

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

Highlights

- New predatory protist (*Tunicaraptor*) with unique morphology is related to animals
- *Tunicaraptor* calls into question many of the well-accepted relationships in Holozoa
- *Tunicaraptor* possesses a unique combination of “animal-specific” gene products
- Eukaryovorous flagellates may represent a major share of unicellular animal relatives

Authors

Denis V. Tikhonenkov,
Kirill V. Mikhailov,
Elisabeth Hehenberger, ...,
Alexander P. Mylnikov,
Vladimir V. Aleoshin, Patrick J. Keeling

Correspondence

kv.mikhailov@belozersky.msu.ru
(K.V.M.),
pkeeling@mail.ubc.ca (P.J.K.),
tikho-denis@yandex.ru (D.V.T.)

In Brief

Tikhonenkov et al. report a new lineage of predatory protists (*Tunicaraptor*) related to animals, challenging the existing phylogenomic framework used to reconstruct the evolution of “animal-specific” genes. This protist highlights the notion that eukaryovorous flagellates may represent a major fraction of the unicellular relatives of animals.



Report

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

Denis V. Tikhonenkov,^{1,2,12,13,*} Kirill V. Mikhailov,^{3,4,12,*} Elisabeth Hehenberger,⁵ Sergei A. Karpov,^{6,7} Kristina I. Prokina,^{1,6} Anton S. Esaulov,⁸ Olga I. Belyakova,⁸ Yuri A. Mazei,^{9,10} Alexander P. Mylnikov,^{1,11} Vladimir V. Aleoshin,^{3,4} and Patrick J. Keeling^{2,*}

¹Papanin Institute for Biology of Inland Waters, Russian Academy of Sciences, Borok 152742, Russia

²Department of Botany, University of British Columbia, Vancouver, BC V6T 1Z4, Canada

³Belozersky Institute for Physico-Chemical Biology, Lomonosov Moscow State University, Moscow 119992, Russia

⁴Kharkevich Institute for Information Transmission Problems, Russian Academy of Sciences, Moscow 127051, Russia

⁵Ocean EcoSystems Biology Unit, RD3, GEOMAR Helmholtz Centre for Ocean Research Kiel, Duesternbrookerweg 20, 24105 Kiel, Germany

⁶Zoological Institute, Russian Academy of Sciences, Saint Petersburg 199034, Russia

⁷Department of Invertebrate Zoology, Faculty of Biology, Saint Petersburg State University, Saint Petersburg 199034, Russia

⁸Department of Zoology and Ecology, Penza State University, Penza 440026, Russia

⁹Department of General Ecology and Hydrobiology, Lomonosov Moscow State University, Moscow 119991, Russia

¹⁰A.N. Severtsov Institute of Ecology and Evolution, Russian Academy of Sciences, Moscow 119071, Russia

¹¹Deceased May 30, 2019

¹²These authors contributed equally

¹³Lead Contact

*Correspondence: tikho-denis@yandex.ru (D.V.T.), kv.mikhailov@belozersky.msu.ru (K.V.M.), pkeeling@mail.ubc.ca (P.J.K.)

<https://doi.org/10.1016/j.cub.2020.08.061>

SUMMARY

The origin of animals is one of the most intensely studied evolutionary events, and our understanding of this transition was greatly advanced by analyses of unicellular relatives of animals, which have shown many “animal-specific” genes actually arose in protistan ancestors long before the emergence of animals [1–3]. These genes have complex distributions, and the protists have diverse lifestyles, so understanding their evolutionary significance requires both a robust phylogeny of animal relatives and a detailed understanding of their biology [4, 5]. But discoveries of new animal-related lineages are rare and historically biased to bacteriovores and parasites. Here, we characterize the morphology and transcriptome content of a new animal-related lineage, predatory flagellate *Tunicaraptor unikontum*. *Tunicaraptor* is an extremely small (3–5 μm) and morphologically simple cell superficially resembling some fungal zoospores, but it survives by preying on other eukaryotes, possibly using a dedicated but transient “mouth,” which is unique for unicellular opisthokonts. The *Tunicaraptor* transcriptome encodes a full complement of flagellar genes and the flagella-associated calcium channel, which is only common to predatory animal relatives and missing in microbial parasites and grazers. *Tunicaraptor* also encodes several major classes of animal cell adhesion molecules, as well as transcription factors and homologs of proteins involved in neurodevelopment that have not been found in other animal-related lineages. Phylogenomics, including *Tunicaraptor*, challenges the existing framework used to reconstruct the evolution of animal-specific genes and emphasizes that the diversity of animal-related lineages may be better understood only once the smaller, more inconspicuous animal-related lineages are better studied.

RESULTS

Novel Unicellular Relative of Animals and Phylogeny of Basal Holozoans

The novel free-living predatory species *Tunicaraptor unikontum* (see Data S3 for formal taxonomic diagnosis) was isolated from coastal marine waters of Chile. *Tunicaraptor* are small (3.5–5.1 μm long), swimming, elongated-oval cells with a single posterior flagellum (Figures 1A–1D, 1I, 1J, and 1M), resembling animal sperm or zoospores of chytrid fungi and aphelids. It rarely

produces short filopodia (Figure 1L), and solitary cells can form temporary aggregates of 3–6 flagellated or non-flagellated cells (Figure 1E). *Tunicaraptor* feeds on eukaryotic prey and could not be sustained on bacteria alone. Cells can feed jointly on a single prey cell (Figures 1F–1H; Video S1) and form aggregations during feeding. Except for the flagellum and a small anterior patch (the “mouth”; Figures 1K and 1Q), the entire cell is covered by external envelope (theca) bearing long (~110 nm) external hairs (Figures 1N, 1O, and 1Q). Dedicated mouth-like feeding structures are common in predatory biflagellate eukaryotes in general



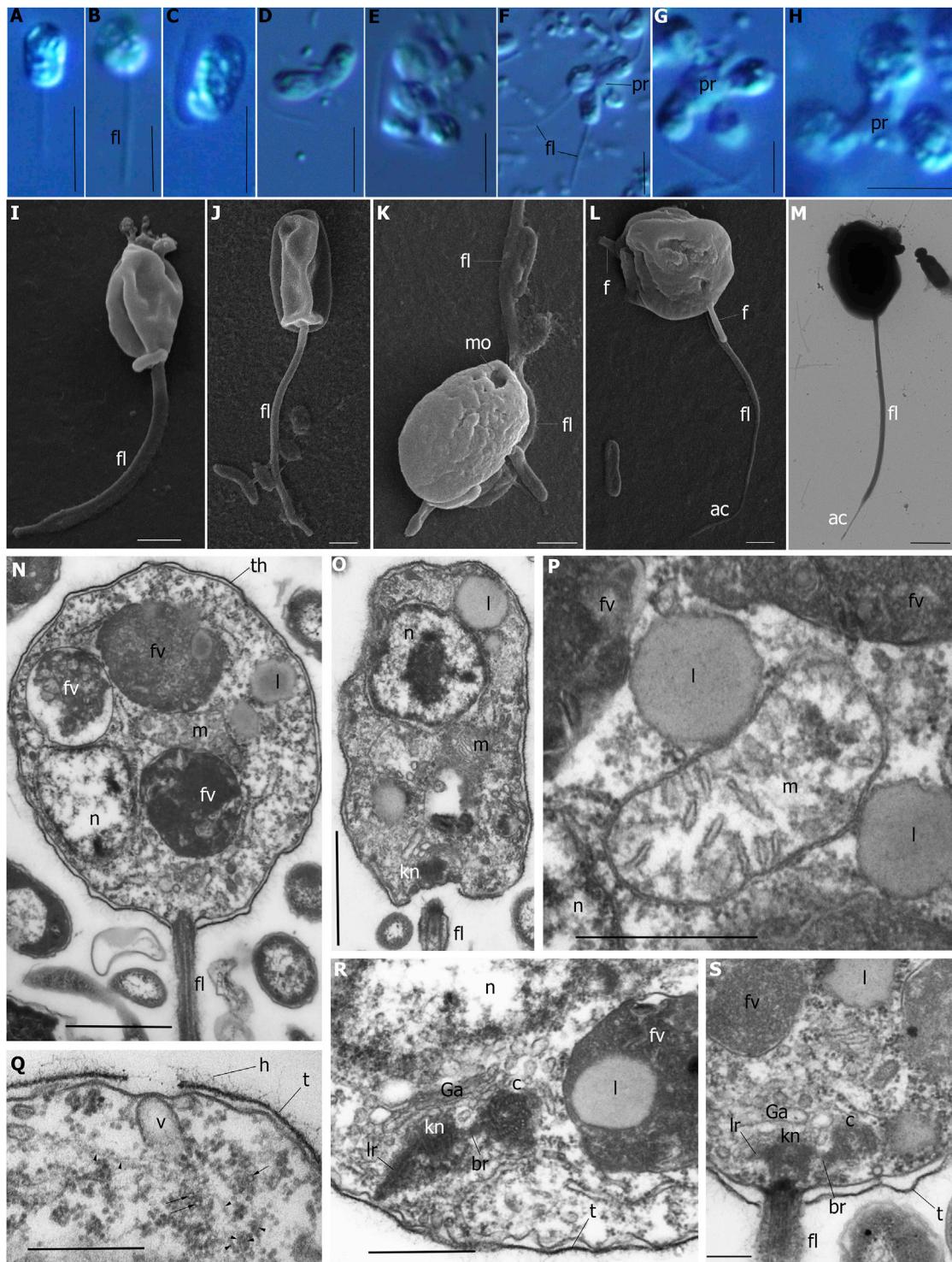


Figure 1. Morphology and Ultrastructure of *Tunicaraptor unikontum*

(A–H) Light microscopy, differential interference contrast (DIC); (I–L) scanning electron microscopy; (M [whole mount] and N–S [cell sections]) transmission electron microscopy.
 (A–C, I, and J) General view of the cell with posterior flagellum (fl), (D) binary fission (cytokinesis), (E) cell aggregation, (F–H) joint feeding on eukaryotic prey (pr) (see also Video S1), (K) cell with small break (mouth [mo]) of the theca at the anterior end, (L) cell with filopodium (f), (M) cell bears acroneme (ac) on the flagellum, and (N and O) longitudinal sections of the vegetative cell showing organelle disposition. Note unstable nuclear position in the cell depending of food uptake: (O) cell before feeding: elongated with anterior nucleus (n) and flagellum and kinetosome (kn) posteriorly; (N) cell after feeding: rounded with food vacuoles (fv) shifting the nucleus (n) backward.

(legend continued on next page)
 Current Biology 30, 4500–4509, November 16, 2020 4501

[6] but unknown in unicellular opisthokonts. Importantly, anterior mouth in *Tunicaraptor* is not associated with the flagellar pocket and apparatus, unlike the ancestral bikont mouth structure. Two orthogonal centrioles are interconnected by a bridge, associated with the Golgi apparatus, posterior to the nucleus (Figure 1R). Centrioles migrate to the cell surface, and one develops the lateral microtubular root and a flagellum, although the other rotates to an acute angle to the kinetosome (Figure 1S). The cell interior also contains prominent food vacuoles (Figures 1N, 1R, and 1S) and a mitochondrion with flat cristae associated with lipid globules (Figure 1P).

We assessed the environmental diversity of *Tunicaraptor* by comparing its small subunit ribosomal RNA (SSU rRNA) to all available data from environmental surveys of microbial diversity. Despite the existence of large SSU rRNA databases from metagenomics and barcode surveys from a wide variety of environments, we found *Tunicaraptor* is not related to any known lineage of environmental sequences, including any of the currently recognized lineages of abundant uncultured opisthokonts (e.g., marine opisthokonts [MAOPs]; Figure S1A) [7–9].

Interpreting the evolutionary significance of animal-specific genes in unicellular relatives depends on a well-resolved phylogeny, and recent advances appear to suggest that phylogenomics is converging on a well-supported framework for the holozoan tree [4, 10, 11]. The phylogenetic position of *Tunicaraptor* in this tree was addressed by adding the transcriptomic data to a novel 200-gene dataset that partially overlaps (Figure S2A) with previously analyzed datasets [4, 11]. The dataset was constructed on the basis of OrthoFinder orthology inference with genomic data and augmented with transcriptomic data using bi-directional similarity searches (STAR Methods). To eliminate contaminating bacterial or eukaryotic prey sequences, the transcriptomic data and each individual alignment were screened and inspected for potential contamination.

Previous phylogenomic studies led to a generally strongly supported opisthokont phylogeny in both Bayesian and maximum likelihood analyses. However, with the addition of *Tunicaraptor*, PhyloBayes reconstruction leads to an unresolved phylogeny because several key nodes lacked convergence between the chains (Figure 2A). Individual chains consistently support a sister relationship of Filasteria to a clade uniting choanoflagellates and animals, but the positions of Pluriformea and *Tunicaraptor* vary between chains. The consensus tree places Pluriformea closer to the filasterian-choanoflagellate-animal group (Filozoa), but Pluriformea is sister to Ichthyosporea in 1 out of 4 chains (both alternatives having been observed previously) [10, 11]. The position of *Tunicaraptor* is even less certain—it is recovered in different positions in all four chains: sister to either Filasteria, Pluriformea, and Filozoa or as the earliest diverging holozoan lineage.

Maximum likelihood (ML) analysis resulted in a different tree topology altogether, uniting Filasteria, Ichthyosporea,

Pluriformea, and *Tunicaraptor* into a monophyletic assemblage (Figure 2B), with Filasteria and *Tunicaraptor* branching with Pluriformea. The ML tree is both weakly supported and at odds with the Bayesian inference. Furthermore, ML trees omitting Pluriformea and *Tunicaraptor* sequences recover the conventional relationship of Filasteria and Ichthyosporea as separate holozoan lineages with high support (Figure 2C).

The lack of phylogenetic congruence prompted us to investigate the position of *T. unikontum* using previously published datasets. Adding *Tunicaraptor* to a 255-gene dataset derived from eukaryote-wide phylogenomic studies [11, 12] and to an 87-gene single-copy protein domain dataset (BVD57) from Grau-Bove et al. [4] both result in *Tunicaraptor* branching with Filasteria (with 71%–85% ultrafast bootstrap support) in ML analyses but disagree on the placement of Pluriformea (Figures S1B and S1C).

Hypothesis testing with 105 possible tree topologies does not fully reconcile the analyzed datasets. Three phylogenetic positions for *T. unikontum* could not be rejected: sister to Filasteria; Filozoa; or to all other Holozoa (Figure S3; Data S1). The latter topology is rejected using the 255-gene dataset after the removal of fast-evolving sites, but not with other datasets. The sister group relationship of Pluriformea and Ichthyosporea is also rejected by the 255-gene dataset following the fast site removal but is favored using the BVD57 dataset. Among the three alignments, the 200-gene dataset is the most permissive in the tests: at least 11 topologies were not rejected at a 5% significance level in any variation of the site removal analysis (Data S1). This includes topologies with a monophyletic union of Filasteria, Ichthyosporea, Pluriformea, and *Tunicaraptor*, which are generally rejected by the other two datasets.

To examine the influence of taxonomic sampling, we expanded both previously published datasets to conform to the taxa represented in the 200-gene alignment. All three alignments show similar degrees of mutational saturation (Figures S2B–S2D), suggesting that the incongruence does not stem from different evolutionary rates. ML reconstruction with the expanded BVD57 alignment does not change the topology but lowers support of several nodes, including the grouping of *Tunicaraptor* with Filasteria (Figure S1D). Taxonomic expansion of the 255-gene alignment, in contrast, results in a different topology, including a highly supported monophyletic assemblage of Filasteria, Ichthyosporea, Pluriformea, and *Tunicaraptor* (Figure S1E), suggesting taxonomic sampling is a significant factor in this dataset.

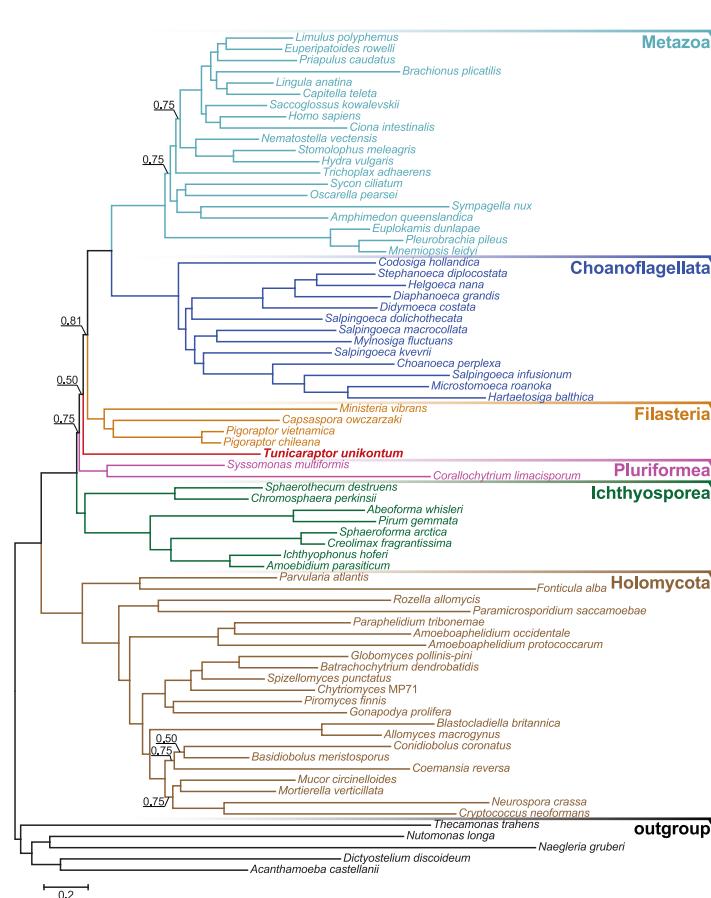
Finally, we attempted to reconcile these differences by combining all three datasets into a 395-gene alignment, spanning a total of 155,000 positions. ML reconstruction recovers the monophyletic union of Filasteria, Ichthyosporea, Pluriformea, and *Tunicaraptor*. The analysis also supports a specific relationship between *Tunicaraptor* and Filasteria and provides moderate

(P) Cell interior showing a mitochondrion (m) associated with lipid globules (l), nucleus (n), and food vacuoles (fv); (Q) anterior end with theca break and proposed cystostomal structures: a vacuole (v) attached to the plasma membrane and associated with microtubules (arrows) and microfilaments (arrowheads). Tunica consists of internal dense layer (t) and long external hairs (h); (R) two centrioles (c) at orthogonal planes to each other with lateral microtubular root (lr) produced by future kinetosome are shown. Centrioles are connected to each other by bridge (br) and associated with Golgi apparatus (Ga), which locates by the nucleus (n). Lipid globule (l) forming in the food vacuole (fv) is visible.

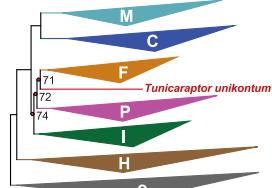
(S) Longitudinal section of flagellar apparatus shows kinetosome (kn) with flagellum (fl) and centriole (c) connected to kinetosome by bridge (br) and lying in one plane at acute angle to it. Golgi cisterns (Ga) underline both the kinetosome with lateral root (lr) and centriole.

See Data S3 for details. Scale bars, (A–H) 5 μm; (I–L) 1 μm; (M) 2 μm; (N and O) 1 μm; (P and R) 500 nm; and (Q and S) 200 nm.

A 200-gene dataset, PhyloBayes analysis



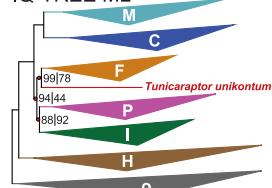
B 200-gene dataset, IQ-TREE ML



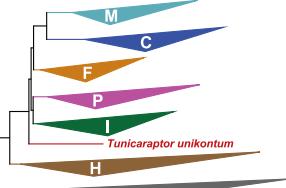
C 200-gene dataset, IQ-TREE ML, excl. Pluriformea, *T.unikontum*



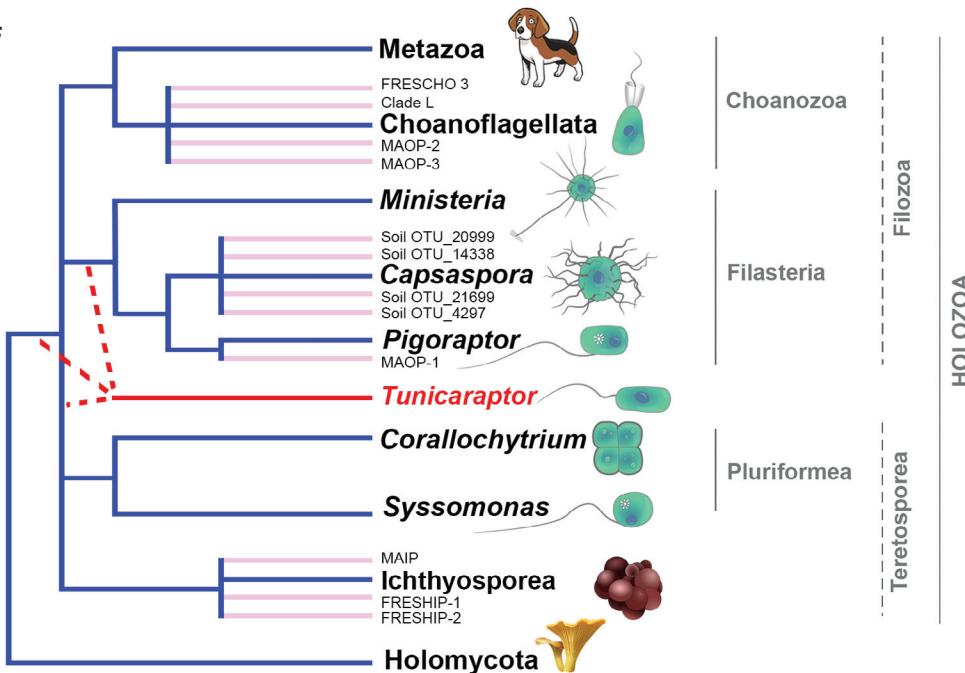
D combined 395-gene dataset, IQ-TREE ML



E combined 395-gene dataset, PhyloBayes, Dayhoff-6 recoded



F



(legend on next page)

support for a relationship between Pluriformea and Ichthyosporea (Figure 2D). Support for the monophyletic union of early holozoans is high in the ML analysis with the full dataset (94% bootstrap support) but decreases drastically when only the holomycotan lineage is used as an outgroup (44% bootstrap support). A similar outgroup effect was reported in previous studies, where restricting the outgroup to Fungi significantly enhanced the support for Filozoa versus the grouping of Filasterea with Ichthyosporea [13]. Interestingly, Bayesian analysis with a recoded dataset recovers *Tunicaraptor* as sister to all other holozoans with complete support (Figure 2E): although it fails to converge across all bipartitions, the phylogenetic relationships between major holozoan lineages are resolved. Topology tests with the combined dataset also failed to reject three possibilities: two variants of a monophyletic union of Filasteria, Ichthyosporea, Pluriformea, and *T. unikontum* and a topology with *T. unikontum* sister to Filasteria and Pluriformea sister to Ichthyosporea (Figure S3). Removing fast-evolving sites leads to the rejection of some topologies but fails to reject additional alternatives, including a topology that places *Tunicaraptor* sister to all other holozoans (Data S1). Additional quality filtering of alignments in the combined dataset using the automated approaches of PREQUAL and Divvier gave essentially the same outcomes for the topology tests, despite notable differences in alignment sizes (Data S1).

Overall, the addition of *Tunicaraptor* demonstrates that holozoan phylogeny is not as well resolved as it appeared in recent analyses but is instead highly sensitive to taxonomic sampling. Given the number of as yet uncharacterized environmental lineages we already know about (Figure 2F), but which are not represented by phylogenomic data, major revisions to the framework have the potential to significantly alter our reconstruction of events leading up to the origin of animals. In addition, all three of the most recently characterized lineages are small, flagellated, eukaryovorous predators (*Syssomonas*, *Pigoraptor*, and *Tunicaraptor*), none of which are well represented in the environmental data. This suggests not only a greater diversity of unicellular animal relatives remains uncharacterized but also that both the methods for isolation and culturing and the methods for environmental sequencing are biased but in different ways. Predators are often widespread, but not abundant, in natural

communities, where they occupy the upper levels of microbial food webs. Although they probably play crucial roles in the flow of energy and nutrients and represent evolutionarily significant branches in the tree of opisthokonts, their real diversity is hard to measure with current environmental survey strategies. Because the actual diversity of unicellular relatives of animals remains poorly characterized, and the topology of trees depends significantly on the taxonomic sampling, we must remain open to future changes to the phylogenetic framework and how these affect our interpretation of the evolution of cellular and genomic innovations leading to the origin of animals.

The Evolution and Distribution of Cell Motility and Adhesion Genes in Holozoa

Unicellular holozoans have diverse motility strategies and associated cytoskeletal adaptations. Even the characteristic opisthokont feature, a single posterior flagellum, has been lost several times independently, each leading to an associated reduction of axonemal dyneins and complexes involved in flagellar assembly and maintenance, including the intraflagellar transport (IFT) complexes, the IFT-associated BBSome complex, and the ciliary transition zone MKS complex (Figure 3; Data S2). These losses can be complete (e.g., some parasitic or commensal ichthyosporeans and the filasterian *Capsaspora*) or partial (e.g., the putatively free-living species with rarely observed flagella like *Corallochytrium*, the ichthyosporean *Chromosphaera perkinsii*, and the filasterian *Ministeria*) [4, 14, 15]. Predatory unicellular holozoans—*Tunicaraptor*, *Syssomonas*, and *Pigoraptor*—all possess complete or nearly complete flagellar complexes, including an ancient cation channel signaling complex, CatSper [16]. This is significant because CatSper has been lost in most unicellular holozoans but retained by all the eukaryovorous predators and the last common ancestor of all animals, pointing to a possible shared evolutionary pressure for regulation of flagellar activity and perhaps a predatory lifestyle linking animals and their unicellular holozoan ancestors.

In addition to flagellar motility, *Tunicaraptor* is capable of producing filopodia (Figure 1L). The transcriptome of *Tunicaraptor* encodes evidence of WASP and WAVE family proteins—actin-nucleation-promoting factors implicated in the formation of dynamic, actin-filled pseudopods across eukaryotes [17], as well

Figure 2. Reconstructions of Early Holozoan Phylogeny

- (A) PhyloBayes consensus tree based on the four chains of CATGTR analysis with the 200-gene alignment (50% burn in). Nodes with posterior probability (pp) below 1.0 are shown with the corresponding pp values; for all of the indicated bipartitions, the analysis failed to achieve convergence between the chains.
 - (B) IQ-TREE ML inference with the 200-gene alignment using the LG+C60+F+G4 model; support values were computed using the ultrafast bootstrap approach with 1,000 replicates. The established opisthokont lineages all receive 100% support in the analysis and are shown in the tree schematically, colored according to (A), and abbreviated by the first letter of the taxon name. Nodes with bootstrap support below 100% are marked, and the associated support values are written next to the nodes.
 - (C) IQ-TREE ML inference with the 200-gene alignment excluding sequences of *T. unikontum* and Pluriformea species *S. multiformis* and *C. limacisporum*; omission of these three sequences restores the filasterian-choanoflagellate-animal group (Filozoa) in the ML analysis.
 - (D) IQ-TREE ML reconstruction with the 395-gene alignment combining three datasets (200-gene, 255-gene, and BVD57); support for the inferred phylogeny of early holozoans is dependent on the outgroup sampling: node support values on the left correspond to reconstruction with the full dataset and values on the right to the dataset that excludes non-opisthokont species.
 - (E) PhyloBayes inference with the combined 395-gene dataset recoded using the Dayhoff-6 scheme. The inner nodes of early holozoan phylogeny receive 1.0 pp in the analysis, but the chains fail to converge for several other bipartitions (maxdiff = 1.0).
 - (F) Schematic summary of the phylogenetic relationships of Opisthokonta, illustrating the currently accepted relationships between major lineages with known taxa and lineages of purely environmental sequences (which are mapped onto the phylogenomic tree topology according to their position in small subunit rRNA gene phylogeny).
- See other phylogenetic reconstructions of the position of *T. unikontum* using previously published datasets and datasets characteristics in Figures S1 and S2. Results of phylogenetic hypothesis testing are present in Figure S3 and Data S1.

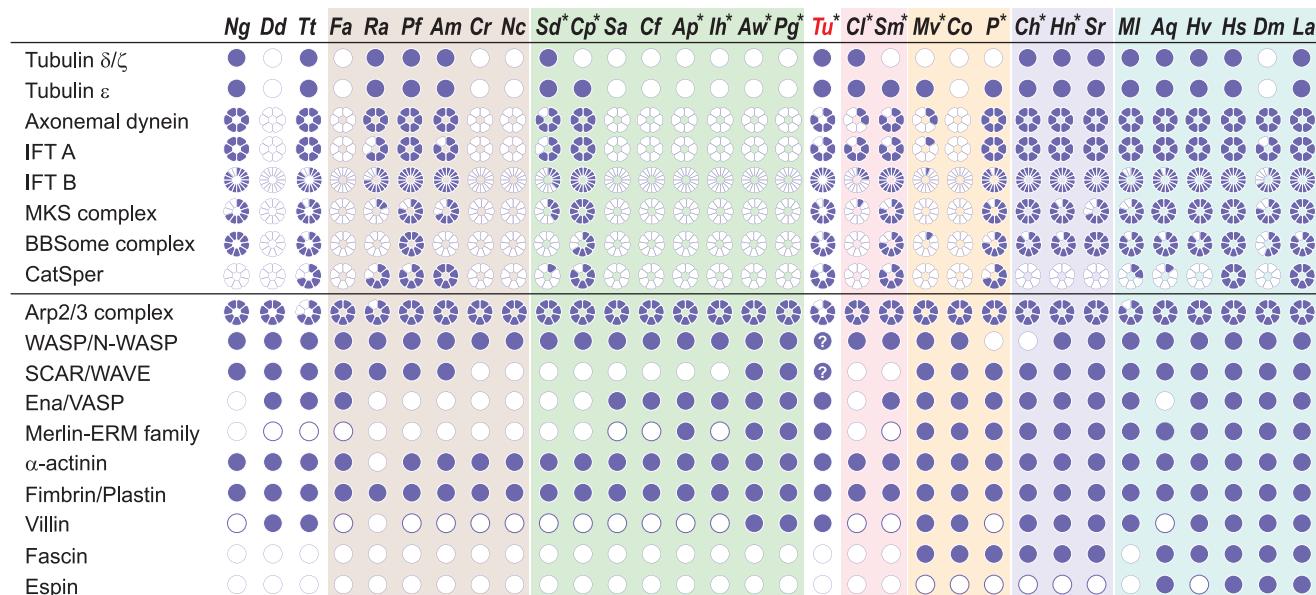


Figure 3. Occurrences of Cytoskeletal Proteins Associated with Genesis and Maintenance of Flagella or Filopodia

Filled shapes mark presence of proteins or complex components as determined by KEGG KAAS, InterProScan or Pfam searches; completeness of protein complexes is represented by fills of circular diagrams (see Data S2 for details). Species abbreviations: outgroup species: Ng, *Naegleria gruberi*; Dd, *Dictyostelium discoideum*; Tt, *Thecamonas trahens*; Holomycota: Fa, *Fonticula alba*; Ra, *Rozella allomyces*; Pf, *Piromyces finnis*; Am, *Allomyces macrogynus*; Cr, *Coemansia reversa*; Nc, *Neurospora crassa*; Ichthyosporea: Sd, *Sphaerothecum destruens*; Cp, *Chromosphaera perkinsii*; Sa, *Sphaeroforma arctica*; Cf, *Creolimax fragrantissima*; Ap, *Amoebidium parasiticum*; Ih, *Ichthyophorus hoferi*; Aw, *Abeoforma whisleri*; Pg, *Pirum gemmata*; Pluriformea: Cl, *Corallochytrium limacisporum*; Sm, *Syssomonas multiformis*; Filasterians: Mv, *Ministeria vibrans*; Co, *Capsaspora owczarzaki*; P, *Pigoraptor spp.*; Choanoflagellata: Ch, *Codosiga hollandica*; Hn, *Helgoeca nana*; Sr, *Salpingoeca rosetta*; Metazoa: MI, *Mnemiopsis leidyi*; Aq, *Amphimedon queenslandica*; Hv, *Hydra vulgaris*; Hs, *Homo sapiens*; Dm, *Drosophila melanogaster*; La, *Lingula anatina*; Tu, *Tunicaraptor unikontum*. Question marks at the WASP and WAVE families for *T. unikontum* signify presence of partial transcripts that support the suggested orthology through reciprocal BLAST searches but lack complete domain architectures for confident identification. Asterisks (*) on names indicate the data come from transcriptomes as opposed to a whole genome.

as Ena/VASP family proteins, which participate in the formation of filopodia by facilitating actin polymerization [18] (Figure 3). A member of the Merlin-ERM family of actin filament and plasma membrane crosslinkers and an actin-bundling protein Villin are also found in the transcriptome. The pseudopodia of *Tunicaraptor* are short, poorly defined, and distinct from the metazoan-type microvilli seen in filasterians. Consistent with the morphological distinction, we find no evidence of actin crosslinking protein fascin, recognized as a conserved filopodia marker protein in filasterians, choanoflagellates, and animals [19].

Unicellular holozoans generally possess several components of animal cell-matrix adhesion complex and the associated elements of extracellular matrix (ECM), despite their lack of multicellular tissues [20, 21]. The crucial components of the integrin adhesome are also conserved in *Tunicaraptor*, which encodes at least two integrin-alpha and two integrin-beta family proteins, suggesting loss of adhesome components is restricted to some choanoflagellates and ichthyosporeans (Figure 4A).

In addition to integrins, *Tunicaraptor* encodes domains of other major classes of animal cell adhesion molecules. Four partial C-type lectin domain proteins and six proteins containing domains of immunoglobulin-like (Ig) superfamily were found. Outside of the choanoflagellate-animal group, the C-type lectin domains have patchy distribution in holozoans, and none are found in filasterians (Figure 4A). *Tunicaraptor* Ig-like domains display highest similarity to the animal I-set, Ig2, and Ig3 immunoglobulin-like

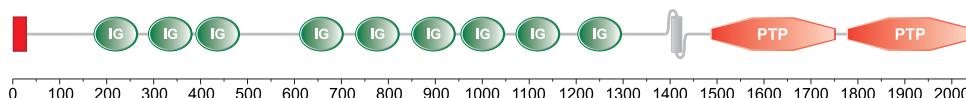
domain families (Data S2). Among holozoans, domains with affinity for the animal Ig-like domain families have only previously been reported in choanoflagellates [22]. Two of *Tunicaraptor* Ig-domain-containing transcripts are predicted to contain a signal peptide sequence; however, only one of these transcripts likely represents a complete protein. The putatively complete transcript encodes a protein tyrosine phosphatase (PTP) that conforms to the classical structure of metazoan transmembrane receptor-like PTP proteins [23]. The extracellular ligand-binding segment of the protein is composed of an array of tandem Ig domains, and the cytoplasmic segment contains two PTP domains (Figure 4B), suggesting that its activity is regulated through dimer formation, similarly to animal-receptor-like PTPs. Another Ig-domain-containing transcript encodes a transmembrane protein with homology to a phosphatidylcholine biosynthesis enzyme lysophosphatidylcholine acyltransferase (LPCAT). Both PTP and LPCAT proteins are common to unicellular holozoans, but the fusion with Ig domains is specific to *Tunicaraptor*. No cadherins were identified in *Tunicaraptor*. The earliest instance of holozoan cadherin domains corresponds to the divergence of filasterians; subsequently, cadherins underwent significant expansions in choanoflagellates and animals [24].

Metazoan Innovations in *Tunicaraptor*

Protein domain searches in *Tunicaraptor* identify several domains otherwise exclusive to animals. *Tunicaraptor* encodes a beta subunit of core-binding factor (CBF β)—an allosteric regulator of Runx

A

	<i>Ng</i>	<i>Dd</i>	<i>Tt</i>	<i>Fa</i>	<i>Ra</i>	<i>Pf</i>	<i>Am</i>	<i>Cr</i>	<i>Nc</i>	<i>Sd*</i>	<i>Cp*</i>	<i>Sa</i>	<i>Cf</i>	<i>Ap*</i>	<i>Ih*</i>	<i>Aw*</i>	<i>Pg*</i>	<i>Tu*</i>	<i>Cl*</i>	<i>Sm*</i>	<i>Mv*</i>	<i>Co</i>	<i>P*</i>	<i>Ch*</i>	<i>Hn*</i>	<i>Sr</i>	<i>MI</i>	<i>Aq</i>	<i>Hv</i>	<i>Hs</i>	<i>Dm</i>	<i>La</i>
Integrin α	●	●	●																													
Integrin β	●	●	●																													
ILK	●	●	●		●	●	●	●	●																							
Paxillin	●	●	●	●	●	●	●	●	●																							
Talin	●	●	●	●	●	●	●	●	●																							
PINCH	●	●	●	●	●	●	●	●	●																							
Parvin	●	●	●	●	●	●	●	●	●																							
Vinculin	●	●	●	●	●	●	●	●	●																							
FAK	●	●	●	●	●	●	●	●	●																							
Laminin G	●	●	●	●	●	●	●	●	●				●																			
Laminin N-terminal (VI)	●	●	●	●	●	●	●	●	●																							
Laminin B (IV)	●	●	●	●	●	●	●	●	●																							
Fibronectin type II	●	●	●	●	●	●	●	●	●				●																			
Fibronectin type III	●	●	●	●	●	●	●	●	●				●																			
Ig-like superfamily	●	●	●	●	●	●	●	●	●									●														
Cadherin	●	●	●	●	●	●	●	●	●																							
Lectin C-type	●	●	●	●	●	●	●	●	●				●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●		

B**Figure 4. Occurrences of Cell Adhesion Domains and Proteins**

(A) Occurrences of integrin adhesome components, ECM domains, and domains of major cell adhesion molecule classes. Filled circles mark presence of domains or proteins; species abbreviations are as in Figure 3.

(B) Domain architecture of Ig-domain-containing protein tyrosine phosphatase in *T. unikontum* with a predicted signal peptide (red), 1.0 probability of Sec signal peptide reported by SignalP.

Domain abbreviations: IG, immunoglobulin-like domain superfamily (IPR036179); PTP, protein tyrosine phosphatase, catalytic domain (SM00194); TMHMM-predicted transmembrane region is shown with a pictogram. Asterisks (*) on names indicate the data come from transcriptomes as opposed to a whole genome. See Figure S4 for phylogeny of ionotropic glutamate receptors and Data S2 for details.

family transcription factors. Runx family transcription factors and the CBF β cooperate as a heterodimeric complex in animal models [25] and were once viewed as a metazoan innovation [26]. Runx proteins have since been reported in unicellular relatives of animals, but CBF β remains absent [27]. *Tunicaraptor* uniquely contains both Runx and CBF β proteins, which suggests cooperative CBF β and Runx interactions also evolved prior to animals.

Several other transcripts in *Tunicaraptor* are homologous to animal-specific proteins implicated in the development and function of the nervous system. *T. unikontum* contains a transcript sharing domains with the repulsive guidance molecule (RGM)—a protein with critical roles in neurodevelopment and axon guidance in animals [28]. The transcript encodes a domain with weak similarity to the N-terminal domain of RGM and is truncated at the C terminus, lacking the C-terminal RGM domain. The C-terminal portion of RGM or the Ig-domain-containing protein Neogenin, a receptor for RGM in animals [29], was not found in the transcriptome. However, a module of tandem Ig domains in one *Tunicaraptor* transcript shows similarity to the structurally related families of animal Ig-domain-containing cell adhesion molecules, including Robo, DSCAM, and contactin, which are involved in axonogenesis [30]. Another transcript encodes an atypical FYVE-type zinc finger that occurs at the N terminus of animal rabphilin and synaptotagmin-like Rab effector proteins, with roles in synaptic vesicle

exocytosis [31]. The zinc finger is part of a Rab-binding domain with a Rab specificity motif SGAWFF [32] conserved in the rabphilin family and the *Tunicaraptor* protein. The protein lacks the commonly associated phospholipid-binding C2 domains of the rabphilin membrane trafficking proteins and thus resembles the structure of Rab effector protein Noc2 [33]. *Tunicaraptor* also has two members of an ionotropic glutamate receptor (iGluR) family—ligand-gated ion channels that mediate excitatory neurotransmission in animals. Animal iGluRs are classified into four subfamilies (AKDF, NMDA, Epsilon, and Lambda) that emerged through duplication events before the divergence of animal phyla [34]. Reconstruction of the iGluR phylogeny places the *T. unikontum* sequences after the divergence of Lambda and NMDA subfamilies and sister to the branch uniting Epsilon and AKDF subfamilies, although alternative positions, including branching of *T. unikontum* iGluRs prior to the NMDA divergence or sister to the Epsilon subfamily, could not be rejected (Figure S4).

DISCUSSION

Tunicaraptor represents a new lineage of holozoan predators, with unique features at the structural and molecular levels. Its phylogenetic position within the holozoans is not yet clear: *Tunicaraptor* may be sister to filasterians, Filozoa, or it may be the

earliest branching holozoan known. What is clear, however, is that including this lineage has a substantial confounding effect on the resolution of conventional views of early holozoan phylogeny. The branching order of unicellular relatives of animals is a critical requirement for accurately reconstructing the steps leading up to the origin of animals. This includes how we interpret all kinds of characteristics, from the complex distribution of “animal-specific” proteins now known to be widespread in their unicellular relatives to the variety of cellular structures, life cycles, and life history strategies of these organisms and how they might have contributed to the origin of multicellularity.

Morphologically, the mouth observed in *Tunicaraptor* raises other interesting questions about early opisthokont evolution, because it is unique among unicellular opisthokonts but common and likely ancestrally present in bikonts, where it is associated with the flagellar pocket at the opposite cell pole. In typical bikonts, beating of the locomotory flagellum pulls them forward through the water. However, if in the ancestor of opisthokonts the main feeding stage was a suspension-feeder attached to a substrate (as in many choanoflagellates, attached *Syssomonas* cells, or *Ministeria*, for example), such flagellar beating is inconsistent with the mode of feeding: flagellar motion would most likely tear the sessile cell from the substrate. Instead, opisthokonts have changed the direction of flagellar beating so that the force propels water away from the cell (creating a current that passes food over the cell). Perhaps this process contributed to the evolutionary extinction of feeding associated with the flagellar pocket and the cytostome in opisthokonts, because the flow of water would make these structures inefficient at best. Remarkably, the mouth in *Tunicaraptor* is not connected with the flagellar apparatus but is instead situated at the opposite cell pole. So the evolutionary history of this structure may be interesting to examine further, as it most likely evolved independently of the canonical mouth structures in the context of opisthokont flagellar motion.

Tunicaraptor also emphasizes the importance of eukaryovorous unicellular holozoans to understanding the transition to multicellularity. Although this was previously unknown, all three of the most recently discovered holozoan lineages are eukaryovorous predators, and they share specific features with the last common ancestor of animals. This raises the possibility of a long-term functional continuity of the predatory lifestyle during holozoan evolution that is not evident from the best-studied unicellular holozoans, which are parasites and bacterial grazers. Such small, inconspicuous, flagellate eukaryovorous predators possess an almost ideal set of characteristics to make them easily overlooked. They are, like predators in general, numerically under-abundant compared with their prey, and physically, they are easily overlooked. More worryingly, these cells closely resemble sperm cells and zoospores of both fungi and aphelid parasites, all of which would lead them to be misidentified and altogether strongly biasing both molecular and morphological observations against their widespread detection and characterization. But collectively, such eukaryovorous organisms now make up almost half the known lineages of unicellular holozoans (Figure 2F), leading us to speculate they may even be a dominant form of life among the key ancestral lineages leading up to the origin of animal multicellularity. Other new holozoan lineages undoubtedly exist and indeed some have even been found in environmental

surveys and remain uncultured (e.g., the marine MAOPs) [7]. Predatory flagellates may represent a major fraction or even the majority of holozoan diversity, and the elucidation of their biological and molecular characteristics will be crucial for understanding of the emergence of multicellular animals.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- **KEY RESOURCES TABLE**
- **RESOURCE AVAILABILITY**
 - Lead Contact
 - Materials Availability
 - Data and Code Availability
- **EXPERIMENTAL MODEL AND SUBJECT DETAILS**
 - Isolation and culturing of novel species and microscopy
- **METHOD DETAILS**
 - Small subunit rRNA gene, RNA-seq, transcriptome assembly and contamination filtering
 - Phylogenetic analyses
 - Transcriptome annotation and protein domain searches
- **QUANTIFICATION AND STATISTICAL ANALYSIS**

SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at <https://doi.org/10.1016/j.cub.2020.08.061>.

A video abstract is available at <https://doi.org/10.1016/j.cub.2020.08.061#mmc7>.

ACKNOWLEDGMENTS

We are thankful to Javier del Campo and Thierry Heger for sharing 18S rRNA sequences and alignments of opisthokont-related environmental lineages and Elena Obushnikova for opisthokont cartoons. This work was supported by the Russian Science Foundation grant no. 18-14-00239 (cell isolation and culturing, sequencing, light and electron microscopy, and analyses) and by the Natural Sciences and Engineering Research Council of Canada grant no. 227301 2014-03994.

AUTHOR CONTRIBUTIONS

Conceptualization, D.V.T., K.V.M., E.H., V.V.A., and P.J.K.; Investigation, K.V.M., D.V.T., S.A.K., K.I.P., A.S.E., O.I.B., Y.A.M., and A.P.M.; Formal Analysis, K.V.M., D.V.T., and S.A.K.; Visualization, K.V.M., D.V.T., S.A.K., K.I.P., A.S.E., O.I.B., and A.P.M.; Supervision, A.P.M., Y.A.M., V.V.A., and P.J.K.; Funding Acquisition, D.V.T. and P.J.K.; Writing – Original Draft, K.V.M., D.V.T., E.H., and P.J.K.; Writing – Review & Editing, all authors.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: May 13, 2020

Revised: July 20, 2020

Accepted: August 17, 2020

Published: September 24, 2020

REFERENCES

1. King, N., Hittinger, C.T., and Carroll, S.B. (2003). Evolution of key cell signaling and adhesion protein families predates animal origins. *Science* 301, 361–363.
2. Ruiz-Trillo, I., Burger, G., Holland, P.W., King, N., Lang, B.F., Roger, A.J., and Gray, M.W. (2007). The origins of multicellularity: a multi-taxon genome initiative. *Trends Genet.* 23, 113–118.
3. Mikhailov, K.V., Konstantinova, A.V., Nikitin, M.A., Troshin, P.V., Rusin, L.Y., Lyubetsky, V.A., Panchin, Y.V., Mylnikov, A.P., Moroz, L.L., Kumar, S., and Aleoshin, V.V. (2009). The origin of Metazoa: a transition from temporal to spatial cell differentiation. *BioEssays* 31, 758–768.
4. 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.
5. Tikhonenkov, D.V., Hehenberger, E., Esaulov, A.S., Belyakova, O.I., Mazei, Y.A., Mylnikov, A.P., and Keeling, P.J. (2020). Insights into the origin of metazoan multicellularity from predatory unicellular relatives of animals. *BMC Biol.* 18, 39.
6. Tikhonenkov, D.V., Strassert, J.F.H., Janouškovec, J., Mylnikov, A.P., Aleoshin, V.V., Burki, F., and Keeling, P.J. (2020). Predatory colponemids are the sister group to all other alveolates. *Mol. Phylogenet. Evol.* 149, 106839.
7. Del Campo, J., Mallo, D., Massana, R., de Vargas, C., Richards, T.A., and Ruiz-Trillo, I. (2015). Diversity and distribution of unicellular opisthokonts along the European coast analysed using high-throughput sequencing. *Environ. Microbiol.* 17, 3195–3207.
8. Arroyo, A.S., López-Escardó, D., Kim, E., Ruiz-Trillo, I., and Najle, S.R. (2018). Novel diversity of deeply branching Holomycota and unicellular holozoans revealed by metabarcoding in Middle Paraná River, Argentina. *Front. Ecol. Evol.* 6, 99.
9. Heger, T.J., Giesbrecht, I.J.W., Gustavsen, J., Del Campo, J., Kellogg, C.T.E., Hoffman, K.M., Lertzman, K., Mohn, W.W., and Keeling, P.J. (2018). High-throughput environmental sequencing reveals high diversity of litter and moss associated protist communities along a gradient of drainage and tree productivity. *Environ. Microbiol.* 20, 1185–1203.
10. López-Escardó, D., Grau-Bové, X., Guillaumet-Adkins, A., Gut, M., Sieracki, M.E., and Ruiz-Trillo, I. (2019). Reconstruction of protein domain evolution using single-cell amplified genomes of uncultured choanoflagellates sheds light on the origin of animals. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 374, 20190088.
11. 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.e6.
12. 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.
13. Torruella, G., Derelle, R., Paps, J., Lang, B.F., Roger, A.J., Shalchian-Tabrizi, K., and Ruiz-Trillo, I. (2012). Phylogenetic relationships within the Opisthokonta based on phylogenomic analyses of conserved single-copy protein domains. *Mol. Biol. Evol.* 29, 531–544.
14. Torruella, G., de Mendoza, A., Grau-Bové, X., Antó, M., Chaplin, M.A., del Campo, J., Eme, L., Pérez-Cordón, G., Whipples, C.M., Nichols, K.M., et al. (2015). Phylogenomics reveals convergent evolution of lifestyles in close relatives of Animals and Fungi. *Curr. Biol.* 25, 2404–2410.
15. Mylnikov, A.P., Tikhonenkov, D.V., Karpov, S.A., and Wylezich, C. (2019). Microscopical studies on *Ministeria vibrans* Tong, 1997 (Filasterea) highlight the cytoskeletal structure of the common ancestor of Filasterea, Metazoa and Choanoflagellata. *Protist* 170, 385–396.
16. Cai, X., Wang, X., Patel, S., and Clapham, D.E. (2015). Insights into the early evolution of animal calcium signaling machinery: a unicellular point of view. *Cell Calcium* 57, 166–173.
17. Fritz-Laylin, L.K., Lord, S.J., and Mullins, R.D. (2017). WASP and SCAR are evolutionarily conserved in actin-filled pseudopod-based motility. *J. Cell Biol.* 216, 1673–1688.
18. Krause, M., Dent, E.W., Bear, J.E., Loureiro, J.J., and Gertler, F.B. (2003). Ena/VASP proteins: regulators of the actin cytoskeleton and cell migration. *Annu. Rev. Cell Dev. Biol.* 19, 541–564.
19. Sebé-Pedrós, A., Burkhardt, P., Sánchez-Pons, N., Fairclough, S.R., Lang, B.F., King, N., and Ruiz-Trillo, I. (2013). Insights into the origin of metazoan filopodia and microvilli. *Mol. Biol. Evol.* 30, 2013–2023.
20. Sebé-Pedrós, A., Roger, A.J., Lang, B.F., King, N., and Ruiz-Trillo, I. (2010). Ancient origin of the integrin-mediated adhesion and signaling machinery. *Proc. Natl. Acad. Sci. USA* 107, 10142–10147.
21. Sebé-Pedrós, A., Irimia, M., Del Campo, J., Parra-Acero, H., Russ, C., Nusbaum, C., Blencowe, B.J., and Ruiz-Trillo, I. (2013). Regulated aggregative multicellularity in a close unicellular relative of metazoa. *eLife* 2, e01287.
22. King, N., Westbrook, M.J., Young, S.L., Kuo, A., Abedin, M., Chapman, J., Fairclough, S., Hellsten, U., Isogai, Y., Letunic, I., et al. (2008). The genome of the choanoflagellate *Monosiga brevicollis* and the origin of metazoans. *Nature* 451, 783–788.
23. Tonks, N.K. (2006). Protein tyrosine phosphatases: from genes, to function, to disease. *Nat. Rev. Mol. Cell Biol.* 7, 833–846.
24. Nichols, S.A., Roberts, B.W., Richter, D.J., Fairclough, S.R., and King, N. (2012). Origin of metazoan cadherin diversity and the antiquity of the classical cadherin/β-catenin complex. *Proc. Natl. Acad. Sci. USA* 109, 13046–13051.
25. Ogawa, E., Inuzuka, M., Maruyama, M., Satake, M., Naito-Fujimoto, M., Ito, Y., and Shigesada, K. (1993). Molecular cloning and characterization of PEBP2 beta, the heterodimeric partner of a novel *Drosophila* runt-related DNA binding protein PEBP2 alpha. *Virology* 194, 314–331.
26. Sullivan, J.C., Sher, D., Eisenstein, M., Shigesada, K., Reitzel, A.M., Marlow, H., Levanon, D., Groner, Y., Finnerty, J.R., and Gat, U. (2008). The evolutionary origin of the Runx/CBFbeta transcription factors—studies of the most basal metazoans. *BMC Evol. Biol.* 8, 228.
27. Sebé-Pedrós, A., de Mendoza, A., Lang, B.F., Degnan, B.M., and Ruiz-Trillo, I. (2011). Unexpected repertoire of metazoan transcription factors in the unicellular holozoan *Capsaspora owczarzaki*. *Mol. Biol. Evol.* 28, 1241–1254.
28. Key, B., and Lah, G.J. (2012). Repulsive guidance molecule A (RGMA): a molecule for all seasons. *Cell Adhes. Migr.* 6, 85–90.
29. Rajagopalan, S., Deitinghoff, L., Davis, D., Conrad, S., Skutella, T., Chedotal, A., Mueller, B.K., and Strittmatter, S.M. (2004). Neogenin mediates the action of repulsive guidance molecule. *Nat. Cell Biol.* 6, 756–762.
30. Liebeskind, B.J., Hofmann, H.A., Hillis, D.M., and Zakon, H.H. (2017). Evolution of animal neural systems. *Annu. Rev. Ecol. Evol. Syst.* 48, 377–398.
31. Fukuda, M. (2007). The role of synaptotagmin and synaptotagmin-like protein (Slp) in regulated exocytosis. In *Molecular Mechanisms of Exocytosis*, R. Regazzi, ed. (Springer New York), pp. 42–61.
32. Ostermeier, C., and Brunger, A.T. (1999). Structural basis of Rab effector specificity: crystal structure of the small G protein Rab3A complexed with the effector domain of rabphilin-3A. *Cell* 96, 363–374.
33. Kotake, K., Ozaki, N., Mizuta, M., Sekiya, S., Inagaki, N., and Seino, S. (1997). Noc2, a putative zinc finger protein involved in exocytosis in endocrine cells. *J. Biol. Chem.* 272, 29407–29410.
34. Ramos-Vicente, D., Ji, J., Gratacós-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.
35. Luft, J.H. (1961). Improvements in epoxy resin embedding methods. *J. Biophys. Biochem. Cytol.* 9, 409–414.
36. Tikhonenkov, D.V., Janouškovec, J., Keeling, P.J., and Mylnikov, A.P. (2016). The morphology, ultrastructure and SSU rRNA gene sequence of a new freshwater flagellate, *Neobodo borokensis* n. sp. (Kinetoplastea, Excavata). *J. Eukaryot. Microbiol.* 63, 220–232.

37. Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D.J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* 25, 3389–3402.
38. Eddy, S.R. (2011). Accelerated profile HMM searches. *PLoS Comput. Biol.* 7, e1002195.
39. The UniProt Consortium (2017). UniProt: the universal protein knowledgebase. *Nucleic Acids Res.* 45 (D1), D158–D169.
40. Finn, R.D., Coggill, P., Eberhardt, R.Y., Eddy, S.R., Mistry, J., Mitchell, A.L., Potter, S.C., Punta, M., Qureshi, M., Sangrador-Vegas, A., et al. (2016). The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Res.* 44 (D1), D279–D285.
41. Li, W., and Godzik, A. (2006). Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* 22, 1658–1659.
42. Simão, F.A., Waterhouse, R.M., Ioannidis, P., Kriventseva, E.V., and Zdobnov, E.M. (2015). BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* 31, 3210–3212.
43. Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* 17, 10–12.
44. Capella-Gutiérrez, S., Silla-Martínez, J.M., and Gabaldón, T. (2009). trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* 25, 1972–1973.
45. Ronquist, F., Teslenko, M., van der Mark, P., Ayres, D.L., Darling, A., Höhna, S., Larget, B., Liu, L., Suchard, M.A., and Huelsenbeck, J.P. (2012). MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst. Biol.* 61, 539–542.
46. Emms, D.M., and Kelly, S. (2015). OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. *Genome Biol.* 16, 157.
47. Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30, 1312–1313.
48. Ebersberger, I., Strauss, S., and von Haeseler, A. (2009). HaMStR: profile hidden markov model based search for orthologs in ESTs. *BMC Evol. Biol.* 9, 157.
49. Hall, T.A. (1999). BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symp. Ser.* 41, 95–98.
50. Whelan, S., Irisarri, I., and Burki, F. (2018). PREQUAL: detecting non-homologous characters in sets of unaligned homologous sequences. *Bioinformatics* 34, 3929–3930.
51. Nguyen, L.T., Schmidt, H.A., von Haeseler, A., and Minh, B.Q. (2015). IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* 32, 268–274.
52. Minh, B.Q., Nguyen, M.A., and von Haeseler, A. (2013). Ultrafast approximation for phylogenetic bootstrap. *Mol. Biol. Evol.* 30, 1188–1195.
53. Dayhoff, M., Schwartz, R., and Orcutt, B. (1978). A model of evolutionary change in proteins. In *Atlas of Protein Sequence and Structure*, M. Dayhoff, ed. (National Biomedical Research Foundation), pp. 345–352.
54. Kumar, S., Stecher, G., and Tamura, K. (2016). MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Mol. Biol. Evol.* 33, 1870–1874.
55. Zhong, M., Hansen, B., Nesnidal, M., Golombek, A., Halanych, K.M., and Struck, T.H. (2011). Detecting the symplesiomorphy trap: a multigene phylogenetic analysis of terebelliform annelids. *BMC Evol. Biol.* 11, 369.
56. Kück, P., and Struck, T.H. (2014). BaCoCa—a heuristic software tool for the parallel assessment of sequence biases in hundreds of gene and taxon partitions. *Mol. Phylogenet. Evol.* 70, 94–98.
57. Jones, P., Binns, D., Chang, H.Y., Fraser, M., Li, W., McAnulla, C., McWilliam, H., Maslen, J., Mitchell, A., Nuka, G., et al. (2014). InterProScan 5: genome-scale protein function classification. *Bioinformatics* 30, 1236–1240.
58. Mitchell, A.L., Attwood, T.K., Babbitt, P.C., Blum, M., Bork, P., Bridge, A., Brown, S.D., Chang, H.-Y., El-Gebali, S., Fraser, M.I., et al. (2019). InterPro in 2019: improving coverage, classification and access to protein sequence annotations. *Nucleic Acids Res.* 47 (D1), D351–D360.
59. Lu, S., Wang, J., Chitsaz, F., Derbyshire, M.K., Geer, R.C., Gonzales, N.R., Gwadz, M., Hurwitz, D.I., Marchler, G.H., Song, J.S., et al. (2020). CDD/SPARCLE: the conserved domain database in 2020. *Nucleic Acids Res.* 48 (D1), D265–D268.
60. Letunic, I., and Bork, P. (2018). 20 years of the SMART protein domain annotation resource. *Nucleic Acids Res.* 46 (D1), D493–D496.
61. Bolger, A.M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30, 2114–2120.
62. Grabherr, M.G., Haas, B.J., Yassour, M., Levin, J.Z., Thompson, D.A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., et al. (2011). Full-length transcriptome assembly from RNA-seq data without a reference genome. *Nat. Biotechnol.* 29, 644–652.
63. Haas, B.J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P.D., Bowden, J., Couger, M.B., Eccles, D., Li, B., Lieber, M., et al. (2013). De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat. Protoc.* 8, 1494–1512.
64. O'Leary, N.A., Wright, M.W., Brister, J.R., Ciuffo, S., Haddad, D., McVeigh, R., et al. (2016). Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Res.* 44, D733–D745.
65. Fournet, M., and Gibbs, M.J. (2006). PATRIC: a program for calculating patristic distances and graphically comparing the components of genetic change. *BMC Evol. Biol.* 6, 1.
66. Kanehisa, M., and Goto, S. (2000). KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res.* 28, 27–30.
67. Moriya, Y., Itoh, M., Okuda, S., Yoshizawa, A.C., and Kanehisa, M. (2007). KAAS: an automatic genome annotation and pathway reconstruction server. *Nucleic Acids Res.* 35, W182–W185.
68. Letunic, I., and Bork, P. (2019). Interactive Tree Of Life (iTOL) v4: recent updates and new developments. *Nucleic Acids Res.* 47 (W1), W256–W259.
69. Katoh, K., and Standley, D.M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* 30, 772–780.
70. Roure, B., Rodriguez-Ezpeleta, N., and Philippe, H. (2007). SCaFoS: a tool for selection, concatenation and fusion of sequences for phylogenomics. *BMC Evol. Biol.* 7 (Suppl 1), S2.
71. Ali, R.H., Bogusz, M., and Whelan, S. (2019). Identifying clusters of high confidence homologies in multiple sequence alignments. *Mol. Biol. Evol.* 36, 2340–2351.
72. Lartillot, N., Rodrigue, N., Stubbs, D., and Richer, J. (2013). PhyloBayes MPI: phylogenetic reconstruction with infinite mixtures of profiles in a parallel environment. *Syst. Biol.* 62, 611–615.
73. Krogh, A., Larsson, B., von Heijne, G., and Sonnhammer, E.L. (2001). Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J. Mol. Biol.* 305, 567–580.
74. Almagro Armenteros, J.J., Tsirigos, K.D., Sønderby, C.K., Petersen, T.N., Winther, O., Brunak, S., von Heijne, G., and Nielsen, H. (2019). SignalP 5.0 improves signal peptide predictions using deep neural networks. *Nat. Biotechnol.* 37, 420–423.
75. Buchfink, B., Xie, C., and Huson, D.H. (2015). Fast and sensitive protein alignment using DIAMOND. *Nat. Methods* 12, 59–60.
76. Shen, W., and Xiong, J. (2019). TaxonKit: a cross-platform and efficient NCBI taxonomy toolkit. bioRxiv. <https://doi.org/10.1101/513523>.
77. Shimodaira, H. (2002). An approximately unbiased test of phylogenetic tree selection. *Syst. Biol.* 51, 492–508.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Biological Samples		
Sample of bottom detritus	Marine lagoon Cabeza de Mar, South Patagonia, Chile	Sample #18, IBIW RAS
Chemicals, Peptides, and Recombinant Proteins		
Glutaraldehyde	Carl Roth GmbH	Art.-Nr. 4157.1
Osmium Tetroxide	Electron Microscopy Sciences	Cat# 19100
Cacodylic Acid	Sigma Chemicals Co	Lot# 14H02591
Hexamethyldisilazane	Aldrich Chemistry	Lot# SHBB0080V
Poly-L-lysine	Sigma-Aldrich	Product Number: P8920-100ML
Araldite 502	Electron Microscopy Sciences	Cat# 10900
Epon-812 Substitute EMbed-812	Electron Microscopy Sciences	Cat# 14900
Critical Commercial Assays		
RNAqueous-Micro Kit	Invitrogen	Cat# AM1931
MasterPure Complete DNA and RNA Purification Kit	Epicenter	Cat.# MC85200
SMART-Seq v4 Ultra Low Input RNA Kit for Sequencing	Takara Bio	Cat.# 634889
Nextera XT DNA Library Prep Kit	Illumina	FC-131-1024
Deposited Data		
<i>Tunicaraptor unikontum</i> 18S rRNA	This study	GenBank: MT611055
Phylogenomic datasets	This study	Mendeley data depository https://doi.org/10.17632/dnwzj78w8w
<i>Tunicaraptor unikontum</i> transcriptome sequencing data and assembly	This study	NCBI BioProject: PRJNA638967
Experimental Models: Organisms/Strains		
Strain Opistho-3	Marine lagoon Cabeza de Mar, South Patagonia, Chile	MI-PR423; urn:lsid:zoobank.org:act:A1E21FFC-F8AE-40F9-AEBF-17EC126B67C6
Oligonucleotides		
ATGCTTGTCCTCAAAG	[32]	18SFU
RYTAAGCCATGC		
CWGTTCACCWACGG	[32]	18SRU
AAACCTTGTACG		
Software and Algorithms		
Trimmomatic	[33]	RRID: SCR_011848
Cutadapt	[34]	RRID: SCR_011841
Trinity	[35]	https://github.com/trinityrnaseq/trinityrnaseq
TransDecoder	[36]	RRID: SCR_017647
CD-HIT	[37]	https://github.com/weizhongli/cdhit
BUSCO	[38]	RRID: SCR_015008
MAFFT	[39]	RRID: SCR_011811
MrBayes	[40]	RRID: SCR_012067
OrthoFinder	[41]	RRID: SCR_017118
RAxML	[42]	RRID: SCR_006086
HMMER	[43]	RRID: SCR_005305
BioEdit	[44]	RRID: SCR_007361

(Continued on next page)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
trimAl	[45]	RRID: SCR_017334
SCaFoS	[46]	https://megasun.bch.umontreal.ca/Software/scafos/scafos.html
PREQUAL	[47]	https://github.com/simonwhelan/prequal
Divvier	[48]	https://github.com/simonwhelan/Divvier
IQ-TREE	[49]	RRID: SCR_017254
PhyloBayes MPI	[50]	RRID: SCR_006402
MEGA	[51]	RRID: SCR_000667
iTOL	[52]	RRID: SCR_018174
BaCoCa	[53]	https://github.com/PatrickKueck/BaCoCa
PATRISTIC	[54]	http://www.bioinformatics.org/patristic/
KAAS	[55]	https://www.genome.jp/kegg/kaas/
PfamScan	EMBL-EBI repository	ftp://ftp.ebi.ac.uk/pub/databases/Pfam/Tools/PfamScan.tar.gz
InterProScan	[56]	RRID: SCR_005829
TMHMM	[57]	RRID: SCR_014935
SignalP	[58]	RRID: SCR_015644
DIAMOND	[59]	RRID: SCR_016071
TaxonKit	[60]	https://bioinf.shenwei.me/taxonkit/
NCBI BLAST	[61]	RRID: SCR_004870
Other		
UniProtKB/Swiss-Prot database	[62]	RRID: SCR_004426
Pfam 32.0 database	[63]	RRID: SCR_004726
eukaryota_odb9 dataset	[38]	https://busco-archive.ezlab.org/v3/datasets/eukaryota_odb9.tar.gz
NCBI Microbial RefSeq database	[64]	https://ftp.ncbi.nlm.nih.gov/genomes/refseq/bacteria
InterPro v74.0 database	[65]	RRID: SCR_006695
NCBI CDD	[66]	RRID: SCR_002077
SMART	[67]	RRID: SCR_005026
NCBI	https://www.ncbi.nlm.nih.gov/	RRID: SCR_006472
JGI Genome Portal	https://genome.jgi.doe.gov/portal/	RRID: SCR_002383
NHGRI Research Projects	https://research.nhgri.nih.gov/	N/A
Multicellgenome lab	http://multicellgenome.com/	N/A
datadryad.org	https://dx.doi.org/10.5061/dryad.26bv4	N/A
figshare.com	https://dx.doi.org/10.6084/m9.figshare.5686984	N/A
KEGG	[68]	RRID: SCR_012773

RESOURCE AVAILABILITY

Lead Contact

Further information and requests for resources and data should be directed to and will be fulfilled by the Lead Contact, Denis Tikhonov (tikho-denis@yandex.ru).

Materials Availability

This study did not generate new unique reagents.

Data and Code Availability

Tunicaraptor unikontum transcriptome sequencing data and assembly are available at the NCBI BioProject PRJNA638967. Phylogenomic datasets including single-gene alignments and trees have been deposited to the Mendeley data depository at <https://doi.org/10.17632/dnwzj78w8w>.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Isolation and culturing of novel species and microscopy

Tunicaraptor unikontum (clone Opisto-3) was obtained from the sample of bottom detritus in marine lagoon Cabeza de Mar (salinity = 33 ‰, T = 10.2°C, pH = 8.4; -4 m altitude), S 52°45'50.6," W 70°58'55.1," South Patagonia, Chile on November 9, 2015. The samples were examined on the third, sixth and ninth days of incubation. Following isolation by glass micropipette, clone Opisto-3 was propagated on the bodonid *Procryptobia sorokini* (strain B-69) grown in marine Schmalz–Pratt's medium (20‰; 18.48 g NaCl, 0.44 g KCl, 3.62 g MgCl₂·6H₂O, 0.71 g MgSO₄·7H₂O, 0.95 g CaCl₂·H₂O, 0.07 g KNO₃, 0.007 g K₂HPO₄·3H₂O l⁻¹ water) by using the bacterium *Pseudomonas fluorescens* as food. The clone Opisto-3 is stored in the "Live culture collection of free-living amoebae, heterotrophic flagellates and heliozoans" at the Papanin Institute for Biology of Inland Waters, Russian Academy of Science.

Light microscopy observations were made by using the Zeiss Axio Scope A.1 equipped with a DIC contrast water immersion objective (63x). The images were taken with the AVT HORN MC-1009/S analog video camera and directly digitized by using the Behold TV 409 FM tuner.

For transmission electron microscopy (TEM), cells were centrifuged, fixed at 1°C for 15–60 min in a cocktail of 0.6% glutaraldehyde and 2% OsO₄ (final concentration) prepared using a 0.1 M cacodylate buffer (pH 7.2). Fixed cells were dehydrated in alcohol and acetone series (30, 50, 70, 96, and 100%, 20 minutes in each step). Afterward, the cells were embedded in a mixture of Araldite and Epon [35]. Ultrathin sections (50 nm) were prepared with an Leica EM UC6 ultramicrotome (Leica Microsystems, Germany) and observed by using the JEM 1011 transmission electron microscope (JEOL, Japan).

For scanning electron microscopy, cells from a culture in exponential growth phase were fixed with 2.5% glutaraldehyde (final concentration). The cells were mounted on a glass coverslip coated with poly-L-lysine for 30 min and subsequently rinsed three times with 0.1 M sodium cacodylate buffer (pH 7.34), which was diluted twice with Schmalz–Pratt's medium. Next, cells were fixed in 1% osmium tetroxide for 1 h. The fixed cells were rinsed three times with distilled water, 10 min each time, and dehydrated with a graded ethanol series from 30% to absolute ethanol (10 min per step), followed by 100% hexamethyldisilazane (three times, 15 min each) and dried at 65°C. Dry glass coverslips were mounted on aluminum stubs, coated with gold-palladium, and observed with a JSM-6510LV scanning electron microscope (JEOL, Japan).

METHOD DETAILS

Small subunit rRNA gene, RNA-seq, transcriptome assembly and contamination filtering

Cells grown in clonal laboratory cultures were collected when the cultures had reached peak abundance and after most of the prey cells had been eaten (based on daily light microscopy observations). Cells were collected by centrifugation (1,000g, room temperature) onto a 0.8-μm membrane of a Vivaclear mini column (Sartorius Stedim Biotech, VK01P042). Total RNA was then extracted using an RNAAqueous-Micro Kit (Invitrogen, AM1931). Total DNA was also extracted from the membrane of a Vivaclear mini column using the MasterPure Complete DNA and RNA Purification Kit (Epicenter, Cat. No. MC85200). The SSU rRNA gene of *T. unikontum* was amplified by polymerase chain reaction (PCR) using the general eukaryotic primers 18SFU-18SRU [36]. PCR products were subsequently cloned and sequenced using Sanger dideoxy sequencing.

Two cDNA libraries for *Tunicaraptor unikontum* were prepared from RNA using a Smart Seq 2 and Smart Seq 4 protocols and Nextera XT kit and sequenced with a MiSeq instrument (Illumina), generating 13M read pairs for each library. The reads were processed with Trimmomatic [61] and cutadapt [43] programs to remove sequencing and Nextera XT kit adaptor sequences. Processed reads from both libraries were assembled together with the Trinity software [62], and ORF predictions were carried out with TransDecoder [63]. BLAST [37] and HMMER [38] searches against the UniProtKB/Swiss-Prot [39] and Pfam [40] databases were performed to assist the ORF prediction. Predicted peptide sequences were clustered with CD-HIT [41] using a 90% identity threshold to reduce the sequence redundancy. The assembly completeness was estimated with BUSCO [42] using the eukaryota_odb9 dataset.

The RNA samples of *T. unikontum* were obtained from a culture containing eukaryotic euglenozoan prey *Procryptobia sorokini* and bacterial contamination. To identify potentially contaminating sequences in the assembled set of peptides we performed BLAST [37] search against the NCBI's non-redundant database. The search was done with an e-value cutoff of 1E-03, and the sequences were classified on the basis of the taxonomic assignment of best hit. From the full set of 64,204 predicted peptides 18,585 were removed as likely prokaryotic or viral contamination, and 9,077 were discarded by producing best hit against sequences classified as Euglenozoa. The cleaned set of *T. unikontum* contains 12,994 peptides; additional 23,548 peptides yield no hits in the database. The cleaned set of peptides was used for all downstream analyses. The filtered transcriptome of *T. unikontum* is evaluated by BUSCO to have 78.6% completeness rating with 13.9% fragmented sequences.

Phylogenetic analyses

For small subunit rRNA gene phylogeny, the obtained SSU rRNA gene sequence of *T. unikontum*, along with MAOP sequences [7, 8] and environmental sequences of Filasterea from Heger et al. [9] were added to the previously constructed alignment of rRNA genes

[11] using the MAFFT [69] option to add full length sequences (`-add`) and the L-INS-i refinement method. The resulting alignment was trimmed by trimAI [44] with a 0.1 gap threshold setting. Phylogeny reconstruction for the small subunit rRNA genes was performed with MrBayes [45] utilizing the GTR substitution model with eight categories of Gamma-distributed among site rate variation and calculation of proportion of invariable sites. The analysis was performed with four independent runs of four Metropolis-coupled Markov chains, sampled across 20 million generations and summarized with a 50% burn-in.

Orthologous groups for multigene phylogenetic analysis with opisthokont taxa were initially constructed on the basis of the genomic data of 52 opisthokont and 3 non-opisthokont species. Ortholog clustering with the genomic data was performed with OrthoFinder [46]. The alignments of inferred orthogroups were prepared with MAFFT using the L-INS-i algorithm [69], and the ML trees constructed with RAxML [47]. The candidate alignments were inspected for cases of erroneous or ambiguous orthology and discarded or corrected when necessary. We selected and manually reviewed 200 orthologous groups from the result of OrthoFinder clustering. To extend the taxonomic sampling of alignments we utilized HMMER [38] searches with the alignment profiles of selected orthogroups and reciprocal BLAST [37] searches for accurate orthology assignments, in a pipeline analogous to the HaMStR approach [48]. Additional BLAST searches were performed for candidate orthologs in transcriptomic datasets against databases of prokaryotic sequences and likely eukaryotic contamination to filter out potentially contaminating sequences. Contamination filtering was performed using the NCBI's Microbial RefSeq database [64] and a selection of eukaryotic genomes, including algal and kinetoplastid organisms commonly used as prey. Several transcriptomes were additionally screened for specific contamination identified during the alignment inspection step: *Ministeria vibrans* for *Allomyces*, *Nutomonas longa* for *Ichthyophonus*, *Pirum gemmata* for human sequences. Searches against databases of potential contaminants were performed with an E-value threshold of 1e-05. Only transcripts scoring lower against the databases of contaminating sequences than the orthogroup genomic sequences were kept for the analyses. Selected orthogroups extended with novel sequences were realigned with MAFFT [69] using the L-INS-i algorithm, and inspected manually using BioEdit [49]. The alignments were trimmed with trimAI [44] using the automated trimming heuristic, and concatenated by SCaFoS [70] into a data matrix containing 75 species with 98,717 aligned amino acid sites.

To cross-check the results of phylogenetic reconstructions with the novel 200-gene dataset, we used two datasets employed in the earlier phylogenomic analyses [4, 11]. The additional datasets were reconstructed by identifying the corresponding OrthoFinder orthogroups using shared orthologous sequences. Taxonomic sampling for these datasets was adjusted to conform to the sampling of the 200-gene dataset using the approach relying on alignment profile and reciprocal BLAST searches. For the combined dataset, the union of three analyzed datasets, the constituent orthogroups were also vetted through manual inspection of alignments. Four orthogroups, corresponding to COR01C, gnb3, FAM96B from the Hehenberger et al. 255-gene dataset [11], and PF00215 from the BVD57 dataset [4], were discarded as problematic. The combined 395-gene dataset was concatenated by SCaFoS [70] into a data matrix with 155,102 amino acid sites. To confirm that uncertainties in phylogenies were not stemming from feasibly preventable sequence or alignment artifacts, we assembled and tested a second version of the 395-gene dataset, by applying automated quality filtering procedures implemented in PREQUAL [50] and Divvier [71] programs. Quality filtering for unaligned sets of orthologous sequences was done by PREQUAL with a 0.95 posterior probability filtering threshold. Filtered sequence sets were aligned with MAFFT using the L-INS-i algorithm, and the alignments were processed with Divvier using the divvygap option and requiring a minimum of 4 characters per column for output. Processed alignments were further trimmed by trimAI with a gap threshold of 0.2, and concatenated by SCaFoS into an alignment with 180,785 amino acid sites.

Phylogeny reconstructions were carried out with maximum likelihood (ML) and the Bayesian inference approaches. ML analyses for all datasets were performed with IQ-TREE [51] using the LG+C60+F+G4 profile mixture model, and ultrafast bootstrap approximation [52] with 1000 replicates for estimation of branch support. Bayesian inference was performed with PhyloBayes MPI [72] for the 200-gene alignment and for the 395-gene combined alignment. The combined alignment utilized a special alphabet corresponding to the Dayhoff recoding scheme with 6 amino acid groups: AGPST, DENQ, HKR, MIVL, WFY, C [53]. PhyloBayes inference for both datasets was done using the CAT-GTR model and 4 discrete Gamma rate categories in four independent chains that were run for 10,000 cycles and sampled with a 50% burn-in. Tree visualization was done with the help of MEGA [54] and iTOL [68] software. ML analyses for all datasets were performed with IQ-TREE [51] using the LG+C60+F+G4 profile mixture model, and ultrafast bootstrap approximation [52] with 1000 replicates for estimation of branch support.

Approximately unbiased (AU) tree topology tests were performed by IQ-TREE, with the site-wise likelihood estimates done under the LG+C60+F+G4 model. Tree topologies for all datasets were constructed in MEGA [54]. Topology tests were done in conjunction with the site or partition removal analyses: alignments were subjected to stepwise elimination of either fastest-evolving sites or most compositionally heterogeneous partitions in batches amounting to approximately 10% of the full alignment's length, and topology tests were performed for each of such datasets. Site rates in the alignments were estimated using IQ-TREE, and the compositional heterogeneity of alignment partitions was evaluated using the relative composition frequency variability (RCFV) metric [55] by Ba-CoCa [56].

Mutational saturation in the concatenated datasets was estimated using the relation of patristic distances inferred from the corresponding ML tree and p-distance values calculated from the alignment. Patristic distances were extracted from IQ-TREE trees using the program PATRISTIC [65], and p-distance values were calculated using MEGA.

Transcriptome annotation and protein domain searches

Annotation of the contamination-filtered *T. unikontum* transcriptome was performed with the KEGG [66] orthology assignments, generated by the KEGG Automatic Annotation Server [67] using bi-directional best hit method and the default BLAST bit score

threshold of 60. Protein domain family searches for predicted peptides were done using HMMER [38] and the PfamScan tool (<ftp://ftp.ebi.ac.uk/pub/databases/Pfam/Tools/PfamScan.tar.gz>) with the Pfam 32.0 database [40] and also by InterProScan [57] with the InterPro v74.0 database [58]. Default family-specific gathering thresholds were used for all domain searches. Protein domain architectures were analyzed using CDD [59], Pfam, InterPro, and SMART [60] resources. Transmembrane regions in proteins were predicted with TMHMM2.0 [73] and signal peptides with SignalP-5.0 [74].

For comparative analyses we collected genomic or transcriptomic data of 23 holozoan species, 6 holomycots, and 3 species outside of the Opisthokonta. The following resources were used to collect the data: NCBI Genome, TSA, and SRA databases (<https://www.ncbi.nlm.nih.gov/>), JGI Genome Portal (<https://genome.jgi.doe.gov/portal/>), NHGRI Research Projects (<https://research.nhgri.nih.gov/>), Multicellgenome lab (<http://multicellgenome.com/>), datadryad.org (<https://dx.doi.org/10.5061/dryad.26bv4>), and figshare.com (<https://dx.doi.org/10.6084/m9.figshare.5686984>). Protein orthology assignments and domain searches were performed with the genomic or transcriptomic data for each organism using the KEGG orthologies, Pfam, and InterPro databases, similarly to searches with the transcriptome of *T. unikontum*. The transcriptomic data were filtered prior to searches, in order to remove bacterial and potential eukaryotic contamination. Contamination filtering was done with the help of DIAMOND [75] searches against the NCBI's non-redundant database. Taxonomy assignments for DIAMOND search results were extracted using TaxonKit [76] and screened for discernible cases of contamination.

QUANTIFICATION AND STATISTICAL ANALYSIS

Branch support evaluations for phylogenies were performed in conjunction with the tree reconstruction procedures, using parameters specific to each of the employed programs. Bayesian inference with MrBayes [45] utilized four independent runs of four Metropolis-coupled Markov chains to evaluate convergence, and the chains were sampled across 20 million generations with a frequency of 1 sample per 10,000 generations; the consensus tree and branch support values were obtained following a 50% burn-in of the four runs. Bayesian inference with PhyloBayes [72] used four independent chains that were run for 10,000 cycles each and sampled with a frequency of 1 sample per 50 cycles; the consensus tree and branch support values were obtained following a 50% burn-in of the four chains. The chain convergence details were checked using the bpcomp utility of the PhyloBayes software. Branch support values for ML phylogenies obtained with IQ-TREE [51] were calculated with the ultrafast bootstrap approximation [52] for the phylogenomic datasets or non-parametric bootstrap for the glutamate receptor phylogeny. Bootstrap support calculation was done with 1000 replicates for each analysis; analyses utilizing the ultrafast bootstrap approximation were performed with the nearest neighbor interchange refinement (-bnni setting). The AU tests [77] of tree topologies were carried out with IQ-TREE using 10,000 replicates for multi-scale bootstrap, with the starting tree corresponding to the ML tree obtained for the full alignment (in the case of the data-removal datasets). The alternative tree topologies were evaluated by the respective AU test p values, with the criterion for rejection defined by the 0.05 significance level; all of the AU test p values for the tested topologies and datasets are listed in [Data S1](#).