The origin of Metazoa: a unicellular perspective

Arnau Sebé-Pedrós[1], Bernard M. Degnan[2] and Iñaki Ruiz-Trillo[3–5]

Abstract | The first animals evolved from an unknown single-celled ancestor in the Precambrian period. Recently, the identification and characterization of the genomic and cellular traits of the protists most closely related to animals have shed light on the origin of animals. Comparisons of animals with these unicellular relatives allow us to reconstruct the first evolutionary steps towards animal multicellularity. Here, we review the results of these investigations and discuss their implications for understanding the earliest stages of animal evolution, including the origin of metazoan genes and genome function.

Protist
An informal name that is given to eukaryotes (usually unicellular eukaryotes) that are not included in the fungal, animal or plant lineages. Protists do not form a monophyletic clade.

[1]Department of Computer Science and Applied Mathematics, Weizmann Institute of Science, Herzl Street 234, Rehovot 7610001, Israel.
[2]School of Biological Sciences, University of Queensland, Brisbane, QLD 4072, Australia.
[3]Institut de Biologia Evolutiva (Consejo Superior de Investigaciones Científicas–Universitat Pompeu Fabra), Passeig Marítim de la Barceloneta 37–49, 08003 Barcelona, Spain.
[4]ICREA, Pg Lluís Companys 23, 08010 Barcelona, Spain.
[5]Departament de Genètica, Microbiologia i Estadística, Universitat de Barcelona, Avinguda Diagonal 643, 08028 Barcelona, Spain.

Correspondence to A.S-P. and I.R-T.
arnau.sebe-pedros@weizmann.ac.il; inaki.ruiz@ibe.upf-csic.es

doi:10.1038/nrg.2017.21
Published online 8 May 2017; corrected 10 May 2017

"No direct proof exists of the origin of the Metazoa from the Protozoa, but such [an] origin besides being necessitated by the principle of evolution is strongly indicated by the facts of embryonic development, in which each metazoan passes from an acellular to a cellular condition", wrote Libbie H. Hyman in her 1940 book, *The Invertebrates: Protozoa Through Ctenophora*[1]. More than 75 years after Hyman's words, we now have direct proof that animals did evolve from an ancestor protist. Not only that, but we also now know that multicellularity has been acquired several times independently within the tree of eukaryotes, including in fungi, in plants, in the different types of algae and in slime moulds (BOX 1; FIG. 1). Multicellular organisms not only grow bigger than unicellular organisms, they also have the capability to perform different cellular functions at the same time owing to spatially organized and regulated division of labour[2–4]. Thus, the question is now not whether animals evolved from a protistan ancestor, but when, and whether this unicellular ancestor possessed features that we know are important in the formation and functioning of extant animal cell types and body plans.

Most work on the origin of animals has focused on determining the nature of the shared ancestor of all contemporary animals. Although comparative studies between bilaterian animals and non-bilaterian animals (sponges, ctenophores, placozoans and cnidarians) have provided important insights into the characteristics of the last common ancestor of metazoans (Urmetazoa)[5–10], this approach is insufficient as a means of improving our understanding of animal origins. To decipher how this transition took place, we also need to elucidate the nature of the unicellular ancestor of animals. Only by having some knowledge of the nature of both the unicellular ancestor and the first animal can we fully understand the unicellular-to-multicellular transition. Given that there is no fossil record of either the unicellular ancestor or the initial steps in the evolution of animal multicellularity, the only way we can do this is by studying the closest extant unicellular relatives of animals and comparing them with animals.

In this Review, we examine how comparative studies of the closest unicellular relatives of Metazoa have revolutionized our understanding of the unicellular ancestor of animals and, consequently, of the unicellular-to-multicellular transition. First, we describe the phylogenetic relationships between animals and other eukaryotes. Next, we explain how comparative genomics studies have provided a detailed reconstruction of the gene repertoire of the protistan ancestor of animals. We then review how studies addressing the cell and regulatory biology of animal relatives have enabled us to infer some of the biological traits of this unicellular ancestor. Based on these latest reconstructions, we further develop the hypothesis proposed by Zakhvatkin[11], which has more recently been expanded upon by Mikhailov *et al.*[12], who suggested that temporally regulated cell differentiation existed in the unicellular ancestor of animals and that animals originated through spatiotemporal integration of these pre-existing cell types. Finally, we discuss how these results pose new and exciting questions, and we propose research avenues to tackle them.

Animals among the eukaryotes

A strongly supported phylogenetic framework is key to addressing any evolutionary question. Thus, the first step towards understanding the unicellular ancestor of animals was the identification of the extant unicellular relatives of animals. Until a decade ago, we knew very little about the phylogenetic relationships between animals and other eukaryotes. Based on morphological similarity alone, it was difficult to identify single-celled

Box 1 | **Multicellularity: a common theme among eukaryotes**

Multicellularity has repeatedly evolved within the eukaryotic tree of life to the point that some authors have considered this a 'minor' major evolutionary transition[2]. The transition to multicellularity has occurred within eukaryotes between 16 (REF. 3) and 22 (REF. 4) times. Moreover, we find examples of multicellularity in all major eukaryotic lineages. Some lineages are entirely multicellular (such as Metazoa and Embryophyta), whereas others only have one or a few multicellular species. Similarly, in some lineages, such as fungi, there have been several reversions to the unicellular state.

Multicellularity can be acquired via two major mechanisms: by clonal division and by cell aggregation. In clonal multicellularity, all the cells in the organism arise from a founder cell that undergoes successive rounds of division, resulting in the formation of a cluster of genetically identical cells. By contrast, aggregative multicellularity develops through the attachment, one to another, of genetically distinct cells to form a multicellular entity. In this latter scenario, intra-organismal competition poses strong fitness challenges, and therefore, the aggregate is predicted to be evolutionarily unstable[114]. Indeed, aggregative multicellularity is found only as a transient life stage. Among lineages with clonal multicellularity, those that result from embryonic development are associated with more complex body plans. Multicellular embryonic development is well described in many embryophytes (land plants) and metazoans. This is not the case for brown and red algae, which remain poorly studied, although there is evidence of embryonic development in these lineages as well[115–117]. Finally, despite fungi being a relatively well-characterized group, multicellularity in fungi is not well understood and quite problematic to categorize.

Not only has multicellularity evolved independently multiple times in eukaryotes, but each of these transitions also occurred at different times in the history of life. Fossils of putative macroscopic multicellular organisms have been described from 2,100 million-year-old rocks in Gabon[118], but it is debatable whether these fossils represent true multicellular organisms or colonies, whether they are eukaryotes or prokaryotes, or even whether they represent biological or abiotic structures[4]. The oldest unequivocal multicellular eukaryotes are the *Bangiomorpha* spp. red algae, which appeared around 1,200 million years ago, and have differentiated holdfasts and reproductive cells[119]. However, most of the extant major multicellular lineages appeared later in evolution. These include macrophyte green algae (Coleochaetales, Zygnematales and Charales), which arose 750 million years ago[120,121]; metazoans, which arose ~650 million years ago[122–124]; embryophytes, which arose ~450 million years ago[125]; multicellular fungi, which arose ~300 million years ago[126]; and phaeophytes, which arose ~130 million years ago[127] (FIG. 1).

Finally, different molecular systems have evolved to serve similar functions in different multicellular lineages. An interesting case is cell adhesion, which is mediated by different molecules in different multicellular taxa. For example, whereas cell adhesion in plants is largely mediated by extracellular 'glues' such as pectins and hemicelluloses, in fungi it involves extracellular glycoproteins, in animals it is mediated by transmembrane proteins such as cadherins and integrins, and in brown algae it involves phlorotannins and alginates (which are polymers of D-mannuronic acid and L-guluronic acid)[128,129]. Similarly, signalling pathways are largely distinct in different multicellular lineages and, even though an expansion in transcription factor (TF)-encoding genes is observed in many of them[67], different structural TF families have expanded in each case. In fact, 'phylogenetic inertia' largely determines the multicellularity toolkit in each group. That is, each of the multicellular lineages is more similar in terms of their TFs or their signalling repertoires to their unicellular relatives than to other multicellular groups. Therefore, despite some common trends, there is no universal gene toolkit for multicellularity.

In summary, the overall pan-eukaryotic picture of multicellularity is characterized by multiple independent transitions, at very different times, and by the use of very different molecular toolkits.

protists that could potentially be close relatives of animals and that could, therefore, shed light on animal origins. The exception are the choanoflagellates, which have been considered to be close relatives of animals for more than a century because of their close resemblance to a specific sponge cell type called the choanocyte[3,13–15]. Even in the case of choanoflagellates, however, it was not until the advent of molecular phylogenetics that they were confirmed as the closest relatives of animals.

Phylogenomic studies have considerably changed our understanding of the tree of life (BOX 1; FIG. 1). In addition to confirming the position of choanoflagellates as the unicellular sister group of animals, studies in the past decade have revealed that two additional independent lineages, the filastereans and the ichthyosporeans, are also closely related to Metazoa[16–19]. Consequently, three unicellular lineages (choanoflagellates, filastereans and ichthyosporeans) form a clade with Metazoa (FIG. 2). This clade, called Holozoa[20], thus forms the reference point for studies of the origin of animals. There have been alternative hypotheses with regard to the specific phylogenetic relationships between unicellular holozoans. For example, some earlier analyses indicated that filastereans and ichthyosporeans formed a monophyletic clade[16], while others showed filastereans to be the sister group of choanoflagellates and animals[17]. The position of another candidate holozoan, the enigmatic *Corallochytrium limacisporum*, also remained controversial, as only a few gene sequences were available[21–24]. However, the most recent phylogenomic analyses, which use dozens of phylogenetic markers, have shown that the clade formed by *C. limacisporum* plus Ichthyosporea is the earliest-branching holozoan lineage, and that Filasterea is the sister group of the Choanoflagellatea and Metazoa[18,19] (FIG. 2). It is worth emphasizing that none of these possible alternative topologies within Holozoa has an impact on the conclusions and reconstructions reviewed here.

Interestingly, the three unicellular lineages included in the Holozoa clade have very different morphologies and lifestyles. The choanoflagellates, which are the group most closely related to animals, are free-living single-celled and colonial flagellates that feed on bacteria. They are divided into two major groups (Craspedida and

Bilaterian animals
A monophyletic group that is defined by bilateral symmetry of the body plan and three germ layers, and that comprises most animal phyla.

Choanocyte
A specialized filter-feeding cell type that is characteristic of sponges. The basic cell structure, with a central flagellum surrounded by a microvilli collar, to some extent resembles that of choanoflagellate cells.

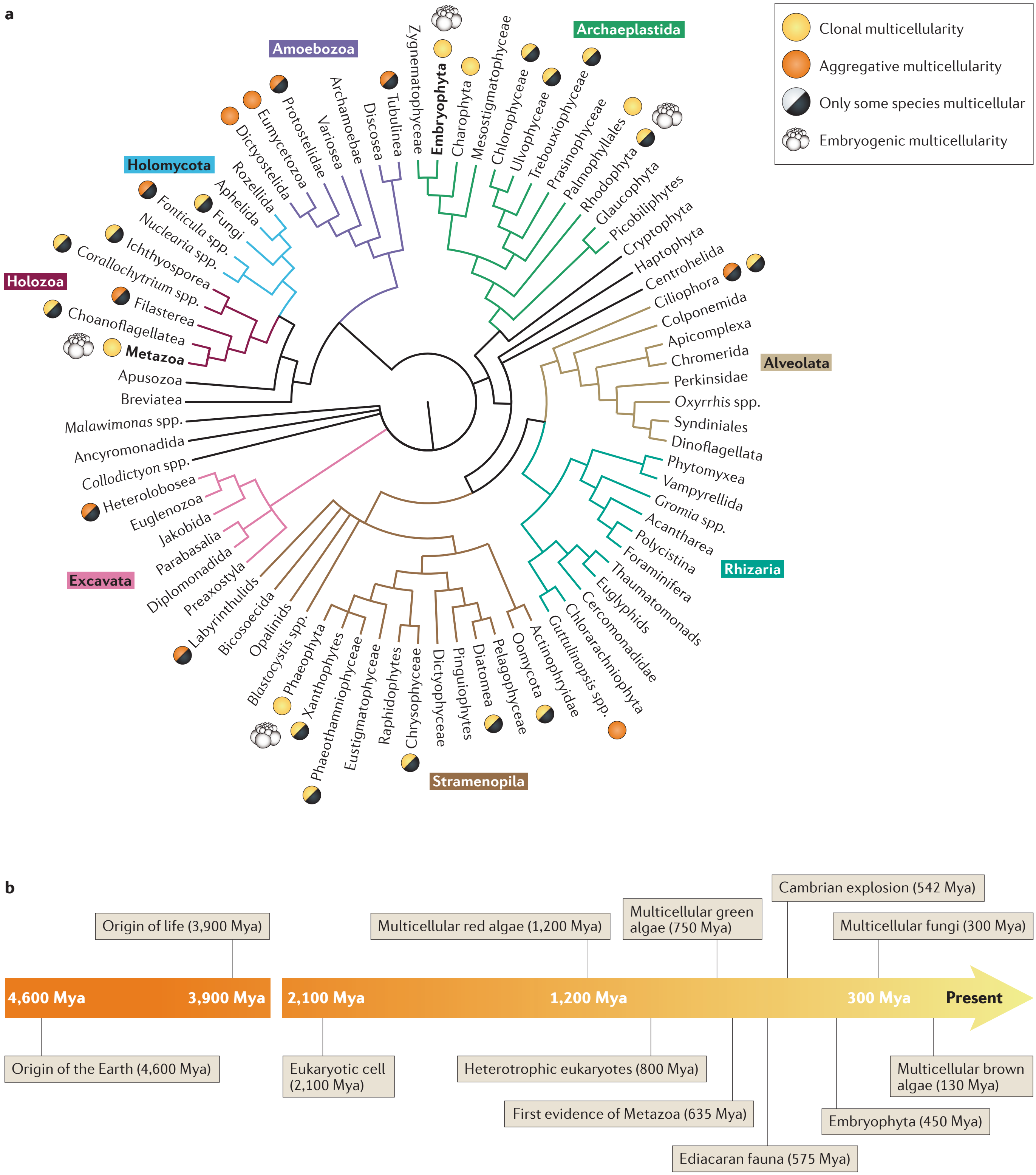


Figure 1 | **The multiple origins of multicellularity. a** | The phylogenetic distribution of multicellularity among eukaryotes. Multicellular forms (clonal or aggregative; see BOX 1) are present in several eukaryotic lineages. Some lineages, such as animals (Metazoa; highlighted in bold) and plants (Embryophyta; highlighted in bold), are entirely multicellular (that is, all species are multicellular), whereas other lineages have only a few multicellular species, with the majority being unicellular. From this widespread distribution, it can be inferred that multicellularity has evolved independently multiple times, although only in four lineages is this multicellularity linked to embryonic development and complex body plans. The tree is a consensus composite based on several recent phylogenomic studies[19,130–138]. **b** | A timeline of the origins of the major multicellular eukaryotic clades showing that transitions to multicellularity have occurred at very different times in the history of life. The estimations are based on fossil record and molecular clock estimates[4,119–123,125–127,139]. Time units are millions of years ago (Mya).

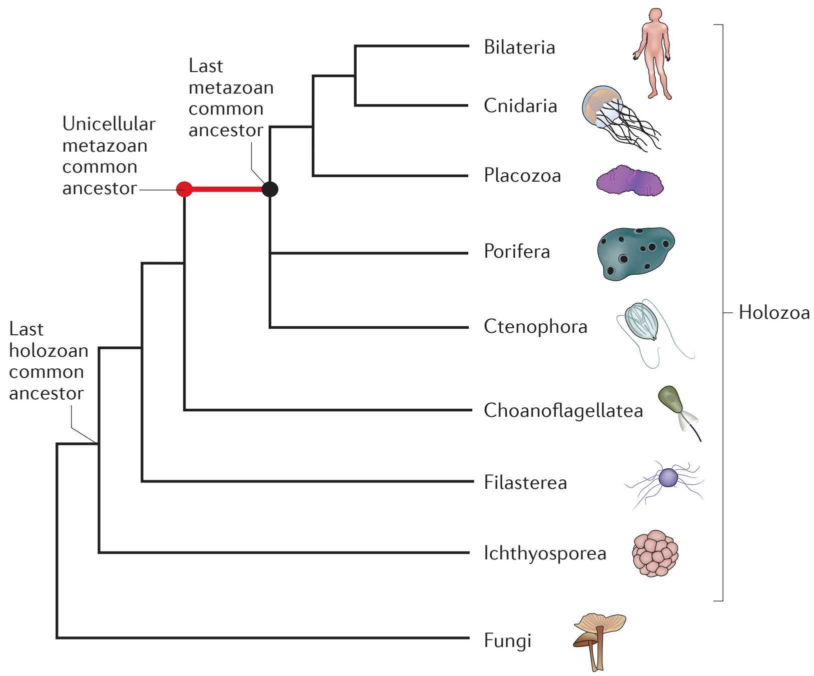


Figure 2 | **Phylogenetic relationships of animals and unicellular Holozoa.** There are three independent unicellular holozoan lineages that are closely related to animals: namely, Ichthyosporea, Filasterea and Choanoflagellatea. The schematic phylogenetic tree is based on the most recent and taxonomically rich phylogenomic studies[18,19]. The position of Ctenophora and Porifera is indicated as a polytomy given the current debate on the branching order of these two lineages[7,140,141].

Acanthoecida) and comprise more than 250 known species[23,25]. Choanoflagellates inhabit marine and freshwater environments, and are widely distributed worldwide. By contrast, the filastereans are amoeboid protists that have filopodia. There are only two described filasterean species so far: the marine free-living *Ministeria vibrans*[17,26]; and *Capsaspora owczarzaki*, which was isolated from a freshwater snail from Puerto Rico and Brazil[27–29]. It was originally suggested that *C. owczarzaki* would provide the snail with resistance to *Schistosoma mansoni* infections by destroying the sporocysts of this parasite. However, further attempts to isolate *C. owczarzaki* from *S. mansoni*-resistant snails were unsuccessful, and the exact interaction of *C. owczarzaki*, if any, with this potential host remains unclear[27,30]. Finally, the ichthyosporeans comprise approximately 40 described species, and they are the parasites or commensals of a wide diversity of animals, including humans and many marine invertebrates[31,32]. Nevertheless, metabarcoding analyses suggest the presence of free-living ichthyosporean species[33]. Most ichthyosporeans reproduce through multinucleated coenocyte colonies and have a wide diversity of dispersal stages, including flagellated and amoeboid forms[34,35].

Based on our latest view of the phylogenetic framework of animals within the eukaryotic clade, it is clear that when investigating the origin of animals, all three of the closest relatives of animals — choanoflagellates, filastereans and ichthyosporeans — should be taken into consideration. Obtaining detailed genomic, regulatory and cell-biological information about representative members of the various groups is likely to yield insights into the transition from unicellularity to metazoan multicellularity.

The unicellular urmetazoan genome

Understanding the transition from unicellularity to multicellularity requires knowledge of the types and extent of genomic innovation that preceded and accompanied this transition. For instance, if the unicellular ancestor of animals contained few genes involved in multicellular processes, it could be inferred that a key event in the evolution of metazoan multicellularity would be the evolution of animal-specific genes. By contrast, if the unicellular ancestor had many of the genes involved in multicellular development and physiology, then the evolution of multicellularity is likely to have involved the co-option of existing genes to perform new functions. To distinguish between these possibilities, however, we need to elucidate the gene content of the unicellular ancestor. By determining which genes and genetic pathways are shared between animals and their relatives, it is then possible to infer which genes and genetic pathways were present in the ancestor.

We now have the complete genome sequences of four unicellular holozoans, which represent each of the three unicellular lineages that are most closely related to animals. These include the genomes of two choanoflagellates (*Monosiga brevicollis* and *Salpingoeca rosetta*), one filasterean (*C. owczarzaki*) and one ichthyosporean (*Creolimax fragrantissima*)[36–39] (FIG. 3a). This rich dataset enables us to reconstruct the gene content of the unicellular ancestor of animals with an unprecedented level of detail. The results have been quite surprising. Although there was gene innovation at the onset of Metazoa, the unicellular ancestor of animals already had a rich repertoire of genes that are required for cell adhesion, cell signalling and transcriptional regulation in modern animals (FIG. 3b).

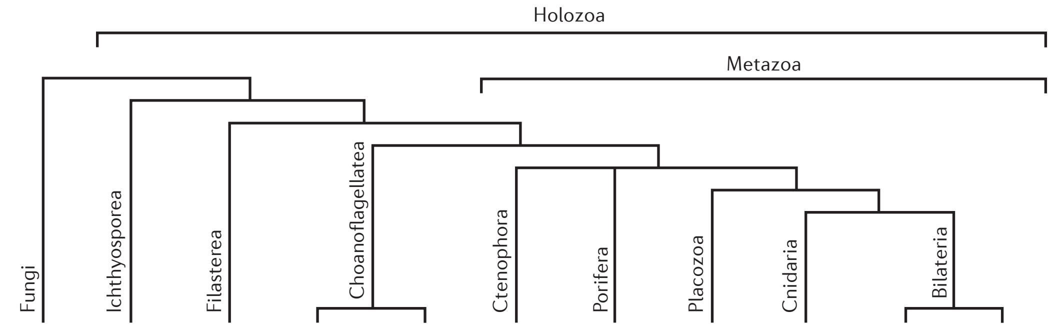
The first example is that of genes encoding cell adhesion proteins, which are necessary for cell–cell and cell–matrix interactions in the formation of cell layers, tissues and the extracellular matrix (ECM) in animals. Genome sequence analysis of unicellular holozoans indicates that the unicellular ancestor of animals already had several mechanisms of cell adhesion, both for cell–cell and cell–ECM adhesion (FIG. 3b). For example, there are approximately 20–30 predicted cadherin domain-containing proteins encoded in the genomes of choanoflagellates[40]. However, classic animal cadherins, which are regulated by β-catenin and are involved in cell–cell adhesion, seem to be metazoan specific. Moreover, *C. owczarzaki* has a complete integrin adhesome, which is a major cell–ECM adhesion system in animals[41]. Other ECM-related proteins are present in unicellular holozoans, and they include several components of the dystrophin-associated glycoprotein complex and multiple ECM-related protein domains, such as laminins, collagens and fibronectins[39]. Finally, C-type lectins, which are involved in cell–cell adhesion, are present in choanoflagellates[36]. Overall, the presence of animal cell

Filopodia
Thin, actin-based cellular projections that are used in environmental sensing and cell motility.

Metabarcoding
Analysis of species or lineage diversity in pooled environmental samples by sequencing of a standardized, common region of DNA, usually the gene encoding the 18S ribosomal RNA.

Coenocyte
A multinucleated cell resulting from successive nuclear divisions (karyokinesis) without associated cytokinesis.

a



	Scer	Cfra	Cowc	Sros	Mbre	Mlei	Aque	Tadh	Nvec	Dmel	Hsap
Genome size (Mb)	12.10	45	28	55.40	41.60	155.90	166.70	98	356.60	168.70	3,095
Number of predicted genes	6,692	8,694	8,657	11,629	9,171	16,548	40,122	11,520	27,273	26,951	22,447
Gene density (genes per Mb)	553	193	309	210	220	106	241	118	76	160	7
Mean intergenic size (bp)	544	1,563	738	885	1,596	3,561	2,481	5,093	6,959	5,542	111,359
% coding sequence in genome	74.58	30.34	58.70	41.69	39.56	14.17	20.98	15.88	6.94	13.23	1.20
Mean number of introns per gene	0.10	6.50	4.00	7.50	6.60	4.50	4.20	7.40	4.30	4.90	6.30
Mean intron size (bp)	271	303	157	252	171	891	326	283	798	1,509	5,923
%GC content	38.10	40.30	53.80	56.00	54.90	38.90	35.80	32.70	40.60	41.70	40.90

b

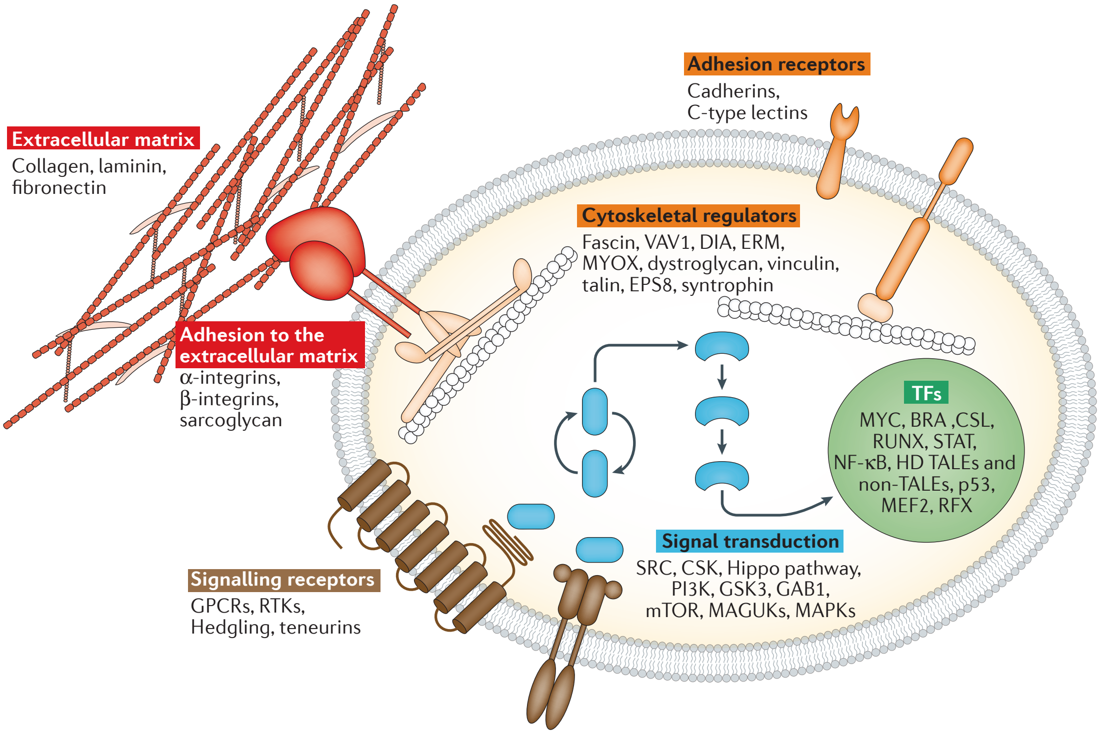


Figure 3 | **Comparative genomics of Holozoa. a** | The genomic features of animals and their unicellular relatives. Animal genomes contain a larger proportion of non-coding sequences than do the genomes of other Holozoa, which might be related to the amount of regulatory information encoded. Other differences are the larger size and higher number of predicted genes in animal genomes relative to the genomes of other Holozoa, although both these features show high variability across species. Data from REFS 38,39,142. **b** | An inferred gene repertoire of the unicellular ancestor of animals. Many genes that are important for metazoan multicellularity-related functions — such as adhesion, signalling and transcriptional regulation — evolved in a unicellular context and were present in the unicellular ancestor of animals. The inference is based on the presence of homologues of these metazoan genes in the genomes of unicellular relatives of animals. The origin of the animal gene repertoire is comprehensively described elsewhere[39,105]. Aque, *Amphimedon queenslandica* (Porifera); BRA, Brachyury; Cfra, *Creolimax fragrantissima* (Ichthyosporea); Cowc, *Capsaspora owczarzaki* (Filasterea); DIA, diaphanous; Dmel, *Drosophila melanogaster* (Bilateria); EPS8, epidermal growth factor receptor kinase substrate 8; GAB, GRB2-associated binding protein; GPCRs, G protein-coupled receptors; GSK3, glycogen synthase kinase 3; HD, homeodomain; Hsap, *Homo sapiens* (Bilateria); MAGUKs, membrane-associated guanylate kinases; MAPKs, mitogen-activated protein kinases; Mbre, *Monosiga brevicollis* (Choanoflagellatea); MEF2, myocyte-specific enhancer factor 2; Mlei, *Mnemiopsis leidyi* (Ctenophora); mTOR, mechanistic target of rapamycin; MYOX, myosin X; NF-κB, nuclear factor-κB; Nvec, *Nematostella vectensis* (Cnidaria); PI3K, phosphatidylinositol 3-kinase; RTKs, receptor tyrosine kinases; Scer, *Saccharomyces cerevisiae* (Fungi); Sros, *Salpingoeca rosetta* (Choanoflagellatea); STAT, signal transducer and activator of transcription; Tadh, *Trichoplax adhaerens* (Placozoa); TALEs, three amino acid loop extensions; TF, transcription factor.

adhesion proteins — including integrins, C-type lectins and cadherins — in extant unicellular holozoans indicates that these adhesion mechanisms were not animal innovations.

Signal transduction is another key requirement for metazoan multicellularity. Several key developmental signalling pathways, such as Hedgehog, WNT, transforming growth factor-β (TGFβ), JAK–STAT and Notch, are highly conserved across Metazoa (with the possible exception of most components of the Notch and Hedgehog pathways, which are absent in ctenophores[6,7]) and are not found in non-metazoans[8,36,39,42]. In other cases, similar signalling receptors are present in the genomes of unicellular holozoans (FIG. 3b). The best-studied case is that of the receptor tyrosine kinases (RTKs). Choanoflagellates, filastereans and ichthyosporeans have dozens of independently evolved RTKs, none of which seems to be homologous to the RTKs of each other or any metazoan RTKs (that is, metazoan RTKs are a fourth independent expansion of RTKs in Holozoa)[43–45]. By contrast, some orthologues of cytoplasmic tyrosine kinases — such as SRC, focal adhesion kinase (FAK) and CSK — are present in both animals and unicellular holozoans. Another conserved premetazoan signalling mechanism is the Hippo signalling pathway, which is present in *C. owczarzaki*[46]. In this case, once again, the intracellular components of the pathway are conserved, whereas the known metazoan upstream receptors, Crumbs and Fat, are absent. Thus, it seems that although some metazoan intracellular signalling components were present in the unicellular ancestor of animals, in most cases their upstream receptors and ligands evolved after metazoans diverged from unicellular holozoans.

Finally, comparative genomics studies have also shown that a considerable proportion of the transcription factor (TF) repertoire of animals was already present in the holozoan unicellular ancestor (FIG. 3b). This includes some TF classes that were previously thought to be metazoan specific (such as nuclear factor-κB (NF-κB), p53, RUNX and T-box)[47,48]. New TF classes — such as the ETS, SMAD, nuclear receptor, Doublesex and interferon-regulatory factor (IRF) families — evolved at the root of Metazoa, and many new families appeared within specific TF classes, including the T-box, SOX, homeobox and basic helix–loop–helix (bHLH) families (reviewed in REF. 49).

Some of these TFs, along with the integrin adhesome[41], were secondarily lost in choanoflagellates[47]. These are two cautionary examples of the limitation of reconstructing ancestral gene content on the basis of very few species and lineages. Therefore, we cannot rule out that secondary loss is not obscuring more ancient evolutionary origins for some gene families.

The finding that many key genes involved in animal multicellularity and development were already present in the unicellular ancestor of animals suggests that the co-option of ancestral genes into new functions was an important mechanism in the evolution of animal multicellularity. That is, many of the genes that currently function within multicellular animals evolved within a unicellular context and were subsequently repurposed for multicellularity. This co-option of the ancestral gene repertoire, together with the evolution of novel animal genes (for example, ligands and receptors), and a substantial expansion and diversification of some ancestral gene families, configured the gene toolkit for animal multicellularity.

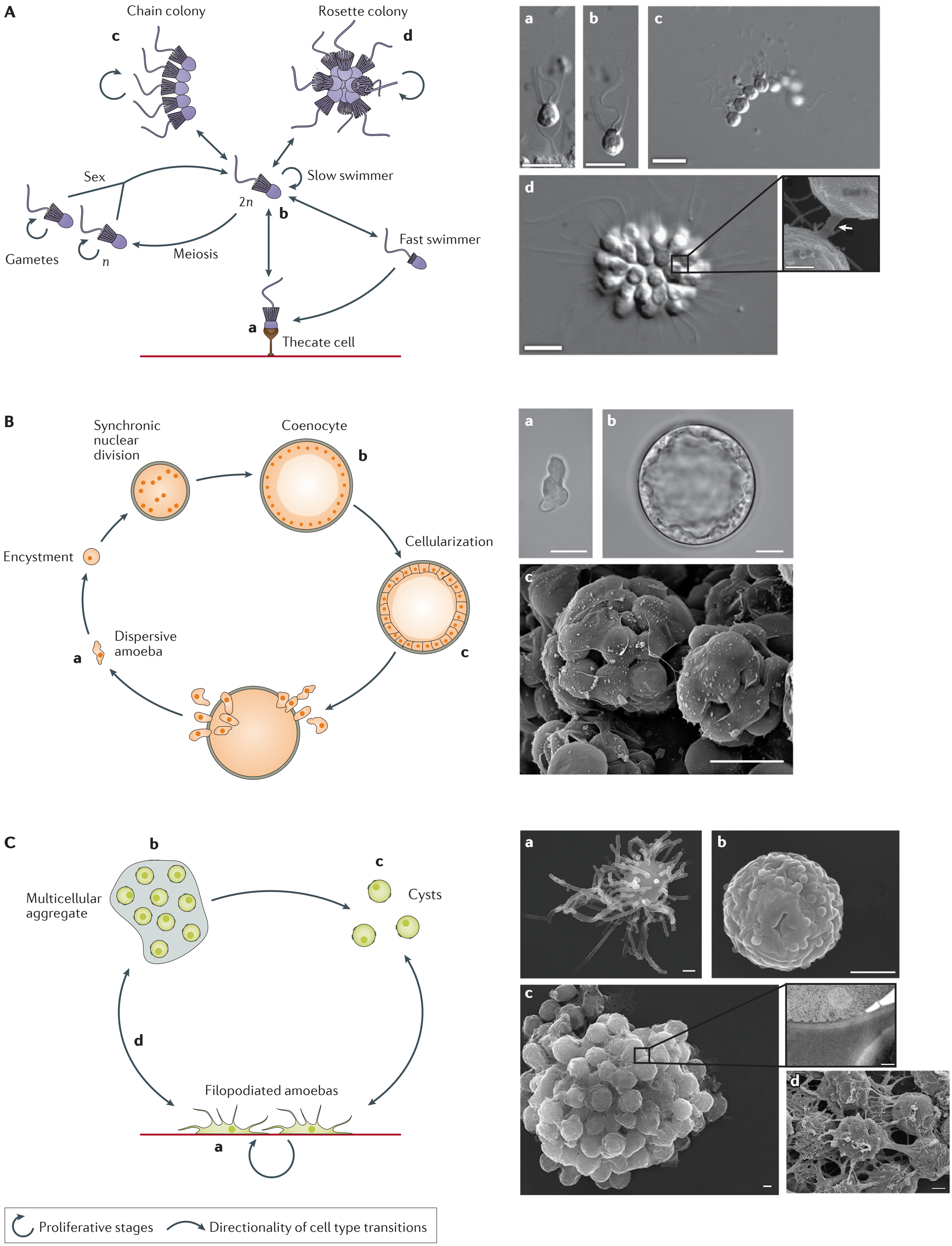
Orthologues
Genes in different species that are descended from a common ancestral gene through a speciation divergence event.

Unicellular holozoan cell types

Although reconstructing the gene content of the protistan ancestor of animals is an important step towards understanding the emergence of animals, gene content alone is not sufficient to provide insights into the cell biology, life cycle and regulation capabilities of the unicellular ancestor. This requires analysis of the biology of the extant unicellular relatives of animals. To this end, a few unicellular holozoan species are being intensively studied and are emerging as candidate model systems in which to address the origin of animals.

Choanoflagellate life stages. *S. rosetta* is the best-studied choanoflagellate species. Analysis of the *S. rosetta* life cycle has revealed that this choanoflagellate forms colonies by clonal cell division[50] and that cells within the colony are linked by cellular bridges[51] (FIG. 4A). Interestingly, colony formation in *S. rosetta* is triggered by

Figure 4 | **Unicellular holozoan life cycles.** Each panel shows the life cycle and temporal cell type transitions in one unicellular holozoan species. For each organism, the life cycle is shown schematically on the left, and microscopy images of the corresponding stages are shown on the right. **A** | The life cycle of the colonial choanoflagellate *Salpingoeca rosetta*. The cycle includes a single-celled sessile thecate stage, slow and fast swimming single-celled stages, and two types of colonial stage — chain and rosette colonies — in which cells are linked by intercellular bridges (indicated by a white arrow in the electron microscopy image in the right panel of part **d**). Starvation triggers the *S. rosetta* sexual cycle, in which diploid cells undergo meiosis and recombination, and the resulting haploid cells (which can also divide asexually) mate anisogamously. Scale bars represent 5 μm, except in the right panel of part **d** where the scale bar represents 1 μm. **B** | The life cycle of the ichthyosporean *Creolimax fragrantissima*. Single-nucleated amoebas disperse until they find a spot to settle and encyst. The cell then undergoes multiple rounds of synchronic nuclear division without intervening cytoplasmic division. The nuclei are placed in the periphery as a large central vacuole grows. Finally, the coenocyte cellularizes, and new amoebas are released. Scale bars represent 10 μm, except in part **c** where the scale bar represents 50 μm. **C** | The life cycle of the filasterean amoeba *Capsaspora owczarzaki*. The trophic proliferative stage is an amoeba with long thin actin-based filopodia. These amoebas can aggregate and produce an extracellular matrix that binds them together. In addition, aggregated cells and amoebas can transform into a cystic resistance form. Scale bars represent 1 μm, except in part **d** where the scale bar represents 200 nm. Part **A** is adapted with permission from REF. 51, Elsevier. Part **B** is adapted with permission from REF. 38, eLife Sciences Publications. Part **C** is adapted with permission from REF. 57, eLife Sciences Publications.


A
Chain colony
c
Rosette colony
d
Sex
Slow swimmer
2n
b
Gametes
n
Meiosis
Fast swimmer
a
Thecate cell
a
b
c
d
B
Synchronic
nuclear
division
Coenocyte
b
Cellularization
Encystment
c
Dispersive
amoeba
a
a
b
c
C
b
Multicellular
aggregate
c
Cysts
d
Filopodiated amoebas
a
a
b
c
d
Proliferative stages
Directionality of cell type transitions

the presence of its bacterial prey, *Algoriphagus machipongonensis*, and more specifically, by a sulfonolipid molecule produced by this bacterium[52]. This observation suggests a deep evolutionary link between bacterial prey capture and early animal multicellularity. Additional life stages of *S. rosetta* include a sessile thecate form, and slow and fast swimmer stages[51] (FIG. 4A). *S. rosetta* also has a sexual life cycle that is triggered by nutrient starvation and involves morphologically differentiated gametes[53]. The presence of sexual reproduction and gametogenesis in choanoflagellates suggests that these processes were present in the unicellular ancestor of animals.

RNA sequencing (RNA-seq) analysis of *S. rosetta* has revealed highly specific transcriptome profiles that are associated with the different life stages[37]. Differentially upregulated genes include those encoding septins in the colonial stage and those encoding different cadherin domain-containing proteins in the colonial and sessile stages. In a recent random mutagenesis screen[54], a C-type lectin was identified as essential for colony formation in *S. rosetta*, providing the first direct gene-to-phenotype link in a unicellular holozoan.

Investigation of another choanoflagellate, *M. brevicollis*, has provided insights into the premetazoan function of cadherins. Two *M. brevicollis* cadherins localize in the microvilli feeding collar and colocalize with the actin cytoskeleton[55]. *M. brevicollis* is a strictly solitary choanoflagellate species, which suggests a role for choanoflagellate cadherins in prey capture. In line with this, none of the studies in the colonial choanoflagellate *S. rosetta* has linked cadherins to colony formation, which further supports the idea of a non-cell–cell adhesion role for cadherins in extant choanoflagellates and, potentially, in the unicellular ancestor of animals.

Ichthyosporean life stages. *C. fragrantissima* is a promising ichthyosporean model system in which tools are available for transient genetic transformation[56]. The life cycle of *C. fragrantissima* (and other ichthyosporeans) is very different from that of choanoflagellates (FIG. 4B). It starts with a mononucleated cell that undergoes multiple rounds of synchronic nuclear division, which results in a large, sessile multinucleated coenocyte (with a diameter of 70–80 μm) that has a cell wall, nuclei localized to the cell periphery and a large central vacuole[35,56]. The rapid cellularization of the coenocyte is followed by the release of mononucleated amoeboid cells, which are highly motile and do not divide. The amoebas disperse until they find a clear spot to settle, where they encyst and begin a new cycle (FIG. 4B). A recent analysis of two life stages of *C. fragrantissima* (multinucleated and amoeba) has shown specific transcriptomic profiles for each stage that involve hundreds of differentially expressed genes[38]. For example, the integrin adhesome and the TF Brachyury are significantly upregulated in the amoeba stage, whereas DNA replication-related and translation-related genes are upregulated in the colonial stage.

Filasterean life stages. Among Filasterea, only the life cycle of *C. owczarzaki* has been described[57] (FIG. 4C). Although *C. owczarzaki* was originally reported as an endosymbiont of a freshwater snail, recent observations suggest that this amoeba may be able to phagocytose bacteria and grow as a bacterivore, similarly to its sister species *M. vibrans* (I.R-T. and colleagues, unpublished observations). The *C. owczarzaki* life cycle includes three different cell stages: an amoeboid stage, a cystic stage and an aggregative multicellular stage. In the amoeboid stage, cells have long thin actin-based filopodia[58], and the amoeba represents the proliferative and phagocytic trophic stage. Upon starvation, *C. owczarzaki* cells retract their filopodia and encapsulate, forming a cystic resistance form. Finally, amoeboid cells can join and form multicellular aggregates, in which cells produce an ECM that holds them together without the need for direct cell–cell contact. This represents the only known case of aggregative multicellularity in Holozoa (FIG. 1).

Transcriptomic analysis of the *C. owczarzaki* life cycle showed that temporally regulated cell differentiation is linked to specific transcriptional profiles[57]. This differential gene regulation involves hundreds of genes, and among them, there are many *C. owczarzaki* homologues of genes that are essential for animal multicellularity. For example, aggregate-stage cells strongly co-express integrin adhesome genes, as well as ECM proteins, including fibronectin domain-containing and laminin domain-containing proteins; by contrast, in the filopodial stage, genes related to actin cytoskeleton, filopodia formation and tyrosine kinase signalling are overexpressed. Moreover, differentially regulated alternative splicing is linked to cell type transitions in *C. owczarzaki* and further contributes to temporally regulated gene expression. A more recent study analysed proteome and phosphosignalling dynamics during the *C. owczarzaki* life cycle using high-throughput proteomics[59]. This study showed that extensive proteome remodelling and hundreds of dynamic phosphosignalling events underlie temporally regulated cell differentiation in *C. owczarzaki*. Interestingly, dozens of tyrosine kinases — including orthologues of several cytoplasmic tyrosine kinases (such as SRC, ABL and TEC), as well as structurally diverse RTKs — were shown to be phospho-activated in a cell type-specific manner. Moreover, multiple TFs seemed to be phosphoregulated, and the Hippo pathway was activated specifically during the aggregative stage. These results further support the idea that elaborate transcriptional, post-transcriptional and phosphosignalling-mediated regulation already existed in the protistan ancestor of animals.

Overall, the diversity of morphologies and cell behaviours in extant unicellular holozoans suggests that the unicellular ancestor of metazoans was a bacterivore that displayed sexual reproduction and multiple temporally differentiated cell types[12]. Most likely, these transitions between the different cell states were tightly regulated by the differential expression of conserved TFs, such as Brachyury, and were triggered by environmental conditions such as nutrient starvation and the presence of bacterial prey.

Regulatory innovation at animal origins

Specific effector gene modules and their immediate upstream regulators define the functional specificity of a

Effector gene
A gene that is related to structural and metabolic cellular functions (for example, enzymes or cytoskeletal proteins), as opposed to a regulatory gene.

given cell type[60] and, as we have described, many of the components of these gene modules evolved in premetazoan lineages. However, as Eric H. Davidson stated in his 2006 book *The Regulatory Genome*[61] "Differentiation gene batteries do not make body plans". Indeed, animal development ultimately depends on the finely regulated spatiotemporal deployment of these effector gene batteries to generate individual cell phenotypes and collective multicellular structures. This dynamic definition of regulatory states is orchestrated by large hierarchical gene regulatory networks (GRNs) and epigenetic mechanisms of cellular memory. The key question here is whether or not these GRNs and epigenetic mechanisms were already present in the unicellular ancestor of animals.

Signalling novelties at the root of Metazoa. Signalling genes are essential elements of metazoan GRNs. In unicellular organisms that display temporally regulated cell type differentiation, environmental cues — such as nutrient deprivation and hypoxia — trigger cell type transformations[62,63]. The control of differentiation through metabolism or chemical gradients is arguably much less prominent in modern metazoans than in unicellular organisms. Instead, cell-to-cell and long-range signalling mechanisms are the most important determinants of cell fate in modern metazoans. Interestingly, most of these signalling pathways originated at the root of Metazoa (see above)[8,36,39] and are essential for body patterning in early-branching animals[10,64]. Thus, these developmental signalling pathways seem to have been important for the evolution of cell type determination and morphogenesis in animals.

The emergence of a metazoan TF toolkit. TFs are the other major players in animal developmental GRNs. Available data suggest that the establishment of the metazoan TF regulatory toolkit resulted from a combination of five processes. First, ancient TF classes — particularly those with highly variable DNA-binding specificities, such as homeoboxes and zinc fingers[65,66] — were vastly expanded by gene duplication at the onset of Metazoa, originating dozens of new families[67]. Second, pre-existing TF classes were co-opted into new functions. This happened to many TFs of holozoan origin, such as T-box TFs, RUNX TFs, p53 and NF-κB. Third, new TF classes — such as the ETS, SMAD, IRF and nuclear receptor families — originated at the root of animals. Fourth, changes in TF–TF and TF–cofactor interactions expanded the number of combinatorial regulatory binding specificities. For example, the basic leucine zipper (bZIP) dimerization preferences of several species, including the choanoflagellate *M. brevicollis*, have been analysed *in vitro*[68]. This study found that the proportion of heterodimeric bZIP interactions (that is, interactions between bZIP TFs belonging to different families) was greatly increased in multicellular species relative to their unicellular relatives. This indicates that the remodelling of TF dimerization networks is a potential way to increase the number of regulatory outputs from a limited set of TFs. Finally, ancient TF-specific downstream regulatory networks were probably co-opted and modified. Whereas *cis*-regulatory element sequence conservation is quickly eroded during evolution, specific regulatory connections tend to be more conserved[69].

What remains unclear is the extent to which such TF regulatory connections might be inferred by comparing extant unicellular relatives with early-branching animals. A recent study suggests that, at least for some TFs, such inference might be possible. In the filasterean *C. owczarzaki*, orthologues of Brachyury and MYC were shown to regulate genes involved in cell motility and cell proliferation, respectively, indicating that they function similarly to their metazoan orthologues[70]. Moreover, MYC from the choanoflagellate *M. brevicollis* has similar interacting partners (the bHLH TF MAX) to those of metazoan MYC and uses the same DNA-binding sites through which metazoan MYC proteins act (sites called E-boxes)[71]. These findings further reinforce the idea of a highly conserved premetazoan MYC regulatory network involved in the control of cell proliferation[70]. In summary, the data suggest that the metazoan TF regulatory toolkit evolved through changes both in the TF gene repertoire and in the *cis*-regulatory and *trans*-regulatory interactions. These changes ultimately resulted in the greatly expanded regulatory capabilities of the metazoan TFs.

Distal enhancer regulation in early Metazoa. TFs bind to specific sequences located at gene promoters and, at least in bilaterian animals, distal enhancer elements. Enhancers are clusters of TF-binding sites that have specific chromatin characteristics such as depletion of nucleosomes (open chromatin sites) and particular histone marks (histone 3 lysine 4 monomethylation (H3K4me1) and histone 3 lysine 27 acetylation (H3K27ac)) in the flanking nucleosomes[72–76]. The presence of p300 (also known as EP300), a histone acetyltransferase of holozoan origin[47], has also been used to predict enhancer regions and activity[77]. High-throughput approaches to identify and validate enhancer candidates and test their functions have shown that most enhancer elements in bilaterian animals are distal to (kilobases up to megabases away from) the gene promoters they regulate, and that they act through the physical looping of the chromatin[78–80]. This chromatin looping is mediated by CTCF, cohesin and other structural proteins[81,82]. Enhancer elements generally reside in intergenic regions and, in more compact genomes, in introns; these intronic enhancer elements are often located in genes neighbouring the genes that they regulate. The estimated number of enhancers is in the order of thousands in animals such as *Drosophila melanogaster*[83] and humans[76] (TABLE 1). Moreover, in *D. melanogaster* the vast majority of enhancers show very restricted spatial and temporal activity during development[83,84], emphasizing the importance of enhancer elements in orchestrating complex regulatory states. Another defining feature of *cis*-regulatory enhancer elements is their combinatorial nature and modularity: multiple binding sites occur in each enhancer[85], and regulatory states are generated by the combined action of multiple enhancers on the same gene, particularly in genes encoding TFs and other

Cis-regulatory element
A genomic segment that regulates the transcription of (usually neighbouring) genes on the same chromosome.

Chromatin looping
Physical folding of the chromatin nucleoprotein fibre. It is often associated with regulatory events that involve physical proximity between distal enhancer elements and gene promoters.

Microsyntenic
Describes small genomic regions in which the physical colocalization of loci is conserved between different species.

developmental regulators[86,87]. Overall, the combined action of both distal enhancers and, to a lesser extent, proximal promoter *cis*-regulatory elements underlies the complex spatiotemporal expression patterns observed during bilaterian development.

Although the evolutionary dynamics of enhancers have been extensively studied in some bilaterians[88], the existence of such regulatory elements in other metazoan or premetazoan lineages has remained a mystery. An indirect hint of the possible existence of distal regulation across all metazoans is the deep evolutionary conservation of microsyntenic blocks across Metazoa[89,90]. These blocks comprise a gene (usually a developmental gene such as one encoding a TF or signalling protein) that is linked to another functionally unrelated neighbouring bystander gene. This linkage is often due to the presence of regulatory elements in the bystander gene. Interestingly, no conserved microsyntenic blocks have been found between animals and their unicellular relatives[89], which suggests that distal regulation evolved at the root of Metazoa.

The first direct experimental evidence for the evolutionary conservation of the epigenetic regulatory landscape beyond Bilateria came from a landmark study in the cnidarian *Nematostella vectensis*[91] (TABLE 1). Approximately 6,000 enhancers were predicted in *N. vectensis*, and they showed similar chromatin signatures (H3K4me1, K3K27ac and the presence of the histone acetyltransferase p300) to those of bilaterian enhancers. Confirming these predictions, 12 of these *N. vectensis* enhancer elements showed activity in *in vivo* reporter assays. Moreover, *N. vectensis* enhancers were found to be particularly enriched close to TF-encoding genes, which suggests the existence of complex TF combinatorial regulatory networks. What remains unknown, however, is whether these *N. vectensis* enhancers work through chromatin looping or through non-looping proximity mechanisms. The latter is suggested by the absence of CTCF, a key protein for chromatin looping, in *N. vectensis* and other non-bilaterian animals[92]. Nevertheless, a non-looping proximity mechanism is inconsistent with the lack of dependence of enhancer activity on enhancer orientation or the position of the enhancer relative to the promoter in *N. vectensis* reporter assays[91]. Moreover, the more ancient cohesin complex (which is present in all animals and most eukaryotes) seems to be key to enhancer looping[81,82], and it has recently been shown that several different structural proteins, but notably not CTCF, are associated with enhancer–promoter chromatin loops in *D. melanogaster*[93,94]. Therefore, it is possible that enhancer–promoter looping in *N. vectensis* occurs, even in the absence of CTCF, via looping mediated instead by cohesin and/or other structural proteins.

Multicellularity and chromosomal architecture. Even though the gene regulatory landscape of *N. vectensis* is similar to that of Bilateria, it remains unclear whether global genome organization may differ between bilaterians and non-bilaterians. In particular, it is unclear whether the genome of non-bilaterians is organized into structural chromatin territories, such as topologically associating domains (TADs), which are found in bilaterian genomes[95–98]. The discretization of the genome into broad structural domains such as TADs allows the existence of autonomous regulatory blocks, which have similar prevalent chromatin features (for example, active

Table 1 | **Genome regulatory features in animals and unicellular relatives**

	Saccharomyces cerevisiae	*Capsaspora owczarzaki*	*Nematostella vectensis*	*Drosophila melanogaster*	*Homo sapiens*
Clade	Fungi	Filasterea	Cnidaria	Bilateria	Bilateria
Multicellularity	Unicellular	Unicellular	Multicellular	Multicellular	Multicellular
% non-coding genome	25.4	58.7	93.1	86.8	98.8
Number of gene deserts (50 kb)	0	1	474	1,033	45,248
Mean intergenic distance (bp)	544	738	6,959	5,542	111,359
Number of transcription factors	95	143	544	497	1,012
H3K27me3/PRC2	–/–	–/–	?/+	+/+	+/+
High-order chromatin structures (TADs and compartments)	–	?	?	+	+
Chromatin loops/CTCF	–/–	?/–	?/–	+/+	+/+
DNA methylation	–	–	+	–	+
Promoter types	II?	II	?	I, II and III	I,II and III
Number of open chromatin sites	4,897	11,927	?	35,507–45,825	2,890,742
Number of enhancer elements	0	0	5,747	50,000–100,000	41,011–399,124
Number of miRNAs	0	0	87	147	677
Number of lincRNA loci	63	652	?	1,119	8,195

+, present; –, absent;?, unknown; H3K27me3, histone H3 lysine 27 trimethylation; lincRNAs, long intergenic non-coding RNAs; miRNAs, microRNAs; PRC2, Polycomb repressive complex 2; TADs, topologically associating domains. Data are derived from REFS 39,67,75,76,83,143–151.

or repressive TADs) and within which looping regulatory interactions can occur[99,100]. CTCF is regarded as an essential factor for TAD formation, and therefore the origin of CTCF in Bilateria might be related to a bilaterian-specific genome structural compartmentalization. Conversely, alternative structural proteins could be contributing to the formation of TADs or similar structures in *N. vectensis* and early-branching metazoans. Thus, deciphering and comparing the 3D genome architectures of non-bilaterian animals will be key to resolving the evolutionary link, if any, between the origin of Metazoa and the emergence of specific mechanisms for genome compartmentalization and folding.

Unicellular holozoan regulatory genomes. Once genomic regulatory features of bilaterians had been observed in non-bilaterian animals, the obvious question was whether this could be further extended to premetazoan lineages. The filasterean *C. owczarzaki* is an ideal candidate in which to investigate this, given that it has a well-described pattern of temporally regulated cell differentiation, and that it has the richest repertoire of metazoan-like TFs (see above) among the unicellular relatives of animals. Analysis of the *C. owczarzaki* regulatory genome showed that temporally differentiated cell types in *C. owczarzaki* are associated with changes in chromatin states and also with thousands of dynamic *cis*-regulatory elements (as defined by accessible chromatin profiling)[70]. These *cis*-regulatory elements are mostly proximal to the transcription start site (in promoter regions and first introns). They are small and, even when they are distal, lack any of the chromatin features associated with animal enhancers. Thus, these results suggest that distal enhancer elements might indeed be an animal innovation. Similar studies are now required in sponges, ctenophores and other unicellular holozoans to precisely define when this innovation occurred.

Another important finding is the absence of repressive marks such as histone H3 lysine 9 trimethylation (H3K9me3; which marks constitutive heterochromatin) and H3K27me3 (which marks inducible developmental enhancers and promoters) in *C. owczarzaki*[70]. Although they remain to be explored in non-bilaterians, these marks are common in bilaterians. By contrast, active promoter chromatin signatures (such as H3K4me3 and H3K27ac) found in *C. owczarzaki* are similar to those observed in *N. vectensis*, bilaterians and also other eukaryotes. Although unidentified *C. owczarzaki*-specific repressive marks might exist, these results suggest that repressive epigenetic states might be an important precondition of animal multicellularity, as they would progressively restrict differentiation potential and maintain the differentiated cell state by 'locking' genomic regions or specific genes.

In summary, a major innovation in the transition to animal multicellularity was the emergence of mechanisms (inductive signals and genome regulation) that allowed for the spatiotemporal deployment of effector gene batteries, many of which already existed in the unicellular ancestor. The specific nature of these mechanisms remains to be fully resolved, but likely candidates include the decoupling of cell differentiation from environmental cues, the expansion of TF regulatory capabilities, the emergence of a combinatorial regulatory lexicon mediated by distal enhancer elements, the evolution of repressive chromatin states and the evolution of a hierarchically organized chromosomal architecture.

Chromatin states
Unique combinations of histone post-translational modifications and chromatin-associated proteins that define different biochemical activities of the genome.

Urmetazoa: a mosaic of premetazoan cell types?

The data generated in recent years have allowed us to reconstruct the nature of the unicellular ancestor of animals in some detail. Based on this, we can now revisit the question of how this ancestor evolved into the first animal. One of the oldest and most enduring evolutionary theories on the origin of animals is Ernst Haeckel's Gastrea theory[101]. Haeckel proposed that the first step in the evolution of animal multicellularity was a hollow ball of identical flagellated cells, which he called a blastea. With some modern adaptations, such as the 'choanoblastea' theory (which emphasizes the resemblance of Haeckel's blastea to a choanoflagellate colony)[102], Haeckel's model is still the most widely accepted explanation of the emergence of animal multicellularity[103–105]. An important assumption of the Gastrea model is that cell differentiation appeared after the origin of multicellularity and, therefore, that there was a single founding cell type in animals (a choanoflagellate-like cell, according to many authors). Some of these interpretations arise from the possible homology between sponge choanocytes and choanoflagellates[106], whereas others simply assume that ontogeny recapitulates phylogeny[103]; that is, that successive developmental stages in an animal ontogeny resemble the evolutionary history of that particular species (for example, early stages would resemble the urmetazoan ancestor).

The recent discoveries reviewed here provide evidence for a potential alternative scenario. First, many genes that are essential to animal multicellularity originated in a unicellular context. Second, several close unicellular animal relatives have complex life cycles that include different cell types and temporally regulated multicellular behaviours[37,56,57]. Moreover, the temporally regulated cell type differentiation of animal relatives is associated with specific transcriptional profiles[37,38,57] that are supported by changing chromatin states[70] and extensive remodelling of signalling states such as phosphosignalling networks[59]. All of these findings indicate that the cell states of premetazoans are the result of bona fide differentiation processes that involve changes in morphology, motility and so on, rather than simply being examples of temporary phenotypic plasticity or changing metabolic states. It is also important to emphasize that these temporally regulated cell transitions in unicellular holozoans are directional (that is, not all cell types can give rise to all cell types) and that only specific cell types are proliferative (for example, only the filopodial stage in *C. owczarzaki*). Finally, there is evidence that major changes in genome regulation occurred at the root of Metazoa, raising the possibility of further expansion and elaboration of regulatory capacity.

Based on the evidence obtained in recent years, we can propose a hypothesis to explain the origin of Metazoa that is rooted in the Synzoospore hypothesis

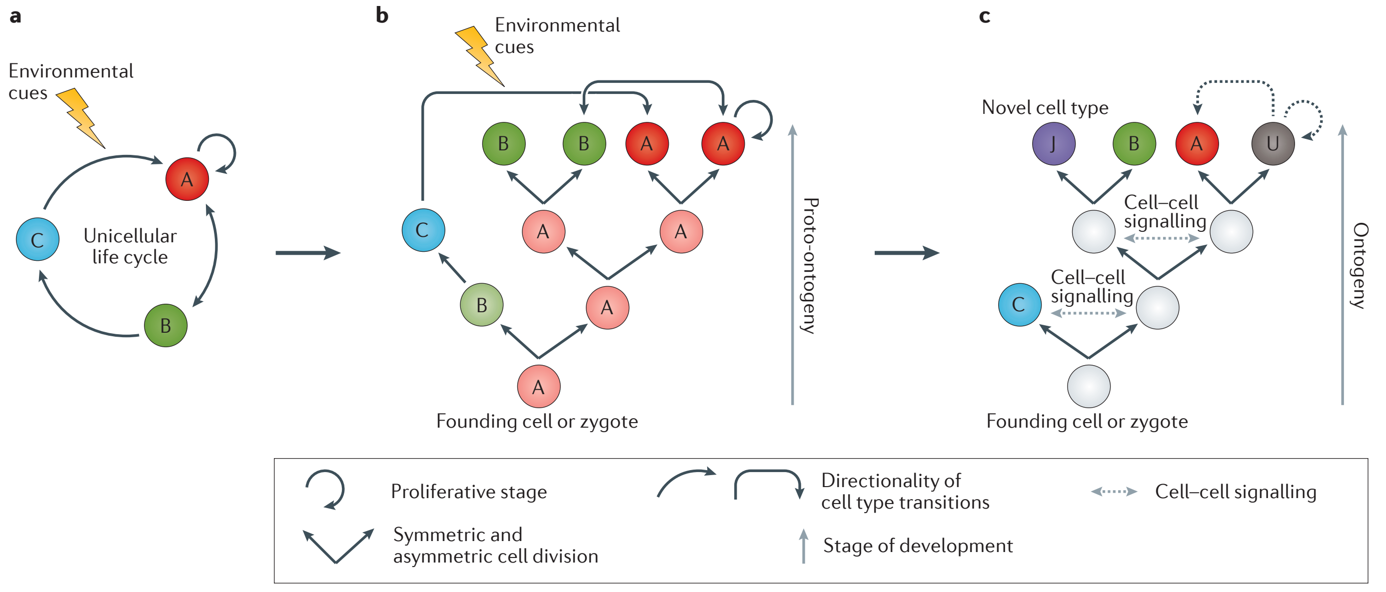


Figure 5 | **A model of the origin of multicellularity: a transition from temporal to spatiotemporal cell differentiation. a** | The complex life cycle of the unicellular ancestor of animals. One cell type (labelled 'A') had proliferative capacity, and both this proliferation and directional cell type transitions were controlled by environmental stimuli. **b** | The pre-existing temporally regulated cell types were integrated into a proto-ontogeny, in which cell differentiation polarities (for example, A differentiates into B, B into C and so on) were the same, and one cell type (in this case A) was proliferative and gave rise to all the others. Environmental cues control terminal cell type (A, B and C) transdifferentiation. **c** | In a later stage, a closed cellular ontogeny controlled by endogenous cell–cell signalling generated stably differentiated cell types, the cell fate of which is maintained by mechanisms of cellular memory and, therefore, no transdifferentiation or very limited transdifferentiation existed. Some cell types (U) retained the ability to proliferate and to differentiate into other cell types (black dashed arrows), but only under endogenously regulated conditions. New cell types progressively evolved (J).

that was originally proposed by Zakhvatkin in 1949 (REF. 11) and was later further developed by Mikhailov *et al.*[12]. In this scenario, Metazoa arose from an ancestral protist with a complex life cycle that involved multiple temporally regulated cell states (FIG. 5a). This life cycle was dependent on environmental stimuli, and probably comprised one or more multicellular life stages (clonal and/or aggregative) and sexual reproduction, as observed in some extant unicellular holozoans. These temporally regulated cell types would become spatially integrated into the first metazoans (FIG. 5b), concomitant with the evolution of additional mechanisms for complex gene regulation. Innovation in signalling pathways, the expansion of the TF repertoire and the evolution of new genome regulatory mechanisms are likely to have been key to providing additional control over the spatial integration of pre-existing cell type-specific gene modules.

The initial coexistence of diverse cell types in the Urmetazoa was associated with limited morphogenetic programmes (for example, simple layering by differential expression of adhesion molecules or the creation of internal spaces for nutrient diffusion) and, probably, labile cell fates; such multicellular organization is observed in extant sponges, in which high rates of cell transdifferentiation exist[107] (FIG. 5b). Also, these cells were probably maintained together within an ECM that was produced by one or more of these cell types. In this initial proto-ontogeny, cell type differentiation polarities were probably similar to those in the unicellular ancestor (that is, not all cell types would be proliferative and not all cell type transitions would be possible), and limited cellular memory existed. Organism size and the total number of cells could be controlled by pre-existing cell proliferation mechanisms that had been co-opted, such as those mediated by MYC, p53 or the Hippo pathway.

Progressively, cell differentiation would become independent of environmental stimuli, and cell identity could instead be established and maintained by developmental regulatory programmes that were initiated by cell–cell communication pathways that evolved in the animal stem lineages (FIG. 5c). These closed ontogenetic cell trajectories would proceed through undifferentiated cell states, and epigenomic mechanisms of cellular memory would be essential to maintain differentiated cell fates, while some cell types might retain broad cellular potential (FIG. 5c). It is likely that the coupling of these ontogenetic differentiation trajectories with morphogenesis (for example, during gastrulation and mesoderm specification in some bilaterians) led to the emergence of the first animal body plans and developmental programmes[108]. New cell types would continuously evolve in the first animal lineages (FIG. 5c), at least in part by a process of subfunctionalization plus innovations that generated sister cell types during evolution[60,109]. This cell type innovation was concomitant with the evolution of new gene modules and TFs[110], and with the co-option of modules usually deployed in other cell types (through the recruitment of their regulatory network).

In summary, we posit that recent results from unicellular holozoans are consistent with regulated cell differentiation existing before the advent of animal multicellularity. Therefore, the first animals probably evolved from a unicellular ancestor that had a complex life cycle, through a transition from temporally regulated to spatiotemporally regulated cell type differentiation[12]. This transition involved the co-option of multiple ancestral gene modules, as well as the evolution of new gene families and genome regulatory mechanisms.

Conclusions and future perspectives

In recent years, the discovery and phylogenetic placing of new unicellular relatives of animals, and the study of their genome content and cell biology, have provided important insights into the nature of the unicellular ancestor of Metazoa. However, vital questions still remain to be answered, and new methodological approaches can help to address them.

The development of tools for nucleic acid delivery (such as electroporation, liposomal transfection, viral vectors and particle bombardment) and genome editing in unicellular holozoans will be a crucial advance if we are to understand the function of multicellularity-related genes in a unicellular context. A first step in that direction has been the development of forward-genetics screens in choanoflagellates[54], which identified a crucial gene for colony formation in *S. rosetta*, and the development of transformation tools in the ichthyosporean *C. fragrantissima*[56]. Future targeted genetic perturbation approaches coupled with systematic phenotypic screens (morphological and behavioural, but also molecular, such as DNA–protein binding or gene expression profiling) will help to elucidate the function of specific gene candidates.

Further insights into the regulatory genome of unicellular holozoan and early-branching metazoan species will be key to pinpointing the evolutionary origins of the mechanisms that are responsible for cell differentiation and cellular memory. Of particular interest will be the study of repressive chromatin states and the analysis of enhancer elements. In addition, the systematic and comparative study of the genome architecture of these lineages (using chromosome conformation capture (3C) techniques) can help to determine when distal regulation through chromatin looping and hierarchical compartmentalization of the genome first evolved. With the steady improvement of chromatin-profiling techniques and analytical tools, we envisage these issues being tackled in the not-too-distant future.

The development of single-cell epigenomic profiling techniques will overcome the problems of population-based approaches that require homogeneous and synchronized cell populations. These approaches will also provide a more refined understanding of epigenomic regulation and its link to gene expression in holozoans[111,112]. It will be more challenging to decipher specific TF regulatory networks and protein–protein interaction networks through direct immunoprecipitation techniques owing to the strong limitation imposed by antibody development in non-model organisms. Several alternatives can help to circumvent this limitation, such as the profiling of open chromatin coupled with *in vitro* TF binding-preference determination[70,113], inference of co-expression networks through single-cell RNA-seq and, if genetic manipulation is available, protein tagging followed by immunoprecipitation, and gene knockout followed by expression profiling. Finally, systematic unbiased characterization of early metazoan cell types through high-throughput single-cell transcriptomics[112] can provide the first glimpses of early multicellular cell type complexity and the regulatory principles that orchestrate it, including master TF regulators and co-expressed gene modules.

Finally, it is important to recognize that our current knowledge of premetazoan lineages is limited to a few species that serendipitously became candidate model systems. However, environmental surveys have revealed that there exist other lineages that are closely related to animals, and that these lineages thrive in marine and freshwater environments[33]. Given that these lineages could provide additional insights into the question of animal origins, efforts to characterize and isolate new holozoan species should continue.

Overall, more than a decade of intense research in unicellular animal relatives has not only yielded important new insights into the question of animal origins but also opened up new research avenues. We expect that these new approaches will help us to further reconstruct the evolutionary path that led from humble protistan beginnings to the complexity of animal life.

1. Hyman, L. H. *The Invertebrates: Protozoa Through Ctenophora* (McGraw-Hill, 1940).
2. Grosberg, R. K. & Strathmann, R. R. The evolution of multicellularity: a minor major transition? *Annu. Rev. Ecol. Evol. Syst.* **38**, 621–654 (2007).
3. King, N. The unicellular ancestry of animal development. *Dev. Cell* **7**, 313–325 (2004).
4. Knoll, A. H. The multiple origins of complex multicellularity. *Annu. Rev. Earth Planet. Sci.* **39**, 217–239 (2011).
5. Putnam, N. H. *et al.* Sea anemone genome reveals ancestral eumetazoan gene repertoire and genomic organization. *Science* **317**, 86–94 (2007).
6. Moroz, L. L. *et al.* The ctenophore genome and the evolutionary origins of neural systems. *Nature* **510**, 109–114 (2014).
7. Ryan, J. F. *et al.* The genome of the ctenophore *Mnemiopsis leidyi* and its implications for cell type evolution. *Science* **342**, 1242592 (2013).
8. Srivastava, M. *et al.* The *Amphimedon queenslandica* genome and the evolution of animal complexity. *Nature* **466**, 720–726 (2010). 9. Srivastava, M. *et al.* The *Trichoplax* genome and the nature of placozoans. *Nature* **454**, 955–960 (2008).
References 5–9 report the genome sequencing and analysis of early animal lineages (Cnidaria, Placozoa, Ctenophora and Porifera). These studies reveal the existence of an extensive gene toolkit that is shared by all animals, and is involved in signalling, adhesion and transcriptional control.
10. Leininger, S. *et al.* Developmental gene expression provides clues to relationships between sponge and eumetazoan body plans. *Nat. Commun.* **5**, 3905 (2014).
11. Zakhvatkin, A. A. *The Comparative Embryology of the Low Invertebrates. Sources and Method of the Origin of Metazoan Development* (Soviet Science, 1949).
12. Mikhailov, K. V. *et al.* The origin of Metazoa: a transition from temporal to spatial cell differentiation. *Bioessays* **31**, 758–768 (2009).
In this seminal review, the authors provide a broad historical perspective on hypotheses about animal origins and, in particular, they support and further extend Zakhvatkin's original Synzoospore hypothesis.
13. James-Clark, H. Note on the *Infusoria flagellata* and the *Spongiae ciliatae*. *Am. J. Sci.* **1**, 113–114 (1866).
14. Saville-Kent, W. *A Manual of the Infusoria, Including a Description of All Known Flagellate, Ciliate, and Tentaculiferous Protozoa, British and Foreign and an Account of the Organization and Affinities of the Sponges* Vol. 1–3 (David Bogue, 1880).
15. Mah, J. L., Christensen-Dalsgaard, K. K. & Leys, S. P. Choanoflagellate and choanocyte collar-flagellar systems and the assumption of homology. *Evol. Dev.* **16**, 25–37 (2014).

16. Ruiz-Trillo, I., Roger, A. J., Burger, G., Gray, M. W. & Lang, B. F. Phylogenomic investigation into the origin of Metazoa. *Mol. Biol. Evol.* **25**, 664–672 (2008).
17. Shalchian-Tabrizi, K. *et al.* Multigene phylogeny of Choanozoa and the origin of animals. *PLoS ONE* **3**, e2098 (2008).
18. Torruella, G. *et al.* Phylogenetic relationships within the Opisthokonta based on phylogenomic analyses of conserved single-copy protein domains. *Mol. Biol. Evol.* **29**, 531–544 (2012).
19. Torruella, G. *et al.* Phylogenomics reveals convergent evolution of lifestyles in close relatives of animals and fungi. *Curr. Biol.* **25**, 2404–2410 (2015).
This is the most comprehensive phylogenomic study on the holozoan clade published to date, both in terms of taxon sampling and the amount of data generated. It provides strong support for the scenario of three independent unicellular lineages close to Metazoa.
20. Lang, B. F., O'Kelly, C., Nerad, T., Gray, M. W. & Burger, G. The closest unicellular relatives of animals. *Curr. Biol.* **12**, 1773–1778 (2002).
21. Zettler, L. A., Nerad, T. A., O'Kelly, C. J. & Sogin, M. L. The nucleariid amoebae: more protists at the animal–fungal boundary. *J. Eukaryot. Microbiol.* **48**, 293–297 (2001).
22. Paps, J., Medina-Chacón, L. A., Marshall, W., Suga, H. & Ruiz-Trillo, I. Molecular phylogeny of unikonts: new insights into the position of apusomonads and ancyromonads and the internal relationships of opisthokonts. *Protist* **164**, 2–12 (2013).
23. Carr, M. & Leadbeater, B. Molecular phylogeny of choanoflagellates, the sister group to Metazoa. *Proc. Natl Acad. Sci. USA* **105**, 16641–16646 (2008).
24. Steenkamp, E. T., Wright, J. & Baldauf, S. L. The protistan origins of animals and fungi. *Mol. Biol. Evol.* **23**, 93–106 (2006).
25. Leadbeater, B. S. C. *The Choanoflagellates: Evolution, Biology and Ecology* (Cambridge Univ. Press, 2015).
26. Tong, S. M. Heterotrophic flagellates and other protists from Southampton Water, UK. *Ophelia* **47**, 71–131 (1997).
27. Hertel, L. A., Bayne, C. J. & Loker, E. S. The symbiont *Capsaspora owczarzaki*, nov. gen. nov. sp., isolated from three strains of the pulmonate snail *Biomphalaria glabrata* is related to members of the Mesomycetozoea. *Int. J. Parasitol.* **32**, 1183–1191 (2002).
28. Stibbs, H. H., Owczarzak, A., Bayne, C. J. & DeWan, P. Schistosome sporocyst-killing amoebae isolated from *Biomphalaria glabrata*. *J. Invertebr. Pathol.* **33**, 159–170 (1979).
29. Owczarzak, A., Stibbs, H. H. & Bayne, C. J. The destruction of *Schistosoma mansoni* mother sporocysts *in vitro* by amoebae isolated from *Biomphalaria glabrata*: an ultrastructural study. *J. Invertebr. Pathol.* **35**, 26–33 (1980).
30. Hertel, L. A., Barbosa, C. S., Santos, R. A. & Loker, E. S. Molecular identification of symbionts from the pulmonate snail *Biomphalaria glabrata* in Brazil. *J. Parasitol.* **90**, 759–763 (2004).
31. Glockling, S. L., Marshall, W. L. & Gleason, F. H. Phylogenetic interpretations and ecological potentials of the Mesomycetozoea (Ichthyosporea). *Fungal Ecol.* **6**, 237–247 (2013).
32. Mendoza, L., Taylor, J. W. & Ajello, L. The class mesomycetozoea: a heterogeneous group of microorganisms at the animal–fungal boundary. *Annu. Rev. Microbiol.* **56**, 315–344 (2002).
33. Del Campo, J. & Ruiz-Trillo, I. Environmental survey meta-analysis reveals hidden diversity among unicellular opisthokonts. *Mol. Biol. Evol.* **30**, 802–805 (2013).
34. Marshall, W. L. & Berbee, M. L. Facing unknowns: living cultures (*Pirum gemmata* gen. nov., sp. nov., and *Abeoforma whisleri*, gen. nov., sp. nov.) from invertebrate digestive tracts represent an undescribed clade within the unicellular Opisthokont lineage Ichthyosporea (Mesomycetozoea). *Protist* **162**, 33–57 (2011).
35. Marshall, W. L., Celio, G., McLaughlin, D. J. & Berbee, M. L. Multiple isolations of a culturable, motile Ichthyosporean (Mesomycetozoa, Opisthokonta), *Creolimax fragrantissima* n. gen., n. sp., from marine invertebrate digestive tracts. *Protist* **159**, 415–433 (2008).
36. King, N. *et al.* The genome of the choanoflagellate *Monosiga brevicollis* and the origin of metazoans. *Nature* **451**, 783–788 (2008).
This paper represents the foundation of comparative genomics approaches to animal origins. It reports the sequencing of the first genome of a unicellular holozoan species: the choanoflagellate *M. brevicollis.* It is the first study to extensively show that many genes that were previously considered to be animal-specific and tightly related to multicellularity evolved in a unicellular context.
37. Fairclough, S. R. *et al.* Premetazoan genome evolution and the regulation of cell differentiation in the choanoflagellate *Salpingoeca rosetta*. *Genome Biol.* **14**, R15 (2013).
38. de Mendoza, A., Suga, H., Permanyer, J., Irimia, M. & Ruiz-Trillo, I. Complex transcriptional regulation and independent evolution of fungal-like traits in a relative of animals. *eLife* **4**, e08904 (2015).
39. Suga, H. *et al.* The *Capsaspora* genome reveals a complex unicellular prehistory of animals. *Nat. Commun.* **4**, 2325 (2013).
40. Nichols, S. A., Roberts, B. W., Richter, D. J., Fairclough, S. R. & King, N. Origin of metazoan cadherin diversity and the antiquity of the classical cadherin/β-catenin complex. *Proc. Natl Acad. Sci. USA* **109**, 13046–13051 (2012).
41. Sebé-Pedrós, A., Roger, A., Lang, B., King, N. & Ruiz-Trillo, I. Ancient origin of the integrin-mediated adhesion and signaling machinery. *Proc. Natl Acad. Sci. USA* **107**, 10142–10147 (2010).
42. Richards, G. S. & Degnan, B. M. The dawn of developmental signaling in the Metazoa. *Cold Spring Harb. Symp. Quant. Biol.* **74**, 81–90 (2009).
43. Suga, H. *et al.* Genomic survey of premetazoans shows deep conservation of cytoplasmic tyrosine kinases and multiple radiations of receptor tyrosine kinases. *Sci. Signal.* **5**, ra35 (2012).
44. Suga, H., Torruella, G., Burger, G., Brown, M. W. & Ruiz-Trillo, I. Earliest holozoan expansion of phosphotyrosine signaling. *Mol. Biol. Evol.* **31**, 517–528 (2014).
45. Manning, G., Young, S. L., Miller, W. T. & Zhai, Y. The protist, *Monosiga brevicollis*, has a tyrosine kinase signaling network more elaborate and diverse than found in any known metazoan. *Proc. Natl Acad. Sci. USA* **105**, 9674–9679 (2008).
46. Sebé-Pedrós, A., Zheng, Y., Ruiz-Trillo, I. & Pan, D. Premetazoan origin of the Hippo signaling pathway. *Cell Rep.* **1**, 13–20 (2012).
47. Sebé-Pedrós, A., de Mendoza, A., Lang, B. F., Degnan, B. M. & Ruiz-Trillo, I. Unexpected repertoire of metazoan transcription factors in the unicellular holozoan *Capsaspora owczarzaki*. *Mol. Biol. Evol.* **28**, 1241–1254 (2011).
48. Sebé-Pedrós, A. *et al.* Early evolution of the T-box transcription factor family. *Proc. Natl Acad. Sci. USA* **110**, 16050–16055 (2013).
49. Sebé-Pedrós, A. & de Mendoza, A. in *Evolutionary Transitions to Multicellular Life* Vol. 2 (eds Ruiz-Trillo, I. & Nedelcu, A. M.) 379–394 (Springer, 2015).
50. Fairclough, S., Dayel, M. & King, N. Multicellular development in a choanoflagellate. *Curr. Biol.* **20**, 875–876 (2010).
51. Dayel, M. J. *et al.* Cell differentiation and morphogenesis in the colony-forming choanoflagellate *Salpingoeca rosetta*. *Dev. Biol.* **357**, 73–82 (2011).
References 37, 50 and 51 describe for the first time in detail the life cycle of a choanoflagellate species and the associated transcriptional regulation.
52. Alegado, R. A. *et al.* A bacterial sulfonolipid triggers multicellular development in the closest living relatives of animals. *eLife* **1**, e00013 (2012).
53. Levin, T. C. & King, N. Evidence for sex and recombination in the choanoflagellate *Salpingoeca rosetta*. *Curr. Biol.* **23**, 2176–2180 (2013).
54. Levin, T. C., Greaney, A. J., Wetzel, L. & King, N. The *rosetteless* gene controls development in the choanoflagellate *S. rosetta*. *eLife* **3**, e04070 (2014).
This is a careful study that, through a forward-genetics screen, provides the first direct evidence of the function of a gene (which encodes a C-type lectin) in a unicellular holozoan.
55. Abedin, M. & King, N. The premetazoan ancestry of cadherins. *Science* **319**, 946–948 (2008).
56. Suga, H. & Ruiz-Trillo, I. Development of ichthyosporeans sheds light on the origin of metazoan multicellularity. *Dev. Biol.* **377**, 284–292 (2013).
This is the first report of the genetic transformation of a unicellular holozoan. It also provides a detailed description of the life cycle of an ichthyosporean species, for which the associated transcriptional regulation is described in reference 38.
57. Sebé-Pedrós, A. *et al.* Regulated aggregative multicellularity in a close unicellular relative of Metazoa. *eLife* **2**, e01287 (2013).
This paper provides the first description of the life cycle of a filasterean and the associated transcriptional regulation, which involves multiple genes that are related to animal multicellularity.
58. Sebé-Pedrós, A. *et al.* Insights into the origin of metazoan filopodia and microvilli. *Mol. Biol. Evol.* **30**, 2013–2023 (2013).
59. Sebé-Pedrós, A. *et al.* High-throughput proteomics reveals the unicellular roots of animal phosphosignaling and cell differentiation. *Dev. Cell* **39**, 186–197 (2016).
60. Arendt, D. *et al.* The origin and evolution of cell types. *Nat. Rev. Genet.* **17**, 744–757 (2016).
61. Davidson, E. *The Regulatory Genome* (Academic Press, 2006).
62. Aguirre, J., Ríos-Momberg, M., Hewitt, D. & Hansberg, W. Reactive oxygen species and development in microbial eukaryotes. *Trends Microbiol.* **13**, 111–118 (2005).
63. Loenarz, C. *et al.* The hypoxia-inducible transcription factor pathway regulates oxygen sensing in the simplest animal, *Trichoplax adhaerens*. *EMBO Rep.* **12**, 63–70 (2011).
64. Adamska, M., Degnan, B. M., Green, K. & Zwafink, C. What sponges can tell us about the evolution of developmental processes. *Zoology (Jena)* **114**, 1–10 (2011).
65. Najafabadi, H. S. *et al.* C2H2 zinc finger proteins greatly expand the human regulatory lexicon. *Nat. Biotechnol.* **33**, 555–562 (2015).
66. Jolma, A. *et al.* DNA-binding specificities of human transcription factors. *Cell* **152**, 327–339 (2013).
67. de Mendoza, A. *et al.* Transcription factor evolution in eukaryotes and the assembly of the regulatory toolkit in multicellular lineages. *Proc. Natl Acad. Sci. USA* **110**, E4858–E4866 (2013).
68. Reinke, A. W., Baek, J., Ashenberg, O. & Keating, A. E. Networks of bZIP protein–protein interactions diversified over a billion years of evolution. *Science* **340**, 730–734 (2013).
69. Stergachis, A. B. *et al.* Conservation of *trans*-acting circuitry during mammalian regulatory evolution. *Nature* **515**, 365–370 (2014).
70. Sebé-Pedrós, A. *et al.* The dynamic regulatory genome of *Capsaspora* and the origin of animal multicellularity. *Cell* **165**, 1224–1237 (2016).
71. Young, S. L. *et al.* Premetazoan ancestry of the Myc–Max network. *Mol. Biol. Evol.* **28**, 2961–2971 (2011).
72. Heintzman, N. D. *et al.* Histone modifications at human enhancers reflect global cell-type-specific gene expression. *Nature* **459**, 108–112 (2009).
73. Rada-Iglesias, A. *et al.* A unique chromatin signature uncovers early developmental enhancers in humans. *Nature* **470**, 279–283 (2011).
74. Corces, M. R. *et al.* Lineage-specific and single-cell chromatin accessibility charts human hematopoiesis and leukemia evolution. *Nat. Genet.* **48**, 1193–1203 (2016).
75. Thurman, R. E. *et al.* The accessible chromatin landscape of the human genome. *Nature* **489**, 75–82 (2012).
76. Andersson, R. *et al.* An atlas of active enhancers across human cell types and tissues. *Nature* **507**, 455–461 (2014).
77. Visel, A. *et al.* ChIP–seq accurately predicts tissue-specific activity of enhancers. *Nature* **457**, 854–858 (2009).
78. Jin, F. *et al.* A high-resolution map of the three-dimensional chromatin interactome in human cells. *Nature* **503**, 290–294 (2013).
79. Deng, W. *et al.* Controlling long-range genomic interactions at a native locus by targeted tethering of a looping factor. *Cell* **149**, 1233–1244 (2012).
80. Shlyueva, D., Stampfel, G. & Stark, A. Transcriptional enhancers: from properties to genome-wide predictions. *Nat. Rev. Genet.* **15**, 272–286 (2014).
81. Phillips-Cremins, J. E. *et al.* Architectural protein subclasses shape 3D organization of genomes during lineage commitment. *Cell* **153**, 1281–1295 (2013).
82. Schmidt, D. *et al.* A CTCF-independent role for cohesin in tissue-specific transcription. *Genome Res.* **20**, 578–588 (2010).
83. Kvon, E. Z. *et al.* Genome-scale functional characterization of *Drosophila* developmental enhancers *in vivo*. *Nature* **512**, 91–95 (2014).

84. Bonn, S. *et al.* Tissue-specific analysis of chromatin state identifies temporal signatures of enhancer activity during embryonic development. *Nat. Genet.* **44**, 148–156 (2012).
85. Schwarzer, W. & Spitz, F. The architecture of gene expression: integrating dispersed *cis*-regulatory modules into coherent regulatory domains. *Curr. Opin. Genet. Dev.* **27**, 74–82 (2014).
86. Levine, M. Transcriptional enhancers in animal development and evolution. *Curr. Biol.* **20**, R754–R763 (2010).
87. Ernst, J. *et al.* Mapping and analysis of chromatin state dynamics in nine human cell types. *Nature* **473**, 43–49 (2011).
88. Villar, D. *et al.* Enhancer evolution across 20 mammalian species. *Cell* **160**, 554–566 (2015).
89. Irimia, M. *et al.* Extensive conservation of ancient microsynteny across metazoans due to *cis*-regulatory constraints. *Genome Res.* **22**, 2356–2367 (2012).
90. Irimia, M., Maeso, I., Roy, S. W. & Fraser, H. B. Ancient *cis*-regulatory constraints and the evolution of genome architecture. *Trends Genet.* **29**, 521–528 (2013).
91. Schwaiger, M. *et al.* Evolutionary conservation of the eumetazoan gene regulatory landscape. *Genome Res.* **24**, 639–650 (2014).
References 70 and 91 are the pioneering studies of the genome regulatory biology of unicellular Holozoa and early Metazoa, respectively. They reveal extensive conservation of epigenomic features within the animal lineages, and important differences between animals and their unicellular relatives.
92. Heger, P., Marin, B., Bartkuhn, M., Schierenberg, E. & Wiehe, T. The chromatin insulator CTCF and the emergence of metazoan diversity. *Proc. Natl Acad. Sci. USA* **109**, 17507–17512 (2012).
93. Eagen, K. P., Lieberman Aiden, E. & Kornberg, R. D. Polycomb-mediated chromatin loops revealed by a sub-kilobase resolution chromatin interaction map. Preprint at *bioRxiv* http://dx.doi.org/10.1101/099804 (2017).
94. Cubeñas-Potts, C. *et al.* Different enhancer classes in *Drosophila* bind distinct architectural proteins and mediate unique chromatin interactions and 3D architecture. *Nucleic Acids Res.* **45**, 39–53 (2016).
95. Nora, E. P. *et al.* Spatial partitioning of the regulatory landscape of the X-inactivation centre. *Nature* **485**, 381–385 (2012).
96. Sexton, T. *et al.* Three-dimensional folding and functional organization principles of the *Drosophila* genome. *Cell* **148**, 458–472 (2012).
97. Dixon, J. R. *et al.* Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* **485**, 376–380 (2012).
98. Bonev, B. & Cavalli, G. Organization and function of the 3D genome. *Nat. Rev. Genet.* **17**, 661–678 (2016).
99. Dixon, J. R., Gorkin, D. U. & Ren, B. Chromatin domains: the unit of chromosome organization. *Mol. Cell* **62**, 668–680 (2016).
100. Tanay, A. & Cavalli, G. Chromosomal domains: epigenetic contexts and functional implications of genomic compartmentalization. *Curr. Opin. Genet. Dev.* **23**, 197–203 (2013).
101. Haeckel, E. Die Gastraea-Theorie, die phylogenetische Klassifikation des Thierreichs und die Homologie der Keimblatter. *Jenaische Z. Naturwiss.* **8**, 1–55 (in German) (1874).
102. Nielsen, C. Six major steps in animal evolution: are we derived sponge larvae? *Evol. Dev.* **10**, 241–257 (2008).
103. Hashimshony, T., Feder, M., Levin, M., Hall, B. K. & Yanai, I. Spatiotemporal transcriptomics reveals the evolutionary history of the endoderm germ layer. *Nature* **519**, 219–222 (2015).
104. Arendt, D., Benito-Gutierrez, E., Brunet, T. & Marlow, H. Gastric pouches and the mucociliary sole: setting the stage for nervous system evolution. *Phil. Trans. R. Soc. B Biol. Sci.* **370**, 20150286 (2015).
105. Richter, D. J. & King, N. The genomic and cellular foundations of animal origins. *Annu. Rev. Genet.* **47**, 509–537 (2013).
106. Adamska, M. Sponges as models to study emergence of complex animals. *Curr. Opin. Genet. Dev.* **39**, 21–28 (2016).
107. Nakanishi, N., Sogabe, S. & Degnan, B. M. Evolutionary origin of gastrulation: insights from sponge development. *BMC Biol.* **12**, 26 (2014).
108. Arenas-Mena, C. The origins of developmental gene regulation. *Evol. Dev.* **19**, 96–107 (2017).
109. Arendt, D. The evolution of cell types in animals: emerging principles from molecular studies. *Nat. Rev. Genet.* **9**, 868–882 (2008).
110. Achim, K. & Arendt, D. Structural evolution of cell types by step-wise assembly of cellular modules. *Curr. Opin. Genet. Dev.* **27**, 102–108 (2014).
111. Schwartzman, O. & Tanay, A. Single-cell epigenomics: techniques and emerging applications. *Nat. Rev. Genet.* **16**, 716–726 (2015).
112. Tanay, A. & Regev, A. Scaling single-cell genomics from phenomenology to mechanism. *Nature* **541**, 331–338 (2017).
113. Neph, S. *et al.* An expansive human regulatory lexicon encoded in transcription factor footprints. *Nature* **489**, 83–90 (2012).
114. Newman, S. A. Physico-genetic determinants in the evolution of development. *Science* **338**, 217–219 (2012).
115. Mshigeni, K. & Lorri, W. Spore germination and early stages of development in *Hypnea musciformis* (Rhodophyta, Gigartinales). *Mar. Biol.* **42**, 161–164 (1977).
116. Bouget, F. Y., Berger, F. & Brownlee, C. Position dependent control of cell fate in the Fucus embryo: role of intercellular communication. *Development* **125**, 1999–2008 (1998).
117. Xie, X., Wang, G., Pan, G. & Gao, S. Variations in morphology and PSII photosynthetic capabilities during the early development of tetraspores of *Gracilaria vermiculophylla* (Ohmi) Papenfuss (Gracilariales, Rhodophyta). *BMC Dev. Biol.* **10**, 43 (2010).
118. El Albani, A. *et al.* Large colonial organisms with coordinated growth in oxygenated environments 2.1 Gyr ago. *Science* **466**, 100–104 (2010).
119. Butterfield, N. J. Modes of pre-Ediacaran multicellularity. *Precambrian Res.* **173**, 201–211 (2009).
120. Becker, B. Snow ball earth and the split of Streptophyta and Chlorophyta. *Trends Plant Sci.* **18**, 180–183 (2013).
121. Laurin-Lemay, S., Brinkmann, H. & Philippe, H. Origin of land plants revisited in the light of sequence contamination and missing data. *Curr. Biol.* **22**, R593–R594 (2012).
122. Love, G. D. *et al.* Fossil steroids record the appearance of Demospongiae during the Cryogenian period. *Nature* **457**, 718–721 (2009).
123. Maloof, A. C. *et al.* Possible animal-body fossils in pre-Marinoan limestones from South Australia. *Nat. Geosci.* **3**, 653–659 (2010).
124. Erwin, D. H. *et al.* The Cambrian conundrum: early divergence and later ecological success in the early history of animals. *Science* **3**, 1091–1097 (2011).
125. Sanderson, M. Molecular data from 27 proteins do not support a Precambrian origin of land plants. *Am. J. Bot.* **90**, 954–956 (2003).
126. Taylor, J. W. & Berbee, M. L. Dating divergences in the Fungal Tree of Life: review and new analyses. *Mycologia* **98**, 838–849 (2006).
127. Silberfeld, T. *et al.* A multi-locus time-calibrated phylogeny of the brown algae (Heterokonta, Ochrophyta, Phaeophyceae): investigating the evolutionary nature of the 'brown algal crown radiation'. *Mol. Phylogenet. Evol.* **56**, 659–674 (2010).
128. Abedin, M. & King, N. Diverse evolutionary paths to cell adhesion. *Trends Cell Biol.* **20**, 734–742 (2010).
129. Niklas, K. J. The evolutionary-developmental origins of multicellularity. *Am. J. Bot.* **101**, 6–25 (2014).
130. Burki, F., Okamoto, N., Pombert, J.-F. & Keeling, P. J. The evolutionary history of haptophytes and cryptophytes: phylogenomic evidence for separate origins. *Proc. Biol. Sci.* **279**, 2246–2254 (2012).
131. Burki, F. The eukaryotic tree of life from a global phylogenomic perspective. *Cold Spring Harb. Perspect. Biol.* **6**, a016147 (2014).
132. Derelle, R., Lopez-Garcia, P., Timpano, H. & Moreira, D. A phylogenomic framework to study the diversity and evolution of stramenopiles (= heterokonts). *Mol. Biol. Evol.* **33**, 2890–2898 (2016).
133. He, D., Sierra, R., Pawlowski, J. & Baldauf, S. L. Reducing long-branch effects in multi-protein data uncovers a close relationship between Alveolata and Rhizaria. *Mol. Phylogenet. Evol.* **101**, 1–7 (2016).
134. Sierra, R. *et al.* Deep relationships of Rhizaria revealed by phylogenomics: a farewell to Haeckel's Radiolaria. *Mol. Phylogenet. Evol.* **67**, 53–59 (2012).
135. Zhao, S. *et al.* Collodictyon — an ancient lineage in the tree of eukaryotes. *Mol. Biol. Evol.* **29**, 1557–1568 (2012).
136. Derelle, R. & Lang, B. F. Rooting the eukaryotic tree with mitochondrial and bacterial proteins. *Mol. Biol. Evol.* **29**, 1277–1289 (2012).
137. Brown, M. W. *et al.* Phylogenomics demonstrates that breviate flagellates are related to opisthokonts and apusomonads. *Proc. Biol. Sci.* **280**, 20131755 (2013).
138. Finet, C., Timme, R. E., Delwiche, C. F. & Marlétaz, F. Multigene phylogeny of the green lineage reveals the origin and diversification of land plants. *Curr. Biol.* **20**, 2217–2222 (2010).
139. Peterson, K. J., Cotton, J. A., Gehling, J. G. & Pisani, D. The Ediacaran emergence of bilaterians: congruence between the genetic and the geological fossil records. *Phil. Trans. R. Soc.* **363**, 1435–1443 (2008).
140. Whelan, N. V., Kocot, K. M., Moroz, L. L. & Halanych, K. M. Error, signal, and the placement of Ctenophora sister to all other animals. *Proc. Natl Acad. Sci. USA* **112**, 5773–5778 (2015).
141. Simion, P. *et al.* A large and consistent phylogenomic dataset supports sponges as the sister group to all other animals. *Curr Biol.* **27**, 958–967 (2017).
142. Fernandez-Valverde, S. L., Calcino, A. D. & Degnan, B. M. Deep developmental transcriptome sequencing uncovers numerous new genes and enhances gene annotation in the sponge *Amphimedon queenslandica*. *BMC Genomics* **16**, 387 (2015).
143. Dunham, I. *et al.* An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57–74 (2012).
144. Moran, Y. *et al.* Cnidarian microRNAs frequently regulate targets by cleavage. *Genome Res.* **24**, 651–663 (2014).
145. Grimson, A. *et al.* Early origins and evolution of microRNAs and Piwi-interacting RNAs in animals. *Nature* **455**, 1193–1197 (2008).
146. Cabili, M. N. *et al.* Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. *Genes Dev.* **25**, 1915–1927 (2011).
147. Young, R. S. *et al.* Identification and properties of 1,119 candidate lincRNA loci in the *Drosophila melanogaster* genome. *Genome Biol. Evol.* **4**, 427–442 (2012).
148. Xie, C. *et al.* NONCODEv4: exploring the world of long non-coding RNA genes. *Nucleic Acids Res.* **42**, D98–D103 (2014).
149. Lee, K. *et al.* Genetic landscape of open chromatin in yeast. *PLoS Genet.* **9**, e1003229 (2013).
150. Thomas, S. *et al.* Dynamic reprogramming of chromatin accessibility during *Drosophila* embryo development. *Genome Biol.* **12**, R43 (2011).
151. Davie, K. *et al.* Discovery of transcription factors and regulatory regions driving *in vivo* tumor development by ATAC-seq and FAIRE–seq open chromatin profiling. *PLOS Genet.* **11**, e1004994 (2015).

Acknowledgements

The authors thank X. Grau-Bové and D. Lara-Astiaso for critical comments on the manuscript, and G. Torruella and A. de Mendoza for discussion and ideas. A.S.-P. is supported by a European Molecular Biology Organization Long-Term Fellowship (ALTF 841-2014). Research by B.M.D. is supported by an Australian Research Council grant. Research by I.R.-T. is supported by an Institució Catalana de Recerca i Estudis Avançats (ICREA) contract, a European Research Council Starting Grant (ERC-2007-StG-206883), a European Research Council Consolidator Grant (ERC-2012-Co -616960) grant, and a grant (BFU2014-57779-P) from Ministerio de Economía y Competitividad (MINECO); his latest research is co-funded by the European Regional Development Fund (fondos FEDER). I.R-T. also acknowledges financial support from Secretaria d'Universitats i Recerca del Departament d'Economia i Coneixement de la Generalitat de Catalunya (project 2014 SGR 619).

Competing interests statement

The authors declare no competing interests.

Publisher's note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Correction

In Figure 1a of the original version of this article, the Choanoflagellatea branch was missing a yellow-black split circle symbolizing that clonal multicellularity occurs in some Choanoflagellatea species. The symbol inadvertently dropped out prior to publication and has now been reinstated in the corrected article. The editors apologize for this error.