

Title: A multicellular developmental program in a close animal relative

Authors: Marine Olivetta¹, Chandni Bhickta¹, Nicolas Chiaruttini², John Burns^{3*}, Omayya Dudin^{1*}

Affiliations:

¹ Swiss Institute for Experimental Cancer Research, School of Life Sciences, Swiss Federal Institute of Technology (EPFL), 1015 Lausanne, Switzerland.

² Bioimaging and Optics Core Facility, Ecole Polytechnique Fédérale de Lausanne (EPFL), 1015 Lausanne, Switzerland.

³ Bigelow Laboratory for Ocean Sciences, East Boothbay, ME

*Corresponding authors. Emails: jburns@bigelow.org , omaya.dudin@epfl.ch

Abstract:

All animals develop from a single-celled zygote into a complex multicellular organism through a series of precisely orchestrated processes. Despite the remarkable conservation of early embryogenesis across animals, the evolutionary origins of this process remain elusive. By combining time-resolved imaging and transcriptomic profiling, we show that single cells of the ichthyosporean *Chromosphaera perkinsii* - a close relative that diverged from animals approximately 1 billion years ago - undergo symmetry breaking and develop through cleavage divisions to produce a prolonged multicellular colony with distinct co-existing cell types. Our findings about the autonomous developmental program of *C. perkinsii*, hint that such animal-like multicellular development is either much older than previously thought or evolved convergently in ichthyosporeans.

One-Sentence Summary:

The ichthyosporean *C. perkinsii* develops via symmetry breaking, cleavage divisions, and forms spatially-organized colonies with distinct cell types.

Introduction

The evolution of multicellular organisms from their unicellular ancestors marks a major transition in the history of life on earth (1–5). This transition was accompanied by fundamental developmental challenges such as generating diverse cell types, forming three-dimensional (3D) tissues and establishing overall coordination to drive body plan formation. Asymmetric cell division contributes to cellular diversity (6–8), while the formation of 3D tissues relies on the precise coordination of cell divisions, adhesion, and signalling (9–14). Compared to other eukaryotic lineages, animal multicellularity stands out by relying on an autonomous developmental program, predominantly driven by intrinsic signals, that drives the emergence of an extensive variety of cell types from a single-celled zygote (15–17). Throughout early animal embryogenesis, several ordered processes take place including cleavage divisions, axis establishment, zygotic genome activation and spatial organization of germ layers. This sequence, while displaying remarkable conservation across species, exhibits a degree of developmental plasticity, highlighting the adaptability of development to distinct ecological pressures (18–20). Despite this plasticity, the fundamental aspects of this developmental sequence suggest that parts of the underlying program may have originated before the emergence of animals themselves (21). This is partly supported by paleontological studies, particularly those focusing on the early Ediacaran Weng'an Biota of the Doushantuo Formation (609 Mya) (22), which have uncovered fossils like *Megasphaera*, *Helicoforamina*, *Spirallicellula*, and *Caveasphaera*, which are tentatively classified animal stem groups (23). These fossils exhibit morphological features characteristic of animal embryos such as cleavages, Y-shaped cell-cell junctions, tetrahedral 4-cell stages, and a proposed open mitotic strategy (23–26). However, their interpretation as stem groups to animals is a subject of ongoing debate (27, 28), as this interpretation strongly relies on the cellular dissimilarities between these fossils and a select few protists, known to be among the closest living relatives of animals.

Animals are closely related to choanoflagellates, filastereans, pluriformeans, and ichthyosporeans (Fig. 1A) (2, 29–32). These lineages not only partly share a genetic toolkit used by animals for development, but can also form transient multicellular structures and display various temporary cell stages in distinct environments (1, 30, 33). Current evidence indicates that the formation of clonal multicellular choanoflagellate colonies (34–36) and the emergence of filastereans aggregates (37–39), both morphologically distinct from any known animal embryonic stage or Ediacaran fossil, occur as a facultative response to external chemical cues (34, 35, 40–43). Moreover, while existing results in the choanoflagellate *Salpingoeca rosetta* suggest the presence of morphologically distinct cells within its transient multicellular colonies (44, 45), definitive evidence for a coordinated developmental program driving cell-type differentiation in any close animal relatives remains lacking. These results argue that the animal intrinsic embryonic program orchestrating 3D-cell architecture and spatial cell-type differentiation likely evolved simultaneously with the emergence of animal multicellularity. However, this assumption is largely based on studies focusing on choanoflagellates and filastereans, often overlooking the life cycles, cellular physiology and development of other close relatives.

Among close animal relatives, the Ichthyosporea, comprising two primary lineages, Dermocystida and Ichthyophonida (Fig. 1A) (46, 47), exhibit a variety of life cycles that combine fungal-like characteristics with transient multicellular stages reminiscent of early animal development (47). Most ichthyophonids, including the model *Sphaeroforma arctica*, undergo coenocytic development, characterized by synchronized nuclear divisions without cytokinesis (48–52). This process depends on an acentriolar microtubule organizing centre (MTOC) that drives a fungal-like closed mitosis (53). Upon reaching a specific nuclear-to-cytoplasmic ratio (54), *S. arctica* undergoes an actomyosin-dependent cellularization, during which a transient multicellular

layer resembling an animal epithelium is formed before the release of new born cells to repeat the cycle (50). In contrast, our recent study reveals that *Chromosphaera perkinsii*, the only cultured free-living dermocystid (55), undergoes centriole-mediated, animal-like open mitosis coupled with cell cleavages (53). This discovery suggests that *C. perkinsii* proliferates through a palintomic developmental mode, highlighting the vast range of cellular and developmental diversity within the Ichthyosporea. However, nothing is currently known about the spatiotemporal dynamics of the ichthyosporean palintomic life cycle, nor is it clear if it is orchestrated by an autonomous developmental program that exhibits any similarities with early animal embryogenesis.

Results

To better characterize the development of *C. perkinsii*, we conducted long-term brightfield live imaging of synchronized cell populations (Fig. 1B to D, S1A to C and Movie S1). In spite of the light sensitivity of *C. perkinsii* (see methods), our results reveal that synchronized single cells undergo ~65 hrs of growth until the first cell cleavage (Fig. 1, B and D, S1B). Following this initial cleavage, the now multicellular colonies maintain a constant size for an additional 30-hour period, which is followed by the release of hundreds of new born cells (Fig. 1B to D, S1E and Movie S1). Interestingly, released cells were of three types. First, cells without flagella and capable of division, termed “mitotic” (Movie S2). Second, flagellated cells that exhibited flagellar motility termed “flagellates” (Fig. 1C and Movie S3). Third, flagellated cells that remained immobile with dynamic membrane protrusions, and were thus designated “amoeboflagellate” (Movie S2). Counting flagellated cells reveals an increase in their number around the time of cell release (>76 hrs), yet they represent less than 15% of all observed single cells (fig. S1D), indicating that only a fraction of cells differentiates into flagellated cells. These observations suggest that *C. perkinsii* differentiates into at least two distinct cell types, with flagellated cells potentially transitioning between flagellate and amoeboflagellate states.

To gain further insights about *C. perkinsii* development, we fixed and stained the nuclei of synchronized *C. perkinsii* cells throughout their entire lifecycle (Fig. 1E to G). We observe that cells exhibit similar size growth dynamics, though with increased variability, which can be attributed to heterogeneity within the cell population (fig. S1F). Upon counting the number of nuclei per cell, we find that after the first cleavage (~65 hrs), the nuclear number rapidly increases, surpassing 500 nuclei per colony before cell release (Fig. 1E to G and S1G). As this significant and rapid increase in nuclear content occurs at constant colony size, we note it is accompanied by a gradual reduction in nuclear diameter (Fig. 1E and G). These results show that *C. perkinsii* develops through successive cleavage division while maintaining a constant overall colony volume (Fig. 1G). This multicellular colony stage is maintained for a substantial duration, accounting for approximately 33% of the life cycle in culture (Fig. 1H). In contrast, the epithelium-like multicellular stage in the ichthyosporean model *S. arctica* accounts for only 2% of its life cycle (50, 54).

To better understand whether *C. perkinsii*'s autonomous program is associated with a distinct transcriptional signature, we employed time-resolved transcriptomic profiling. We isolated and sequenced mRNA from two independent synchronized cultures, covering key time points from the single-cell stage through the multicellular stage until cell release (fig. S2A and Table S1). We identify 7,773 transcripts that are differentially expressed ($FDR < 0.05$) across these stages, representing 62% of *C. perkinsii*'s total genes. These genes clustered into five distinct expression patterns throughout development (Fig. 2A, S2A). These patterns highlight the differences between the unicellular phases, both before cleavages (54 hrs) and after release (104 and 120 hrs), compared to the multicellular phase (72 to 96 hrs) (Fig. 2A, S2B). Gene ontology (GO) enrichment analysis reveals that unicellular phase clusters (1, 2 and 5) are enriched in GO terms related to nutrient and lipid biosynthesis and polysaccharide catabolism (Fig. 2B and Tables S2 and 3), whereas multicellular phase clusters (3 and 4) show enrichment in cell division and flagellar differentiation (Fig. 2B). Therefore, prior to the first cleavage, *C. perkinsii* produces and accumulates key cellular components, such as proteins and lipids in a manner reminiscent of oocyte maturation (56, 57). Following the first division, these biosynthetic and catabolic processes are turned off, indicating a gradual consumption of previously synthesized nutrients as the cleavages progress (Fig. 2A and S2A) and only reappear around cell release.

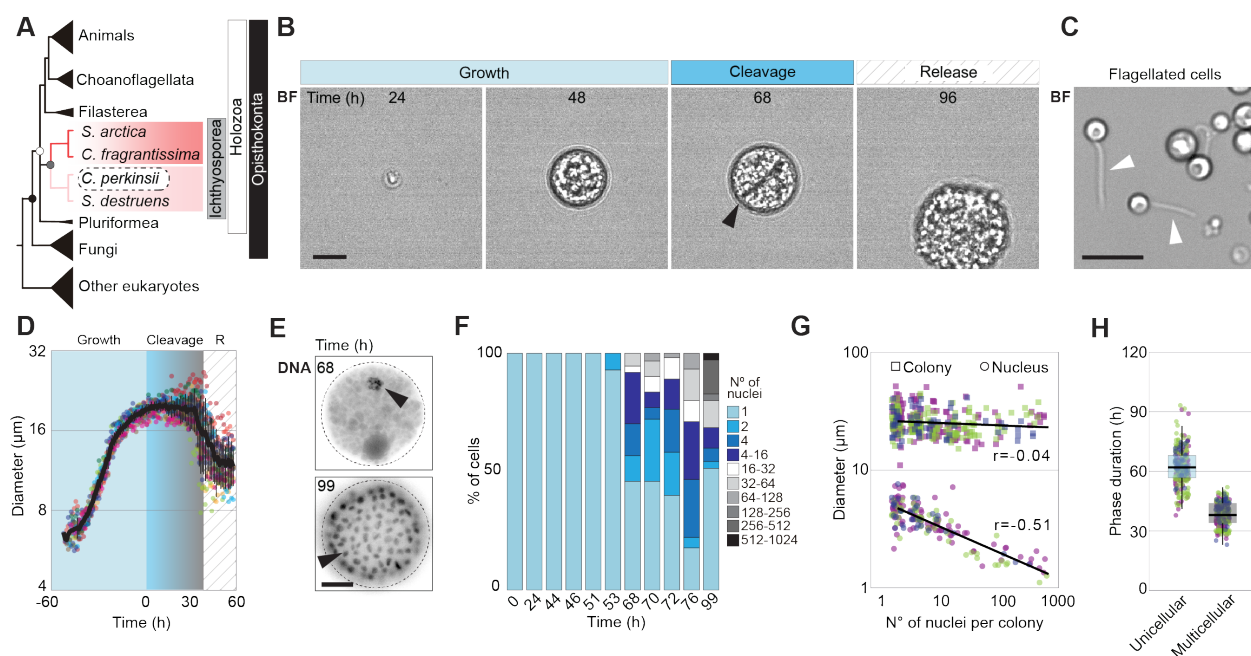


Fig. 1: *C. perkinsii* undergoes a series of cleavages at constant volume

(A) Cladogram representing the position of ichthyosporeans, featuring ichthyophonids such as *S. arctica* (red) and dermocystids, including *C. perkinsii* (pink), within the eukaryotic tree. (B) Time-lapse images of the life-cycle of *C. perkinsii* show cell-size increase, cleavage at the 1-to-2 cell stage (arrow) and release of new-born cells (movie S1). Bar, 10 µm. (C) Microscopy of flagellated cells observed following cell release (arrows). Bar, 10 µm. (D) Mean cell diameter over time of 10 single cell traces aligned to the 1st cell cleavage reveals the duration of distinct cell stages: Growth, cleavage, and cell release (R). (E) Microscopy of DNA-stained cells at 68 and 99 hrs of growth, highlighting the number and size of nuclei at each timepoints (arrows). Bar, 10 µm. (F) Distribution of nuclear content of *C. perkinsii* colonies across the life cycle at 23°C measured by microscopy of Hoechst-stained cells ($n > 96$ /time point). (G) Distribution of colony (squares) and nucleus (circles) diameters relative to the number of nuclei per colony demonstrates that cellular size remains constant, whereas nuclear size decreases ($n = 1571$ colony). The colours represent 3 independent replicates. (H) Average phase duration for the unicellular and multicellular growth phases ($n = 244$ cell/colony). The colours represent 3 independent replicates.

In contrast, the expression of genes driving catabolic processes remains low after release (cluster 2), likely due to nutrient depletion from the medium (Fig. 2A and B, S2B). Towards the end of the multicellular phase a distinct transcriptional pattern emerges (cluster 4) with enrichment in genes for cilia/flagella formation, motility, and left/right patterning in animals (Tables S2 and 3).

To determine if the developmental program of *C. perkinsii* is distinct among ichthyosporeans, we compared it with the transcriptional signature of the ichthyosporean *S. arctica* which undergoes coenocytic growth followed by cellularization. We find that key GO functional categories including the cell cycle, cytokinesis, DNA replication, polarity, and differentiation do not overall correlate between *C. perkinsii* and *S. arctica*, indicating distinct transcriptional programs (Fig. 2C, S2C and D). Despite these differences, certain cytoskeletal regulators previously highlighted and required in *S. arctica* epithelium-like cell layer (50) such as formin 2, critical for actin nucleation, correlates between both species regardless of developmental type, suggesting a housekeeping role (fig. S2E). In contrast, formin 5 exhibits an inverse correlation, implying a species-specific role in actin dynamics (fig. S2E). Next, we sought to examine whether the transcriptional program of *C. perkinsii* exhibits similarities with that of early embryonic development in early-branching animals. To accomplish this, we categorized genes from animals and ichthyosporeans into ortho-groups; clusters of genes that originated from a single gene in their last common ancestor. We then analysed the expression patterns of these genes from the onset of zygote formation (Z), through the various cleavage stages, and up to the blastula stage (B) as previously described (58) (see methods). By calculating the average correlation of gene expression among orthogroups across time points for *C. perkinsii* and the three earliest-branching animals - the ctenophore *Mnemiopsis leidyi*, the sponge *Amphimedon queenslandica*, and the cnidarian *Nematostella vectensis* - we find an overrepresentation of positively correlated expression patterns indicating similarities in the transcriptional program between *C. perkinsii* and these three early-branching animals (fig. S2G). By focusing on a subset of genes important for flagellar formation (POC1, IFT122, IFT52) (59, 60), cellular signalling (Ga, G β , PKAc- α) (61–63), cell adhesion (Talin, β -Catenin, Vinculin, Grancalcin) (64–67), and transcriptional regulation during animal development (Brachyury, CBP/p300, Piwi, Rbl1, P53, RunX) (55, 68–70), we can observe comparable patterns of gene expression between *C. perkinsii* and basal animals. In certain cases, this expression exhibits conserved temporal patterns across most lineages (Fig. 2D, S2H). Notably, Brachyury, a transcription factor regulating gastrulation, and Piwi, a marker of germline formation, show such conserved expression patterns in many, though not all, animals (Fig. 2D). While we cannot exclude the potential impact of noise from bulk culture asynchrony, ortho-group identification, or developmental time averaging, our results show similar patterns of expression of genes critical for animal embryogenesis during the multicellular development of *C. perkinsii*.

To further explore the development of *C. perkinsii* at the cellular level, we tracked the plasma membrane in synchronized cells using long-term live imaging. Our results show that *C. perkinsii* undergoes a series of ordered cleavages with cortical rotations, starting from a single cell, progressing to a 2-cell stage and then beyond (fig. S3A, Movie S4). Ultimately, this yields a spatially organized multicellular colony surrounded by an external cell wall and occasionally exhibiting a central cavity reminiscent of a blastocoel (fig. S3B).

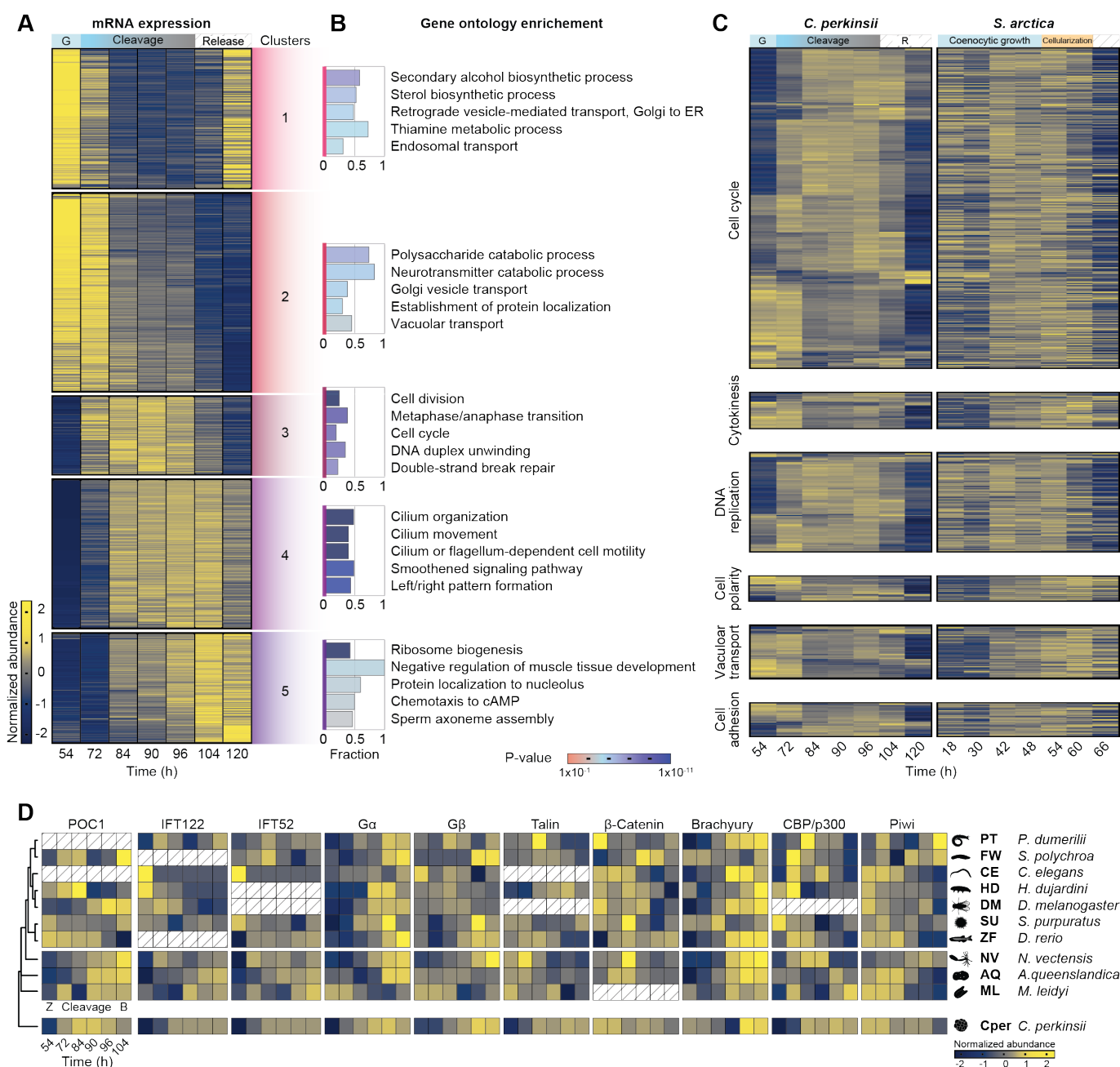


Fig. 2: Transcriptional dynamics across the palintomic life cycle of *C. perkinsii*

(A) A heatmap of 7773 coding genes clustered into 5 dynamic clusters. (B) Gene ontology enrichment of the 5 different dynamic clusters highlighting the patterned transcriptional program of *C. perkinsii* development linked to Tables S2 and 3. (C) A comparative heatmap of gene expression dynamics of orthologs between the palintomic *C. perkinsii* and the coenocytic *S. arctica* (50) life cycles for key GO categories, including cell cycle (GO:0007049), cytokinesis (GO:0000910), DNA replication (GO:0006260), cell polarity (GO:0007163), Vacuolar transport (GO:0007034), and cell adhesion (GO:0098609). It highlights the distinct transcriptional programs between both ichthyosporean species. The heatmaps for *C. perkinsii* and *S. arctica* have the same row order allowing direct comparison of gene expression patterns of orthologs between the two species. (D) A comparative heatmap of gene expression for key regulators involved in flagellar motility (POC1, IFT122, IFT52), cell signalling ($G\alpha$, $G\beta$), adhesion (Talin, β -Catenin), and transcriptional regulation of animal development (Brachyury, CBP/p300, Piwi) dynamics between the *C. perkinsii* (Cper) life cycle and early embryonic development stages of early-branching animals. This comparison spans from the onset of zygote formation (Z) through various cleavage stages and up to the blastula stage (B).

Due to *C. perkinsii* photo-sensitivity and a need for increased resolution, we opted for imaging synchronized cultures at specific developmental times instead of long-term live imaging. We used three approaches: live-imaging of membrane-stained cells to comprehend overall spatial organisation (Fig. 3A, Movie S5), immunostaining of actin and nuclei to characterize nuclear positioning (Fig. 3B), and ultrastructural expansion microscopy (U-ExM) of microtubules, crucial for antibody accessibility, to reveal spindle orientation during early cleavages (Fig. 3C).

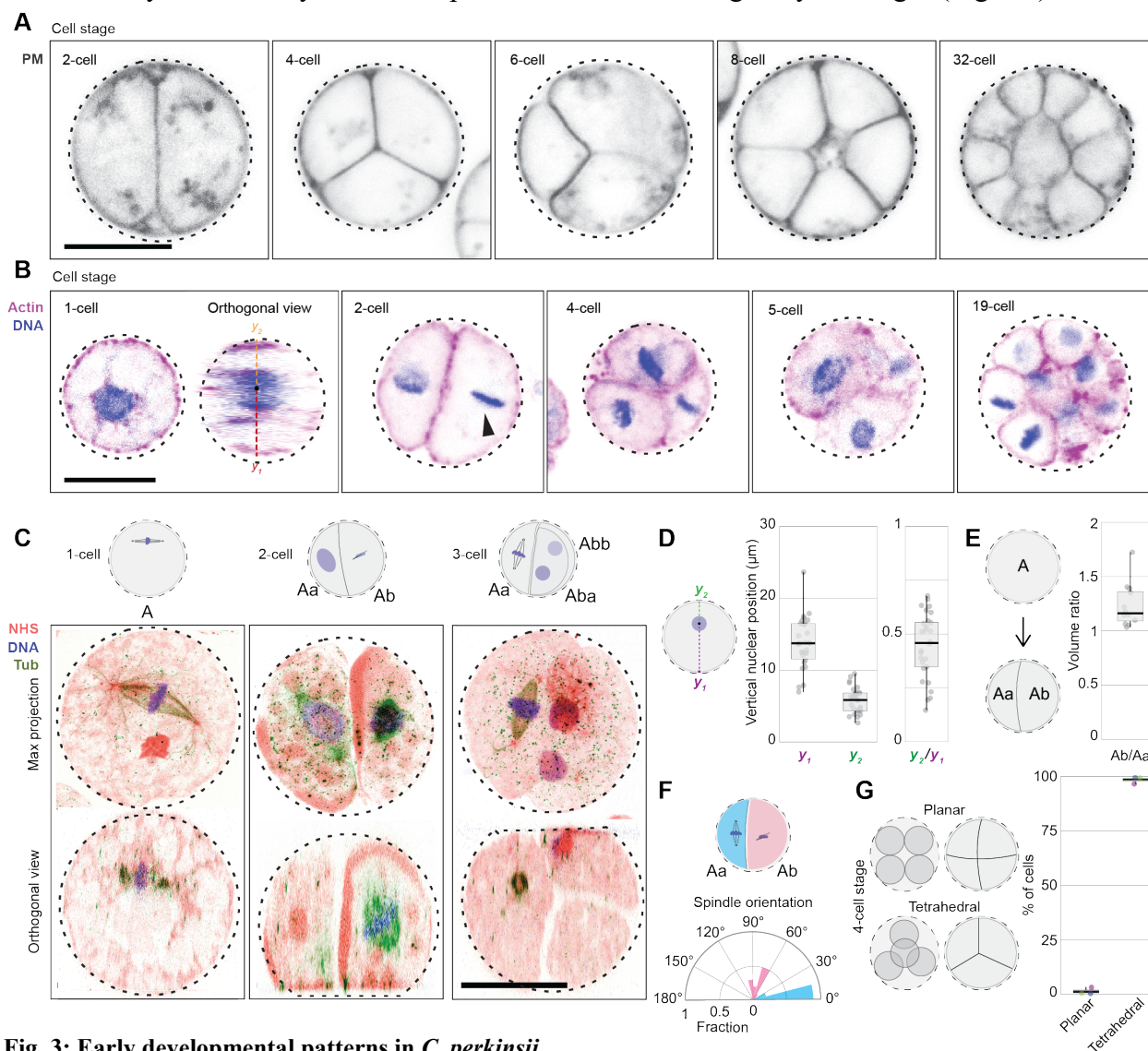


Fig. 3: Early developmental patterns in *C. perkinsii*

(A) PM-stained live colonies at distinct cell stages, highlighting the patterned cleavages, tetrahedral 4-cell stage and formation of spatially organized multicellular colonies (movie S5). Bar, 10 μm. (B) Actin (magenta) and DNA-stained (blue) colonies at distinct cell-stages showcasing the cortical positioning of the nucleus, asymmetrical cell division (in volume and in time), and the formation of a multicellular colony. Bar, 10 μm. (C) U-ExM stained colonies for pan-labelling with NHS-Ester (red), microtubules (green), and DNA (blue), highlighting the first mitotic division at the cortex, perpendicular spindles at the 2-cell stage, and a 3-cell stage (movie S6). Bar, 10 μm. (D) Box-plot showing the mean vertical localisation of the nucleus at the 1-cell stage prior or at the 1st cleavage highlighting the cortical localisation of the nucleus (n = 28 cells). (E) Box-plot illustrating the volumetric ratio following the 1st division resulting in Ab and Aa cells, highlighting the asymmetrical cell division (n=14 colonies). (F) A polar plot representing the spindle angle for the Aa and Ab cells during mitosis, demonstrating that the Ab cells divides perpendicular to the 1st cleavage whereas the Aa cells divides parallel to it (n = 11 colonies). (G) Box-plot showing the percent of 4-cell stage exhibiting a planar or tetrahedral spatial organisation (n = 108 colonies).

Initially, the nucleus of the 1-cell stage (Cell A) is centrally located, before it migrates to the cortex, where it undergoes mitosis (Fig. 3B to 3D and S3C to S3E). This yields two daughter cells (Aa and Ab), with Ab being ~1.3 times larger in volume (Fig. 3A, 3E and S3D). Remarkably, we also note that the Ab cell undergoes mitosis ahead of the Aa cell (Fig. 3B, 3C and S3F), resulting in a transient 3-cell stage (Fig. 3C and S3G, Movie S6), thereby highlighting an asynchrony in mitotic entry. Furthermore, we observe that the Ab cell undergoes mitosis perpendicular to the 1st division plane, while the Aa cell undergoes mitosis parallel to it (Fig. 3C and 3F). Therefore, Aa and Ab orient their spindles and divide perpendicularly to each other, ultimately forming a tetrahedral 4-cell stage (Fig. 3B, 3C and 3G). Tracking subsequent cleavages proved more challenging, however we find that the number of cells within the multicellular colonies appears to be variable (Fig. 1F, 1G, and S1G). This potentially suggests that in our experimental setup, *C. perkinsii* undergoes a rotational cleavage pattern before the 8-cell stage, followed by a more chaotic pattern thereafter. Together, these results reveal that the development of multicellular *C. perkinsii* colonies is marked by early symmetry breaking, a phenomenon known to drive cell differentiation across different multicellular systems, followed by formation of a spatially-organized, blastula-like, multicellular colony.

Next, we sought to better understand the differentiation program that results in the formation of distinct cell types, focusing in particular on the flagellated cells, which are easily identifiable using brightfield microscopy (Fig. 1C), expansion microscopy with pan-labelling and antibodies against both beta-tubulin and acetylated tubulin (Fig. 4A and B). As flagellar gene expression starts much earlier than cell release (Fig. 2A to C), we wondered whether the differentiation into flagellated cells occurs post-release, representing a case of temporal cell-type, or whether it differentiates already within the multicellular colony. To this end, we employed expansion microscopy to label flagella and nuclei in the late multicellular colonies (T78, T84, T96). Remarkably, we detect flagellated cells within the multicellular colonies, confirming a differentiation process before cell release (Fig. 4C, Movies S7 and S8). As development progresses and the total number of cells per colony increases, the number of flagellated cells also rises (Fig. 4C and D). However, this number never exceeds 50% of all cells present within the colony, exhibits significant heterogeneity between colonies; moreover, we noted that flagellated cells often cluster spatially within the colony (Fig. 4C to E, Movies S7 and S8). These results suggest the coexistence of at least two distinct cell types within multicellular *C. perkinsii* colonies.

Discussion

Exploring the evolutionary history that led to the emergence of animal multicellularity has been challenging. This stems notably from the scarcity of fossils predating modern-day animals and the limitations stemming from studying only a few closely related protists, primarily focusing on their genomic content and often neglecting their cellular physiology (2, 33, 71, 72). Here, we characterize the life cycle of the ichthyosporean *C. perkinsii*, a close animal relative, providing evidence for an intrinsic and clonal multicellular developmental program. Among all known animal relatives, this program displays unique morphological similarities with early animal embryogenesis. Our results show that *C. perkinsii* undergoes palintomic division (Fig. 4F), involving cell cleavages coupled to open mitosis, diverging from the coenocytic development and cellularization in other studied ichthyosporeans (52). Notably, the development of *C. perkinsii* is stereotypical and occurs independently of apparent external trigger, leading to the formation of colonies harboring hundreds of cells (Fig. 4F).

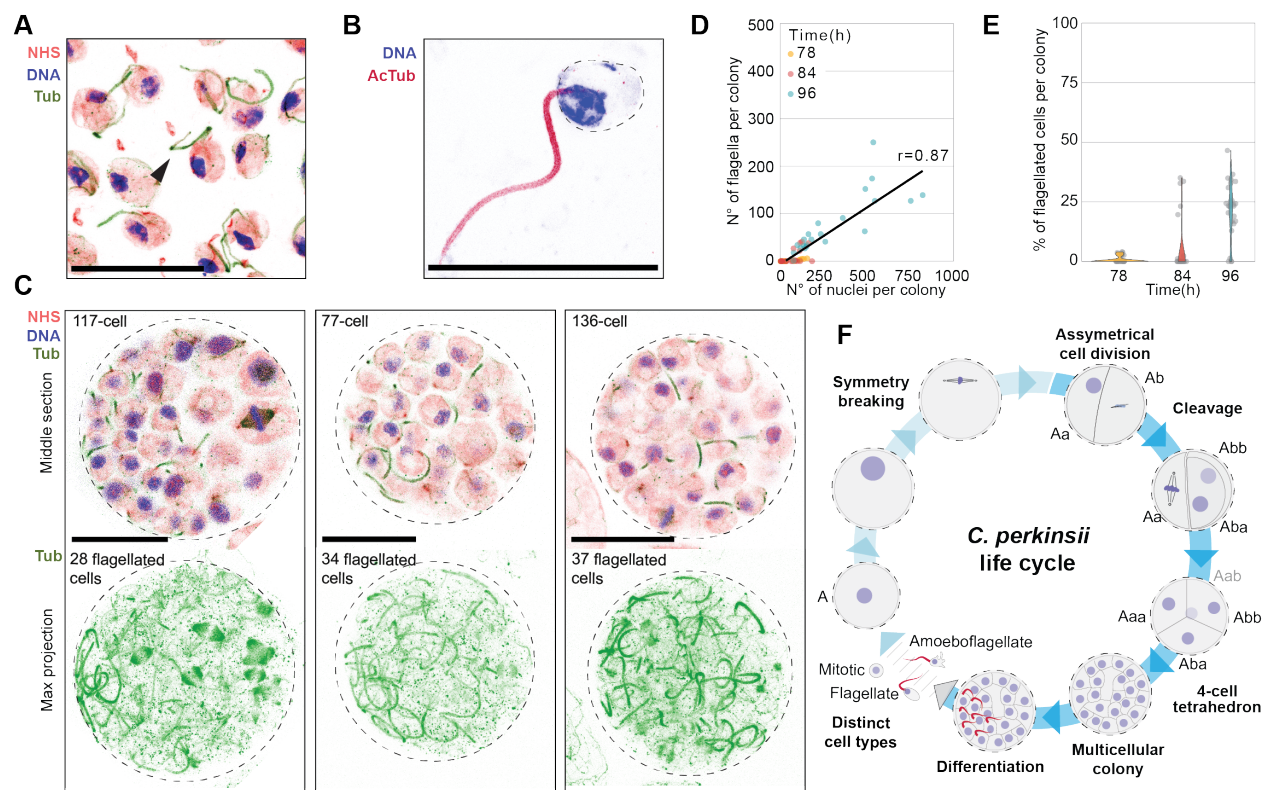


Fig. 4: Cell differentiation within *C. perkinsii*'s colonies

(A) U-ExM stained cells for pan-labelling with NHS-Ester (red), microtubules (green), and DNA (blue), identifying flagellated cells (arrow). Bar, 10 μ m. (B) U-ExM stained flagellated cell for Acetylated tubulin (dark red) and DNA (blue). Bar, 10 μ m. (C) U-ExM stained late colonies for pan-labelling with NHS-Ester (red), microtubules (green), and DNA (blue), highlighting the co-existence of flagellated and non-flagellated cells within the multicellular colony (movies S7 and S8). Bar, 10 μ m. (D) Dot plot illustrating the number of flagellated cells per colony, which increases with the overall number of cells per colony. Notably, although there is a positive correlation, the increase is not linear, showing that distinct cell types coexist within the multicellular colonies ($n > 27$ colonies/time point). (E) Violin plot showing the percent of flagellated cells within a multicellular colony at distinct developmental times ($n > 27$ colonies/time point). (F) A model representing the developmental program of *C. perkinsii*, beginning with the relocation of the nucleus to the cortex, symmetry breaking at the 1-to-2 cell stage, formation of a tetrahedral 4-cell stage, development into a multicellular colony with intricate 3D architecture, differentiation of a subset of cells into flagellated cells, and release of distinct cell types.

The observation that *C. perkinsii* positions the nucleus cortically in the one-cell stage, undergoes asymmetric cell division, forms spatially organized multicellular colonies, and differentiates into distinct cell types, demonstrates striking morphological and transcriptional parallels with the early stages of animal embryogenesis, which occur following fertilization and precede gastrulation (20). This developmental program in *C. perkinsii* challenges the traditional distinctions separating 'simple' from 'complex' multicellularity (5). 'Complex' multicellularity is characterized by traits such as cell-cell adhesion, intercellular communication, and a sophisticated developmental program guiding tissue differentiation, traits typically associated with animals, land plants, red and brown algae, Basidiomycota, and Ascomycota. By exhibiting some of these characteristics, *C. perkinsii* emerges as a potential intermediary, blurring the lines between the two classifications.

Although the transcriptional signature of *C. perkinsii*'s development correlates positively with that of early-diverging animals, determining whether it is ancestral to all animals or whether it evolved independently requires still further investigation, including the spatial identification of

cell-type signatures at the single cell level. Nonetheless, the ability of *C. perkinsii* to organize into a 4-cell tetrahedral stage and form late blastula-like colonies with clustered flagellated cells within the colony (Fig. 4F), sets *C. perkinsii* apart from previously studied animal relatives including other ichthyosporeans, filastereans and choanoflagellates (2, 30, 33). Moreover, *C. perkinsii*, which is free-living, exhibits morphological similarities with Ediacaran embryonic-like fossil such as *Megasphaera* (~600 Mya). These fossils are characterized by Y-shaped cell junctions (24), known to require rotating sister cell division planes, flexible cell membranes not limited by separating cell walls, and cell-cell adhesion. While these similarities do not fully elucidate the evolutionary path from Ichthyosporea to animals, they favour the interpretation of *Megasphaera* as an ichthyosporean-like fossil rather than a stem group to animals (27).

Altogether, our study highlights the multicellular developmental diversity within the Ichthyosporea, as illustrated by the palintomic program in *C. perkinsii* described here and the coenocytic program of *S. arctica*, reminiscent of *Drosophila*'s early embryos. With both species representing opposite ends of the developmental spectrum, the Ichthyosporea provide a unique opportunity for a form of "comparative embryology" that can reveal both shared and species-specific mechanisms driving multicellularity at the root of animals.

References

1. A. Sebé-Pedrós, B. M. Degnan, I. Ruiz-Trillo, The origin of Metazoa: a unicellular perspective. doi: 10.1038/nrg.2017.21 (2017).
2. T. Brunet, N. King, The Single-Celled Ancestors of Animals: A History of Hypotheses. *The Evolution of Multicellularity*, 251–278 (2022).
3. T. Brunet, N. King, The Origin of Animal Multicellularity and Cell Differentiation. [Preprint] (2017). <https://doi.org/10.1016/j.devcel.2017.09.016>.
4. R. K. Grosberg, R. R. Strathmann, The Evolution of Multicellularity: A Minor Major Transition? <https://doi.org/10.1146/annurev.ecolsys.36.102403.114735> **38**, 621–654 (2007).
5. A. H. Knoll, The Multiple Origins of Complex Multicellularity. <https://doi.org/10.1146/annurev.earth.031208.100209> **39**, 217–239 (2011).
6. I. De Smet, T. Beeckman, Asymmetric cell division in land plants and algae: the driving force for differentiation. *Nature Reviews Molecular Cell Biology* 2011 12:3 **12**, 177–188 (2011).
7. P. Gönczy, Mechanisms of asymmetric cell division: flies and worms pave the way. *Nature Reviews Molecular Cell Biology* 2008 9:5 **9**, 355–366 (2008).
8. J. A. Knoblich, Asymmetric cell division during animal development. *Nature Reviews Molecular Cell Biology* 2001 2:1 **2**, 11–20 (2001).
9. T. Lechler, M. Mapelli, Spindle positioning and its impact on vertebrate tissue architecture and cell fate. *Nature Reviews Molecular Cell Biology* 2021 22:10 **22**, 691–708 (2021).
10. S. L. Prosser, L. Pelletier, Mitotic spindle assembly in animal cells: a fine balancing act. *Nat. Rev. Mol. Cell Biol.* **18**, 187–201 (2017).
11. T. Y. C. Tsai, R. M. Garner, S. G. Megason, Adhesion-Based Self-Organization in Tissue Patterning. <https://doi.org/10.1146/annurev-cellbio-120420-100215> **38**, 349–374 (2022).
12. M. J. Wheelock, K. R. Johnson, Cadherins as Modulators of Cellular Phenotype. *Annu Rev Cell Dev Biol* **19**, 207–235 (2003).
13. T. Y. C. Tsai, R. M. Garner, S. G. Megason, Adhesion-Based Self-Organization in Tissue Patterning. *Annu Rev Cell Dev Biol* **38**, 349–374 (2022).
14. D. Wu, K. M. Yamada, S. Wang, Tissue Morphogenesis Through Dynamic Cell and Matrix Interactions. <https://doi.org/10.1146/annurev-cellbio-020223-031019> **39**, 123–144 (2023).
15. D. Arendt, J. M. Musser, C. V. H. Baker, A. Bergman, C. Cepko, D. H. Erwin, M. Pavlicev, G. Schlosser, S. Widder, M. D. Laubichler, G. P. Wagner, The origin and evolution of cell types. *Nature Reviews Genetics* 2016 17:12 **17**, 744–757 (2016).
16. G. Bell, A. O. Mooers, Size and complexity among multicellular organisms. *Biol. J. Linn. Soc.* **60**, 345–363 (1997).
17. P. Márquez-Zacarias, R. M. Pineau, M. Gomez, A. Veliz-Cuba, D. Murrugarra, W. C. Ratcliff, K. J. Niklas, Evolution of Cellular Differentiation: From Hypotheses to Models. *Trends Ecol Evol* **36**, 49–60 (2021).
18. L. Wolpert, Do we understand development? *Science* (1979) **266**, 571–572 (1994).

19. I. Salazar-Ciudad, Morphological evolution and embryonic developmental diversity in metazoa. *Development* **137**, 531–539 (2010).
20. A. T. Kalinka, P. Tomancak, The evolution of early animal embryos: conservation or divergence? *Trends Ecol Evol* **27**, 385–393 (2012).
- 5 21. I. Ruiz-Trillo, A. de Mendoza, Towards understanding the origin of animal development. *Development*, doi: 10.1242/dev.192575 (2020).
22. S. Xiao, A. D. Muscente, L. Chen, C. Zhou, J. D. Schifftbauer, A. D. Wood, N. F. Polys, X. Yuan, The Weng'an biota and the Ediacaran radiation of multicellular eukaryotes. *Natl Sci Rev* **1**, 498–520 (2014).
23. Z. Yin, W. Sun, P. Liu, J. Chen, D. J. Bottjer, J. Li, M. Zhu, Diverse and complex developmental mechanisms of early Ediacaran embryo-like fossils from the Weng'an Biota, southwest China. *Philosophical Transactions of the Royal Society B: Biological Sciences* **377** (2022).
- 10 24. Mitotic topologies and mechanics of Neoproterozoic algae and animal embryos | Paleobiology | Cambridge Core. <https://www.cambridge.org/core/journals/paleobiology/article/abs/mitotic-topologies-and-mechanics-of-neoproterozoic-algae-and-animal-embryos/C2EC4ADB10047F12E176A3370AD6B627>.
- 15 25. X. Shuhai, A. H. And, L. Knoll, S. & Xiao, A. H. Knoll, S. Xiao, S. H. Edu], Fossil preservation in the Neoproterozoic Doushantuo phosphorite Lagerstätte, South China. *Lethaia* **32**, 219–238 (1999).
26. L. Chen, S. Xiao, K. Pang, C. Zhou, X. Yuan, Cell differentiation and germ-soma separation in Ediacaran animal embryo-like fossils. *Nature* **516**, 238–241 (2014).
- 20 27. T. Hultgren, J. A. Cunningham, C. Yin, M. Stamparoni, F. Marone, P. C. J. Donoghue, S. Bengtson, Fossilized nuclei and germination structures identify ediacaran “animal embryos” as encysting protists. *Science* (1979) **334**, 1696–1699 (2011).
28. Z. Yin, K. Vargas, J. Cunningham, S. Bengtson, M. Zhu, F. Marone, P. Donoghue, The Early Ediacaran Caveasphaera Foreshadows the Evolutionary Origin of Animal-like Embryology. *Current Biology* **29**, 4307–4314.e2 (2019).
- 25 29. K. J. Niklas, S. A. Newman, The origins of multicellular organisms. *Evol Dev* **15**, 41–52 (2013).
- 30 30. N. Ros-Rocher, A. Pérez-Posada, M. M. Leger, I. Ruiz-Trillo, The origin of animals: an ancestral reconstruction of the unicellular-to-multicellular transition. *Open Biol* **11** (2021).
31. H. Liu, J. L. Steenwyk, X. Zhou, D. T. Schultz, K. M. Kocot, X.-X. Shen, A. Rokas, Y. Li, A genome-scale Opisthokonta tree of life: toward phylogenomic resolution of ancient divergences. *bioRxiv*, 2023.09.20.556338 (2023).
32. E. Ocaña-Pallarès, T. A. Williams, D. López-Escardó, A. S. Arroyo, J. S. Pathmanathan, E. Baptiste, D. V. Tikhonenkov, P. J. Keeling, G. J. Szöllösi, I. Ruiz-Trillo, Divergent genomic trajectories predate the origin of animals and fungi. *Nature* **2022** 609:7928 **609**, 747–753 (2022).
- 35 33. T. Brunet, N. King, The Origin of Animal Multicellularity and Cell Differentiation. doi: 10.1016/j.devcel.2017.09.016 (2017).
34. T. Brunet, M. Albert, W. Roman, M. C. Coyle, D. C. Spitzer, N. King, A flagellate-to-amoeboid switch in the closest living relatives of animals. *Elife* **10**, 1–73 (2021).
- 35 35. T. Brunet, B. T. Larson, T. A. Linden, M. J. A. Vermeij, K. McDonald, N. King, Light-regulated collective contractility in a multicellular choanoflagellate. *Science*, doi: 10.1126/science.aay2346 (2019).
- 40 36. S. R. Fairclough, M. J. Dayel, N. King, Multicellular development in a choanoflagellate. [Preprint] (2010). <https://doi.org/10.1016/j.cub.2010.09.014>.
37. H. Parra-Acero, M. Harcet, N. Sánchez-Pons, E. Casacuberta, N. H. Brown, O. Dudin, I. Ruiz-Trillo, Integrin-Mediated Adhesion in the Unicellular Holozoan Capsaspora owczarzaki. *Current Biology*, doi: 10.1016/j.cub.2020.08.015 (2020).
- 45 38. A. Sebé-Pedrós, M. Irimia, J. del Campo, H. Parra-Acero, C. Russ, C. Nusbaum, B. J. Blencowe, I. Ruiz-Trillo, Regulated aggregative multicellularity in a close unicellular relative of metazoa. *Elife*, doi: 10.7554/eLife.01287 (2013).
39. A. Pérez-Posada, O. Dudin, E. Ocaña-Pallarès, I. Ruiz-Trillo, A. Ondracka, Cell cycle transcriptomics of Capsaspora provides insights into the evolution of cyclin-CDK machinery. *PLoS Genet*, doi: 10.1371/journal.pgen.1008584 (2020).
- 50 40. N. Ros-Rocher, R. Q. Kidner, C. Gerdts, W. S. Davidson, I. Ruiz-Trillo, J. P. Gerdts, Chemical factors induce aggregative multicellularity in a close unicellular relative of animals. *bioRxiv*, 2022.12.01.517680 (2022).
41. A. Woznica, A. M. Cantley, C. Beemelmanns, E. Freinkman, J. Clardy, N. King, Bacterial lipids activate, synergize, and inhibit a developmental switch in choanoflagellates. *Proc Natl Acad Sci U S A*, doi: 10.1073/pnas.1605015113 (2016).
- 55

42. T. C. Levin, A. J. Greaney, L. Wetzel, N. King, The Rosetteless gene controls development in the choanoflagellate *S. rosetta*. *Elife*, doi: 10.7554/eLife.04070 (2014).
43. T. C. Levin, N. King, Evidence for Sex and Recombination in the Choanoflagellate *Salpingoeca rosetta*. *Current Biology*, doi: 10.1016/j.cub.2013.08.061 (2013).
- 5 44. M. J. Dayel, R. A. Alegado, S. R. Fairclough, T. C. Levin, S. A. Nichols, K. McDonald, N. King, Cell differentiation and morphogenesis in the colony-forming choanoflagellate *Salpingoeca rosetta*. *Dev Biol* **357**, 73–82 (2011).
45. D. Laundon, B. T. Larson, K. McDonald, N. King, P. Burkhardt, The architecture of cell differentiation in choanoflagellates and sponge choanocytes. *PLoS Biol* **17**, e3000226 (2019).
- 10 46. S. L. Glockling, W. L. Marshall, F. H. Gleason, Phylogenetic interpretations and ecological potentials of the Mesomycetozoa (Ichthyosporea). *Fungal Ecol* **6**, 237–247 (2013).
47. L. Mendoza, J. W. Taylor, L. Ajello, The class Mesomycetozoa: A heterogeneous group of microorganisms at the animal-fungal boundary. *Annu Rev Microbiol* [Preprint] (2002). <https://doi.org/10.1146/annurev.micro.56.012302.160950>.
- 15 48. A. Ondracka, O. Dudin, I. Ruiz-Trillo, Decoupling of Nuclear Division Cycles and Cell Size during the Coenocytic Growth of the Ichthyosporean *Sphaeroforma arctica*. *Current Biology* **28** (2018).
49. O. Dudin, S. Wielgoss, A. M. New, I. Ruiz-Trillo, Regulation of sedimentation rate shapes the evolution of multicellularity in a close unicellular relative of animals. *PLoS Biol* **20**, e3001551 (2022).
- 20 50. O. Dudin, A. Ondracka, X. Grau-Bové, A. A. B. Haraldsen, A. Toyoda, H. Suga, J. Bråte, I. Ruiz-Trillo, A unicellular relative of animals generates a layer of polarized cells by actomyosin-dependent cellularization. *Elife* **8** (2019).
51. L. A. Sarre, I. V. Kim, V. Ovchinnikov, M. Olivetta, H. Suga, O. Dudin, A. Sebé-Pedrós, A. de Mendoza, DNA methylation enables recurrent endogenization of giant viruses in an animal relative. *bioRxiv*, 2024.01.08.574619 (2024).
- 25 52. B. McCartney, O. Dudin, Cellularization across eukaryotes: Conserved mechanisms and novel strategies. *Curr Opin Cell Biol* **80**, 102157 (2023).
53. H. Shah, M. Olivetta, C. Bhickta, P. Ronchi, M. Trupinić, E. C. Tromer, I. M. Tolić, Y. Schwab, O. Dudin, G. Dey, Life cycle-coupled evolution of mitosis in close relatives of animals. *bioRxiv*, 2023.05.10.540163 (2023).
- 30 54. M. Olivetta, O. Dudin, The nuclear-to-cytoplasmic ratio drives cellularization in the close animal relative *Sphaeroforma arctica*. *Current Biology* **33**, 1597–1605.e3 (2023).
55. X. Grau-Bové, G. Torruella, S. Donachie, H. Suga, G. Leonard, T. A. Richards, I. Ruiz-Trillo, Dynamics of genomic innovation in the unicellular ancestry of animals. *Elife* **6** (2017).
- 35 56. M. Binyam, K. Scheffler, M. Schuh, Assembly and Positioning of the Oocyte Meiotic Spindle. <https://doi.org/10.1146/annurev-cellbio-100616-060553> **34**, 381–403 (2018).
57. S. Vasudevan, E. Seli, J. A. Steitz, Metazoan oocyte and early embryo development program: a progression through translation regulatory cascades. *Genes Dev* **20**, 138–146 (2006).
58. M. Levin, L. Anavy, A. G. Cole, E. Winter, N. Mostov, S. Khair, N. Senderovich, E. Kovalev, D. H. Silver, M. Feder, S. L. Fernandez-Valverde, N. Nakanishi, D. Simmons, O. Simakov, T. Larsson, S.-Y. Liu, A. Jeraffi-Vider, K. Yaniv, J. F. Ryan, M. Q. Martindale, J. C. Rink, D. Arendt, S. M. Degnan, B. M. Degnan, T. Hashimshony, I. Yanai, The mid-developmental transition and the evolution of animal body plans. doi: 10.1038/nature16994 (2016).
- 40 59. N. Klena, G. Pigino, Structural Biology of Cilia and Intraflagellar Transport. <https://doi.org/10.1146/annurev-cellbio-120219-034238> **38**, 103–123 (2022).
- 45 60. Z. Carvalho-Santos, J. Azimzadeh, J. B. Pereira-Leal, M. Bettencourt-Dias, Tracing the origins of centrioles, cilia, and flagella. *Journal of Cell Biology* **194**, 165–175 (2011).
61. A. De Mendoza, A. Sebé-Pedrós, I. Ruiz-Trillo, The Evolution of the GPCR Signaling System in Eukaryotes: Modularity, Conservation, and the Transition to Metazoan Multicellularity. *Genome Biol Evol* **6**, 606 (2014).
62. T. Brunet, D. S. Booth, Cell polarity in the protist-to-animal transition. *Curr Top Dev Biol* **154**, 1–36 (2023).
- 50 63. B. A. Wright, M. Kvansakul, B. Schierwater, P. O. Humbert, Cell polarity signalling at the birth of multicellularity: What can we learn from the first animals. *Front Cell Dev Biol* **10**, 1024489 (2022).
64. B. Klapholz, N. H. Brown, Talin - the master of integrin adhesions. *J Cell Sci* **130**, 2435–2446 (2017).
65. J. L. Bays, K. A. DeMali, Vinculin in cell–cell and cell–matrix adhesions. *Cellular and Molecular Life Sciences* **74**, 2999–3009 (2017).
- 55 66. B. M. Gumbiner, Cell Adhesion: The Molecular Basis of Tissue Architecture and Morphogenesis. *Cell* **84**, 345–357 (1996).

67. J. H. Shawky, L. A. Davidson, Tissue Mechanics and Adhesion during Embryo Development. *Dev Biol* **401**, 152 (2015).
68. A. Sebé-Pedrós, A. De Mendoza, B. F. Lang, B. M. Degnan, I. Ruiz-Trillo, Unexpected repertoire of metazoan transcription factors in the unicellular holozoan capsaspora owczarzaki. *Mol Biol Evol*, doi: 10.1093/molbev/msq309 (2011).
69. A. De Mendoza, A. Sebé-Pedrós, M. S. Šestak, M. Matejčić, G. Torruella, T. Domazet-Lošo, I. Ruiz-Trillo, Transcription factor evolution in eukaryotes and the assembly of the regulatory toolkit in multicellular lineages. *Proc Natl Acad Sci U S A*, doi: 10.1073/pnas.1311818110 (2013).
70. A. Sebé-Pedrós, I. Ruiz-Trillo, “Evolution and Classification of the T-Box Transcription Factor Family” in *Current Topics in Developmental Biology* (2017).
71. E. G. Leigh, J. M. Smith, E. Szathmari, The Major Transitions of Evolution. *Evolution (N Y)*, doi: 10.2307/2410462 (1995).
72. A. Sebé-Pedrós, B. M. Degnan, I. Ruiz-Trillo, The origin of Metazoa: A unicellular perspective. [Preprint] (2017). <https://doi.org/10.1038/nrg.2017.21>.
73. D. Gambarotto, F. U. Zwettler, M. Le Guennec, M. Schmidt-Cernohorska, D. Fortun, S. Borgers, J. Heine, J. G. Schloetel, M. Reuss, M. Unser, E. S. Boyden, M. Sauer, V. Hamel, P. Guichard, Imaging cellular ultrastructures using expansion microscopy (U-ExM). *Nature Methods* **16**, 71–74 (2018).
74. D. Gambarotto, V. Hamel, P. Guichard, Ultrastructure expansion microscopy (U-ExM). *Methods Cell Biol* **161**, 57–81 (2021).
75. S. Machado, V. Mercier, N. Chiaruttini, LimeSeg: A coarse-grained lipid membrane simulation for 3D image segmentation. *BMC Bioinformatics* **20**, 1–12 (2019).
76. J. Schindelin, I. Arganda-Carreras, E. Frise, V. Kaynig, M. Longair, T. Pietzsch, S. Preibisch, C. Rueden, S. Saalfeld, B. Schmid, J. Y. Tinevez, D. J. White, V. Hartenstein, K. Eliceiri, P. Tomancak, A. Cardona, Fiji: an open-source platform for biological-image analysis. *Nature Methods* **9**, 676–682 (2012).
77. Multicellgenome Lab. https://figshare.com/authors/Multicellgenome_Lab/2628379.
78. R. Patro, G. Duggal, M. I. Love, R. A. Irizarry, C. Kingsford, Salmon provides fast and bias-aware quantification of transcript expression. *Nature Methods* **14**, 417–419 (2017).
79. R: The R Project for Statistical Computing. <https://www.r-project.org/>.
80. C. Soneson, M. I. Love, M. D. Robinson, Differential analyses for RNA-seq: Transcript-level estimates improve gene-level inferences. *F1000Res* **4** (2016).
81. M. D. Robinson, D. J. McCarthy, G. K. Smyth, edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140 (2010).
82. D. J. McCarthy, Y. Chen, G. K. Smyth, Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation. *Nucleic Acids Res* **40**, 4288 (2012).
83. Y. Chen, A. T. L. Lun, G. K. Smyth, From reads to genes to pathways: Differential expression analysis of RNA-Seq experiments using Rsubread and the edgeR quasi-likelihood pipeline. *F1000Res* **5** (2016).
84. e1071: Misc Functions of the Department of Statistics, Probability Theory Group (Formerly: E1071), TU Wien version 1.7-15 from R-Forge. <https://rdrr.io/rforge/e1071/>.
85. viridis: Colorblind-Friendly Color Maps for R version 0.6.4 from CRAN. <https://rdrr.io/cran/viridis/>.
86. superheat: A Graphical Tool for Exploring Complex Datasets Using Heatmaps version 0.1.0 from CRAN. <https://rdrr.io/cran/superheat/>.
87. C. Camacho, G. Coulouris, V. Avagyan, N. Ma, J. Papadopoulos, K. Bealer, T. L. Madden, BLAST+: Architecture and applications. *BMC Bioinformatics* **10**, 1–9 (2009).
88. E. Boutet, D. Lieberherr, M. Tognolli, M. Schneider, P. Bansal, A. J. Bridge, S. Poux, L. Bougueleret, I. Xenarios, UniProtKB/Swiss-Prot, the Manually Annotated Section of the UniProt KnowledgeBase: How to Use the Entry View. *Methods Mol Biol* **1374**, 23–54 (2016).
89. B. Rost, Enzyme function less conserved than anticipated. *J Mol Biol* **318**, 595–608 (2002).
90. J. A. Burns, A. Paasch, A. Narechania, E. Kim, Comparative Genomics of a Bacterivorous Green Alga Reveals Evolutionary Causalities and Consequences of Phago-Mixotrophic Mode of Nutrition. *Genome Biol Evol* **7**, 3047–3061 (2015).
91. topGO: Enrichment Analysis for Gene Ontology version 2.42.0 from Bioconductor. <https://rdrr.io/bioc/topGO/>.
92. D. M. Emms, S. Kelly, OrthoFinder: Phylogenetic orthology inference for comparative genomics. *Genome Biol* **20**, 1–14 (2019).
93. GitHub - TransDecoder/TransDecoder: TransDecoder source. <https://github.com/TransDecoder/TransDecoder>.

94. S. R. Engel, E. D. Wong, R. S. Nash, S. Aleksander, M. Alexander, E. Douglass, K. Karra, S. R. Miyasato, M. Simison, M. S. Skrzypek, S. Weng, J. M. Cherry, New data and collaborations at the Saccharomyces Genome Database: updated reference genome, alleles, and the Alliance of Genome Resources. *Genetics* **220** (2022).
95. V. Wood, R. Gwilliam, M. A. Rajandream, M. Lyne, R. Lyne, A. Stewart, J. Sgouros, N. Peat, J. Hayles, S. Baker, D. Basham, S. Bowman, K. Brooks, D. Brown, S. Brown, T. Chillingworth, C. Churcher, M. Collins, R. Connor, A. Cronin, P. Davis, T. Feltwell, A. Fraser, S. Gentles, A. Goble, N. Hamlin, D. Harris, J. Hidalgo, G. Hodgson, S. Holroyd, T. Hornsby, S. Howarth, E. J. Huckle, S. Hunt, K. Jagels, K. James, L. Jones, M. Jones, S. Leather, S. McDonald, J. McLean, P. Mooney, S. Moule, K. Mungall, L. Murphy, D. Niblett, C. Odell, K. Oliver, S. O'Neil, D. Pearson, M. A. Quail, E. Rabinowitsch, K. Rutherford, S. Rutter, D. Saunders, K. Seeger, S. Sharp, J. Skelton, M. Simmonds, R. Squares, S. Squares, K. Stevens, K. Taylor, R. G. Taylor, A. Tivey, S. Walsh, T. Warren, S. Whitehead, J. Woodward, G. Volckaert, R. Aert, J. Robben, B. Grymonprez, I. Weltjens, E. Vanstreels, M. Rieger, M. Schäfer, S. Müller-Auer, C. Gabel, M. Fuchs, C. Fritze, E. Holzer, D. Moestl, H. Hilbert, K. Borzym, I. Langer, A. Beck, H. Lehrach, R. Reinhardt, T. M. Pohl, P. Eger, W. Zimmermann, H. Wedler, R. Wambutt, B. Purnelle, A. Goffeau, E. Cadieu, S. Dréano, S. Gloux, V. Lelaure, S. Mottier, F. Galibert, S. J. Aves, Z. Xiang, C. Hunt, K. Moore, S. M. Hurst, M. Lucas, M. Rochet, C. Gaillardin, V. A. Tallada, A. Garzon, G. Thode, R. R. Daga, L. Cruzado, J. Jimenez, M. Sánchez, F. del Rey, J. Benito, A. Domínguez, J. L. Revuelta, S. Moreno, J. Armstrong, S. L. Forsburg, L. Cerrutti, T. Lowe, W. R. McCombie, I. Paulsen, J. Potashkin, G. V. Shpakovski, D. Ussery, B. G. Barrell, P. Nurse, The genome sequence of *Schizosaccharomyces pombe*. *Nature* **2002** *415*:6874 **415**, 871–880 (2002).
96. I. Eichinger, J. A. Pachebat, G. Glöckner, M. A. Rajandream, R. Sugang, M. Berriman, J. Song, R. Olsen, K. Szafranski, Q. Xu, B. Tunggal, S. Kummerfeld, M. Madera, B. A. Konfortov, F. Rivero, A. T. Bankier, R. Lehmann, N. Hamlin, R. Davies, P. Gaudet, P. Fey, K. Pilcher, G. Chen, D. Saunders, E. Sodergren, P. Davis, A. Kerhornou, X. Nie, N. Hall, C. Anjard, L. Hemphill, N. Bason, P. Farbrother, B. Desany, E. Just, T. Morio, R. Rost, C. Churcher, J. Cooper, S. Haydock, N. Van Driessche, A. Cronin, I. Goodhead, D. Muzny, T. Mourier, A. Pain, M. Lu, D. Harper, R. Lindsay, H. Hauser, K. James, M. Quiles, M. Madan Babu, T. Saito, C. Buchrieser, A. Wardroper, M. Felder, M. Thangavelu, D. Johnson, A. Knights, H. Loulseged, K. Mungall, K. Oliver, C. Price, M. A. Quail, H. Urushihara, J. Hernandez, E. Rabinowitsch, D. Steffen, M. Sanders, J. Ma, Y. Kohara, S. Sharp, M. Simmonds, S. Spiegler, A. Tivey, S. Sugano, B. White, D. Walker, J. Woodward, T. Winckler, Y. Tanaka, G. Shaulsky, M. Schleicher, G. Weinstock, A. Rosenthal, E. C. Cox, R. L. Chisholm, R. Gibbs, W. F. Loomis, M. Platzer, R. R. Kay, J. Williams, P. H. Dear, A. A. Noegel, B. Barrell, A. Kuspa, The genome of the social amoeba *Dictyostelium discoideum*. *Nature* **2005** *435*:7038 **435**, 43–57 (2005).
97. K. Katoh, D. M. Standley, MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol* **30**, 772–780 (2013).
98. HMMER. <http://hmmer.org/>.

Acknowledgments: Pierre Gönczy, Gautam Dey, Alex de Mendoza, Chema Martín-Durán, Hiral Shah, Alexander Woglar for comments on manuscript and general feedback; Hiroshi Suga for *C. perkinsii* cultures; and both the EPFL BioImaging and Optics Platform (BIOP) and the EPFL Gene Expression Core Facility for their support.

Funding:

U.S. National Science Foundation Grant (OIA-1826734) (JB).

Swiss National Science Foundation Ambizione Grant (PZ00P3_185859) (OD, MO).

Author contributions:

Conceptualization: OD

Methodology: MO, CB, NC, JB, OD

Investigation: MO, CB, NC, JB, OD

Visualization: MO, NC, JB, OD

Funding acquisition: JB, OD

Project administration: OD

Supervision: OD

Writing – original draft: OD

Writing – review & editing: MO, NC, JB, OD

Competing interests: Authors declare that they have no competing interests.

Data and materials availability: All data used for quantifications as well as movies are available on figshare (10.6084/m9.figshare.25351078) with the raw images available upon request.

The generated code for transcriptomic analysis is accessible on github

https://github.com/burnsajohn/Chromosphaera-perkinsii_Transcriptomics and raw RNAseq data are on <https://www.ncbi.nlm.nih.gov/sra/PRJNA1091032>. Note that the data and code archives are embargoed while the manuscript is under review.

Supplementary Materials

Materials and Methods

Culture conditions

C. perkinsii cultures, previously described in (53, 55), were maintained at 23°C in CpM medium (Yeast extract 3g/L, Malt extract 3g/L, Peptone 5g/L, Glucose 10g/L, NaCl 20g/L) protected from light. Cultures were either streaked onto CpM Agar plates (2% Agar) or grown into liquid CpM medium in rectangular canted neck cell culture flask with vented cap (Falcon; ref: 353108). For maintenance, liquid cultures are propagated monthly from single colonies growing on CpM Agar, and always protected from light. Liquid cultures were refreshed every two weeks (1:1000) dilution and restarted from a cryopreserved stock every 6 months. For synchronization, 10ml of a 6-day-old culture is filtered using a 5µm filter to obtain a homogenous small-celled population which is then diluted 1:100 in CpM medium and tracked for an entire life-cycle (120hrs).

Microscopy

Imaging of live and fixed *C. perkinsii* cells was conducted utilizing a fully motorized Nikon Ti2-E epifluorescence inverted microscope, which was equipped with a hardware autofocus PFS4 system, a Lumencor SOLA SMII illumination system, and a Hamamatsu ORCA-spark Digital CMOS camera. Imaging was performed using a CFI Plan Fluor 20X objective with a numerical aperture of 0.50, a CFI Plan Fluor 40X Air objective, and a CFI Plan Fluor 60X Oil objective with a numerical aperture ranging from 0.5 to 1.25.

To maintain the temperature at a consistent 23°C, a cooling/heating P Lab-Tek S1 insert (Pecon GmbH) connected to a Lauda Loop 100 circulating water bath was employed. To minimize light-induced phototoxicity, the light source beam was manually adjusted to its lowest setting, and a 715nm LongPass Color Filter (Thorlabs; ref: FGL715S) was utilized to selectively allow far-red wavelengths to pass through. Moreover, we meticulously controlled the imaging frequency and the number of Z-stacks during long-term imaging to prevent cell lysis. Our observations revealed that an imaging interval of 15 minutes, combined with 7 Z-slices over 24 hours, represented the upper limit for optimal imaging conditions, beyond which most cells exhibited abnormal behaviour and lysis. For confocal imaging of membrane-stained, phalloidin-stained and expanded samples, an upright Leica SP8 confocal microscope with an HC PL APO 40X/1.25 Glycerol objective was used.

Cell fixation and staining

Cell fixation was carried out either using 4% formaldehyde and 250mM Sorbitol for 20 minutes or using pre-chilled Methanol (100%) at -20°C for 7 minutes, followed by two PBS washes. To stain the nuclei, cultures were allowed to settle for 15 minutes at room temperature (RT) before fixation, and Hoechst 33342 nuclear stain (ThermoFisher; 62249) was added at a concentration of 20mM. For actin staining, fixed cells underwent a single PBS wash before adding the F-actin stain Alexa Fluor 488-Phalloidin (ThermoFisher; A12379) at a final concentration of 0.165mM. Before imaging, the fixed and stained cells were concentrated and placed between a slide and coverslip. For live-cell imaging, saturated cultures were diluted 500X in CpM medium inside an 8-well ibidi chamber (Ibidi; ref:80826) or a 35mm dish (Mattek; P35G-1.5-14-C). The plastic cover was removed to ensure oxygenation during the whole experiment period. To reduce evaporation, Silicon oil 100 cSt was added on top (Sigma; 378364). For plasma membrane live staining, FM4-

64 (ThermoFisher; T13320) at a final concentration of 10mM from 100X DMSO diluted stock solution was directly added to the medium before imaging.

Ultrastructural Expansion Microscopy (U-ExM)

U-ExM was adapted from (53, 73, 74). Initially, cells were fixed in a 4% formaldehyde solution with 250mM Sorbitol for 10 minutes. Following fixation, they underwent two washes with 1X PBS and were subsequently resuspended in 1ml PBS. Coverslips were pre-coated with Poly-l-lysine for an hour, and the fixed cells were then added and allowed to adhere for 1 hour. Anchoring in AA/FA (1% Acrylamide (AA)/ 0.7% Formaldehyde (FA)) was performed for 12 hrs at 37°C. Gelation relied on a monomer solution (19% (wt/wt) sodium acrylate (Chem Cruz, AKSci ref: 7446-81-3), 10% (wt/wt) Acrylamide (Sigma-Aldrich; A4058), 0.1% (wt/wt) N, N'-methylenbisacrylamide (Sigma-Aldrich; M1533) in PBS and was performed at 37°C for 1 hr in a moist chamber. For denaturation, gels were transferred to the denaturation buffer for 15 min at RT and then shifted to 95°C for 1.5 hours. Following denaturation, expansion was performed with water exchanges as previously described(53, 73). Post expansion, gel diameter was measured and used to determine the expansion factor. For microtubule staining of U-ExM gels, mouse primary antibodies targeting beta-tubulin (DSHB; E7) (1/300 dilution) or acetylated-tubulin (Sigma; T6793) and Goat anti-Mouse IgG (H+L) Cross-Adsorbed Secondary Antibodies coupled to Alexa Fluor 488 (Thermofisher; A-11001) (1/500 dilution) were used. For protein pan-labelling, gels were incubated with Alexa Fluor 568 NHS-Ester (Thermofisher; A20003) in NaHCO₃ (pH=8.28) for 1.5 hrs. For nuclear staining, Hoechst 33352 was incubated in PBS for 10 minutes before re-expansion in water. For gel mounting, gels were cut to appropriate sizes and attached to pre-coated Poly-l-lysine coverslips and sealed using i-Spacers (Sunjin Lab; #IS013).

Image analysis

Image analysis was done using ImageJ software (version 1.52) and Imaris (version 10.0.1) (Bitplane). For cell and nuclear diameter measurements (Fig. 1D and G, S1E and S1F), brightfield (live-imaging) and immunostained images were transformed into a binary before using the particle analysis function in ImageJ with a circularity parameter set to 0.15 to 1 to measure the perimeter. As cells are spherical, we computed the cell diameter as $D = 2(C/2\pi)$. For nuclear content distribution (Fig. 1F and S1G), fixed and Hoechst-stained cells and colonies were imaged and the number of nuclei per cell was either counted using ObjectJ Plugin in ImageJ, or by segmenting the nuclei using the Imaris “surface” tool when the number of nuclei was difficult to count manually. For nuclear position measurements (Fig. 3D) nuclear-stained and U-ExM cells were first oriented with the nucleus positioned at the top. Using the line tool in imageJ we measured the distance between the cortex, either from the bottom (y1) or from the top (y2), and the center of the metaphase plate. For volume ratio measurements of the 2-cell stage (Fig. 3E), we used LimeSeg (75) in ImageJ/Fiji (76) for semi-automated segmentation of nuclei and phalloidin-stained cells. Briefly, after manual visual inspection, a spherical seed was manually added to the ROI manager for each detected nucleus. These spherical seeds were then inflated with LimeSeg, using two main parameters: 0.8 microns between each surface element and applying a pressure of 0.009 a.u. on the phalloidin channel. Typically, segmentation convergence to the 3D cell shape was achieved in a matter of seconds (Fig. S3D). For spindle orientation measurements (Fig. 3F), U-ExM cells stained for microtubules where the entirety of the spindle, including both spindle poles, were visible were used. The angles were then measured between the plane of the first cleavage and the plane defined by the positions of both spindle poles. To quantify the number of flagellated cells released in the environment (Fig. 4C), we used brightfield or U-ExM images and counted the single cells harbouring a flagellum that were not localized within any multicellular colonies. For the

analysis of flagellated cells within multicellular colonies (Fig. 4D and 4E), U-ExM cells stained for Microtubules or flagella (Acetylated Tubulin) and DNA (Hoechst) were examined. Initially, we measured nuclear content using the surface tool in Imaris. Subsequently, flagellated cells were counted using ObjectJ in ImageJ or the surface tool in Imaris, depending on the number of flagellated cells. In instances where there were more than 150 flagellated cells per colony, a manual approach combining the skeletonize tool and ObjectJ in ImageJ was employed. All figures were assembled with Illustrator CC 2020 (Adobe). Several figures were generated using ggplot2 in R version 4.0.5.53.

RNA isolation, library preparation and sequencing

Synchronized cultures of *C. perkinsii* at 23°C protected from light were sampled at critical time points throughout its entire life cycle (120 hrs). Total RNA was extracted by Trizol and purified using RNA Clean & Concentrator™-5 kit (Zymo; ref: R1013) from 300 to 600ml of culture at each time point. Libraries were prepared using 400ng of RNA using the Illumina stranded mRNA ligation prep (ISML). Libraries were sequenced on NovaS6000 in a PE60 at the EPFL gene expression core facility. We obtained between 51.2 and 104.5 M reads per sample. Reads were trimmed for ISML (Nextera) adapters.

Gene expression analysis in *C. perkinsii*

RNAseq reads were mapped to the *C. perkinsii* transcriptome (77) using Salmon (v1.1.0) (78) in mapping-based mode and the full genome (55) to generate decoys for accurate read mapping. Salmon quant.sf files were read into the R programming environment (v4.3.1) (79) using the tximport (v1.30.0)(80) tool. Differential expression analyses within *C. perkinsii* were conducted with edgeR (v4.0.3) (81–83). RNA was collected and sequenced for three replicate cultures, however upon examination of the data, one replicate was discarded due to anomalous expression of all time points from that culture as observed in a multiple dimensional scaling plot (fig. S2A), and examination of expression patterns from that culture indicative of a possible infection or contamination of the culture (Table S1). The remaining two replicates (A and C, fig. S2A) were examined for differential gene expression across time by considering differential expression of each time point relative to expression in the first time point using genewise negative binomial generalized linear models with quasi-likelihood tests (glmQLFTest) (83) in edgeR. Differentially expressed genes across all time points (false discovery rate (FDR) adjusted p-value < 0.05) were collected for further analyses of expression patterns. Expression patterns were extracted from the differentially expressed genes by fuzzy c-means clustering in R from package e1071 (v1.7.14) (84). Considering within-cluster sum of squares for different values of k-clusters, 5 clusters were chosen as the optimal cluster number. Gene expression patterns across time were plotted using the viridis (v0.6.4) (85) cividis palette and superheat (v1.0.0) plotting functions (86).

GO functional enrichment analyses

Functional annotation of the *C. perkinsii* transcriptome was conducted by BLASTP (v2.10.0+) (87) analysis of *C. perkinsii* predicted peptides against the UniProtKB/Swiss-Prot expertly curated protein database (downloaded Jun 16, 2021, containing 565,254 entries) (88). BLASTP hits were filtered by best “homology-derived structure of proteins” (HSSP) distance score (89) as in Burns et al. 2016 (90). The GO annotations for best Swiss-Prot hits to each *C. perkinsii* peptide were converted to a .goa format file and used for functional enrichment analyses of expression clusters with topGO (v2.54.0) (91).

Comparison of *C. perkinsii* gene expression to *S. arctica* and 10 animal developmental time courses.

To compare gene expression patterns across evolutionary divides, we used OrthoFinder (v 2.5.4) (92) to cluster genes into orthogroups across ten animals found in Levin et al. 2016 (58): *Caenorhabditis elegans*; *Platynereis dumerilii*; *Drosophila melanogaster*; *Strongylocentrotus purpuratus*; *Danio rerio*; *Mnemiopsis leidyi*; *Nematostella vectensis*; *Schmidtea polychroa*; *Hypsibius dujardini*; *Amphimedon queenslandica*; as well as *C. perkinsii*, and *S. arctica* (50), and four additional ichthyosporeans: *Amoebidium parasiticum*, *Creolimax fragrantissima*, *Ichthyophonus hoferi*, *Sphaerothecum destruens*; one Corallochytrean: *Corallochytrium limacisporum*; two fungi: *Saccharomyces cerevisiae*, and *Schizosaccharomyces pombe*; and one amoebozoan: *Dictyostelium discoideum*. Transcriptomes and read count data for the ten animals from Levin et al. 2016 (58) were downloaded from the associated NCBI GEO database (accession number GSE70185). Transcriptomes were translated using TransDecoder (v5.5.0) (93). Peptide files for additional ichthyosporeans and the corallochytrean were downloaded from the Ruiz-Trillo Multicellgenome lab resources on FigShare (77). The yeast and amoebozoan predicted peptide data (outgroups) were downloaded from their NCBI genome pages (94–96). Orthogroups from the predicted peptides of the 20 organisms were predicted using the tool OrthoFinder (92).

To compare developmental transcriptional patterns among animals and *C. perkinsii*, we selected timepoints from animal development for each animal that went from the one or two cell stage through the blastula stage, as detailed in Table S4, based on the developmental time courses detailed in Levin et al. 2016 (58). We used sliding windows to create an equivalent number of timepoints for each animal's development, reducing the number of timepoints to match those of *C. perkinsii*. We used the first 6 timepoints of *C. perkinsii* development for comparison with animals (54h, 72h, 84h, 90h, 96h, 104h). The final timepoint (120h), aligning with cell release, was considered to be unique to *C. perkinsii* development and not likely shared with multicellular animals whose multicellular developmental programs continue beyond the blastula stage.

Orthogroups were used for comparing expression patterns across animals. Given that an orthogroup may encompass multiple genes, each exhibiting unique expression patterns, we employed network analysis to segregate these genes based on their expression profiles before conducting cross-species comparisons (fig. S2F). Specifically, we calculated the Euclidean distance between all gene pairs within an orthogroup to quantify expression dissimilarities. This analysis facilitated the construction of a graph where genes, represented as nodes, were interconnected by edges if their expression distance fell below a predefined threshold (threshold of 1.61), thereby indicating similarity. Conversely, genes with distances surpassing this threshold remained unlinked. Subsequently, isolated nodes were eliminated from the graph, which was then subjected to a connected component analysis using the igraph package (version 1.6.0) in R. This step identified clusters of three or more interconnected genes from within or across organisms as distinct expression patterns within the orthogroup, which were then extracted for further analyses. In instances where an organism lacked a gene corresponding to a specific expression pattern within a connected component yet possessed other genes from the same orthogroup, the expression values of these genes were averaged to represent that organism's contribution to the orthogroup's overall expression profile. The network procedure allowed analyses of all distinct expression patterns per orthogroup across animals and is illustrated in Fig. 2D and S2H.

Functional annotation of orthogroups was achieved by 1) aligning all proteins within an orthogroup using mafft (v7.487) (97); 2) converting that alignment to a hidden markov model using hmmer (v3.3.2) (98), and emitting a strict consensus sequence from that hidden markov model. Then 3) using BLASTP of each consensus sequence against the UniProtKB/Swiss-Prot

database (downloaded Jun 16, 2021, containing 565,254 entries) (88), and collecting the best hit by hssp score as above for *C. perkinsii* genes. The UniProtKB/Swiss-Prot gene names and annotations were transferred to the orthogroups in this way and used for functional inference and GO analyses.

Figs. S1 to S3

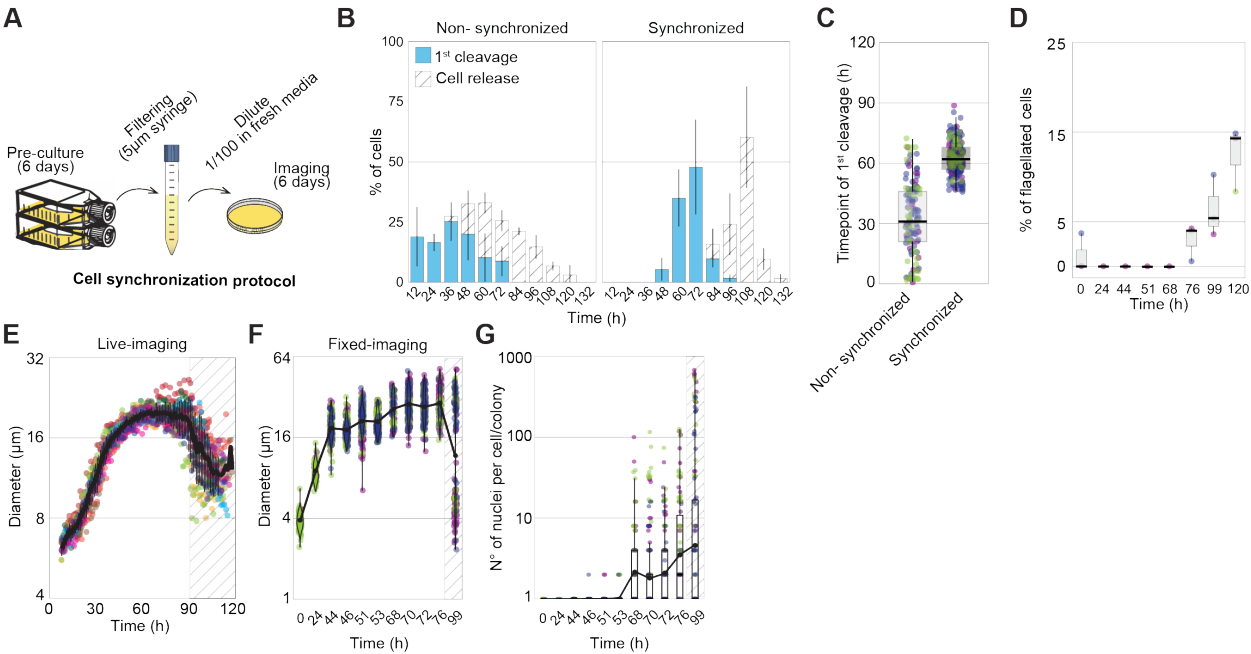


Fig. S1: Culture synchronization and growth dynamics of *C. perkinsii*.

(A) Schematic of the synchronization protocol of *C. perkinsii*.

(B) Distribution of cells undergoing their first cleavage or cell release throughout the life cycle at 23°C, following either synchronization or not ($n > 129$ cell/colony per time point).

(C) The average time at which non-synchronized and synchronized cells undergo their 1st cleavage in a bulk culture ($n > 129$ cell/colony per time point).

(D) Percentage of free-swimming flagellated cells over time ($n > 67$ cells/time point).

(E) Cell diameter over time of 10 single cell traces aligned to time. Variability increases from 75 hrs onwards due to asynchronous cell release.

(F) Cell diameter over time of fixed cells ($n = 1571$ cell/colony). The colours represent 3 independent replicates.

(G) Distribution of number of nuclei per cell/colony across the life-cycle ($n = 1571$ cell/colony). The colours represent 3 independent replicates.

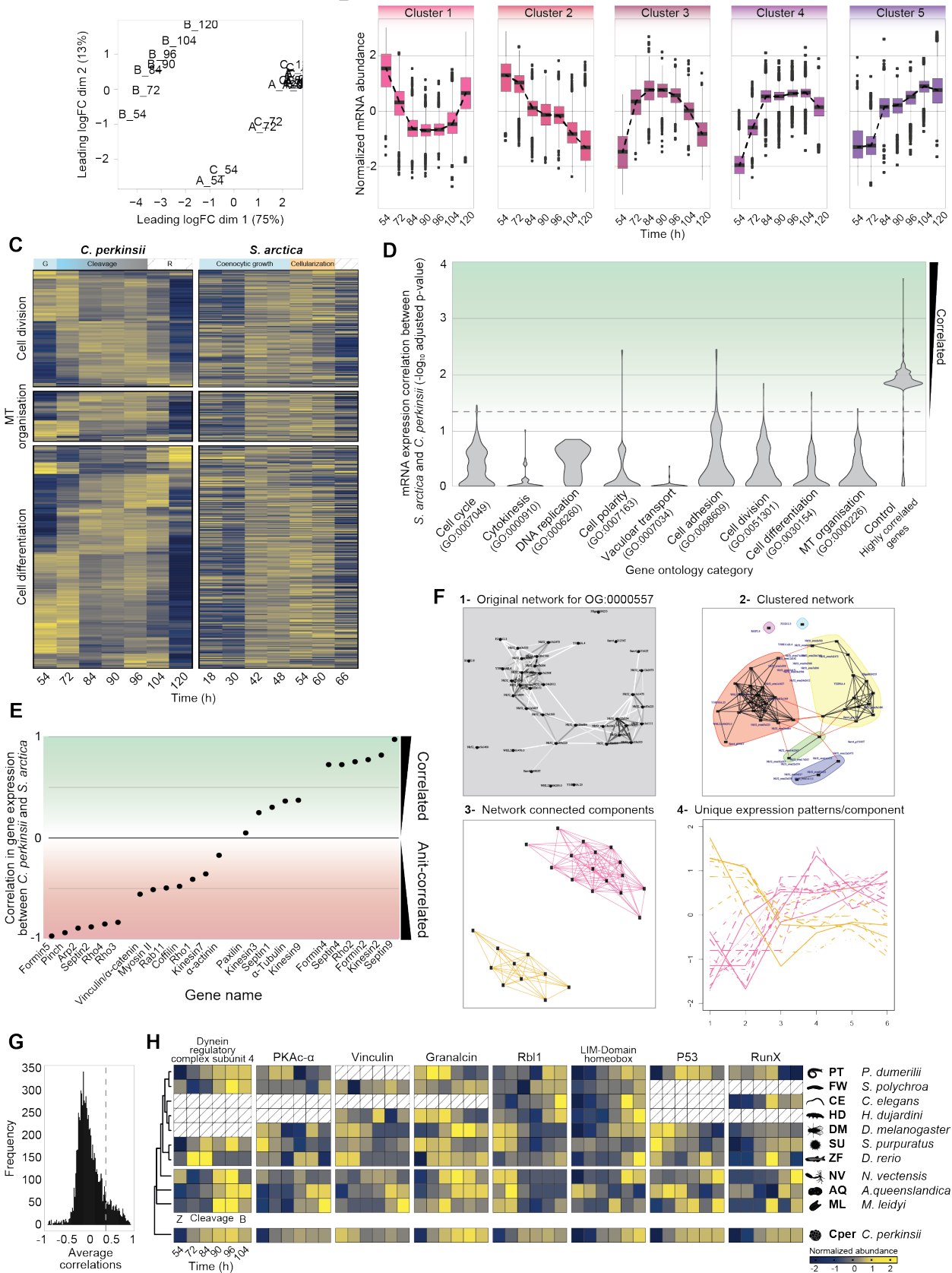


Fig. S2: Transcriptional dynamics in *C. perkinsii*'s palintomic lifecycle and comparative analysis with *S. arctica* and early-branching animals.

(A) Multiple dimensional scaling (MDS) plot of RNAseq counts of three biological replicate developmental time course experiments with *C. perkinsii*. Timepoints for replicates A and C clustered together, while the entire time course of replicate B was set apart indicating markedly different RNAseq counts in all timepoints of replicate B. Differential expression analyses between B, and A and C replicates indicated evidence of an infection in the “B” culture (see Table S1). “B” replicates were discarded for further analyses due to the evidence of infection and outlier expression values.

(B) Mean expression profile for the 5 distinct gene expression clusters for both replicates.

(C) A comparative heatmap of gene expression dynamics between *C. perkinsii* and *S. arctica* for key GO categories, including cell division (GO:0051301), microtubule cytoskeleton organisation (GO:0000226), and cell differentiation (GO:0030154). It highlights the distinct transcriptional programs between both ichthyosporean species.

(D) Statistical evaluation of rowwise correlations between *C. perkinsii* and *S. arctica* for each GO category. For each row, a correlation test was run and the p-values were collected and corrected for multiple testing using the Benjamini-Hochberg procedure. The distribution of adjusted p-values is plotted for each GO category. The horizontal line represents a p-value of 0.05. Genes with a p-value above the $p=0.05$ line have significantly correlated expression patterns between *C. perkinsii* and *S. arctica*. The plot shows that few genes are correlated for the indicated GO categories. The final category “Control Highly Correlated Genes” shows the p-value distribution when the 300 genes with the highest correlation values between *C. perkinsii* and *S. arctica* are subjected to this analysis—it is the expected distribution if any GO category had genes with highly correlated expression patterns between the two organisms.

(E) Correlation values of *C. perkinsii* compared to *S. arctica* expression patterns for genes that were determined to be important during the life cycle/development of *S. arctica* in Dudin et al, 2019 (50). Values range from strongly anti-correlated to strongly correlated with no discernable functional coherence.

(F) Demonstration of the network strategy to decouple distinct expression patterns that occur within an orthogroup, across animals. 1) For each OG, a network is constructed with all genes from all organisms within that OG represented as nodes, and edges representing the distance between their developmental expression vectors. 2) The network is subjected to edge-betweenness clustering and only those clusters containing three or more nodes, each with at least three connections, are retained. 3) The resultant clusters can be represented as network connected components. 4) Expression patterns for each connected component are similar within a connected component, but different between connected components indicating separation of distinct expression patterns within an orthogroup. The distinct expression patterns are analyzed separately for each OG across animals for detection of functions with shared gene expression. The procedure of separating distinct expression patterns per OG is more sensitive than the alternative of averaging gene expression within an OG for each animal.

(G) The distribution of average correlation values of all pairwise correlations between three basal animals (*Mnemiopsis leidyi*, *Nematostella vectensis*, *Amphimedon queenslandica*) and *C. perkinsii* for each orthogroup (gene). Orthogroups with an average correlation value greater than 0.5 (vertical dashed line) were subjected to GO enrichment analyses to determine which biological processes are enriched among genes with highly correlated expression in the development of basal animals and *C. perkinsii*.

(H) A comparative heatmap of gene expression for key regulators involved in flagellar motility (Dynein regulatory complex subunit 4), cell signalling (PKAc- α), adhesion (Vinculin, Grancalcin), and transcriptional regulation of animal development (Rbl1, LIM-Domain homeobox, P53, RunX) dynamics between the *C. perkinsii* (Cper) life cycle and early embryonic development stages of early-branching animals. This comparison spans from the onset of zygote formation (Z) through various cleavage stages and up to the blastula stage (B).

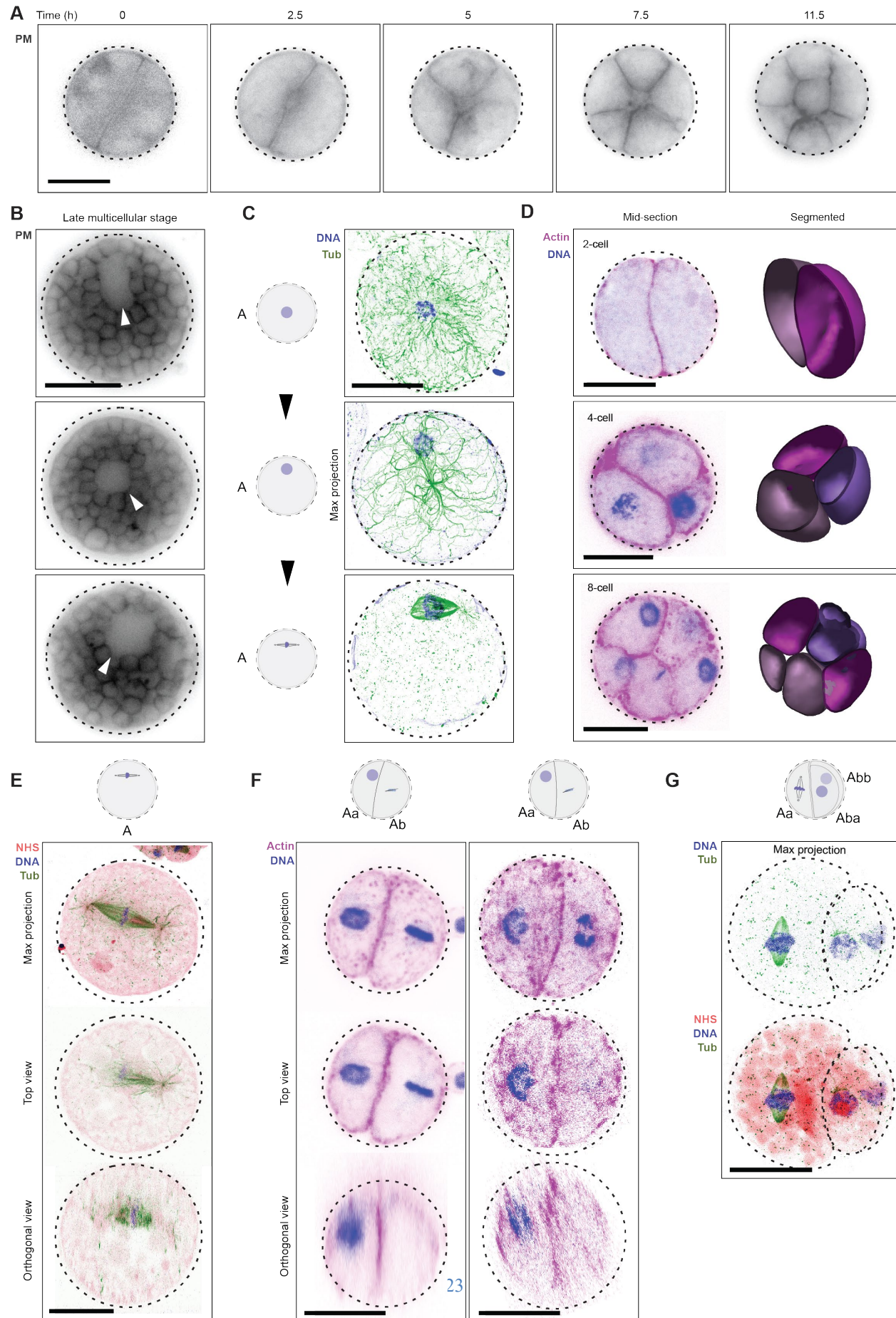


Fig. S3: Symmetry breaking during early development of *C. perkinsii*.

(A) Cleavage dynamics at constant volume over time, visualized using a plasma membrane (PM) marker FM4-64 (movie S4). Bar, 10 μ m.

(B) PM-stained live late colonies exhibiting internal cavities (arrows). Bar, 10 μ m.

(C) U-ExM stained cells for microtubules (green), and DNA (blue), highlighting nuclear migration to the cortex prior to 1st mitotic division. Bar, 10 μ m.

(D) Actin (magenta) and DNA-stained (blue) colonies at various cell stages, accompanied by volumetric segmentation of cells, highlighting the asymmetry in volume following the first cleavage. Scale bar: 10 μ m.

(E) U-ExM stained cell for pan-labelling with NHS-Ester (red), microtubules (green), and DNA (blue), illustrating the first mitotic division at the cortex. Bar, 10 μ m.

(F) Actin (magenta) and DNA-stained (blue) colonies at distinct cell-stages showing the asymmetrical cell division in volume and time between the Aa and Ab cells. Bar, 10 μ m.

(G) U-ExM stained colonies for pan-labelling with NHS-Ester (red), microtubules (green), and DNA (blue), highlighting a three-cell stage. Bar, 10 μ m.

Movies. S1 to S8

Movie S1: Time lapse of two synchronized cells of *C. perkinsii* obtained with epifluorescent microscopy. Time interval between frames is 1 hr. The movie is played at 7fps. Both cells can be seen undergoing a full life-cycle with the release of new born cells. Bar, 15 μ m.

Movie S2: Time lapse of distinct cell types: Ameboflagellate (left) harbouring a flagellum and exhibiting cellular protrusions and a mitotic cell without a flagellum (right) undergoing seemingly cell division. Time interval between frames is 5 sec. The movie is played at 7fps. Bar, 5 μ m.

Movie S3: Flagellated cells freely moving around following cell release among other non-flagellated cells. Time interval between frames is 1 sec. The movie is played at 7fps. Bar, 10 μ m.

Movie S4: Time lapse of a PM-stained *C. perkinsii* cell undergoing a series of cleavages at constant volume and showcasing cortical rotations. Time interval between frames is 15 min. The movie is played at 7fps. Bar, 10 μ m.

Movie S5: Z-stacks of PM-stained live colonies at distinct cell stages, highlighting the patterned cleavages, tetrahedral 4 cell stage and formation of spatially organized multicellular colonies. Bar, 10 μ m.

Movie S6: Z-stacks of two U-ExM stained colonies for pan-labelling with NHS-Ester (red), microtubules (green), and DNA (blue) at the three-cell stage. Scale bar: 10 μ m.

Movies S7 and S8: 3D visualisation of U-ExM stained late colonies for pan-labelling with NHS-Ester (grey), microtubules (red), and DNA (cyan), highlighting the co-existence and clustering of flagellated and non-flagellated cells within the multicellular colony. Bar is uncorrected for expansion factor of 4.2X.

Tables. S1 to S4

Table S1: Enriched GO terms in differentially expressed genes in *C. perkinsii* replicate B compared to A and C replicates. GO term enrichment is indicative of a viral infection, which explains the distinct RNAseq profile and exclusion of B replicate compared to that of A and C.

Table S2: GO term enrichment for the 5 different gene expression clusters throughout the life cycle of *C. perkinsii*.

Table S3: *C. perkinsii* genes associated with the GO term enrichment for the 5 different gene expression clusters.

TwoLetterCode	Species	Blastula timepoint cutoff	Timepoint windows for averaging
PT	<i>P. dumerilii</i>	14	1-4, 4-6, 6-8, 8-10,10-12,12-14
FW	<i>S. polychroa</i>	9	1-3,3-5,5-6,6-7,7-8,8-9
CE	<i>C. elegans</i>	20	1-5, 5-8, 8-11,11-14.14-17,17-20
HD	<i>H. dujardini</i>	15	1-4,4-7,7-9,9-11,11-13,13-15
DM	<i>D. melanogaster</i>	9	1-3,3-5,5-6,6-7,7-8,8-9
SU	<i>S. purpuratus</i>	22	1-5,5-9,9-13,13-16,16-19,19-22
ZF	<i>D. rerio</i>	6	1,2,3,4,5,6
NV	<i>N. vectensis</i>	10	1-3,3-5,5-7,7-8,8-9,9-10
AQ	<i>A. queenslandica</i>	11	1-3,3-5,5-7,7-9,9-10,10-11
ML	<i>M. leidy</i>	11	1-3,3-5,5-7,7-9,9-10,10-11

Table S4: Developmental time points and time averaging windows for animal gene expression comparisons from the one or two cell stage through the blastula stage. The number of time points was selected based on developmental information in Levin et al. 2016(58).

5 Supplementary references

1. A. Seb  Pedr  s, B. M. Degnan, I. Ruiz-Trillo, The origin of Metazoa: a unicellular perspective. doi: 10.1038/nrg.2017.21 (2017).
2. T. Brunet, N. King, The Single-Celled Ancestors of Animals : A History of Hypotheses. *The Evolution of Multicellularity*, 251–278 (2022).
3. T. Brunet, N. King, The Origin of Animal Multicellularity and Cell Differentiation. [Preprint] (2017). <https://doi.org/10.1016/j.devcel.2017.09.016>.
4. R. K. Grosberg, R. R. Strathmann, The Evolution of Multicellularity: A Minor Major Transition? <https://doi.org/10.1146/annurev.ecolsys.36.102403.114735> **38**, 621–654 (2007).
5. A. H. Knoll, The Multiple Origins of Complex Multicellularity. <https://doi.org/10.1146/annurev.earth.031208.100209> **39**, 217–239 (2011).
6. I. De Smet, T. Beeckman, Asymmetric cell division in land plants and algae: the driving force for differentiation. *Nature Reviews Molecular Cell Biology* 2011 12:3 **12**, 177–188 (2011).
7. P. G  nczy, Mechanisms of asymmetric cell division: flies and worms pave the way. *Nature Reviews Molecular Cell Biology* 2008 9:5 **9**, 355–366 (2008).
8. J. A. Knoblich, Asymmetric cell division during animal development. *Nature Reviews Molecular Cell Biology* 2001 2:1 **2**, 11–20 (2001).
9. T. Lechler, M. Mapelli, Spindle positioning and its impact on vertebrate tissue architecture and cell fate. *Nature Reviews Molecular Cell Biology* 2021 22:10 **22**, 691–708 (2021).
10. S. L. Prosser, L. Pelletier, Mitotic spindle assembly in animal cells: a fine balancing act. *Nat. Rev. Mol. Cell Biol.* **18**, 187–201 (2017).

11. T. Y. C. Tsai, R. M. Garner, S. G. Megason, Adhesion-Based Self-Organization in Tissue Patterning. <https://doi.org/10.1146/annurev-cellbio-120420-100215> **38**, 349–374 (2022).
12. M. J. Wheelock, K. R. Johnson, Cadherins as Modulators of Cellular Phenotype. *Annu Rev Cell Dev Biol* **19**, 207–235 (2003).
- 5 13. T. Y. C. Tsai, R. M. Garner, S. G. Megason, Adhesion-Based Self-Organization in Tissue Patterning. *Annu Rev Cell Dev Biol* **38**, 349–374 (2022).
14. D. Wu, K. M. Yamada, S. Wang, Tissue Morphogenesis Through Dynamic Cell and Matrix Interactions. <https://doi.org/10.1146/annurev-cellbio-020223-031019> **39**, 123–144 (2023).
- 10 15. D. Arendt, J. M. Musser, C. V. H. Baker, A. Bergman, C. Cepko, D. H. Erwin, M. Pavlicev, G. Schlosser, S. Widder, M. D. Laubichler, G. P. Wagner, The origin and evolution of cell types. *Nature Reviews Genetics* **2016 17:12** **17**, 744–757 (2016).
16. G. Bell, A. O. Mooers, Size and complexity among multicellular organisms. *Biol. J. Linn. Soc.* **60**, 345–363 (1997).
- 15 17. P. Márquez-Zacarias, R. M. Pineau, M. Gomez, A. Veliz-Cuba, D. Murrugarra, W. C. Ratcliff, K. J. Niklas, Evolution of Cellular Differentiation: From Hypotheses to Models. *Trends Ecol Evol* **36**, 49–60 (2021).
18. L. Wolpert, Do we understand development? *Science (1979)* **266**, 571–572 (1994).
19. I. Salazar-Ciudad, Morphological evolution and embryonic developmental diversity in metazoa. *Development* **137**, 531–539 (2010).
- 20 20. A. T. Kalinka, P. Tomancak, The evolution of early animal embryos: conservation or divergence? *Trends Ecol Evol* **27**, 385–393 (2012).
21. I. Ruiz-Trillo, A. de Mendoza, Towards understanding the origin of animal development. *Development*, doi: 10.1242/dev.192575 (2020).
22. S. Xiao, A. D. Muscente, L. Chen, C. Zhou, J. D. Schiffbauer, A. D. Wood, N. F. Polys, X. Yuan, The Weng'an biota and the Ediacaran radiation of multicellular eukaryotes. *Natl Sci Rev* **1**, 498–520 (2014).
- 25 23. Z. Yin, W. Sun, P. Liu, J. Chen, D. J. Bottjer, J. Li, M. Zhu, Diverse and complex developmental mechanisms of early Ediacaran embryo-like fossils from the Weng'an Biota, southwest China. *Philosophical Transactions of the Royal Society B: Biological Sciences* **377** (2022).
24. Mitotic topologies and mechanics of Neoproterozoic algae and animal embryos | Paleobiology | Cambridge Core. <https://www.cambridge.org/core/journals/paleobiology/article/abs/mitotic-topologies-and-mechanics-of-neoproterozoic-algae-and-animal-embryos/C2EC4ADB10047F12E176A3370AD6B627>.
- 30 25. X. Shuhai, A. H. And, L. Knoll, S. & Xiao, A. H. Knoll, S. Xiao, S. H. Edu] Q, A. H. Knoll, H. Edu], Fossil preservation in the Neoproterozoic Doushantuo phosphorite Lagerstätte, South China. *Lethaia* **32**, 219–238 (1999).
26. L. Chen, S. Xiao, K. Pang, C. Zhou, X. Yuan, Cell differentiation and germ-soma separation in Ediacaran animal embryo-like fossils. *Nature* **516**, 238–241 (2014).
- 35 27. T. Hultgren, J. A. Cunningham, C. Yin, M. Stamparoni, F. Marone, P. C. J. Donoghue, S. Bengtson, Fossilized nuclei and germination structures identify ediacaran “animal embryos” as encysting protists. *Science (1979)* **334**, 1696–1699 (2011).
28. Z. Yin, K. Vargas, J. Cunningham, S. Bengtson, M. Zhu, F. Marone, P. Donoghue, The Early Ediacaran Caveasphaera Foreshadows the Evolutionary Origin of Animal-like Embryology. *Current Biology* **29**, 4307–4314.e2 (2019).
- 40 29. K. J. Niklas, S. A. Newman, The origins of multicellular organisms. *Evol Dev* **15**, 41–52 (2013).
30. N. Ros-Rocher, A. Pérez-Posada, M. M. Leger, I. Ruiz-Trillo, The origin of animals: an ancestral reconstruction of the unicellular-to-multicellular transition. *Open Biol* **11** (2021).
- 45 31. H. Liu, J. L. Steenwyk, X. Zhou, D. T. Schultz, K. M. Kocot, X.-X. Shen, A. Rokas, Y. Li, A genome-scale Opisthokonta tree of life: toward phylogenomic resolution of ancient divergences. *bioRxiv*, 2023.09.20.556338 (2023).
32. E. Ocaña-Pallarès, T. A. Williams, D. López-Escardó, A. S. Arroyo, J. S. Pathmanathan, E. Baptiste, D. V. Tikhonenkov, P. J. Keeling, G. J. Szöllösi, I. Ruiz-Trillo, Divergent genomic trajectories predate the origin of animals and fungi. *Nature* **2022 609:7928** **609**, 747–753 (2022).
- 50 33. T. Brunet, N. King, The Origin of Animal Multicellularity and Cell Differentiation. doi: 10.1016/j.devcel.2017.09.016 (2017).
34. T. Brunet, M. Albert, W. Roman, M. C. Coyle, D. C. Spitzer, N. King, A flagellate-to-amoeboid switch in the closest living relatives of animals. *Elife* **10**, 1–73 (2021).
- 55 35. T. Brunet, B. T. Larson, T. A. Linden, M. J. A. Vermeij, K. McDonald, N. King, Light-regulated collective contractility in a multicellular choanoflagellate. *Science*, doi: 10.1126/science.aay2346 (2019).

36. S. R. Fairclough, M. J. Dayel, N. King, Multicellular development in a choanoflagellate. [Preprint] (2010). <https://doi.org/10.1016/j.cub.2010.09.014>.
37. H. Parra-Acero, M. Haret, N. Sánchez-Pons, E. Casacuberta, N. H. Brown, O. Dudin, I. Ruiz-Trillo, Integrin-Mediated Adhesion in the Unicellular Holozoan *Capsaspora owczarzaki*. *Current Biology*, doi: 10.1016/j.cub.2020.08.015 (2020).
38. A. Sebé-Pedrós, M. Irimia, J. del Campo, H. Parra-Acero, C. Russ, C. Nusbaum, B. J. Blencowe, I. Ruiz-Trillo, Regulated aggregative multicellularity in a close unicellular relative of metazoa. *Elife*, doi: 10.7554/eLife.01287 (2013).
39. A. Pérez-Posada, O. Dudin, E. Ocaña-Pallarès, I. Ruiz-Trillo, A. Ondracka, Cell cycle transcriptomics of *Capsaspora* provides insights into the evolution of cyclin-CDK machinery. *PLoS Genet*, doi: 10.1371/journal.pgen.1008584 (2020).
40. N. Ros-Rocher, R. Q. Kidner, C. Gerdt, W. S. Davidson, I. Ruiz-Trillo, J. P. Gerdt, Chemical factors induce aggregative multicellularity in a close unicellular relative of animals. *bioRxiv*, 2022.12.01.517680 (2022).
41. A. Woznica, A. M. Cantley, C. Beemelmans, E. Freinkman, J. Clardy, N. King, Bacterial lipids activate, synergize, and inhibit a developmental switch in choanoflagellates. *Proc Natl Acad Sci U S A*, doi: 10.1073/pnas.1605015113 (2016).
42. T. C. Levin, A. J. Greaney, L. Wetzel, N. King, The Rosetteless gene controls development in the choanoflagellate *S. rosetta*. *Elife*, doi: 10.7554/eLife.04070 (2014).
43. T. C. Levin, N. King, Evidence for Sex and Recombination in the Choanoflagellate *Salpingoeca rosetta*. *Current Biology*, doi: 10.1016/j.cub.2013.08.061 (2013).
44. M. J. Dayel, R. A. Alegado, S. R. Fairclough, T. C. Levin, S. A. Nichols, K. McDonald, N. King, Cell differentiation and morphogenesis in the colony-forming choanoflagellate *Salpingoeca rosetta*. *Dev Biol* **357**, 73–82 (2011).
45. D. Laundon, B. T. Larson, K. McDonald, N. King, P. Burkhardt, The architecture of cell differentiation in choanoflagellates and sponge choanocytes. *PLoS Biol* **17**, e3000226 (2019).
46. S. L. Glockling, W. L. Marshall, F. H. Gleason, Phylogenetic interpretations and ecological potentials of the Mesomycetozoa (Ichthyosporea). *Fungal Ecol* **6**, 237–247 (2013).
47. L. Mendoza, J. W. Taylor, L. Ajello, The class Mesomycetozoa: A heterogeneous group of microorganisms at the animal-fungal boundary. *Annu Rev Microbiol* [Preprint] (2002). <https://doi.org/10.1146/annurev.micro.56.012302.160950>.
48. A. Ondracka, O. Dudin, I. Ruiz-Trillo, Decoupling of Nuclear Division Cycles and Cell Size during the Coenocytic Growth of the Ichthyosporean *Sphaeroforma arctica*. *Current Biology* **28** (2018).
49. O. Dudin, S. Wielgoss, A. M. New, I. Ruiz-Trillo, Regulation of sedimentation rate shapes the evolution of multicellularity in a close unicellular relative of animals. *PLoS Biol* **20**, e3001551 (2022).
50. O. Dudin, A. Ondracka, X. Grau-Bové, A. A. B. Haraldsen, A. Toyoda, H. Suga, J. Bråte, I. Ruiz-Trillo, A unicellular relative of animals generates a layer of polarized cells by actomyosin-dependent cellularization. *Elife* **8** (2019).
51. L. A. Sarre, I. V. Kim, V. Ovchinnikov, M. Olivetta, H. Suga, O. Dudin, A. Sebé-Pedrós, A. de Mendoza, DNA methylation enables recurrent endogenization of giant viruses in an animal relative. *bioRxiv*, 2024.01.08.574619 (2024).
52. B. McCartney, O. Dudin, Cellularization across eukaryotes: Conserved mechanisms and novel strategies. *Curr Opin Cell Biol* **80**, 102157 (2023).
53. H. Shah, M. Olivetta, C. Bhickta, P. Ronchi, M. Trupinić, E. C. Tromer, I. M. Tolić, Y. Schwab, O. Dudin, G. Dey, Life cycle-coupled evolution of mitosis in close relatives of animals. *bioRxiv*, 2023.05.10.540163 (2023).
54. M. Olivetta, O. Dudin, The nuclear-to-cytoplasmic ratio drives cellularization in the close animal relative *Sphaeroforma arctica*. *Current Biology* **33**, 1597–1605.e3 (2023).
55. X. Grau-Bové, G. Torruella, S. Donachie, H. Suga, G. Leonard, T. A. Richards, I. Ruiz-Trillo, Dynamics of genomic innovation in the unicellular ancestry of animals. *Elife* **6** (2017).
56. M. Binyam, K. Scheffler, M. Schuh, Assembly and Positioning of the Oocyte Meiotic Spindle. <https://doi.org/10.1146/annurev-cellbio-100616-060553> **34**, 381–403 (2018).
57. S. Vasudevan, E. Seli, J. A. Steitz, Metazoan oocyte and early embryo development program: a progression through translation regulatory cascades. *Genes Dev* **20**, 138–146 (2006).
58. M. Levin, L. Anavy, A. G. Cole, E. Winter, N. Mostov, S. Khair, N. Senderovich, E. Kovalev, D. H. Silver, M. Feder, S. L. Fernandez-Valverde, N. Nakanishi, D. Simmons, O. Simakov, T. Larsson, S.-Y. Liu, A. Jeraf-Vider, K. Yaniv, J. F. Ryan, M. Q. Martindale, J. C. Rink, D. Arendt, S. M. Degnan, B. M. Degnan, T.

- Hashimshony, I. Yanai, The mid-developmental transition and the evolution of animal body plans. doi: 10.1038/nature16994 (2016).
59. N. Klena, G. Pigino, Structural Biology of Cilia and Intraflagellar Transport. <https://doi.org/10.1146/annurev-cellbio-120219-034238> **38**, 103–123 (2022).
60. Z. Carvalho-Santos, J. Azimzadeh, J. B. Pereira-Leal, M. Bettencourt-Dias, Tracing the origins of centrioles, cilia, and flagella. *Journal of Cell Biology* **194**, 165–175 (2011).
61. A. De Mendoza, A. Seb  -Pedr  s, I. Ruiz-Trillo, The Evolution of the GPCR Signaling System in Eukaryotes: Modularity, Conservation, and the Transition to Metazoan Multicellularity. *Genome Biol Evol* **6**, 606 (2014).
62. T. Brunet, D. S. Booth, Cell polarity in the protist-to-animal transition. *Curr Top Dev Biol* **154**, 1–36 (2023).
63. B. A. Wright, M. Kvansakul, B. Schierwater, P. O. Humbert, Cell polarity signalling at the birth of multicellularity: What can we learn from the first animals. *Front Cell Dev Biol* **10**, 1024489 (2022).
64. B. Klapholz, N. H. Brown, Talin - the master of integrin adhesions. *J Cell Sci* **130**, 2435–2446 (2017).
65. J. L. Bays, K. A. DeMali, Vinculin in cell–cell and cell–matrix adhesions. *Cellular and Molecular Life Sciences* **74**, 2999–3009 (2017).
66. B. M. Gumbiner, Cell Adhesion: The Molecular Basis of Tissue Architecture and Morphogenesis. *Cell* **84**, 345–357 (1996).
67. J. H. Shawky, L. A. Davidson, Tissue Mechanics and Adhesion during Embryo Development. *Dev Biol* **401**, 152 (2015).
68. A. Seb  -Pedr  s, A. De Mendoza, B. F. Lang, B. M. Degnan, I. Ruiz-Trillo, Unexpected repertoire of metazoan transcription factors in the unicellular holozoan capsaspora owczarzaki. *Mol Biol Evol*, doi: 10.1093/molbev/msq309 (2011).
69. A. De Mendoza, A. Seb  -Pedr  s, M. S.   stak, M. Matej    , G. Torruella, T. Domazet-Lo  o, I. Ruiz-Trillo, Transcription factor evolution in eukaryotes and the assembly of the regulatory toolkit in multicellular lineages. *Proc Natl Acad Sci U S A*, doi: 10.1073/pnas.1311818110 (2013).
70. A. Seb  -Pedr  s, I. Ruiz-Trillo, “Evolution and Classification of the T-Box Transcription Factor Family” in *Current Topics in Developmental Biology* (2017).
71. E. G. Leigh, J. M. Smith, E. Szathmary, The Major Transitions of Evolution. *Evolution (N Y)*, doi: 10.2307/2410462 (1995).
72. A. Seb  -Pedr  s, B. M. Degnan, I. Ruiz-Trillo, The origin of Metazoa: A unicellular perspective. [Preprint] (2017). <https://doi.org/10.1038/nrg.2017.21>.
73. D. Gambarotto, F. U. Zwettler, M. Le Guennec, M. Schmidt-Cernohorska, D. Fortun, S. Borgers, J. Heine, J. G. Schloetel, M. Reuss, M. Unser, E. S. Boyden, M. Sauer, V. Hamel, P. Guichard, Imaging cellular ultrastructures using expansion microscopy (U-ExM). *Nature Methods* **16**, 71–74 (2018).
74. D. Gambarotto, V. Hamel, P. Guichard, Ultrastructure expansion microscopy (U-ExM). *Methods Cell Biol* **161**, 57–81 (2021).
75. S. Machado, V. Mercier, N. Chiaruttini, LimeSeg: A coarse-grained lipid membrane simulation for 3D image segmentation. *BMC Bioinformatics* **20**, 1–12 (2019).
76. J. Schindelin, I. Arganda-Carreras, E. Frise, V. Kaynig, M. Longair, T. Pietzsch, S. Preibisch, C. Rueden, S. Saalfeld, B. Schmid, J. Y. Tinevez, D. J. White, V. Hartenstein, K. Eliceiri, P. Tomancak, A. Cardona, Fiji: an open-source platform for biological-image analysis. *Nature Methods* **9**, 676–682 (2012).
77. Multicellgenome Lab. https://figshare.com/authors/Multicellgenome_Lab/2628379.
78. R. Patro, G. Duggal, M. I. Love, R. A. Irizarry, C. Kingsford, Salmon provides fast and bias-aware quantification of transcript expression. *Nature Methods* **14**, 417–419 (2017).
79. R: The R Project for Statistical Computing. <https://www.r-project.org/>.
80. C. Soneson, M. I. Love, M. D. Robinson, Differential analyses for RNA-seq: Transcript-level estimates improve gene-level inferences. *F1000Res* **4** (2016).
81. M. D. Robinson, D. J. McCarthy, G. K. Smyth, edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140 (2010).
82. D. J. McCarthy, Y. Chen, G. K. Smyth, Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation. *Nucleic Acids Res* **40**, 4288 (2012).
83. Y. Chen, A. T. L. Lun, G. K. Smyth, From reads to genes to pathways: Differential expression analysis of RNA-Seq experiments using Rsubread and the edgeR quasi-likelihood pipeline. *F1000Res* **5** (2016).
84. e1071: Misc Functions of the Department of Statistics, Probability Theory Group (Formerly: E1071), TU Wien version 1.7-15 from R-Forge. <https://rdrr.io/rforge/e1071/>.
85. viridis: Colorblind-Friendly Color Maps for R version 0.6.4 from CRAN. <https://rdrr.io/cran/viridis/>.
86. superheat: A Graphical Tool for Exploring Complex Datasets Using Heatmaps version 0.1.0 from CRAN. <https://rdrr.io/cran/superheat/>.

87. C. Camacho, G. Coulouris, V. Avagyan, N. Ma, J. Papadopoulos, K. Bealer, T. L. Madden, BLAST+: Architecture and applications. *BMC Bioinformatics* **10**, 1–9 (2009).
88. E. Boutet, D. Lieberherr, M. Tognolli, M. Schneider, P. Bansal, A. J. Bridge, S. Poux, L. Bougueleret, I. Xenarios, UniProtKB/Swiss-Prot, the Manually Annotated Section of the UniProt KnowledgeBase: How to Use the Entry View. *Methods Mol Biol* **1374**, 23–54 (2016).
89. B. Rost, Enzyme function less conserved than anticipated. *J Mol Biol* **318**, 595–608 (2002).
90. J. A. Burns, A. Paasch, A. Narechania, E. Kim, Comparative Genomics of a Bacterivorous Green Alga Reveals Evolutionary Causalities and Consequences of Phago-Mixotrophic Mode of Nutrition. *Genome Biol Evol* **7**, 3047–3061 (2015).
91. topGO: Enrichment Analysis for Gene Ontology version 2.42.0 from Bioconductor. <https://rdrr.io/bioc/topGO/>.
92. D. M. Emms, S. Kelly, OrthoFinder: Phylogenetic orthology inference for comparative genomics. *Genome Biol* **20**, 1–14 (2019).
93. GitHub - TransDecoder/TransDecoder: TransDecoder source. <https://github.com/TransDecoder/TransDecoder>.
94. S. R. Engel, E. D. Wong, R. S. Nash, S. Aleksander, M. Alexander, E. Douglass, K. Karra, S. R. Miyasato, M. Simison, M. S. Skrzypek, S. Weng, J. M. Cherry, New data and collaborations at the Saccharomyces Genome Database: updated reference genome, alleles, and the Alliance of Genome Resources. *Genetics* **220** (2022).
95. V. Wood, R. Gwilliam, M. A. Rajandream, M. Lyne, R. Lyne, A. Stewart, J. Sgouros, N. Peat, J. Hayles, S. Baker, D. Basham, S. Bowman, K. Brooks, D. Brown, S. Brown, T. Chillingworth, C. Churcher, M. Collins, R. Connor, A. Cronin, P. Davis, T. Feltwell, A. Fraser, S. Gentles, A. Goble, N. Hamlin, D. Harris, J. Hidalgo, G. Hodgson, S. Holroyd, T. Hornsby, S. Howarth, E. J. Huckle, S. Hunt, K. Jagels, K. James, L. Jones, M. Jones, S. Leather, S. McDonald, J. McLean, P. Mooney, S. Moule, K. Mungall, L. Murphy, D. Niblett, C. Odell, K. Oliver, S. O’Neil, D. Pearson, M. A. Quail, E. Rabinowitsch, K. Rutherford, S. Rutter, D. Saunders, K. Seeger, S. Sharp, J. Skelton, M. Simmonds, R. Squares, S. Squares, K. Stevens, K. Taylor, R. G. Taylor, A. Tivey, S. Walsh, T. Warren, S. Whitehead, J. Woodward, G. Volckaert, R. Aert, J. Robben, B. Grymonprez, I. Weltjens, E. Vanstreels, M. Rieger, M. Schäfer, S. Müller-Auer, C. Gabel, M. Fuchs, C. Fritze, E. Holzer, D. Moestl, H. Hilbert, K. Borzym, I. Langer, A. Beck, H. Lehrach, R. Reinhardt, T. M. Pohl, P. Eger, W. Zimmermann, H. Wedler, R. Wambutt, B. Purnelle, A. Goffeau, E. Cadieu, S. Dréano, S. Gloux, V. Lelaure, S. Mottier, F. Galibert, S. J. Aves, Z. Xiang, C. Hunt, K. Moore, S. M. Hurst, M. Lucas, M. Rochet, C. Gaillardin, V. A. Tallada, A. Garzon, G. Thode, R. R. Daga, L. Cruzado, J. Jimenez, M. Sánchez, F. del Rey, J. Benito, A. Domínguez, J. L. Revuelta, S. Moreno, J. Armstrong, S. L. Forsburg, L. Cerrutti, T. Lowe, W. R. McCombie, I. Paulsen, J. Potashkin, G. V. Shpakovski, D. Ussery, B. G. Barrell, P. Nurse, The genome sequence of *Schizosaccharomyces pombe*. *Nature* **2002 415:6874** **415**, 871–880 (2002).
96. I. Eichinger, J. A. Pachebat, G. Glöckner, M. A. Rajandream, R. Sugang, M. Berriman, J. Song, R. Olsen, K. Szafranski, Q. Xu, B. Tunggal, S. Kummerfeld, M. Madera, B. A. Konfortov, F. Rivero, A. T. Bankier, R. Lehmann, N. Hamlin, R. Davies, P. Gaudet, P. Fey, K. Pilcher, G. Chen, D. Saunders, E. Sodergren, P. Davis, A. Kerhornou, X. Nie, N. Hall, C. Anjard, L. Hemphill, N. Bason, P. Farbrother, B. Desany, E. Just, T. Morio, R. Rost, C. Churcher, J. Cooper, S. Haydock, N. Van Driessche, A. Cronin, I. Goodhead, D. Muzny, T. Mourier, A. Pain, M. Lu, D. Harper, R. Lindsay, H. Hauser, K. James, M. Quiles, M. Madan Babu, T. Saito, C. Buchrieser, A. Wardroper, M. Felder, M. Thangavelu, D. Johnson, A. Knights, H. Louseged, K. Mungall, K. Oliver, C. Price, M. A. Quail, H. Urushihara, J. Hernandez, E. Rabinowitsch, D. Steffen, M. Sanders, J. Ma, Y. Kohara, S. Sharp, M. Simmonds, S. Spiegler, A. Tivey, S. Sugano, B. White, D. Walker, J. Woodward, T. Winckler, Y. Tanaka, G. Shaulsky, M. Schleicher, G. Weinstock, A. Rosenthal, E. C. Cox, R. L. Chisholm, R. Gibbs, W. F. Loomis, M. Platzer, R. R. Kay, J. Williams, P. H. Dear, A. A. Noegel, B. Barrell, A. Kuspa, The genome of the social amoeba *Dictyostelium discoideum*. *Nature* **2005 435:7038** **435**, 43–57 (2005).
97. K. Katoh, D. M. Standley, MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol* **30**, 772–780 (2013).
98. HMMER. <http://hmmer.org/>.