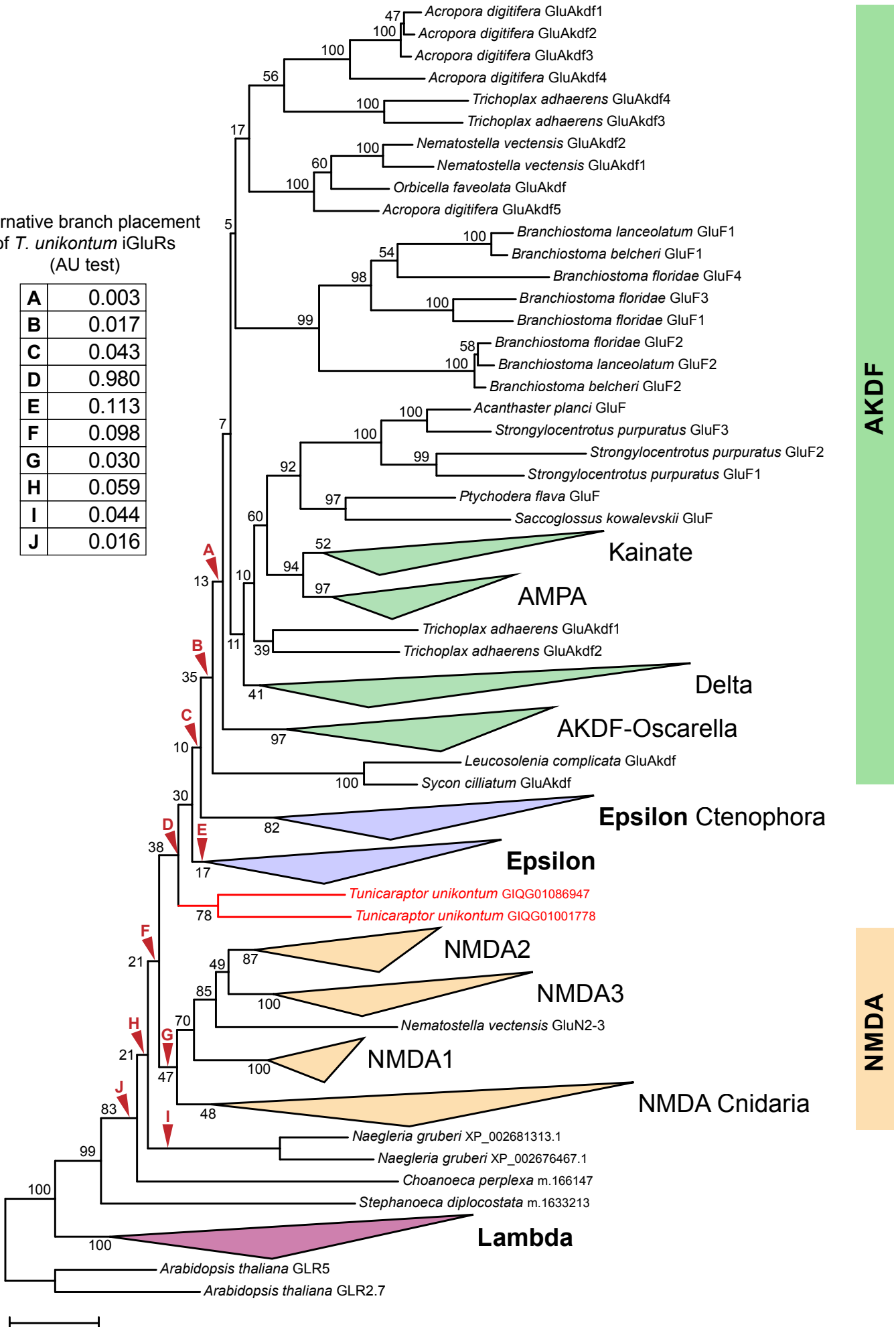

Figure S4. Phylogeny of ionotropic glutamate receptors, Related to Figure 4. The phylogeny was reconstructed with IQ-TREE using the LG+R7 evolutionary model, selected as best-fit by ModelFinder. The alignment comprising 250 iGluR sequences is based on the sequence set from the study by Ramos-Vicente et al. [S4], with the addition of iGluR sequences from unicellular holozoans and *Naegleria*. Monophyletic classes or subfamilies of iGluRs (AMPA, Kainate, Delta, etc) are collapsed in the tree. The node support values were estimated using standard nonparametric bootstrap with 1000 replicates. Table on the left features AU test p-values for the tree topologies, exploring several alternative positions of *T. unikontum* iGluR sequences; the variants are labeled with letters and the positions are indicated with arrowheads in the tree.

Supplemental References

S1 Hehenberger, E., Tikhonenkov, D.V., Kolisko, M., del Campo, J., Esaulov, A.S., Mylnikov, A.P., and Keeling, P.J. (2017). Novel predators reshape holozoan phylogeny and reveal the presence of a two-component signaling system in the ancestor of Animals. Curr. Biol. *27*, 2043–2050.

S2 Grau-Bové, X., Torruella, G., Donachie, S., Suga, H., Leonard, G., Richards, T.A., and Ruiz-Trillo, I. (2017). Dynamics of genomic innovation in the unicellular ancestry of animals. Elife *6*, e26036.

S3 Burki, F., Kaplan, M., Tikhonenkov, D.V., Zlatogursky, V., Minh, B.Q., Radaykina, L.V., Smirnov, A., Mylnikov, A.P., and Keeling, P.J. (2016). Untangling the early diversification of eukaryotes: a phylogenomic study of the evolutionary origins of Centrohelida, Haptophyta and Cryptista. Proc. Biol. Sci. *283*:20152802.

S4 Ramos-Vicente, D., Ji, J., Gratacòs-Batlle, E., Gou, G., Reig-Viader, R., Luís, J., Burguera, D., Navas-Perez, E., García-Fernández, J., Fuentes-Prior, P., et al. (2018). Metazoan evolution of glutamate receptors reveals unreported phylogenetic groups and divergent lineage-specific events. Elife *7*, e35774.