Report

Current Biology

Novel Predators Reshape Holozoan Phylogeny and Reveal the Presence of a Two-Component Signaling System in the Ancestor of Animals

Highlights

- New taxa restructure and stabilize the phylogenomic framework of Holozoa
- A two-component signaling system is present in the ancestor of animals

Authors

Elisabeth Hehenberger, Denis V. Tikhonenkov, Martin Kolisko, Javier del Campo, Anton S. Esaulov, Alexander P. Mylnikov, Patrick J. Keeling

Correspondence

helisabe@mail.ubc.ca

In Brief

Hehenberger et al. describe three novel Holozoa and infer an updated phylogenomic framework for this group, in the course recovering a new lineage. Using comparative genomics, they show the presence of a two-component signaling system across all unicellular Holozoa and suggest the loss of this signaling system with the origin of animals.

Hehenberger et al., 2017, Current Biology *27*, 2043–2050
July 10, 2017
http://dx.doi.org/10.1016/j.cub.2017.06.006

CellPress

Current Biology
Report

Novel Predators Reshape Holozoan Phylogeny and Reveal the Presence of a Two-Component Signaling System in the Ancestor of Animals

Elisabeth Hehenberger,[1,5,*] Denis V. Tikhonenkov,[2] Martin Kolisko,[3] Javier del Campo,[1] Anton S. Esaulov,[4] Alexander P. Mylnikov,[2] and Patrick J. Keeling[1]
[1]Department of Botany, University of British Columbia, 3529–6270 University Boulevard, Vancouver, BC V6T 1Z4, Canada
[2]Laboratory of Microbiology, Institute for Biology of Inland Waters, Russian Academy of Sciences, Yaroslavl Region, Borok 152742, Russia
[3]Biology Centre, Institute of Parasitology, Czech Academy of Sciences, Branisovska 31, 370 05 Ceske Budejovice, Czech Republic
[4]Department of Microbiology, Epidemiology and Infectious Diseases, Penza State University, Lermontov Street 37, Penza 440026, Russia
[5]Lead Contact
*Correspondence: helisabe@mail.ubc.ca
http://dx.doi.org/10.1016/j.cub.2017.06.006

SUMMARY

Our understanding of the origin of animals has been transformed by characterizing their most closely related, unicellular sisters: the choanoflagellates, filastereans, and ichthyosporeans. Together with animals, these lineages make up the Holozoa [1, 2]. Many traits previously considered "animal specific" were subsequently found in other holozoans [3, 4], showing that they evolved before animals, although exactly when is currently uncertain because several key relationships remain unresolved [2, 5]. Here we report the morphology and transcriptome sequencing from three novel unicellular holozoans: *Pigoraptor vietnamica* and *Pigoraptor chileana*, which are related to filastereans, and *Syssomonas multiformis*, which forms a new lineage with *Corallochytrium* in phylogenomic analyses. All three species are predatory flagellates that feed on large eukaryotic prey, and all three also appear to exhibit complex life histories with several distinct stages, including multicellular clusters. Examination of genes associated with multicellularity in animals showed that the new filastereans contain a cell-adhesion gene repertoire similar to those of other species in this group. *Syssomonas multiformis* possessed a smaller complement overall but does encode genes absent from the earlier-branching ichthyosporeans. Analysis of the T-box transcription factor domain showed expansion of T-box transcription factors based on combination with a non-T-box domain (a receiver domain), which has not been described outside of vertebrates. This domain and other domains we identified in all unicellular holozoans are part of the two-component signaling system that has been lost in animals, suggesting the continued use of this system in the closest relatives of animals and emphasizing the importance of studying loss of function as well as gain in major evolutionary transitions.

RESULTS AND DISCUSSION

Expanding the Phylogenomic Framework for Holozoa

Three novel holozoan taxa were isolated from freshwater and/or freshwater sediments: *Syssomonas multiformis*, *Pigoraptor vietnamica*, and *Pigoraptor chileana* (see STAR Methods for formal taxonomic diagnosis). The most conspicuous characteristic of all of the three is their ability to readily feed on eukaryotic prey similar in size to themselves, an unusual if not unique feeding mechanism in unicellular holozoans (Figure 1C). Although they capture only the cytoplasmic content of eukaryotic cells, reminiscent of the phagocytotic uptake of *Schistosoma mansoni* sporocyst content by the filasterean *Capsaspora owczarzaki* in lab conditions [6], they can also engulf bacteria and small detritus, similar to the phagocytotic uptake of bacteria in choanoflagellates [7]. These taxa are also morphologically plastic: all three form flagellates (Figures 1A–1C, 1I, 1J, 1M, and 1N), but also cysts, and even multicellular clusters (Figures 1F–1H, 1K, 1L, 1O, and 1P). *Syssomonas multiformis* exists in amoeboflagellate and amoeboid stages (Figures 1D and 1E). The many forms of *S. multiformis* in particular indicate highly complex life cycles.

To place the new species into a phylogenetic framework, we built on a previously published dataset of 263 genes [8], extensively expanding the taxon sampling for unicellular relatives of animals and also adding several early-branching animals. The final curated data matrix contained 255 genes (81,495 aa positions) from 38 taxa, with 12% missing data overall (Table S1). Using this 255-gene dataset, we resolved the position of the new species and also the branching order of holozoan subgroups as a whole with moderate to high support. Overall, this tree shows that filastereans are sister group to a clade containing animals and choanoflagellates, with high-confidence posterior probabilities and non-parametric maximum-likelihood (ML) bootstrap values (Figure 2A). Similarly, significant support for the dermocystid *Sphaerothecum destruens* as sister to the Ichythophonida was found (Figure 2A). Most importantly, however, we recovered a new

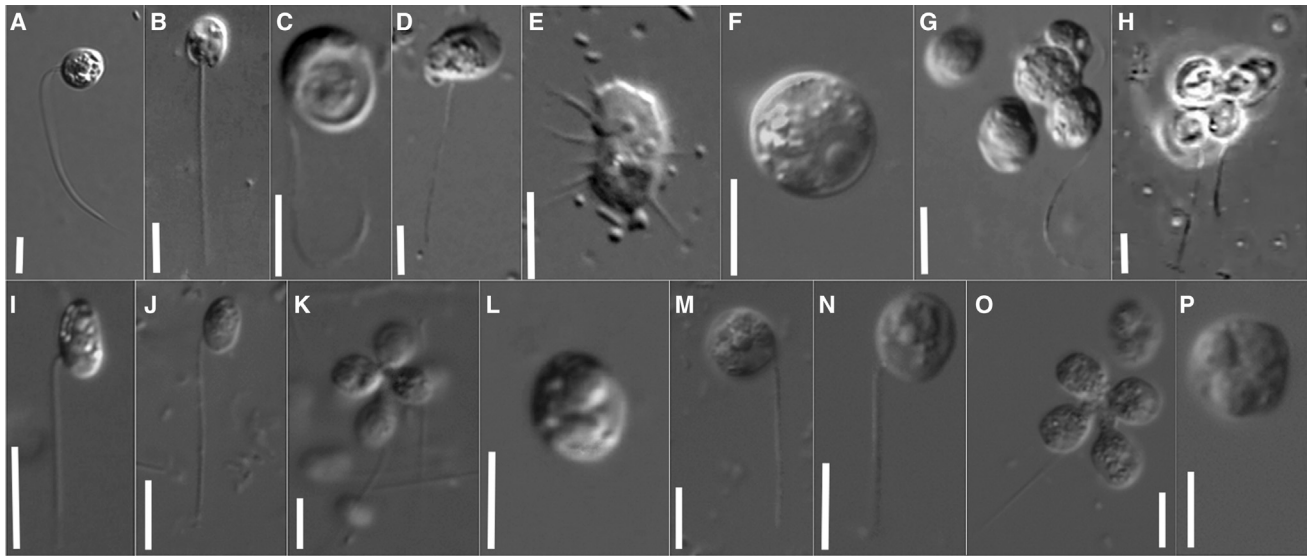


Figure 1. Morphology of New Unicellular Holozoans Showing Multiple Life-Cycle Stages

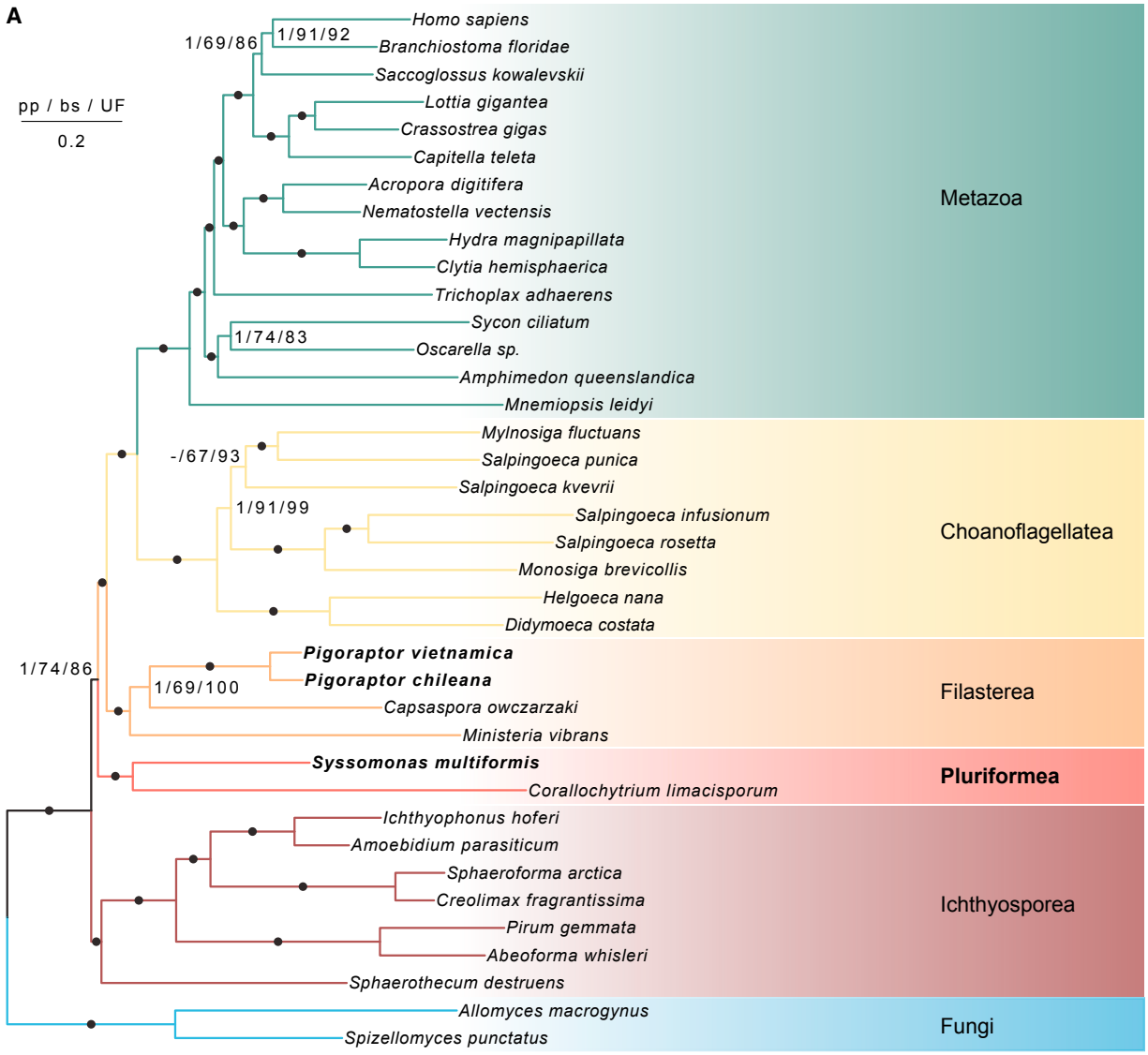
(A–H) Light micrographs of *Syssomonas multiformis* stages: flagellated cell with laterally emerging flagellum (A), flagellated cell with posteriorly orientated flagellum (B), flagellated cell with large food vacuole (C), amoeboflagellate (D), amoeboid stage (E), cyst (F), and two different cell clusters (G and H).
(I–L) Light micrographs of *Pigoraptor vietnamica* stages: two different flagellated cells (I and J), cell cluster (K), and cyst (L).
(M–P) Light micrographs of *Pigoraptor chileana* stages: two different flagellated cells (M and N), cell cluster (O), and cyst (P).
Scale bars represent 10 μm.

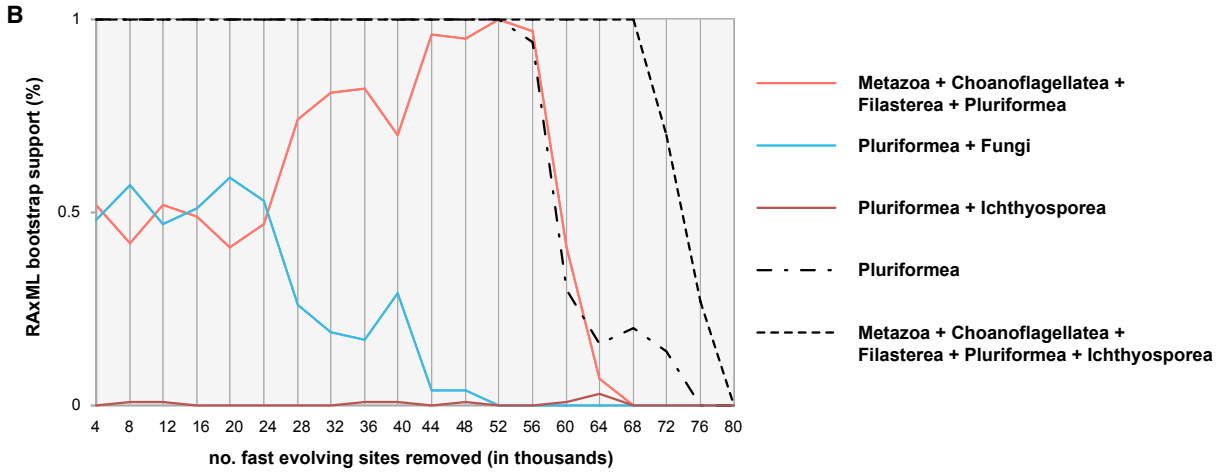
holozoan group composed of *S. multiformis*, which is a freshwater-dwelling predator with a preference for other eukaryotes, and *Corallochytrium limacisporum*, which is an enigmatic osmotrophic marine organism described in association with corals [9, 10] (Figure 2A). We provisionally name this new group "Pluriformea," indicating the variety of forms found in these species. The relationship of *Corallochytrium* to other holozoans has been controversial, and recently it was proposed to group with ichthyosporeans in the "Teretosporea," as the earliest-branching holozoans [5], which is not supported by our analyses. The node placing ichthyosporeans deepest among holozoans is only moderately supported by ML bootstrap values (74% ML bootstrap support), but it is recovered in both ML and Bayesian inference analyses (Figures 2A, S1A, and S1B). Additionally, stepwise removal of fast-evolving sites, which are the most common cause of long branch attraction (LBA), steadily increases the ML bootstrap support for the node in question (Figure 2B), indicating that the recovered topology is not an artifact resulting from LBA. Indeed, fast-evolving site removal suggests that either the Pluriformea or one of the taxa within that group in particular might be artifactually drawn toward the fungi (Figure 2B), which might in turn lead to lower bootstrap support for the basal position of ichthyosporeans. Approximately unbiased (AU) tests do not reject the Teretosporea (Table S2); however, no support for this topology is found in any of the fast-evolving sites removal datasets (Figure 2B), and a phylogenomic analysis based on the same 255-gene dataset but excluding the novel taxa did not recover this group either (see "Phylogenomic reconstruction without novel taxa" in the Dryad data depository [http://datadryad.org] at http://dx.doi.org/10.5061/dryad.26bv4). Finally, to investigate a possible effect of the outgroup on the tree topology, we expanded our outgroup taxon selection to 21 fungal and nucleariid taxa and observed the same topology as in Figure 2A that was reconstructed with a much more reduced outgroup (Figure S1C). While differing in habitat and feeding, *S. multiformis* and *C. limacisporum* do share a number of morphological features: both form amoebae at some point in their life cycle and can also exist as flagellated cells or, in case of *C. limacisporum*, at least possess the genes to do so [5]. The seemingly complex life cycle of *S. multiformis* needs to be fully investigated, but preliminary evidence suggests that it shares with *C. limacisporum* the ability to form a syncytium before dividing into progeny.

Pigoraptor vietnamica and *Pigoraptor chileana* cluster together and with filastereans, which were so far represented by only two species with distinct lifestyles: the well-investigated endosymbiont of a freshwater snail, *Capsaspora owczarzaki*, and the free-living marine heterotroph *Ministeria vibrans* [11]. However, using the small subunit (SSU) rRNA genes from the new species enabled us to determine that an abundant environmental clade, Marine Opisthokonts 1 [12, 13], also branches with filastereans (Figure S2). The new filastereans species are clearly flagellated (Figures 1I, 1J, 1M, and 1N), confirming the previously inferred loss of flagella in *C. owczarzaki* [5]. Like *C. owczarzaki*, *P. vietnamica* and *P. chileana* can encyst and also form clusters of multiple cells. All three new species feed on eukaryotic prey, suggesting that this feeding mechanism might also be widespread within the unicellular Holozoa and that the ancestor was able to take up prey distinctly larger than bacteria.

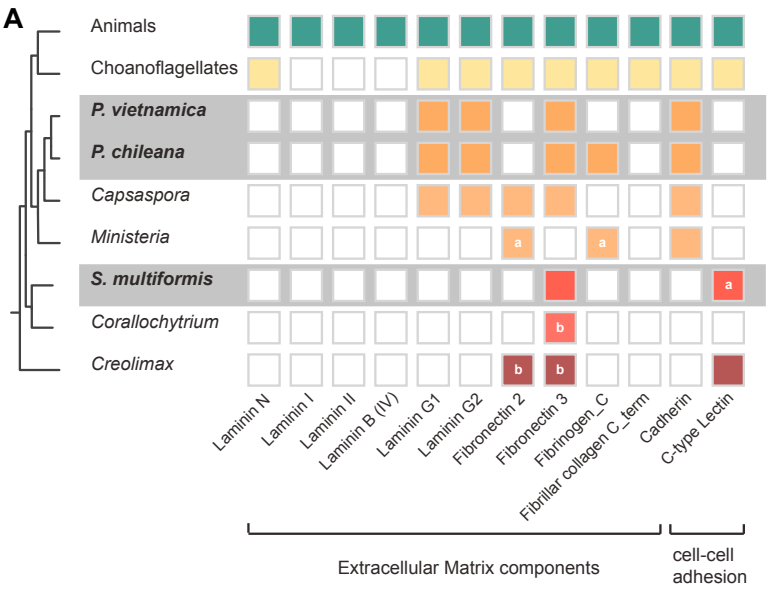
Evolution of Cell-Cell Adhesion and Cell-ECM Adhesion Components across Holozoa

Stabilization and expansion of the phylogenomic framework of Holozoa allows us to further clarify the evolutionary histories of





(legend on next page)



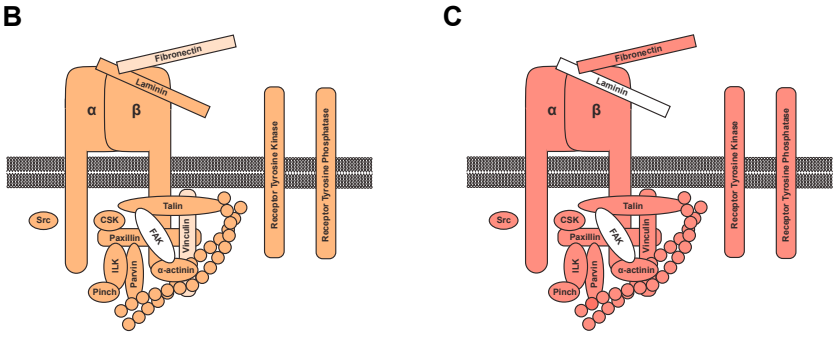


Figure 3. ECM, Cell-Cell Adhesion, and Integrin Adhesome Components in Unicellular Holozoa

(A) Repertoire of animal ECM and cell-cell adhesion domains in the four unicellular holozoan lineages; color indicates presence, and white indicates absence. The letter "a" indicates short sequence fragments containing the domain in question, but not assigned to a homolog with confidence; the letter "b" indicates no signal peptide. Different lineages are indicated by distinct colors (same as in Figure 2A). The transcriptomes of *Pigoraptor vietnamica*, *Pigoraptor chileana*, *Syssomonas multiformis*, *Ministeria vibrans*, and *Corallochytrium limacisporum* were analyzed in this study; the results indicated for animals, choanoflagellates, *Capsaspora owczarzaki*, and *Creolimax fragrantissima* are taken from [14]. See the domain architectures of 5′-complete transcripts of *P. vietnamica*, *P. chileana*, and *S. multiformis* in Figure S3A.

(B and C) Integrin adhesome components of *P. vietnamica* and *P. chileana* (B) and of *S. multiformis* (C); color indicates presence (tinted indicates presence only in *P. chileana*), and white indicates absence. Protein-protein interactions are based on a consensus adhesome from [15], displaying only interactions with medium or high level of experimental evidence. See also a presence/absence-matrix of the integrin adhesome components of the same unicellular holozoans analyzed in (A) in Figure S3B.

the molecular components required for multicellularity. To that end, we performed homology searches and phylogenetic analyses of genes involved in extracellular matrix (ECM) adhesion and cell-cell adhesion. Both new filastereans contain a nearly complete integrin adhesome and the associated signaling and cell-adhesion proteins, such as the IPP complex, scaffolding proteins, receptor tyrosine kinases, and at least one receptor tyrosine phosphatase (Figures 3A, 3B, S3A, and S3B).

The presence of Fibronectin-3 domains in several receptor tyrosine kinases and one *P. chileana* receptor tyrosine phosphatase (see "Receptor protein tyrosine kinase phylogeny" and "Receptor protein tyrosine phosphatase phylogeny" in the Dryad data depository [http://datadryad.org] at http://dx.doi.org/10.5061/dryad.26bv4), known to interact with integrins [16], suggests a possible association between the integrin adhesome and tyrosine kinase/phosphatase signaling in *Pigoraptor*. A similar association was suggested for the integrin adhesome and tyrosine kinase signaling in their closest relative, *C. owczarzaki*, as it also contains receptor tyrosine kinases possessing Fibronectin-3 domains [17]. In *C. owczarzaki*, it was further shown that the components of the integrin adhesome and tyrosine signaling are strongly upregulated in the aggregative stage, strongly suggesting a role in aggregate formation for these components in *Capsaspora* [17]. In *P. chileana*, we also found a secreted protein (containing a signal peptide) consisting of Fibronectin-3 domains only, and both *Pigoraptor* species express several secreted Laminin G domain-containing proteins (Figure S3A), all putative components of the ECM [16, 18]. The integrin-based adhesion-related complement in *S. multiformis* reflects its intermediary position in the tree in that it is reduced compared to the filastereans but includes components not found in the more basal ichthyosporean, *Creolimax fragrantissima*. For example, no Laminin G (or other Laminin) domains were found in our data, whereas Laminin G domains are known in filastereans and choanoflagellates (Figure 3A). However, the presence of secreted

Figure 2. Phylogeny of Holozoa and Fast-Evolving Site Removal Analysis

(A) Phylogenomic tree based on 255 concatenated proteins inferred by IQ-TREE under the LG+C40+F+G4 model (see Tables S1 and S2 and STAR Methods). Novel species and lineages are shown in bold. Node supports are Bayesian posterior probabilities (pp) from a PhyloBayes consensus tree, nonparametric ML bootstrap (bp) values obtained from 1,000 ML replicates using the LG+F+Γ model implemented in RAxML, and ultrafast bootstrap (UF) values from IQ-TREE. Black dots on branches correspond to 1.0 Bayesian pp and >95% bp and UF. See the Bayesian CAT+GTR+Γ4 topologies in Figures S1A and S1B, a phylogeny with expanded outgroup taxon sampling in Figure S1C, and an rRNA SSU phylogeny constrained by the IQ-TREE in Figure S2.

(B) Fast-evolving site removal sorted alignment positions based on their rates of evolution from fastest to slowest evolving. Sites were then sequentially removed, 4,000 sites at a time, generating alternative datasets at each step. Rapid bootstrap values for each bipartition of interest are plotted.

Fibronectin-3 domain-containing transcripts in *S. multiformis* (Figures 3A and S3A) indicates an interaction of the integrin adhesome with an animal-like ECM, standing in contrast to *Creolimax*, which has no secreted ECM-related domains [14]. Altogether, these data suggest an emergence of an animal-like ECM structure in the ancestor of Pluriformea, filastereans, choanoflagellates, and metazoans after its divergence from ichthyosporeans.

Focal adhesion kinase (FAK) is a cytoplasmic (non-receptor) protein tyrosine kinase that functions in integrin signal transduction together with SRC, another cytoplasmic protein tyrosine kinase, and plays a central role in animal development [19, 20]. Outside of the Metazoa, a bona fide FAK was to date only described in the filastereans *C. owczarzaki* and *M. vibrans* [21, 22]. We found a *S. multiformis* protein clustering with *C. owczarzaki* and *M. vibrans* in a phylogenetic analysis, but no *Pigoraptor* representatives were recovered and choanoflagellate candidates present in that tree did not contain a focal adhesion targeting region indicative for FAKs (see "CSK and FAK protein tyrosine kinase phylogenies" in the Dryad data depository [http://datadryad.org] at http://dx.doi.org/10.5061/dryad.26bv4). The *S. multiformis* sequence contained a transmembrane domain, suggesting that it encodes for a receptor protein tyrosine kinase, and we could not identify a focal adhesion targeting region either. As RNA-sequencing data do not represent a complete transcriptome, it is difficult to clearly state the absence of any given protein. But the fact that we could not find even a fragment, despite the presence of a large number of tyrosine kinases in general (see protein tyrosine kinase trees in the Dryad data depository [http://datadryad.org] at http://dx.doi.org/10.5061/dryad.26bv4), suggests that FAK is at most present at a low abundance in these organisms and perhaps is even absent (although this cannot be concluded from transcriptome data).

Besides integrin adhesome-associated proteins and signaling pathways, we also searched for proteins particularly involved in cell-cell adhesion. Both *Pigoraptor* lineages have cadherin domain-containing transcripts similar to the cadherin protein found in *Capsaspora* and, also like *Capsaspora*, no evidence for C-type lectin domain-containing proteins. Conversely, in *S. multiformis*, no cadherin domains were found, but truncated transcripts containing one or more C-type lectin domains were identified. Both cadherin-like and C-type lectin domains have been found outside the Holozoa [23–25], suggesting that either cadherins or C-type lectins may have been specifically lost in the ichthyosporeans, *S. multiformis*, and filastereans. Whether this distribution is due to lineage-specific roles of those proteins or whether functional redundancy has led to loss of one type of adhesion domain is currently unclear.

Novel Brachyury-Type T-Box Transcription Factors Integrated with a Holozoan Two-Component System

The T-box proteins are an essential family of transcription factors in metazoan development [26, 27] and are thus of key interest in efforts to understand the origin of animals. We identified several T-box domain-containing transcripts in the *Pigoraptor* lineages and one in *S. multiformis*, adding to the complement of premetazoan T-box genes already described in filastereans, ichthyosporeans, and fungi [28, 29]. Phylogenetic analysis placed all new T-box domains within the Brachyury family, the most ancient T-box class (Figure S4). We found a transcript containing two T-box domains, as already described in *C. owczarzaki*, in both new filastereans. However, unlike previous analyses [28], the *C. owczarzaki* double T-box homolog does not cluster with the other premetaozan T-box class Tbx7 but instead clusters together with the novel double T-box genes within the Brachyury clade (Figure S4).

Most interestingly, however, we identified transcripts with a T-box domain fused to a Response regulator receiver domain. T-box transcription factors containing other domains than the T-box domain have so far only been described in vertebrates, and in that case the domain in question is another DNA-binding domain [30]. The Response regulator receiver domain functions as receiver of phosphorylation signals in two-component signal transduction systems [31]. In its simplest form, this signaling pathway consists of a histidine kinase that regulates the activity of a response regulator and its attached output domain by phosphotransfer. Although these two-component system regulators are predominantly found in prokaryotes, the system has also been co-opted by eukaryotes [32, 33], concomitant with a prevalent use of multi-step phosphorelays performing more than one phosphotransfer during signal transmission [34]. Two-component signaling elements have been identified in a variety of eukaryotes, such as plants, green algae, diatoms, amoebozoans, and fungi, but not in metazoans [34, 35]. A search for these elements (hybrid histidine kinases, phosphotransfer proteins, and response regulators) in our data, as well as in published genomes representing all known unicellular holozoan lineages, revealed the presence of proteins involved in two-component signaling across all investigated groups (Figure 4). While the genomes of choanoflagellates and ichthyosporeans encode a maximum of one protein per component, filastereans and *S. multiformis* contain a diverse array of hybrid histidine kinases associated with additional sensory domains (Table S3). The response regulator is either associated with a T-box (*Pigoraptor* and *S. multiformis*) or an HLH domain (choanoflagellates, *C. owczarzaki,* and *S. multiformis*), another motif characteristic for transcription factors, suggesting that in those lineages, two-component signaling is integrated with transcriptional regulation. To exclude the possibility of a loss of this system after the origin of animals, we also searched for two-component system associated domains in non-bilaterian metazoan lineages represented in our tree and investigated all sequences annotated as metazoan in the pfam database containing such domains. All putative candidates were identified as bacterial contaminations, strongly indicating that the two-component system was lost around the same times as the emergence of animals. The loss of a major signaling system concurrently with the origin of multicellularity represents an interesting new dimension to this important evolutionary transition.

Conclusions

An expanded and stabilized phylogenetic framework for the Holozoa allows for a clearer interpretation of the evolutionary history of specific traits leading up to the origin of multicellularity in Metazoa. Phylogenomic analysis revealed a new holozoan clade,

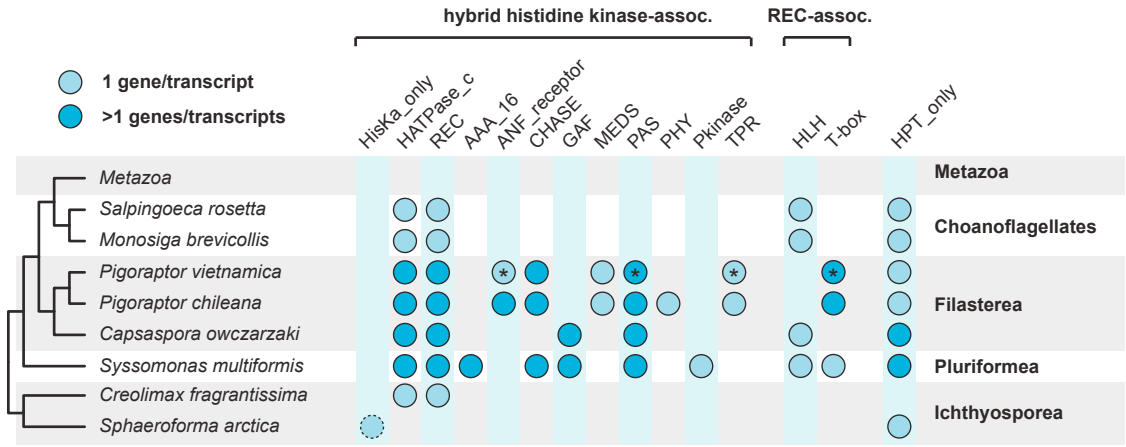


Figure 4. Domain Composition and Distribution of Holozoan Histidine Kinase, Receiver, and Phosphotransfer Domain-Containing Proteins

A schematic tree of Holozoa is shown on the left. To the right, circles indicates the presence of a domain in that species; a light-blue circle means only one gene/transcript could be identified, a dark-blue circle means that more than one gene/transcript was identified, a dotted circle means partial domain only and an asterisk indicates partial transcripts missing one or more domain(s) in question but clustering with a corresponding *P. chileana* homolog in phylogenetic reconstructions. The names above indicate the domains present in either hybrid histidine kinase proteins or in proteins containing the response regulator receiver domain (REC). All sequences containing the histidine kinase acceptor domain (HisKA), REC domain, and histidine-containing phosphotransfer domain (HPT) were obtained using hmmscan against the PfamA database. Such identified sequences were analyzed for their phylogenetic association and then submitted to Pfam websearch and the SMART database to verify the presence of those domains as well as to identify other domains present. All datasets used for analysis, but the novel species (*P. vietnamica*, *P. chileana*, and *S. multiformis*) were annotated proteins based on genome assemblies. See also Figure S4 for a T-box domain phylogeny containing transcripts with REC domains and Table S3 for accession numbers and peptide identifiers.

Pluriformea, including *S. multiformis* and *C. limacisporum*. The distribution of ECM adhesion and cell-cell adhesion molecular components suggests the early emergence of an animal-like ECM in the last common ancestor of pluriformeans, filastereans, choanoflagellates, and metazoans and indicates that *C. owczarzaki* and *M. vibrans* are the only unicellular holozoan species to have retained a FAK tyrosine kinase, central to animal development. Most significantly, we identify T-box transcription factors with unusual domain architecture in holozoans, showing that transcriptional regulation is controlled by a two-component signaling system. Two-component signaling was identified not only in the novel species, but across all lineages of unicellular Holozoa, confirming that the loss of two-component signaling took place around the time of the origin of animals.

STAR★METHODS

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

- KEY RESOURCES TABLE
- CONTACT FOR REAGENT AND RESOURCE SHARING
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Formal taxonomic diagnosis
- METHOD DETAILS
  - Culturing of novel species and RNA-seq
  - Assembly of datasets used in phylogenomic analysis
  - Single-gene tree preparation and concatenation
  - Phylogenomic analyses, fast-evolving site removal analysis and AU test
  - 18S rRNA phylogenetic analysis
  - T-box domain phylogeny
  - Gene and domain analyses
- DATA AND SOFTWARE AVAILABILITY

SUPPLEMENTAL INFORMATION

Supplemental Information includes four figures and three tables and can be found with this article online at http://dx.doi.org/10.1016/j.cub.2017.06.006.

AUTHOR CONTRIBUTIONS

Conceptualization, E.H., D.V.T., and P.J.K.; Resources, P.J.K., D.V.T., and A.P.M.; Investigation, E.H., D.V.T., M.K., J.d.C., and A.S.E.; Visualization, E.H., D.V.T., and J.d.C.; Writing – Original Draft, E.H. and P.J.K.; Writing – Review & Editing, E.H., P.J.K., J.d.C., D.V.T., and M.K.

ACKNOWLEDGMENTS

This work was supported by a grant from the Canadian Institutes for Health Research (MOP-42517) to P.J.K. E.H., M.K., and J.d.C. were supported by a grant to the Centre for Microbial Diversity and Evolution from the Tula Foundation. J.d.C. was supported by a Marie Curie International Outgoing Fellowship grant (FP7-PEOPLE-2012-IOF-331450 CAARL). D.V.T. was supported by a grant from the Russian Science Foundation (1414-00515). This work was also supported by grants from the Russian Foundation for Basic Research (15-34-20065 and 17-04-00899) and a grant from the President of the Russian Federation (MK-7436.2015.4). We thank Iñaki Ruiz-Trillo (Institut de Biologia Evolutiva) for sharing data and in particular Guifré Torruella (now Université Paris-Sud) and Xavier Grau-Bové (Institut de Biologia Evolutiva) for providing data of several unicellular holozaons. We thank Daniel Richter (Station Biologique de Roscoff) for providing choanoflagellate data. We thank Fabien Burki (Uppsala University) for providing phylogenomic gene sets, scripts, and methodological help. We thank Dr. Hoan Q. Tran, Tran Duc Dien, and Kristina I. Prokina for help with sample collection in Vietnam and Chile, as well as the staff of the Vietnam-Russian Tropical Centre, Coastal Branch (Nha Trang, Vietnam), especially Nguyen Thị Hai Thanh and Tran Thanh Quang, for their assistance

with trip management and sampling. We thank Compute/Calcul Canada for computing resources and assistance.

Received: March 14, 2017
Revised: April 24, 2017
Accepted: June 1, 2017
Published: June 22, 2017

REFERENCES

1. Lang, B.F., O'Kelly, C., Nerad, T., Gray, M.W., and Burger, G. (2002). The closest unicellular relatives of animals. Curr. Biol. *12*, 1773–1778.
2. 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.
3. 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.
4. Suga, H., Chen, Z., de Mendoza, A., Sebé-Pedrós, A., Brown, M.W., Kramer, E., Carr, M., Kerner, P., Vervoort, M., Sánchez-Pons, N., et al. (2013). The *Capsaspora* genome reveals a complex unicellular prehistory of animals. Nat. Commun. *4*, 2325.
5. Torruella, G., de Mendoza, A., Grau-Bové, X., Antó, M., Chaplin, M.A., del Campo, J., Eme, L., Pérez-Cordón, G., Whipps, 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.
6. Owczarzak, A., Stibbs, H.H., and Bayne, C.J. (1980). The destruction of *Schistosoma mansoni* mother sporocysts in vitro by amoebae isolated from *Biomphalaria glabrata*: an ultrastructural study. J. Invertebr. Pathol. *35*, 26–33.
7. Dayel, M.J., and King, N. (2014). Prey capture and phagocytosis in the choanoflagellate *Salpingoeca rosetta*. PLoS ONE *9*, e95577.
8. 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.
9. Raghukumar, S. (1987). Occurrence of the Thraustochytrid, *Corallochytrium limacisporum* gen. et sp. nov. in the coral reef lagoons of the Lakshadweep Islands in the Arabian Sea. Bot. Mar. *30*, 83–89.
10. Raghukumar, S., and Balasubramanian, R. (1991). Occurrence of Thraustochytrid fungi in corals and coral mucus. Indian J. Mar. Sci. *20*, 176–181.
11. Tong, S.M. (1997). Heterotrophic flagellates and other protists from Southampton Water, UK. Ophelia *42*, 71–131.
12. del Campo, J., and Ruiz-Trillo, I. (2013). Environmental survey meta-analysis reveals hidden diversity among unicellular opisthokonts. Mol. Biol. Evol. *30*, 802–805.
13. 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.
14. de Mendoza, A., Suga, H., Permanyer, J., Irimia, M., and Ruiz-Trillo, I. (2015). Complex transcriptional regulation and independent evolution of fungal-like traits in a relative of animals. eLife *4*, e08904.
15. Horton, E.R., Byron, A., Askari, J.A., Ng, D.H.J., Millon-Frémillon, A., Robertson, J., Koper, E.J., Paul, N.R., Warwood, S., Knight, D., et al. (2015). Definition of a consensus integrin adhesome and its dynamics during adhesion complex assembly and disassembly. Nat. Cell Biol. *17*, 1577–1587.
16. Pankov, R., and Yamada, K.M. (2002). Fibronectin at a glance. J. Cell Sci. *115*, 3861–3863.
17. 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.
18. Bosman, F.T., and Stamenkovic, I. (2003). Functional structure and composition of the extracellular matrix. J. Pathol. *200*, 423–428.
19. Mitra, S.K., and Schlaepfer, D.D. (2006). Integrin-regulated FAK-Src signaling in normal and cancer cells. Curr. Opin. Cell Biol. *18*, 516–523.
20. Schaller, M.D. (2010). Cellular functions of FAK kinases: insight into molecular mechanisms and novel functions. J. Cell Sci. *123*, 1007–1013.
21. Sebé-Pedrós, A., Roger, A.J., Lang, F.B., 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.
22. Suga, H., Dacre, M., de Mendoza, A., Shalchian-Tabrizi, K., Manning, G., and Ruiz-Trillo, I. (2012). Genomic survey of premetazoans shows deep conservation of cytoplasmic tyrosine kinases and multiple radiations of receptor tyrosine kinases. Sci. Signal. *5*, ra35.
23. Cao, L., Yan, X., Borysenko, C.W., Blair, H.C., Wu, C., and Yu, L. (2005). CHDL: a cadherin-like domain in Proteobacteria and Cyanobacteria. FEMS Microbiol. Lett. *251*, 203–209.
24. Zelensky, A.N., and Gready, J.E. (2005). The C-type lectin-like domain superfamily. FEBS J. *272*, 6179–6217.
25. Fletcher, K.I.G., van West, P., and Gachon, C.M.M. (2016). Nonagonal cadherins: A new protein family found within the Stramenopiles. Gene *593*, 64–75.
26. Papaioannou, V.E., and Silver, L.M. (1998). The T-box gene family. BioEssays *20*, 9–19.
27. Papaioannou, V.E. (2001). T-box genes in development: from hydra to humans. Int. Rev. Cytol. *207*, 1–70.
28. Sebé-Pedrós, A., Ariza-Cosano, A., Weirauch, M.T., Leininger, S., Yang, A., Torruella, G., Adamski, M., Adamska, M., Hughes, T.R., Gómez-Skarmeta, J.L., and Ruiz-Trillo, I. (2013). Early evolution of the T-box transcription factor family. Proc. Natl. Acad. Sci. USA *110*, 16050–16055.
29. Sebé-Pedrós, A., and Ruiz-Trillo, I. (2017). Evolution and classification of the T-Box transcription factor family. Curr. Top. Dev. Biol. *122*, 1–26.
30. Hurlin, P.J., Steingrìmsson, E., Copeland, N.G., Jenkins, N.A., and Eisenman, R.N. (1999). Mga, a dual-specificity transcription factor that interacts with Max and contains a T-domain DNA-binding motif. EMBO J. *18*, 7019–7028.
31. Stock, A.M., Robinson, V.L., and Goudreau, P.N. (2000). Two-component signal transduction. Annu. Rev. Biochem. *69*, 183–215.
32. Ota, I.M., and Varshavsky, A. (1993). A yeast protein similar to bacterial two-component regulators. Science *262*, 566–569.
33. Chang, C., Kwok, S.F., Bleecker, A.B., and Meyerowitz, E.M. (1993). *Arabidopsis* ethylene-response gene ETR1: similarity of product to two-component regulators. Science *262*, 539–544.
34. Schaller, G.E., Shiu, S.-H., and Armitage, J.P. (2011). Two-component systems and their co-option for eukaryotic signal transduction. Curr. Biol. *21*, R320–R330.
35. Schaap, P., Barrantes, I., Minx, P., Sasaki, N., Anderson, R.W., Bénard, M., Biggar, K.K., Buchler, N.E., Bundschuh, R., Chen, X., et al. (2015). The *Physarum polycephalum* genome reveals extensive use of prokaryotic two-component and metazoan-type tyrosine kinase signaling. Genome Biol. Evol. *8*, 109–125.
36. Medlin, L., Elwood, H.J., Stickel, S., and Sogin, M.L. (1988). The characterization of enzymatically amplified eukaryotic 16S-like rRNA-coding regions. Gene *71*, 491–499.
37. Tikhonenkov, D.V., Janouškovec, J., Mylnikov, A.P., Mikhailov, K.V., Simdyanov, T.G., Aleoshin, V.V., and Keeling, P.J. (2014). Description of *Colponema vietnamica* sp.n. and *Acavomonas peruviana* n. gen. n. sp., two new alveolate phyla (Colponemidia nom. nov. and Acavomonidia nom. nov.) and their contributions to reconstructing the ancestral state of alveolates and eukaryotes. PLoS ONE *9*, e95467.

38. 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.

39. 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.

40. 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.

41. 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.

42. 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.

43. Susko, E., Field, C., Blouin, C., and Roger, A.J. (2003). Estimation of rates-across-sites distributions in phylogenetic substitution models. Syst. Biol. *52*, 594–603.

44. Shimodaira, H., and Hasegawa, M. (2001). CONSEL: for assessing the confidence of phylogenetic tree selection. Bioinformatics *17*, 1246–1247.

45. Letunic, I., and Bork, P. (2016). Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. Nucleic Acids Res. *44* (W1), W242–W245.

46. Price, M.N., Dehal, P.S., and Arkin, A.P. (2010). FastTree 2–approximately maximum-likelihood trees for large alignments. PLoS ONE *5*, e9490.

47. Finn, R.D., Bateman, A., Clements, J., Coggill, P., Eberhardt, R.Y., Eddy, S.R., Heger, A., Hetherington, K., Holm, L., Mistry, J., et al. (2014). Pfam: the protein families database. Nucleic Acids Res. *42*, D222–D230.

48. Keeling, P.J., Burki, F., Wilcox, H.M., Allam, B., Allen, E.E., Amaral-Zettler, L.A., Armbrust, E.V., Archibald, J.M., Bharti, A.K., Bell, C.J., et al. (2014). The Marine Microbial Eukaryote Transcriptome Sequencing Project (MMETSP): illuminating the functional diversity of eukaryotic life in the oceans through transcriptome sequencing. PLoS Biol. *12*, e1001889.

49. Ruiz-Trillo, I., Burger, G., Holland, P.W.H., 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.

50. Hemmrich, G., and Bosch, T.C.G. (2008). Compagen, a comparative genomics platform for early branching metazoan animals, reveals early origins of genes regulating stem-cell differentiation. BioEssays *30*, 1010–1018.

51. Picelli, S., Faridani, O.R., Björklund, A.K., Winberg, G., Sagasser, S., and Sandberg, R. (2014). Full-length RNA-seq from single cells using Smart-seq2. Nat. Protoc. *9*, 171–181.

52. Zhang, J., Kobert, K., Flouri, T., and Stamatakis, A. (2014). PEAR: a fast and accurate Illumina Paired-End reAd mergeR. Bioinformatics *30*, 614–620.

53. Bolger, A.M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics *30*, 2114–2120.

54. Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nat. Methods *9*, 357–359.

55. Schmieder, R., and Edwards, R. (2011). Fast identification and removal of sequence contamination from genomic and metagenomic datasets. PLoS ONE *6*, e17288.

56. Bateman, A., Martin, M.J., O'Donovan, C., Magrane, M., Alpi, E., Antunes, R., Bely, B., Bingley, M., Bonilla, C., Britto, R., et al.; The UniProt Consortium (2017). UniProt: the universal protein knowledgebase. Nucleic Acids Res. *45* (D1), D158–D169.

57. Carr, M., Richter, D.J., Fozouni, P., Smith, T.J., Jeuck, A., Leadbeater, B.S.C., and Nitsche, F. (2017). A six-gene phylogeny provides new insights into choanoflagellate evolution. Mol. Phylogenet. Evol. *107*, 166–178.

58. Burki, F., Corradi, N., Sierra, R., Pawlowski, J., Meyer, G.R., Abbott, C.L., and Keeling, P.J. (2013). Phylogenomics of the intracellular parasite *Mikrocytos mackini* reveals evidence for a mitosome in rhizaria. Curr. Biol. *23*, 1541–1547.

59. Burki, F., Okamoto, N., Pombert, J.-F., and Keeling, P.J. (2012). The evolutionary history of haptophytes and cryptophytes: phylogenomic evidence for separate origins. Proc. Biol. Sci. *279*, 2246–2254.

60. Janouškovec, J., Tikhonenkov, D.V., Burki, F., Howe, A.T., Kolísko, M., Mylnikov, A.P., and Keeling, P.J. (2015). Factors mediating plastid dependency and the origins of parasitism in apicomplexans and their close relatives. Proc. Natl. Acad. Sci. USA *112*, 10200–10207.

61. Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990). Basic local alignment search tool. J. Mol. Biol. *215*, 403–410.

62. 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.

63. Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics *30*, 1312–1313.

64. Quang, S., Gascuel, O., and Lartillot, N. (2008). Empirical profile mixture models for phylogenetic reconstruction. Bioinformatics *24*, 2317–2323.

65. Lartillot, N., Lepage, T., and Blanquart, S. (2009). PhyloBayes 3: a Bayesian software package for phylogenetic reconstruction and molecular dating. Bioinformatics *25*, 2286–2288.

66. Grigoriev, I.V., Nikitin, R., Haridas, S., Kuo, A., Ohm, R., Otillar, R., Riley, R., Salamov, A., Zhao, X., Korzeniewski, F., et al. (2014). MycoCosm portal: gearing up for 1000 fungal genomes. Nucleic Acids Res. *42*, D699–D704.

67. Larsson, A. (2014). AliView: a fast and lightweight alignment viewer and editor for large datasets. Bioinformatics *30*, 3276–3278.

68. Stamatakis, A. (2006). RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics *22*, 2688–2690.

69. 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.

70. Robinson, J.T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E.S., Getz, G., and Mesirov, J.P. (2011). Integrative genomics viewer. Nat. Biotechnol. *29*, 24–26.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Biological Samples		
Colp-12 strain	Freshwater pool, Vietnam	MI-PR205
Opistho-1 strain	Freshwater Lake, Vietnam	MI-PR206
Opistho-2 strain	Freshwater temporary water body, Chile	MI-PR207
Critical Commercial Assays		
RNAqueous Kit	Ambion	Cat# AM1912
SMARTer Pico PCR cDNA Synthesis Kit	Clontech	Cat# 634928
Nextera XT DNA Library Prep Kit	Illumina	FC-131-1024
Deposited Data		
Syssomonas multiformis 18S	This study	GenBank: MF190551
Pigoraptor vietnamica 18S	This study	GenBank: MF190552
Pigoraptor chileana 18S	This study	GenBank: MF190553
Syssomonas multiformis predicted peptides (culture/sorted)	This study	Dryad data depository (http://datadryad.org): http://dx.doi.org/10.5061/dryad.26bv4
Pigoraptor vietnamica predicted peptides	This study	Dryad data depository (http://datadryad.org): http://dx.doi.org/10.5061/dryad.26bv4
Pigoraptor chileana predicted peptides	This study	Dryad data depository (http://datadryad.org): http://dx.doi.org/10.5061/dryad.26bv4
Single-gene alignments (trimmed) for phylogenomic analysis	This study	Dryad data depository (http://datadryad.org): http://dx.doi.org/10.5061/dryad.26bv4
Phylogenomic reconstruction without novel taxa	This study	Dryad data depository (http://datadryad.org): http://dx.doi.org/10.5061/dryad.26bv4
Protein tyrosine kinase/phosphatase trees	This study	Dryad data depository (http://datadryad.org): http://dx.doi.org/10.5061/dryad.26bv4
Phylogenies from figures in newick format	This study	Dryad data depository (http://datadryad.org): http://dx.doi.org/10.5061/dryad.26bv4
Oligonucleotides		
AAC CTG GTT GAT CCT GCC AGT	[36]	EukA
TGA TCC TTC TGC AGG TTC ACC TAC	[36]	EukB
GCGCTACCTGGTTGATCCTGCC	[37]	PF-1
TGATCCTTCTGCAGGTTCACCTAC	[37]	FAD4
Software and Algorithms		
pear-0.9.6	http://www.exelixis-lab.org/software.htm	RRID: SCR_003776
Trimmomatic v0.36	http://www.usadellab.org/cms/index.php?page=trimmomatic	RRID: SCR_011848
Bowtie 2.1.0	http://bowtie-bio.sourceforge.net/bowtie2/index.shtml	RRID: SCR_005476
DeconSeq v0.4.3	http://deconseq.sourceforge.net	RRID: SCR_007006
Trinity v 2.0.6	[38, 39]	https://github.com/trinityrnaseq/trinityrnaseq/wiki
BLAST 2.2.30+	http://blast.ncbi.nlm.nih.gov/Blast.cgi	RRID: SCR_004870
MAFFT v7.212	http://mafft.cbrc.jp/alignment/server/	RRID: SCR_011811
trimAl 1.2rev59	[40]	http://trimal.cgenomics.org/
RAxML v 8.0.2	https://github.com/stamatak/standard-RAxML	RRID: SCR_006086
RAxML v7.2.6	https://github.com/stamatak/standard-RAxML	RRID: SCR_006086

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
SCaFoS v1.2.5	[41]	http://megasun.bch.umontreal.ca/Software/scafos/scafos.html
iqtree-omp v1.4.3	[42]	http://www.iqtree.org/
dist_est v1.0	[43]	http://www.mathstat.dal.ca/~tsusko/software.html
CONSEL 0.20	[44]	http://www.sigmath.es.osaka-u.ac.jp/shimo-lab/prog/consel/
PhyloBayes MPI v1.4e	http://megasun.bch.umontreal.ca/People/lartillot/www/	RRID: SCR_006402
AliView v1.17.1	https://github.com/AliView	RRID: SCR_002780
iTOL v3	[45]	http://itol.embl.de/
HMMER3.1	http://hmmer.janelia.org	RRID: SCR_005305
InterProScan 5	http://www.ebi.ac.uk/Tools/pfa/iprscan/	RRID: SCR_005829
FastTree v2.1.7	[46]	http://www.microbesonline.org/fasttree/
IGV_2.3.79	http://www.broadinstitute.org/igv/	RRID: SCR_011793
Other		
Swiss-Prot Database	http://www.uniprot.org/help/uniprotkb	RRID: SCR_004426
Pfam-A database v29.0	[47]	ftp://ftp.ebi.ac.uk/pub/databases/Pfam/releases/
Marine Microbial Eukaryote Transcriptome Sequencing Project Database	[48]	https://www.imicrobe.us/project/view/104
Broad Institute Microbial Eukaryote Genomes	[49]	https://www.broadinstitute.org/scientific-community/science/programs/genome-sequencing-and-analysis/update-our-microbial-eukaryotes
Compagen Database	[50]	http://compagen.org/terms.html
NCBI Sequence Read Archive	https://www.ncbi.nlm.nih.gov/sra	RRID: SCR_004891
Ensembl Genomes Database	http://ensemblgenomes.org/	RRID: SCR_006773
NCBI BioProject Database	https://www.ncbi.nlm.nih.gov/bioproject	RRID: SCR_004801
JGI Genome Portal	http://genome.jgi.doe.gov/	RRID: SCR_002383
Broad Institute Fungal Genome Initiative	https://www.broadinstitute.org/fungal-genome-initiative	RRID: SCR_003169

CONTACT FOR REAGENT AND RESOURCE SHARING

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Elisabeth Hehenberger (helisabe@mail.ubc.ca).

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Formal taxonomic diagnosis

Syssomonas *n. gen. Tikhonenkov, Hehenberger, Mylnikov et Keeling 2017*

Assignment. Eukaryota; Opisthokonta; Holozoa.

Diagnosis. Predominantly unicellular, uniflagellar protist. Motile swimming cells are roundish. The smooth flagellum emerges from the middle-lateral point of the cell, ended by short acroneme and directs backward. The cells are naked. Protist can form clusters of multiple cells. Predator of heterotrophic chrysomonads and bodonids.

Etymology. The genus name means "aggregative monad," Greek.

Zoobank Registration. urn:lsid:zoobank.org:act:9B4DACD4-3744-43F5-A954-7E4E085C861E.

Type species. Syssomonas multiformis.

Syssomonas multiformis *n. sp. Tikhonenkov, Hehenberger, Mylnikov et Keeling 2017*

Diagnosis. Diagnosis as for genus *Syssomonas.* The life cycle consists of uniflagellar roundish motile swimming cells, amoeboflagellates, amoeboid non-flagellar cells, spherical cysts. The flagellated cells are 7-14 μm in diameter.

Type material. A block of chemically fixed resin-embedded cells of the type strain, Colp-12, is deposited in Marine Invertebrate Collection, Beaty Biodiversity Museum, University of British Columbia as MI-PR205. This constitutes the name-bearing type of the new species (a hapantotype).

Figure 1A illustrates a live cell of strain Colp-12.

Type locality. Freshwater pool, Tà Lài, Cát Tiên National Park, Dong Nai Province, S.R. Vietnam.

Etymology. From *multus* (''many'') + *-fōrmis* (''having the form of''), Latin.

Gene sequence. The 18S rRNA gene sequence has the accession number GenBank: MF190551.

Zoobank Registration. urn:lsid:zoobank.org:act:E81DD41E-449C-446E-A0CC-877ADD953E32.

Pigoraptor *n. gen. Tikhonenkov, Hehenberger, Mylnikov et Keeling 2017*

Assignment. Eukaryota; Opisthokonta; Holozoa; Filasterea.

Diagnosis. Predominantly unicellular, uniflagellar protist. Motile swimming cells are elongated-oval or roundish. The smooth flagellum emerges from the middle-lateral point of the cell, ended by long acroneme and directs backward. The cells are naked, sometimes produce short, thin pseudopodia. Protist can form clusters of multiple cells. Predator of heterotrophic bodonids, especially slow moving and immobile cells or cysts.

Etymology. The genus name means ''lazy predator,'' Latin.

Zoobank Registration. urn:lsid:zoobank.org:act:DA209900-ECC7-449D-8ED2-8A8B38B41505.

Type species. Pigoraptor vietnamica.

Pigoraptor vietnamica *n. sp. Tikhonenkov, Hehenberger, Mylnikov et Keeling 2017*

Diagnosis. Diagnosis as for genus *Pigoraptor.* The life cycle consists of uniflagellar motile swimming cells and spherical cysts. The flagellated elongated-oval cells are 5-12 μm in diameter.

Type material. A block of chemically fixed resin-embedded cells of the type strain, Opistho-1, is deposited in Marine Invertebrate Collection, Beaty Biodiversity Museum, University of British Columbia as MI-PR206. This constitutes the name-bearing type of the new species (a hapantotype).

Figure 1I illustrates a live cell of strain Opistho-1.

Type locality. Freshwater Lake Dak Minh, silty sand on the littoral, Dak Lak Province, S.R. Vietnam.

Etymology. The species name means ''Vietnam-dwelling.''

Gene sequence. The 18S rRNA gene sequence has the accession number GenBank: MF190552.

Zoobank Registration. urn:lsid:zoobank.org:act:DDF46222-8924-40AF-95C7-24CC1A8FB9B6.

Pigoraptor chileana *n. sp. Tikhonenkov, Hehenberger, Mylnikov et Keeling 2017*

Diagnosis. Diagnosis as for genus *Pigoraptor.* The life cycle consists of uniflagellar motile swimming cells and spherical cysts. The flagellated roundish cells are 6-14 μm in diameter.

Type material. A block of chemically fixed resin-embedded cells of the type strain, Opistho-2, is deposited in Marine Invertebrate Collection, Beaty Biodiversity Museum, University of British Columbia as MI-PR207. This constitutes the name-bearing type of the new species (a hapantotype).

Figure 1M illustrates a live cell of strain Opistho-2.

Type locality. Bottom sediments from freshwater temporary water body (*submerged meadow*) near the Lake Lago Blanca, Tierra del Fuego, Chile.

Etymology. The species name means ''Chile-dwelling.''

Gene sequence. The 18S rRNA gene sequence has the accession number GenBank: MF190553.

Zoobank Registration. urn:lsid:zoobank.org:act:B384FA84-7C46-44EA-884A-0195EDF731B2.

METHOD DETAILS

Culturing of novel species and RNA-seq

To establish clonal cultures, single cells were isolated using a micromanipulator fitted with a glass micropipette. Single cells were transferred to a Petri dish containing a clonal culture of a bacteriotrophic flagellate, either the chrysomonad *Spumella* sp. (for clone Colp-12) or the freshwater bodonid *Parabodo caudatus* (for clones Opistho-1 and Opistho-2). Prey cells were cultivated at room temperature in Pratt medium with addition of *Pseudomonas fluorescens* as food [37]. The clones of opisthokonts Colp-12, Opistho-1, Opistho-2 are stored in the ''Live culture collection of free-living amoebae, heterotrophic flagellates and heliozoans'' at the Institute for Biology of Inland Waters, Russian Academy of Science.

Clone Colp-12 (*Syssomonas multiformis*): Cells sorted on FACS Area Cell Sorter (200 cells) as well as cells from total culture (~7 mL, 7000 cells) were collected by centrifugation (2000 x g, room temperature) on Vivaclear Mini columns 0.8 μm membranes (Sartorium Stedim Biotech Gmng, Germany, Cat. No VK01P042). Total RNA was extracted using the RNAqueous kit (Ambion, lot # 1304062) and was converted into cDNA prior to sequencing using the SMARTer technology (SMARTer Pico PCR cDNA Synthesis Kit, Clontech, lot # 1308018A). Transcriptomic paired-end libraries were prepared with the Nextera XT DNA Library Prep Kit (Illumina). Sequencing was performed on the Illumina MiSeq platform with read lengths of 250bp.

Clones Opistho-1 and Opistho-2 (*Pigoraptor vietnamica* and *Pigoraptor chileana*): Twenty single cells of each clone were transferred to 0.2 mL thin-walled PCR tubes containing 2 μL of cell lysis buffer (0.2% Triton X-100 solution and RNase inhibitor). cDNA libraries were synthetized from single cells employing reverse transcription and the Smart-Seq2 protocol [51]. Transcriptomic

paired-end libraries were prepared with the Nextera XT DNA Library Prep Kit (Illumina). Sequencing was performed on the Illumina MiSeq platform with read lengths of 300bp. Clone Opistho-1 was also sequenced from 11 single cells in an initial analysis (same protocol as above) and the resulting data were used to complement the 20-cell data in the phylogenomic analysis.

Assembly of datasets used in phylogenomic analysis

The *Syssomonas multiformis* raw read datasets, hereafter referred to as "Sm_culture" and "Sm_sorted," were subjected to several cleaning steps as we suspected prey contamination particularly in Sm_culture as well as human contamination in Sm_sorted due to handling during sorting. After removal of transposon sequences resulting from library preparation, reads were merged with PEAR using default options [52]. Merged reads were trimmed with Trimmomatic v 0.36 [53] using the default Trinity settings and then mapped to the transcriptome of *Spumella elongata,* available from the Marine Microbial Eukaryote Transcriptome Sequencing Project [48], using Bowtie 2.1.0 [54] and allowing no mismatches. The Sm_sorted and the Sm_culture datasets were further compared against *Homo sapiens* and *S. elongata*, respectively, with the stand-alone version of DeconSeq [55] using experimentally defined parameters. Thus merged and cleaned reads were assembled using Trinity and coding regions were predicted by TransDecoder [38, 39] including a similarity search against the Swiss-Prot database [56].

The *Pigoraptor vietnamica* and *Pigoraptor chileana* raw read datasets were merged using PEAR, then trimmed and simultaneously cleaned from possible primers/transposon sequences with Trimmomatic. As the data were generated from manually isolated cells from a prey-depleted culture, from experience we expected a lower prey contamination. Merged and trimmed reads were assembled and coding regions in the assembly were predicted as described above.

Raw reads for *Ministeria vibrans* (SRR343051.sra) were downloaded from NCBI's short reads archive and merged, trimmed, assembled and translated as described above. Assembled transcripts for *Abeoforma whisleri*, *Corallochytrium limacisporum*, *Pirum gemmata*, *Sphaerothecum destruens* (all kindly provided by Guifré Torruella), *Ichthyophonus hoferi* (kindly provided by Xavier Grau-Bové), *Amoebidium parasiticum* (https://www.ncbi.nlm.nih.gov/bioproject/PRJNA189477), *Oscarella* sp., *Acropora digitifera* and *Clytia hemisphaerica* (http://compagen.org/datasets.html) were either translated into all six frames using the EMBOSS Transeq tool or peptides were predicted using TransDecoder and a similarity search as described above.

Single-gene tree preparation and concatenation

In addition to the taxa described above we added to or updated the following available transcriptome or genome data in an existing phylogenomic framework: *Saccoglossus kowalevskii* and *Hydra magnipapillata* (ftp://ftp.ncbi.nlm.nih.gov/genomes/), *Trichoplax adhaerens* (https://www.ncbi.nlm.nih.gov/bioproject/12874), *Amphimedon queenslandica*, *Capitella teleta* and *Mnemiopsis leidyi* (ftp://ftp.ensemblgenomes.org/), *Sycon ciliatum* (http://compagen.org/datasets.html [50]), *Salpingoeca rosetta*, *Capsaspora owczarzaki*, *Sphaeroforma arctica, Spizellomyes punctatus* and *Allomyces macrogynus* (https://www.broadinstitute.org/annotation/genome/multicellularity_project [49]), *Creolimax fragrantissima* (https://figshare.com/articles/Creolimax_fragrantissima_genome_data/1403592) and then unpublished transcriptomic data of six additional choanoflagellates (kindly provided by Daniel Richter, now published in [57], see Table 1 within the reference for nomenclature changes): *Didymoeca costata*, *Helgoeca nana*, *Mylnosiga fluctuans*, *Salpingoeca infusionum*, *Salpingoeca kvevrii* and *Salpingoeca punica*.

All listed datasets were searched for 263 genes used in previous phylogenomic analyses [58–60] to generate single gene trees as described in [8], using in-house scripts and available software as listed. In brief: first all sequences in the 263-gene set, representing a wide eukaryotic diversity, were used as queries in BLASTP [61] searches against the datasets listed above. After a parsing step, a maximum of four non-redundant hits was added to the initial 263-gene set used for the search. The expanded gene sets were aligned with MAFFT L-ins-i v7 [62] and ambiguously aligned positions were trimmed off with trimAL v1.2 ([40], maximum number of gaps allowed per site: 20%). To identify paralogs and/or contaminations, we performed ML single tree reconstructions under the LG+Γ4 model using RAxML v8 [63] in combination with 100 rapid bootstraps. Single-gene trees were manually screened to flag contaminating and paralogous sequences. Of note, particularly the *S. multiformis* transcripts derived from culture contained a high fraction of *Spumella sp*. transcripts and we frequently found bacterial transcripts in *P. vietnamica*, whereas the *S. multiformis* transcripts derived from cell sorting were successfully cleaned from the human contamination as shown by the absence of human transcripts in the single-gene trees. The datasets of *A. digitifera*, *C. hemisphaerica* and *S. ciliatum* also contained transcripts of *Symbiodinium* sp., a member of the Viridiplantae and a ciliate, respectively. All such sequences as well as other less frequent contaminants, paralogs and sequences of uncertain identity were removed from the expanded gene sets, resulting in a final 255-gene dataset. The gene sets were aligned and trimmed as before, followed by taxon selection and concatenation with SCaFoS v 1.2.5 [41], resulting in a 38-taxa matrix of 81,495 aa positions used for phylogenomic analysis.

Phylogenomic analyses, fast-evolving site removal analysis and AU test

The ML phylogeny (see Figure 2A) was inferred by IQ-TREE [42] using the C40 empirical mixture model in combination with the LG matrix, amino acid frequencies computed from the data and four gamma categories for handling the rate heterogeneity across sites (LG+C40+F+G4 model) plus 1000 ultrafast (UF) bootstrap approximations. Empirical profile mixture models represent the true evolutionary processes more realistically than traditional empirical matrices (e.g., LG [64]). Additionally, nonparametric ML bootstrap support was obtained from 1000 ML replicates using the LG+F+Γ model implemented in RAxML. Per site evolutionary rates were estimated using the program DistEst [43] and sites were then sorted from fastest to slowest rate. Sites with highest rate were then sequentially removed 4000 sites at a time, generating an alternative dataset at each step.

We created the topologies for AU test by generating all possible relationships between Ichthyosporea, Pluriformea, Filasterea, Choanoflagellatea and Metazoa, while the topologies within these groups were fixed to our IQ-TREE topology. Branch lengths and other parameters were optimized by IQ-TREE under the LG+C40+F+G4 model and site likelihoods were generated. To perform AU test, the program CONSEL [44] was used with default settings. All topologies were rejected (p value < 0.05); with the exception of our main topology shown in Figure 2, the Teretosporea topology (i.e., the Pluriformea and Ichthyosporea monophyly) and the topology where Pluriformea form the earliest-branching group of the Holozoa (see Table S2).

Bayesian Inference was performed with PhyloBayes MPI 1.4 [65] using the GTR matrix in combination with the CAT infinite mixture model and four gamma rate categories. We ran four MCMC chains for ~9,800 generations and saved every second tree. We tested for convergence (maxdiff < 0.1) using bpcomp, implemented in PhyloBayes, with default parameters (burn-in of 20%). Two chains each converged (see Figures S1A and S1B) with the topologies differing within the metazoa.

We also tested the effect of an expanded outgroup on our tree topology by computing an ML tree with 400 nonparametric bootstrap approximations using the LG+F+Γ model as implemented in RAxML (see Figure S1C). The additional taxa are: *Aspergillus flavus* (Aspergillus Comparative Sequencing Project, Broad Institute), *Batrachochytrium dendrobatidis* (Batrachochytrium dendrobatidis Sequencing Project, Broad Institute), *Coprinopsis cinerea* (Coprinopsis cinerea Sequencing Project, Broad Institute), *Fonticula alba* and *Mortierella verticillata* (https://www.broadinstitute.org/annotation/genome/multicellularity_project), *Mucor circinelloides* and *Rhizopus delemar* (Mucorales Sequencing Project, Broad Institute), *Schizosaccharomyces pombe* (Schizo*Saccharomyces* group Sequencing Project, Broad Institute), *Catenaria anguillulae*, *Conidiobolus coronatus*, *Gonapodya prolifera*, *Phycomyces blakesleeanus*, *Piromyces* sp. and *Rhizophagus irregularis* (http://genome.jgi.doe.gov/programs/fungi/index.jsf [66]), *Blastocladiella emersonii* (http://www.ncbi.nlm.nih.gov/nucest/?term=txid4808[Organism:noexp]), *Nuclearia simplex* strain 2 and 4 (TBestDB), *Rozella allomycis* (https://www.ncbi.nlm.nih.gov/bioproject/81749) and Ustilago maydis (http://bioinformatik.wzw.tum.de/index.php?id=63).

18S rRNA phylogenetic analysis

The 18S rRNA sequence for strain Colp-12 was amplified using the primer pairs PF1-FAD4 [37] and EukA-EukB [36], for strain Opistho-1 the primer pair EukA-EukB was used and the 18S rRNA sequence for strain Opistho-2 was extracted from the transcriptome data.

The taxon sampling for the 18S rRNA tree was based on the ones used in del Campo and Ruiz-Trillo 2013 [12]. Sequences were aligned using MAFFT auto mode (-auto). Alignments were checked using AliView [67] and highly variable regions of the alignment were removed using trimAl 1.2 (-gt 0.3 -st 0.001). Maximum likelihood (ML) phylogenetic trees were constructed with the fastML method of RAxML 7.2.6 [68] assuming the GTRCAT substitution model. The resulting ML tree came out of 1000 independent tree searches (starting from distinct randomized maximum parsimony trees). In order to assess tree uncertainty, a non-parametric bootstrap analysis was performed with 1000 replicates. A constrained tree (see Figure S2) was built using the same RAxML parameters plus the g function and using our multigene tree as constrain. Tree figures were edited with iTol [45].

T-box domain phylogeny

To identify T-box domains in the transcriptomes of *S. multiformis*, *P. vietnamica* and *P. chileana* we used hmmscan (HMMER3.1; hmmer.org) to search our data against the Pfam-A database v29.0 [47], requiring all possible hits to fall within the stringent manually curated Pfam gathering threshold. Thus identified transcripts were submitted to the InterProScan [69] web tool to determine the domain coordinates. The extracted domain sequences were added to an alignment from [28], realigned with MAFFT L-ins-i and trimmed using trimAl as described above. We reconstructed the best of 50 ML trees using the LG+Γ model implemented in RAxML v8 and calculated 1000 ML nonparametric bootstrap trees (see Figure S4). We also performed Bayesian Inference using PhyloBayes, running 4 chains until convergence (bpcomp-calculated maxdiff < 0.1) and plotted the resulting posterior probabilities onto the ML tree.

Gene and domain analyses

Genes associated with multicellularity (see Figure S3B) were identified by a bi-directional BLASTP search using the corresponding *H. sapiens* protein sequences as query and in turn using the best hit in a reverse search against the GenBank nr database to again recover the protein in question. In case of unclear phylogenetic association or doubts about the transcript identity, candidates were further analyzed by phylogenetic reconstruction using FastTree v2.1.7 [46] and a custom database that expands on the taxa listed in Table S1 to account for the contaminations detected during phylogenomic analysis. Due to the large size of the tyrosine kinase family, we prepared high quality trees using RAxML v8 for the cytoplasmic tyrosine kinases SRC, CSK and FAK as well as for the receptor tyrosine kinases and phosphatases discussed in the manuscript. After an initial BLASTP step of the query against our custom database, were parsed for hits with an e-value threshold $\leq$ 1e-25 or less, to yield a reasonable alignment size. After alignment with MAFFT L-ins-i, poorly aligned regions were eliminated using trimAl with a gap threshold set at 80%. After initial tree reconstruction with FastTree and manual alignment and tree inspection, ML tree reconstructions were performed with RAxML v8, calculating the best of 50 trees and 1000 nonparametric bootstrap replicates using the LG+Γ model implemented in RAxML for each query candidate.

We performed domain analysis (see Figures 3A and 4) by searching the datasets of interest against the Pfam-A database using hmmscan as implemented in hmmer as described above (using the Pfam gathering threshold). Transcripts containing domains of interest were further investigated with the InterProScan web tools as well as submitted to the SMART sequences search. Software

integrated into InterProScan and the SMART sequences search was also used to predict signal peptides (SignalP) and transmembrane regions (TMHMM) to test whether a peptide is putatively secreted and/or integrated into a membrane or not. To exclude possible contaminations, transcripts were also submitted to a BLASTP search against the GenBank nr database and, in case of doubt concerning their phylogenetic association, were used in phylogenetic reconstructions using FastTree 2.1.7 and our custom database. Transcripts with unexpected domain combinations, such as combinations of eukaryotic with typical bacterial domains, were also subjected to read coverage analysis. For that purpose, we mapped the raw reads of the novel taxa onto their respective Trinity assemblies, using Bowtie 2.1.0 with the non-deterministic option and inspected the read coverage of transcripts of interest using the Integrative Genomics Viewer (IGV) [70].

DATA AND SOFTWARE AVAILABILITY

The "*Syssomonas multiformis* predicted peptides (culture)", the "*Syssomonas multiformis* predicted peptides (sorted)", the "*Pigoraptor vietnamica* predicted peptides", the "*Pigoraptor chileana* predicted peptides ", the "Single-gene alignments (trimmed) for phylogenomic analysis", the "Phylogenomic reconstruction without novel taxa", all protein tyrosine kinase/phosphatase trees and all trees from the main and supplemental figures of the manuscript in newick format have been deposited to the Dryad data depository (http://datadryad.org) at http://dx.doi.org/10.5061/dryad.26bv4.