



UNIVERSITAT<sub>DE</sub>  
BARCELONA

## ***Capsaspora owczarzaki* as a unicellular model to study the pre-metazoan integrin adhesome**

Helena Parra Acero



Aquesta tesi doctoral està subjecta a la llicència **Reconeixement- NoComercial – Compartir Igual 4.0. Espanya de Creative Commons.**

Esta tesis doctoral está sujeta a la licencia **Reconocimiento - NoComercial – Compartir Igual 4.0. España de Creative Commons.**

This doctoral thesis is licensed under the **Creative Commons Attribution-NonCommercial-ShareAlike 4.0. Spain License.**

Programa de Doctorado de Genética  
Departamento de Genética, Microbiología y Estadística  
Facultad de Biología  
Universidad de Barcelona

***Capsaspora owczarzaki* as a unicellular model to study the  
pre-metazoan integrin adhesome**

*Capsaspora owczarzaki* como organismo modelo unicelular  
para estudiar el adhesoma de integrina

Memoria presentada por **Helena Parra Acero** para optar al grado de  
doctora por la Universidad de Barcelona

Iñaki Ruiz Trillo  
Director

Pedro Martínez Serra  
Tutor

Helena Parra Acero  
Autora

Cover: *Capsaspora owczarzaki* staining of nucleus and filopodia

## Agradecimientos

Al mirar atrás, los años que ha durado esta tesis me parecen una gran película (drama/documental/thriller/comedia) en la que todas las personas que me han rodeado han sido grandes actores y actrices. Así, os agradezco vuestras mejores interpretaciones, por orden de aparición:

Iñaki, por acogerme en tu laboratorio y darme la oportunidad de jugar y descubrir cosas que nunca había imaginado. La tarjeta de entrada no era para entrar a un edificio, sino a un mundo brillante y absurdo de bichos, ADN y proteínas, que hacen cosas inexplicables y maravillosas que me habría perdido si no fuera por tu confianza en mí desde el principio hasta el final.

“Old Generation”, por cuidarme a vuestra manera y darme ejemplo para experimentar sin miedo. Arnau en especial, por pedirme que te acompañase en tu “punk-science” primero, y más tarde, en tu “science” de verdad. Hiroshi, por invitarme educadamente a jugar con los bichos, y por animarme a toquetear los microscopios. A Núria S.P. por la enormísima herencia que me dejaste y los sabios consejos, aunque éstos hayan hecho efecto tarde. Siempre llevé la peluca de Ponsy con orgullo.

Xavi G., hermanito menor, aunque en realidad has sido mi hermanito mayor, por contarme tus movidas comparativas que me transportaban de la poyata a lo abstracto, y me hacían amar un poco más la poyata al volver a ella sana y salva. También por tu infinidad de consejos acertados.

David, por divertirme proponiéndome experimentos. Nos emocionaba pensar que en realidad tenían todo el sentido del mundo y que ojalá tuviéramos un saco de dinero infinito para realizarlos. Y por tu manera de capear los problemas, que he invocado en los malos momentos.

Mati, for your love and for saving me from the most painful part of the immunoparties. It was cool to share the fun with you!

Meri, por asegurarte de que estuvieran siempre todos los ingredientes a mano listos para divertirme; y por tu extra par de manos, siempre firmes, que me han ayudado cuando las mías temblaban.

Nurieta la Grande, que pasaste de indicarme dónde se guardaba el kit de purificar ADN, a ser la mujer investigadora que me ha enseñado organización, disciplina, y a programar en R; que sin quererlo me has desmontado la paciencia y me la has reconstruido más fuerte, y que has tirado del carro la primera en cada vez que patinábamos en el barro. Mi Núria con tilde.

Sebas, por ayudarme a pensar mejor, con esa manera especial de no dar respuestas sino preguntas, y esa manía de hablarle al aire.

Alberto, por enseñarme mis limitaciones y compartir conmigo tus trucos.

Xavi F., por enseñarme otra manera de hacer poyata, y el microscopio que me faltaba.

Alberto, Edu, Alicia, Kons y Ola por ofrecerme vuestra visión particular de “new-new generation” en los aspectos científicos y no científicos. Ola, también por hacerme disfrutar explicando las cosas.

Nick, you and all your lab, for taking care of me and showing me how life is in a completely different world. A Elena, por mi enlace con- y tu visión desde- el mundo animal en este lab.

“Mariscurís” threesome (Omi-Mich-Andrej), who, representing Sensibility, have given me good advice. Specially, Omayá for pushing when, instead of a muddy road, we were travelling on a road full of pebbles, sometimes so interesting that I stepped down to check “what is it here?”. Dan, por tu optimismo extravagante .

Y a las personas desde Administración, expertos salva-vidas.

Y sin olvidar que una tira “palante” acompañada, gracias al equipo de apoyo:

Pilar A. y Jose Antonio P. (también llamados mamá y papá), por apoyarme en todas mis decisiones, por vuestra paciencia, también por vuestro excelente servicio de catering. Pilar también, por ser mi ejemplo de fortaleza, y Jose Antonio por ser mi ejemplo bromista. Alicia, hermana biológica, por inspirarme con tus decisiones admirables en tu propia aventura vital. Familia toda, por la resignación ante mis ausencias por culpa de estos “bichitos”. También por ignorar mis malos humores y por el ánimo que me habéis mandado.

Gissela, Helena, Cris de Miguel y todas las demás personas del IBE y de fuera, con las que he disfrutado del campo, de la música y de la pintura. Por vuestra compañía desde el principio. También a todos los que habéis sido parte del equipo de emergencia.

Maria R., que me has cuidado siempre desde cerca, por tu brujería y tus garrafas de optimismo y alegría. Consejo de sabias/o, que me habéis bajado a la realidad con vuestras correas mágicas cuando os gritaba desde lejos. Rocío y Paula por acompañarme en las tormentas.

Xavi, por tus magníficas ramas de *Quercus*.

## Table of contents

<b>INTRODUCTION .....</b>	<b>1</b>
1. ORIGIN OF ANIMALS.....	3
1.1. <i>Holozoa: animals and their unicellular relatives</i> .....	3
1.2. <i>The last unicellular ancestor of animals and co-option as the mechanism to drive the emergence of animals</i> .....	4
2. CELL ADHESION IN ANIMALS .....	6
2.1. <i>Integrin adhesions in animals</i> .....	6
2.1.1 Adhesome components .....	6
2.1.1.1. Integrin and ligands .....	7
2.1.1.2. Adaptor proteins.....	9
2.2. <i>Integrin adhesion structure</i> .....	10
3. PRE-METAZOAN INTEGRIN ADHESOME AND ECM COMPONENTS .....	12
4. TOOLS TO STUDY PROTEIN FUNCTION .....	13
4.1- <i>Tools to study IAPs function in animal models</i> .....	13
4.2- <i>Genetic tools developed for unicellular eukaryotes</i> .....	13
5. <i>CAPSASPORA OWCZARZAKI</i> .....	15
<b>OBJECTIVES .....</b>	<b>17</b>
<b>RESULTS .....</b>	<b>21</b>
R1: REGULATED AGGREGATIVE MULTICELLULARITY IN A CLOSE UNICELLULAR RELATIVE OF METAZOA .....	25
R2: THE DYNAMIC REGULATORY GENOME OF <i>CAPSASPORA</i> AND THE ORIGIN OF ANIMAL MULTICELLULARITY.....	51
R3: TRANSFECTION OF <i>CAPSASPORA OWCZARZAKI</i> , A CLOSE UNICELLULAR RELATIVE OF ANIMALS .....	93
R4: A FIRST APPROACH INTO THE FUNCTION OF THE INTEGRIN ADHESOME OF <i>CAPSASPORA OWCZARZAKI</i> ....	123
<b>DISCUSSION .....</b>	<b>149</b>
1. <i>CAPSASPORA OWCZARZAKI</i> IS A KEY ORGANISM TO UNDERSTAND THE ORIGIN OF ANIMALS .....	151
2. TURNING <i>CAPSASPORA</i> INTO A MODEL ORGANISM TO STUDY THE ORIGIN OF ANIMALS .....	152
3. OUTLOOK.....	156
<b>CONCLUSIONS .....</b>	<b>159</b>
<b>REFERENCES .....</b>	<b>163</b>



# **INTRODUCTION**





## 1. Origin of animals

Animals form a monophyletic clade, meaning they all share a common ancestor (Baldauf 1999; Cavalier-Smith et al. 1996). To understand how animals evolved from their unicellular ancestor, we can compare their genomes to those of their extant unicellular relatives (Rokas 2008; Sebé-Pedrós et al. 2017).

### 1.1. Holozoa: animals and their unicellular relatives

Animals or metazoans are part of the supergroup called Opisthokonta. The Opisthokonta includes Holozoa (metazoans and their unicellular relatives) and Holomycota (fungi and their unicellular relatives) (Adl et al. 2012) (Fig. 1A).

The closest sister group of Metazoa is Choanoflagellata (King & Carroll 2001; Lang et al. 2002).

Choanoflagellata is a monophyletic group with many species described (Carr et al. 2008; Leadbeater 2015; Carr et al. 2017). They are free-living organisms that prey on bacteria and are ubiquitously found in marine and freshwater environments. Their morphology is highly conserved among the different choanoflagellates, consisting on a spherical cell with an apical flagellum surrounded by a collar of actin-based microvilli (Leadbeater 2015). Some species also produce an inorganic covering called lorica. They are usually solitary but some species can form colonies.

The sister clade to Metazoa and Choanoflagellata, is Filasterea. It includes only four known species, *Capsaspora owczarzaki* (Ruiz-Trillo et al. 2004), the newly discovered pigoraptors, *Pigoraptor chileana* and *Pigoraptor vietnamica*, (Hehenberger et al. 2017), and *Ministeria vibrans* (Shalchian-Tabrizi et al. 2008). *Capsaspora* is a filopodiated amoeba that was found as a symbiont from a freshwater snail (Hertel et al. 2002). The life cycle in culture conditions includes an amoeba stage (growth stage), a cystic stage and an aggregative stage (Sebé-Pedrós, Irimia, et al. 2013). *Ministeria* is a marine free-living heterotrophic protist that preys on bacteria. It has multiple filopodia and also presents a flagellum (Tong 1997; Torruella et al. 2015). During the time I worked in this thesis, two new filastereans (*Pigoraptor*) were described. They are freshwater heterotrophic protists that feed on other eukaryotes and can also engulf bacteria. They are flagellated and can form multicellular clusters (Hehenberger et al. 2017).

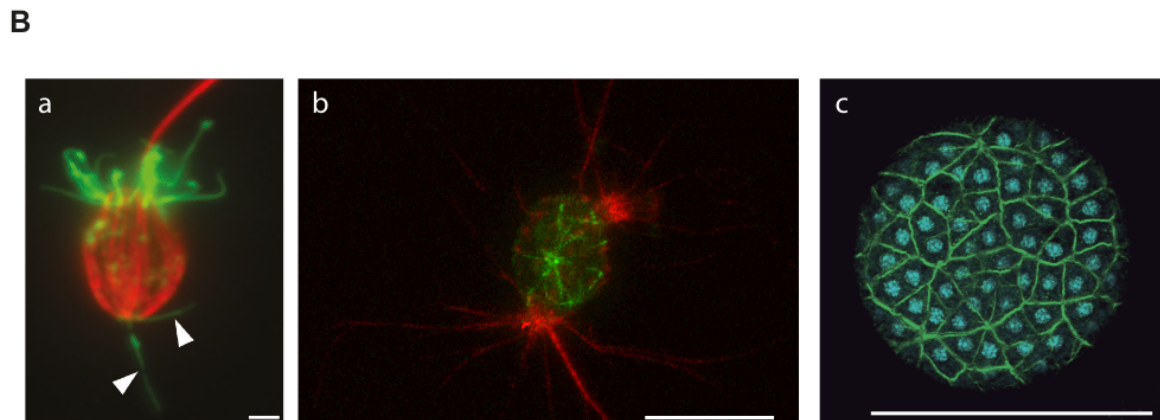
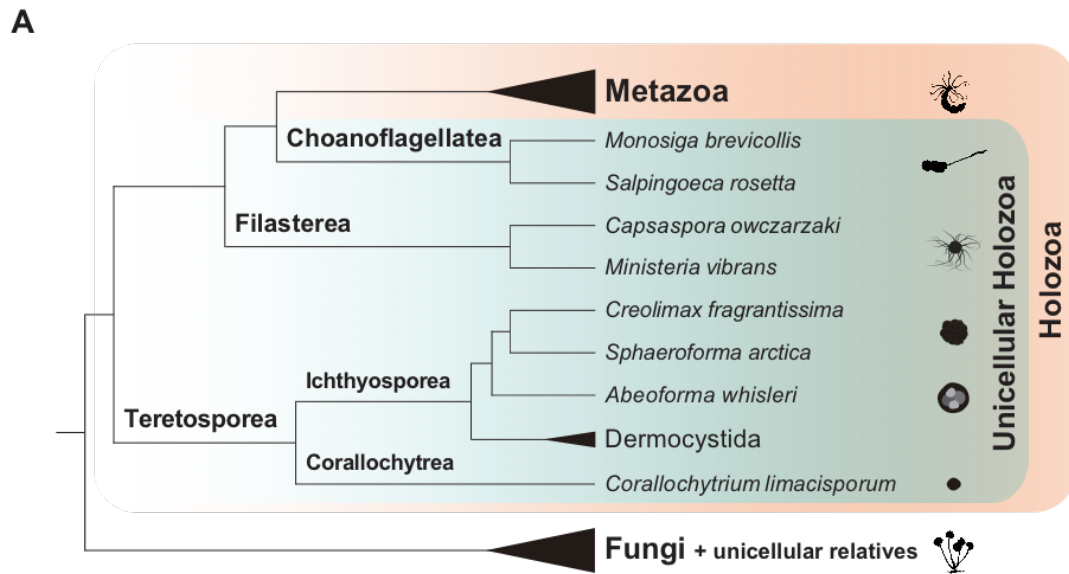
The clade sister to the previous group is Ichthyosporea. This clade contains numerous species described, most of them parasites or commensals of animals (Mendoza et al. 2002; Glockling et al. 2013), although few species are free-living (Hassett et al. 2015). Their cell cycle usually includes a multinucleate coenocytic stage that cellularises and many amoebas are released, although in some cases, the dispersive form has a flagellum (Mendoza et al. 2002).

*Corallochytrium limacisporum* is an organism whose morphology resembles that of ichthyosporeans, but its phylogenetic position remains to be fully discerned. The last two phylogenomic analyses place *C. limacisporum* either as sister-group to ichthyosporeans, forming the Teretosporea clade (Grau-Bové et al. 2017), or as sister-group to *Syssomonas multiformis*, forming the Pluriformea clade (Hehenberger et al. 2017).

## **1.2. The last unicellular ancestor of animals and co-option as the mechanism to drive the emergence of animals**

Analyses of genomes and transcriptomes of unicellular holozoans (i.e. unicellular relative of animals) have provided a more complete view on the elements that were already present in the unicellular ancestor of animals (Fig. 1B). The gene repertoire of this ancestor includes many genes that have a role in multicellular functions in animals, which suggest they were co-opted from a unicellular context into their new function within a multicellular entity (King 2004; King et al. 2008; Richter & King 2013; Sebé-Pedrós et al. 2017). Moreover, we have also learnt that not only co-option to new functions might explain the emergence of animals, but also a change in the regulatory mechanisms might have been key, providing a richer and finer regulation (Sebé-Pedrós, Ballaré, et al. 2016; Sebé-Pedrós, Peña, et al. 2016), as well as appearance of new genes (Richter et al. 2018; Paps & Holland 2018).

How co-option worked at the onset of Metazoa, however, remains unclear. To answer this, we need, first, to unravel the the function of premetazoan genes within a unicellular context. The main set of genes that have called our attention are those directly related to essential multicellular characteristics of animals: signalling receptors and elements from signal transduction pathways, like Src, Csk, and Hippo; transcription factors, like Brachyury, NF- $\kappa$ B, Myc and Runx; and cell-cell adhesion and cell-matrix adhesion receptors (i.e. cadherins and integrins) (Sebé-Pedrós et al. 2010; Sebé-Pedros et al. 2011; Suga et al. 2012; Nichols et al. 2012; Sebé-Pedrós et al. 2013). Knowing the function of premetazoan genes in a unicellular context bring us a bit closer to infer the ancestral function they could have had in the ancestor of animals.



**Figure 1. Holozoans**

(A) Phylogenetic relationships among animals and their unicellular relatives. Adapted from Torruella et al., 2015 (B) Staining of unicellular holozoans. (a) *Salpingoeca rosetta* with microtubules (red), (note flagellum), microvilli collar (green) and basal actin microfilaments (arrowheads). Scale bar = 1µm. (Sebé-Pedrós, Burkhardt et al. 2013) (b) *Capsaspora owczarzaki* amoeba with microtubules (green) and filopodia (red). Scale bar = 5µm. (c) *Sphaeroforma arctica* coenocyte with multiple nuclei (blue) and cellularised (actin in green). Scale bar = 50 µm (shared by Omayya Dudin).

## **2. Cell adhesion in animals**

When describing an animal, one of the more obvious characteristic is that it is formed by multiple cells that are sticking together. But not only to stick together, adhesion mechanisms are crucial for its development. Forces are exerted, and these shapes the tissues. Also some cells migrate to different localisations within the body and stay in particular places. Adhesion mechanisms play also an essential role in maintenance of the organism, for example in cells from the immune system cells, which move around in the body, or in the epithelia that closes a wound (Gumbiner 1996).

Cell adhesion mechanisms in animals can be roughly divided into two groups: mechanisms that mediate adhesion between neighbouring cells, from which cadherins are the most widely studied (Halbleib & Nelson 2006); and mechanisms that mediate adhesion of cells to the extracellular matrix (ECM), in which those mediated by integrins are the major player (Hynes 2002; Maartens & Brown 2015).

### **2.1. Integrin adhesions in animals**

Integrin adhesions are structures formed by a receptor (integrin) that recognises a ligand in the extracellular matrix, and adaptor proteins in the cytoplasm that connect integrin to the cytoskeleton (Hynes 2002; Zamir & Geiger 2001). In cells in culture, these structures are formed in regions of the membrane in close contact with the substrate (BurrIDGE 1988; Zamir & Geiger 2001). The structure that is thus formed by the linkage created among integrins, adaptor proteins and actin cytoskeleton, is called the integrin adhesion complex (IAC) (Zaidel-Bar et al. 2007; Geiger & Zaidel-Bar 2012; Horton et al. 2015; Klapholz & Brown 2017).

Integrin adhesions provide an anchor to the cell that is essential for a stable connection to the ECM as well as for migration (Hynes 2002; Bökel & Brown 2002). They are dynamic structures (they assemble and disassemble) and their composition also is different according to the cell type or to the ECM molecule that is recognised (Zamir & Geiger 2001; Geiger & Yamada 2011; Horton et al. 2015). Many different experiments have increased the number of proteins that localise in these structures and defined the interactions among them, building up the set of proteins involved in the cell-matrix adhesion mediated by integrins, which is named the “integrin adhesome” (Zaidel-Bar et al. 2007; Horton et al. 2016).

#### **2.1.1 Adhesome components**

The integrin adhesome is thus the collective name for the proteins that take part in integrin adhesions (Zaidel-Bar et al. 2007; Geiger & Zaidel-Bar 2012). According to their biological function, proteins from the integrin adhesome are classified into receptors (integrins), adaptor proteins, actin regulators and proteins with catalytic

activity, involved in signalling and the regulation of the adhesion itself (Zaidel-Bar et al. 2007).

### **2.1.1.1. Integrin and ligands**

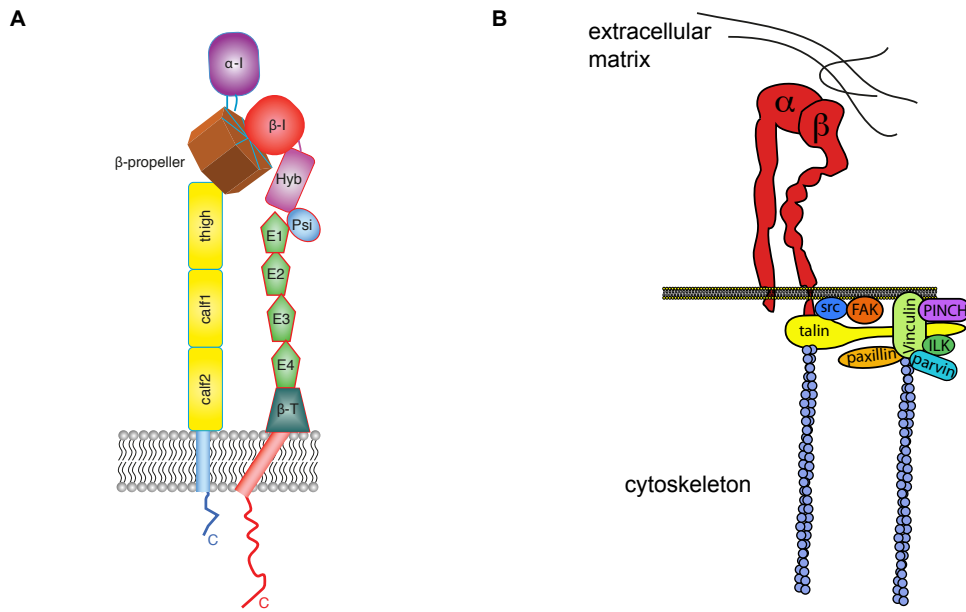
#### Integrin

An integrin is a transmembrane glycoprotein heterodimer, formed by an alpha and a beta subunit (Hynes 1992). There are different alpha and beta subunits, which can combine to form diverse heterodimers, and this combination determines the specificity for different ligands (Barczyk et al. 2010; Maartens & Brown 2015).

Integrins have a large extracellular portion, a transmembrane domain and a cytoplasmic domain (Fig. 2A). The extracellular portion of integrin alpha subunit contains a beta-propeller and a “thigh” and “calf” region, that contains Immunoglobulin (Ig)-like domains (Campbell & Humphries 2011). The extracellular portion of integrin beta subunit has a domain similar to von Willebrand A domain, which is called “ $\beta$ -I-domain” (Campbell & Humphries 2011; Whittaker & Hynes 2002), followed by repeated cysteine-rich EGF domains (Campbell & Humphries 2011). Ligand recognition occurs at the interface of the beta-propeller in the alpha subunit and the “ $\beta$ -I-domain” of the beta subunit (Xiong 2002; Springer et al. 2008; Nagae et al. 2012). Additionally, in chordates, some alpha subunits also have a homolog of this domain, called “ $\alpha$ -I-domain”. In integrins containing this subunit, ligand recognition occurs at the “ $\alpha$ -I-domain”, and the site in “ $\beta$ -I-domain” is thought to have a regulatory role (Campbell & Humphries 2011; Whittaker & Hynes 2002). The “ $\alpha$ -I-domain” contains a MIDAS site (metal-ion-dependent-adhesion site), where a  $Mg^{2+}$  ion coordinates with an acidic residue of the ligand. The “ $\beta$ -I-domain” also contains a MIDAS site, that binds  $Mg^{2+}$ , and two additional sites for  $Ca^{2+}$  binding, ADMIDAS and SyMBS (formerly known as LIMBS, Zhu 2008) which are essential for ligand binding. Alpha subunits contain additional  $Ca^{2+}$  binding sites that are shown to influence ligand binding. The cytoplasmic domain of integrin beta contains NPXY motifs that are known to bind several cytoplasmic proteins (Campbell & Humphries 2011).

#### Integrin ligands

Integrins can recognise different molecules in the ECM. The ligand specificity is provided by the different combinations of alpha and beta subunits. For instance, in vertebrates there are 18 alpha and 8 beta subunits that form 24 different integrin heterodimers which have affinity for different ligands. Yet, an integrin can recognise different ligands, and a ligand can be recognised by different integrins (Barczyk et al. 2010; Campbell & Humphries 2011).



**Figure 2. Integrins and integrin adhesion in animals**

(A) Integrin structure. Integrin  $\alpha$  contains an  $\alpha$ -I-domain, a  $\beta$ -propeller, a calf and thigh region, a transmembrane domain and a cytoplasmic region. Integrin  $\beta$  contains a  $\beta$ -I domain, a Hybrid (Hyb) and Psi domain, EGF modules (E),  $\beta$ -tail, a transmembrane domain and a cytoplasmic region. (adapted from Campbell et al. 2011) (B) Schematic representation of an integrin adhesion complex.

### Fibronectin

In vertebrates, one of the best-characterised integrin ligands is fibronectin, a component of the extracellular matrix and plasma. It is a glycoprotein dimer composed of two monomers which are formed by repetitive units named type-I, type-II and type-III domains (Pankov & Yamada 2002). Outside vertebrates, fibronectin-like proteins have been identified, which contain a different domain architecture (Adams et al. 2015).

Within fibronectin, multiple motifs have been identified as integrin binding motifs (Pankov & Yamada 2002). The most studied integrin binding motif is the tripeptide Arg-Gly-Asp (RGD) found in the type-III (10) domain (Pierschbacher & Ruoslahti 1984; Pankov & Yamada 2002), which is recognised by several integrins (Humphries et al. 2006). The RGD motif has also been found in other ECM proteins which are recognised by integrins, like vitronectin and fibrinogen (Barczyk et al. 2010). An RGD binding integrin also recognises other motifs in fibronectin (Pankov & Yamada 2002). Other motifs recognised by integrins within fibronectin are the LDV motif and the REDV, localised in an alternative spliced site (Pankov & Yamada 2002). There are additional sites found in vertebrate fibronectin that are recognised by LDV-recognising integrins: IDAPS and KALDAPT (in other type-III repeats), and EDGIGEL (in a spliced site) (Pankov & Yamada 2002).

### Other ECM proteins

Collagens are also a well known components of the animal ECM (Rozario & DeSimone 2010). The minimal recognition sequence for integrin binding has been described as GFOGER, although not all collagens contain it (Barczyk et al. 2010).

Laminins are another major component of animal ECM (Rozario & DeSimone 2010). They consist on 3 chains (alpha, beta and gamma) that are structured in distinct domains. These domains are N-terminal domains, LE domains, L4 domains, LF domains, and five globular domains (LG) which are only present in the C-terminal region of the alpha chain. It has been found that some integrins recognise the N-terminal domains whereas others recognise the C-terminal region, especially implicating the LG domains 1-3 (Yamada & Sekiguchi 2015). The mechanism underlying this recognition and the specific motifs that mediate it are not known yet (Yamada & Sekiguchi 2015).

#### **2.1.1.2. Adaptor proteins**

The link with the cytoskeleton involves a group of cytoplasmic proteins, called adaptors (Fig. 2B). Integrins can bind several adaptor proteins through their cytoplasmic tail (Legate & Fassler 2009). These proteins can bind as well to actin filaments, thus mediating a direct link between integrin and actin cytoskeleton. This is the case of talin and alpha-actinin. Others take part in interconnecting the structure by binding to one another, like paxillin and parvin, or to actin, like vinculin. Moreover, some of the adaptors have catalytic activity (FAK, ILK, Src), so they are called catalytic adaptors or signalling proteins (Legate & Fassler 2009; Zaidel-Bar et al. 2007). Below, some of the adaptor proteins are explained.

#### Talin

Talin is an adaptor protein that provides a direct link between integrins and the cytoskeleton as well as a platform for other adaptor proteins, whose function is essential in all adhesions (Klapholz et al. 2015). Its structure is defined as a head and a rod linked by an unstructured region. The head is a FERM domain that bind the cytoplasmic talin of integrin B subunit, and the tyrosine kinase FAK. The rod is formed by helical bundles, and they contain a second binding site for integrin B, binding sites for actin, multiple vinculin binding sites (VBSs), and sites to bind RIAM, a protein involved in binding talin to membrane (Calderwood et al. 2013).

#### Vinculin

Vinculin is an adaptor protein involved in adhesion formation and stability. It provides a link to the cytoskeleton, recruits other proteins and can regulate actin cytoskeleton (Bays & DeMali 2017). It is a protein formed by alpha-helical bundles that form a “head” and a “tail” domain, separated by a “neck” region or linker. The head domain contains binding sites for talin and alpha-actinin among other



proteins. The linker can bind actin regulatory proteins (VASP and Arp2/3). The tail contains binding sites for paxillin and actin (Ziegler et al. 2006; Bays & DeMali 2017).

### Paxillin

Paxillin is an adaptor protein that provides a platform for numerous protein interactions (Deakin & Turner 2008). It contains LD (leu-asp rich) domains in its N-terminal region and LIM domains in its C-terminal region. LD domains are binding sites for several proteins, like vinculin, and FAK (Deakin & Turner 2008). LIM domains are essential for localisation of Paxillin to adhesion sites and they also serve as binding sites for other proteins (Deakin & Turner 2008). Paxillin also contains numerous phosphorylation sites, some provide binding sites for other proteins, and others (in LIM domains) are thought to regulate its localisation to adhesions (Deakin & Turner 2008; Webb 2005). Some of the proteins recruited are kinases (like Src) and phosphatases. This makes paxillin a critical adaptor where signalling converges (Deakin & Turner 2008).

## **2.2. Integrin adhesion structure**

Depending on the size of the adhesion visualised by microscopy, different names have been used to define adhesions: nascent adhesions, focal complexes, and focal adhesions, which refer to different moments in the IAC dynamics in a migrating cell (Parsons et al. 2010; Geiger & Yamada 2011) (Fig. 3A). This way, nascent adhesions elongate and grow into focal complexes, which are later stabilised and grow into even bigger adhesions, called focal adhesions (FAs).

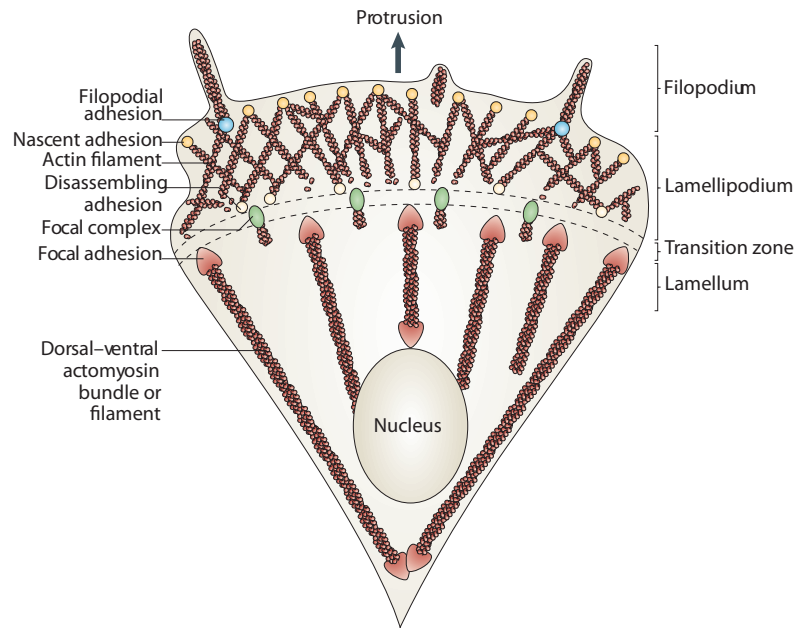
Each of these adhesions are found in different parts of a migrating cell. The front of a migrating cell is called leading edge, formed by the lamellipodium, and filopodia. The lamellipodium is a thin sheet-like structure which contains many actin filaments polymerised by formins and nucleated by Arp2/3, and filopodia emerge from the lamellipodium as thin protrusions formed by parallel bundles of actin filaments (Mattila & Lappalainen 2008; Mellor 2010).

Nascent adhesions are the firstly formed adhesions, visualised as dots in the leading edge of migrating cells. They are transient, they can either disassemble or turn into focal complexes, which are found more centrally and associated with large actin bundles (Vicente-Manzanares & Horwitz 2011; Huttenlocher & Horwitz 2011). These adhesions already contain integrins, talin, vinculin and paxillin (Vicente-Manzanares & Horwitz 2011; Deakin & Turner 2008), but the precise mechanism by which these proteins are recruited to assemble an IAC is still being elucidated (Wehrle-Haller 2012; Klapholz & Brown 2017). Although the mechanisms that regulate assembly of IACs, remain to be fully understood, its architecture has been elucidated. The architecture of FAs has been described as a scaffold with three

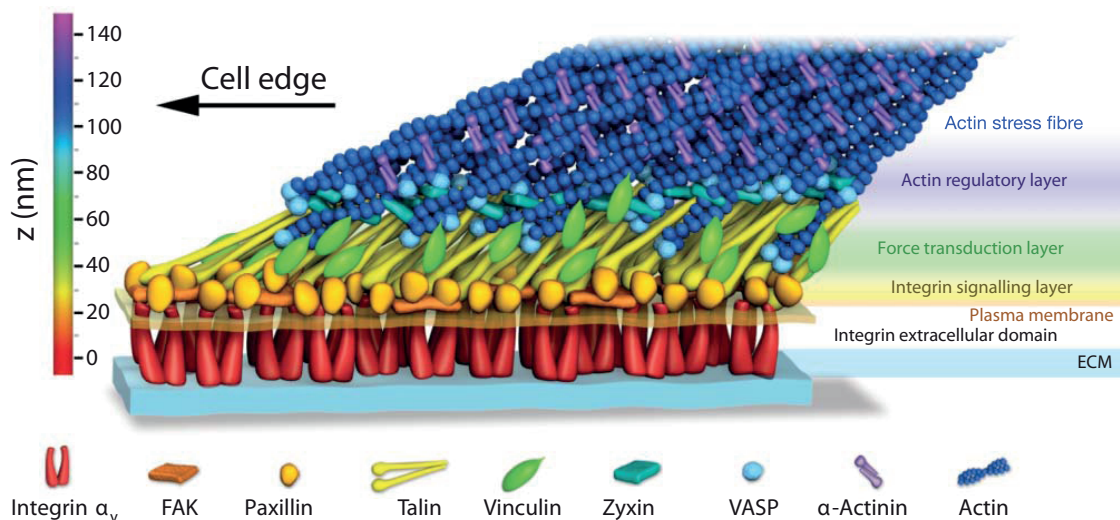
organised levels, which persist along the maturation steps (Kanchanawong et al. 2010) (Fig. 3B).

As a summary, we can understand an IAC as a structure that on the extracellular part can recognise different ligands and on the intracellular part links to the actin cytoskeleton, providing an anchor to the cell.

A



B



**Figure 3. Integrin adhesion structure**

(A) Representation of the integrin adhesion types in a migrating cell (Parsons et al. 2010). (B) Molecular architecture of a focal adhesion modelled from super-resolution microscopy data (Kanchanawong et al. 2010).

### 3. Pre-metazoan integrin adhesome and ECM components

We now know that some of the key proteins from the integrin adhesome predate the origin of animals, as their homologs were found in their unicellular relatives (Sebé-Pedrós et al. 2010). Among the holozoan organisms that have been studied, integrins (both alpha and beta subunits) were identified so far in filastereans [*Capsaspora* and *Pigoraptor* (Sebé-Pedrós et al. 2010; Hehenberger et al. 2017)], and in ichthyosporeans (*Creolimax*) (de Mendoza et al. 2015). It seems that integrins have been secondarily lost in the choanoflagellates (Sebé-Pedrós et al. 2010) except for *Didymoeca costata*, which contains a protein with an integrin beta domain (Richter et al. 2018). Integrins were also found in other lineages outside holozoans. *Thecamonas trahens*, an apusomonad, and *Pygsuia biforma*, a breviatean, contain alpha and beta integrins (Sebé-Pedrós et al. 2010; Brown et al. 2013). These findings mean that integrins appeared not just before animals emerged, but much earlier during the evolution of eukaryotes.

More detailed analysis of the sequence of integrins in *Capsaspora*, *Thecamonas* and *Pygsuia* revealed that they also conserve the overall functional domain architecture from animal integrins (Sebé-Pedrós et al. 2010; Brown et al. 2013). Beta integrins, besides having the b-I-domain, they also have predicted signal peptides, a cysteine-rich stalk, a transmembrane domain and a cytoplasmic region. Beta integrins from *Capsaspora* and *Thecamonas* were analysed more thoroughly and the cation binding motifs MIDAS; ADMIDAS and SyMBS, plus the NPXY motif in the cytoplasmic region of integrin B, are also found in all of them. All these features are however missing in one of the four beta integrin homologs in *Capsaspora* (B4) (Sebé-Pedrós et al. 2010).

Alpha integrins of *Capsaspora*, *Thecamonas* and *Pygsuia* also contain also predicted signal peptides (except A4 of *Capsaspora*) and transmembrane domains (Sebé-Pedrós et al. 2010; Brown et al. 2013). Conservation of the “i-a-domain” has not been reported in these proteins. However, the FG-GAP domains that form the beta-propeller in animal alpha integrins, are conserved. Furthermore, the three identified cation binding sites in alpha integrins in animals are also present in *Capsaspora* and *Thecamonas* alpha homologs (although not in *Capsaspora\_A4*, which contains just two).

Some of the adaptor proteins such as talin, vinculin and paxillin appeared before the Amoebozoa clade diverged, as they are found among amoebozoan species as well as in opisthokonts (Sebé-Pedrós et al. 2010). Some of the signalling proteins, (i.e. Src and FAK, seem to have appeared later during evolution, in the Holozoa clade (Sebé-Pedrós et al. 2010; Hehenberger et al. 2017; Brown et al. 2013). Interestingly, elements of the adhesome, including integrins, seem to have been lost independently in Fungi lineages (Sebé-Pedrós et al. 2010).

Homologs of the animal ECM proteins were not identified in these relative of animals. However, the domains that constitute fibronectin, laminin and collagen are surprisingly present in some of their proteins (Hehenberger et al. 2017; de Mendoza et al. 2015). *Capsaspora* specifically contain proteins with fibronectin type-II and type-III domain, and laminin domains G1 and G2, which are also present in some other unicellular holozoans. In contrast, a collagen domain (fibrillar collagen c-term domain) is only found in choanoflagellates.

In order to find out the original function of these proteins before animals emerged, studying the function in their unicellular relatives represents a good approach (Sebé-Pedrós et al. 2017; Richter & King 2013).

#### **4. Tools to study protein function**

##### **4.1- Tools to study IAPs function in animal models**

The initial steps in elucidating the function of the different proteins of the integrin adhesome were done using antibodies, adhesion assays and protein assays, which deciphered their localisation and their interaction with ECM proteins (Hynes 1987; Burridge 1988). To identify new adhesome components, efforts were later directed into proteomic studies on isolated IACs from different cell types (Horton et al. 2015). Their function in an intact organism has been elucidated by genetic experiments in animals for which tools have been long time available (Takada et al. 2007; Bulgakova et al. 2012).

Functional experiments thus rely on well established methodologies for obtaining mutants and transgenic organisms/cells, the latest depending on cell transfection techniques.

##### **4.2- Genetic tools developed for unicellular eukaryotes**

There are many techniques used to transfer DNA into eukaryotic cells, that have been successful in cell types or organisms that are very diverse. The most widely used are the following:

-Calcium-phosphate precipitation: DNA is delivered to the cells in particles formed by precipitation of calcium and phosphate from a buffered solution, and are thought to be endocytosed (Grosjean et al. 2006). It was first used to transfect adenovirus 5 DNA into human cells (Graham & van der Eb 1973). It is widely used for transfecting mammalian cells (Kingston et al. 2003; Kim & Eberwine 2010), especially sensitive cells like neurons (Jiang & Chen 2006), and for transfecting the amoebozoan *Dictyostelium discoideum* (Nellen et al. 1984; Gaudet et al. 2007). It has been found that several conditions are determinant in the efficiency of transfection. These include the concentration of calcium and phosphate in the solution, the pH, the

temperature at which the precipitation occurs and the time. DNA concentration also showed to affect (Jordan et al. 1996; Jordan & Wurm 2004). On the other side, an osmotic shock with glycerol or DMSO applied after incubation of cells with the DNA precipitates has proven to increase the efficiency of transfection (Jordan & Wurm 2004; Grosjean et al. 2006).

-Electroporation: DNA is delivered to the cells after by affecting the membrane permeability under the application of high voltage (Kotnik et al. 2015). This method was first used in mouse cells (Neumann et al. 1982), and was soon tried in many different animal cell types and some unicellular organisms (Potter 1988). For instance, it was used to successfully transform *Leishmania* (Kapler et al. 1990), the amoebozoan *Dictyostelium discoideum* (Howard et al. 1988) and the unicellular green algae *Chlamydomonas reinhardtii* (Brown et al. 1991). Electroporation is a generalised method for transfecting mammalian cells (Kingston et al. 2003; Kim & Eberwine 2010) and also for *Dictyostelium* (Gaudet et al. 2007). It is still tried with success on other unicellular organisms, for example the diatom *Phaeodactylum tricorutum* (Niu et al. 2012) or the excavate *Parabodo caudatus* (Gomaa et al. 2017). Transfection efficiency depends on many variables, such as the magnitude of the electric pulse, the number of pulses and the conductivity of the medium, which is dependent on the buffer, all parameters that need to be optimised for the organism to transfect (Kingston et al. 2003; Gomaa et al. 2017; Miyahara et al. 2013). There are some electroporation devices commercially available that allow to choose different electrical parameters to adjust the method to the cell type. Examples are Neon (Thermofisher) and NEPA21 (Nepagene).

-Liposome-mediated transfection: DNA molecules are delivered to the cells forming a complex with cationic lipids, which are thought to interact with the membrane and allow its internalisation by endocytosis, although the mechanism is unclear (Felgner et al. 1987; Liu et al. 2003). This technique is widely used for mammalian cells (Kingston et al. 2003; Kim & Eberwine 2010) and several commercial products have been developed, like Lipofectamine (Thermofisher).

-Magnetofection: consists on targeted delivery of the DNA to the cells by coupling it to magnetic particles and applying a magnetic field (Scherer et al. 2002). This method is used in mammalian cells (Kim & Eberwine 2010), for instance in neurons, which are cells hard to transfect (Buerli et al. 2007). Several products are available (OzBioscience), which can also be combined with lipid molecules. The parameters to adjust are the DNA concentration, the ratio of DNA and magnetic particles and the time of incubation with the cells (Plank et al. 2003).

Among Holozoans, genetic tools available are very recent. So far, a forward genetics approach has been developed in the choanoflagellate *Salpingoeca rosetta* leading to the discovery of *rosetteless*, a gene related to colony formation in this species

(Levin et al. 2014). Transfection has already been developed for choanoflagellates (Booth et al. 2018) and the ichthyosporean *Creolimax fragrantissima*, where it allowed the description of synchronous nuclear division during coenocytic development (Suga & Ruiz-Trillo 2013) and in *Corallochytrium limacisporum* (characterisation of stable transfectants is underway). By the beginning of this work, there were no genetic tools reported in filastereans. This thesis included the development of a protocol for transfecting *Capsaspora owczarzaki* (Parra-Acero et al. 2018).

## 5. *Capsaspora owczarzaki*

*Capsaspora* is a unicellular ameboid organism of 3-7  $\mu\text{m}$  diameter, with long filopodia that extend from the body (Stibbs et al. 1979; Hertel et al. 2002) (Fig. 4A). *Capsaspora* was originally found from explants of a strain of *Biomphalaria glabrata*, a fresh water snail, cultivated in a laboratory for studies on the parasite *Schistosoma mansoni*, for which *Biomphalaria* is an intermediate host (Stibbs et al. 1979; Hertel et al. 2002). *Capsaspora* was described as killing and phagocytosing the sporocyst of this parasite *in vitro* via a specialised protrusion (Stibbs et al. 1979; Owczarzak et al. 1980; Hertel et al. 2002), although it is not known whether this also occurs *in vivo* or whether it confers the snail any resistance to the parasite (Hertel et al. 2002).

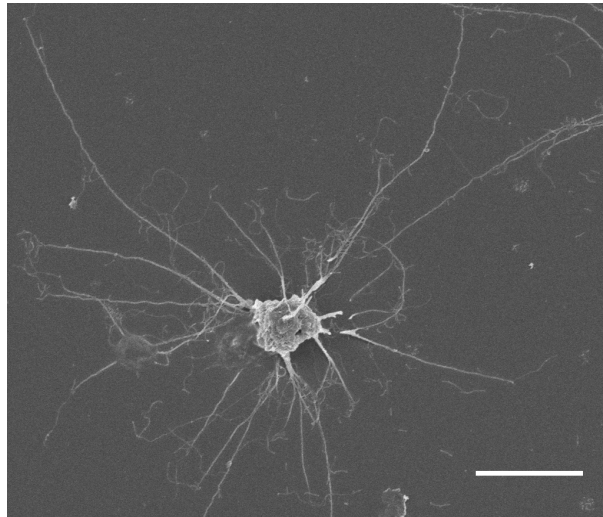
All *Capsaspora* isolations were performed from snails cultivated in laboratory conditions (Stibbs et al. 1979; Hertel et al. 2002), but the slight differences in the 18S (ribosomal DNA) sequence observed between two isolates from snails with different origin suggested the possibility that *Capsaspora* is also a symbiont from natural populations of this snail (Hertel et al. 2002). However, new environmental surveys have not shown any sign of this organism (Ferrer-Bonet & Ruiz-Trillo 2017).

The study of *Capsaspora* is performed on organisms maintained in culture conditions. In this conditions, *Capsaspora* presents an ameboid filopodial stage with exponential growth (Sebé-Pedrós, Irimia, et al. 2013), a cystic stage that is observed after crowding (Hertel et al. 2002; Sebé-Pedrós, Irimia, et al. 2013) and an aggregative stage in which cells gather together and form clusters, which happen spontaneously during normal conditions and can also be induced by agitation (Sebé-Pedrós, Irimia, et al. 2013) (Fig. 4B). Electron microscopy images have shown the presence of an extracellular material of unknown composition around the cells forming aggregates, tentatively called extracellular matrix (Sebé-Pedrós, Irimia, et al. 2013) (Fig. 4B f).

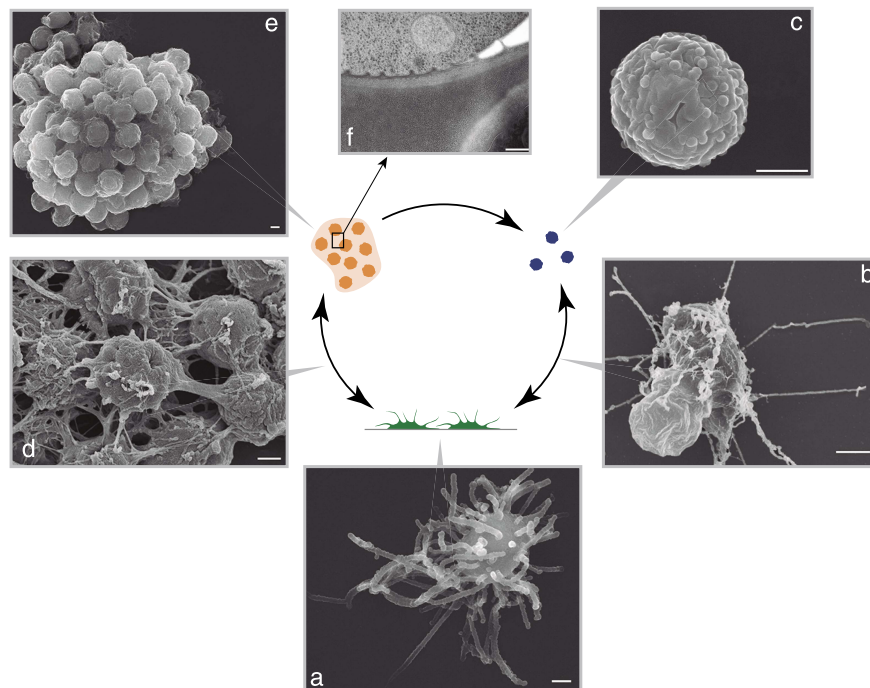
*Capsaspora* has been studied extensively at the molecular level. The genome of *Capsaspora* and a transcriptome of each of the three life stages in culture have been sequenced (Suga et al. 2013; Sebé-Pedrós, Irimia, et al. 2013). Its proteome and

phosphoproteome have been characterised across the three life stages as well as its epigenomic features (Sebé-Pedrós, Ballaré, et al. 2016; Sebé-Pedrós, Peña, et al. 2016). All this knowledge makes *Capsaspora* an ideal organism for studies directed towards deciphering origin of animals.

A



B



**Figure 4. *Capsaspora owczarzaki***

(A) SEM image of a filopodiated *Capsaspora owczarzaki*. scale bar = 5 $\mu$ m. (B) Life stages of *Capsaspora owczarzaki* in culture conditions. (a) filopodial amoeba. (b) transition from filopodial to cystic. (c) rounded cyst. (d) transition from adherent to aggregative stage. (e) mature aggregate. (f) transmission-EM showing adjacent cells in the aggregate separated by extracellular matrix. Scales bar= 1 $\mu$ m, except panel d, = 200 nm (Sebé-Pedrós, Irimia, et al. 2013).

## **OBJECTIVES**





The analysis of *Capsaspora owczarzaki* genome revealed it has a complex gene repertoire that includes many genes involved in animal multicellularity, like the most complete transcription factor repertoire among unicellular relatives of animals, and the core elements of the integrin adhesome. These results and its phylogenetic position highlighted *Capsaspora* as a promising organism to study functionally the genes related to animal multicellularity as a proxy to infer their ancestral function, with the ultimate goal of understanding how multicellularity in animals evolved from a unicellular organism.

Understanding the behaviour of *Capsaspora* in culture conditions is essential to design experiments with the objective of studying a particular feature. Given that no functional studies have been carried out to date, setting up standard protocols for manipulating *Capsaspora* as well as developing molecular tools such as antibodies, is also necessary. Genetic tools that would allow the functional analysis of selected genes are also not available. In this framework, and focusing on *Capsaspora*, I developed this Ph.D. thesis with the following main objectives:

1. Describing the cell cycle in culture conditions of *Capsaspora owczarzaki*
2. Setting up molecular tools to perform future functional analysis
3. Producing a protocol to transfect *Capsaspora owczarzaki*
4. Gaining insight into the function of the protein adhesome of *Capsaspora owczarzaki*



## **RESULTS**



## Impact factor and authorship report of the publications

Director: Dr. Iñaki Ruiz-Trillo

Three out of the four articles that conform this thesis dissertation (Results R1-4) have been published in high impact journals covering the fields of Cell and Molecular Biology, Genetics and Genomics, and Developmental Biology, including both field-specific and multidisciplinary journals. The fourth article is an unpublished manuscript, still in preparation, which is intended to be submitted to a high impact journal in the Cell Biology area.

Among the four manuscripts here presented, Helena Parra-Acero has been the sole first author of R4 and co-first author of the manuscript presented in R3. The two remaining articles are collaborations with fellow researchers in the laboratory of Iñaki Ruiz-Trillo.

The specific contributions of Helena Parra-Acero to each publication are indicated below, together with the yearly impact factor and ranking of each journal (as per the ISI proprietary ranking).

### **Publication R1 – Aggregative multicellularity in *Capsaspora***

Sebé-Pedrós, A., Irimia, M., del Campo, J., **Parra-Acero, H.**, Russ, C., Nusbaum, C., Blencowe, B. J., Ruiz-Trillo, I. 2013. Regulated aggregative multicellularity in a close unicellular relative of metazoa. *eLife* 2:e01287.

Impact Factor (2013): 8.519

Journal ranking: Biology Q1 (4/85)

Authorship: Collaboration in an article authored by ASP, who designed the study jointly with MI, CN and BJB and IRT. This article was part of the thesis of ASP. Helena assisted ASP in the collection and flow cytometry analysis of samples to describe the cell cycle of *Capsaspora*. Helena was also involved in developing the staining assay, performing several dye testing and acquiring the images with confocal microscopy, which demonstrated that aggregates are essentially non-clonal.

### **Publication R2 – Regulatory genome of *Capsaspora***

Sebé-Pedrós A., Ballaré C., **Parra-Acero, H.**, Chiva, C., Tena, J. J., Sabidó, E., Gómez-Skarmeta, J. L., Di Croce L., Ruiz-Trillo I. 2016. The dynamic regulatory genome of *Capsaspora* and the origin of animal multicellularity. *Cell* 165: 1224–1237.

Impact Factor (2016): 30.410

Journal ranking: Biochemistry and Molecular Biology Q1 (1/290); Cell Biology Q1 (2/190)

Authorship: Collaboration in an article authored by ASP, who designed the project with JLGS, IRT, and LD. ASP, CB, and Helena performed the experiments. Helena was in charge of Brachyury antibody design and testing. Helena was also in charge of deacetylase drug testing, deacetylase inhibition experiments and microscopy imaging. Moreover, Helena participated in chromatin extraction, preparation and testing with Brachyury and several PolII antibodies for subsequent ChiP-seq experiments.

### **Publication R3 – Transfection of *Capsaspora***

**Parra-Acero, H.**, Ros-Rocher, N., Perez-Posada, A., Kożyczkowska, A., Sánchez-Pons, N., Nakata, A., Suga, H., Najle, S. R., Ruiz-Trillo, I. *Development* dev162107.

Impact Factor (2018): 5.413

Journal ranking: Developmental Biology Q1 (5/42)

Authorship: Co-first authorship with NRR. This article is part of the thesis of NRR. Helena played a vital role in the design of the experiments. She also performed an important percentage of the experiments (was involved in the cloning of the vectors, transfection experiments, cytometry analysis and imaging). Helena also lead (with NRR) the analyses and interpretation of the results; as well as the writing the manuscript.

### **Unpublished result R4 (in preparation) – Integrin adhesome of *Capsaspora***

**Parra-Acero, H.**, Dudin, O., Harcet, M., Ruiz-Trillo, I. A first approach into the function of the integrin adhesome of *Capsaspora owczarzaki*

Authorship: Helena and OD designed the study. Helena designed and tested the antibodies, performed immunostaining, developed the adhesion assay, and performed live microscopy. Helena also analysed the results, with OD, and wrote the manuscript.

Iñaki Ruiz-Trillo

Director

## **R1: Regulated aggregative multicellularity in a close unicellular relative of metazoa**

### *Abstract*

The evolution of metazoans from their unicellular ancestors was one of the most important events in the history of life. However, the cellular and genetic changes that ultimately led to the evolution of multicellularity are not known. In this study, we describe an aggregative multicellular stage in the protist *Capsaspora owczarzaki*, a close unicellular relative of metazoans. Remarkably, transition to the aggregative stage is associated with significant upregulation of orthologs of genes known to establish multicellularity and tissue architecture in metazoans. We further observe transitions in regulated alternative splicing during the *C. owczarzaki* life cycle, including the deployment of an exon network associated with signaling, a feature of splicing regulation so far only observed in metazoans. Our results reveal the existence of a highly regulated aggregative stage in *C. owczarzaki* and further suggest that features of aggregative behavior in an ancestral protist may have been co-opted to develop some multicellular properties currently seen in metazoans.





# Regulated aggregative multicellularity in a close unicellular relative of metazoa

Arnau Sebé-Pedrós<sup>1,2</sup>, Manuel Irimia<sup>3</sup>, Javier del Campo<sup>1</sup>, Helena Parra-Acero<sup>1</sup>, Carsten Russ<sup>4</sup>, Chad Nusbaum<sup>4</sup>, Benjamin J Blencowe<sup>3\*</sup>, Iñaki Ruiz-Trillo<sup>1,2,5\*</sup>

<sup>1</sup>Department of Functional Genomics and Evolution, Institut de Biologia Evolutiva (CSIC–Universitat Pompeu Fabra), Barcelona, Spain; <sup>2</sup>Departament de Genètica, Facultat de Biologia, Universitat de Barcelona, Barcelona, Spain; <sup>3</sup>Donnelly Centre, University of Toronto, Toronto, Canada; <sup>4</sup>Genomics Platform, Broad Institute of Harvard and the Massachusetts Institute of Technology, Cambridge, United States; <sup>5</sup>Institució Catalana de Recerca i Estudis Avançats (ICREA), Barcelona, Spain

**Abstract** The evolution of metazoans from their unicellular ancestors was one of the most important events in the history of life. However, the cellular and genetic changes that ultimately led to the evolution of multicellularity are not known. In this study, we describe an aggregative multicellular stage in the protist *Capsaspora owczarzaki*, a close unicellular relative of metazoans. Remarkably, transition to the aggregative stage is associated with significant upregulation of orthologs of genes known to establish multicellularity and tissue architecture in metazoans. We further observe transitions in regulated alternative splicing during the *C. owczarzaki* life cycle, including the deployment of an exon network associated with signaling, a feature of splicing regulation so far only observed in metazoans. Our results reveal the existence of a highly regulated aggregative stage in *C. owczarzaki* and further suggest that features of aggregative behavior in an ancestral protist may have been co-opted to develop some multicellular properties currently seen in metazoans.

DOI: [10.7554/eLife.01287.001](https://doi.org/10.7554/eLife.01287.001)

**\*For correspondence:**

b.blencowe@utoronto.ca (BJB);  
inaki.ruiz@multicellgenome.org  
(IR-T)

**Competing interests:** See page 17


**Funding:** See page 17

**Received:** 25 July 2013

**Accepted:** 06 November 2013

**Published:** 24 December 2013

**Reviewing editor:** Diethard Tautz, Max Planck Institute for Evolutionary Biology, Germany

 Copyright Sebé-Pedrós et al. This article is distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use and redistribution provided that the original author and source are credited.

## Introduction

Living organisms emerged from the integration of multiple levels of organization. These levels were shaped by both physiochemical constraints and historical circumstances, the latter being more important in more complex systems (Jacob, 1977). Therefore, it is important to identify the phylogenetic inertia (sensu Burt, 2001) imposed by the raw starting material in order to properly understand major evolutionary transitions, such as the origin of metazoan multicellularity (Knoll, 2011). Examination of both the genetic repertoire (King, 2004; Ruiz-Trillo et al., 2007; Rokas, 2008) and the cell types present in the immediate unicellular relatives of metazoans can provide insights into this evolutionary transition, as they reveal the historical constraints in early metazoan evolution. In this regard, the analyses of unicellular holozoan genomes, that is choanoflagellates and filastereans, have shown that the genetic repertoire of the metazoan unicellular ancestor was much more complex than previously thought (Abedin and King, 2008; King et al., 2008; Sebé-Pedrós et al., 2010; Suga et al., 2013).

Multicellularity has been independently acquired multiple times during the evolution of eukaryotes, in more than 20 different lineages including animals, plants, fungi, slime molds, green and brown algae, and several other eukaryotes (King, 2004; Parfrey and Lahr, 2013). Multicellular organisms evolved through two major strategies: aggregation of different cells or clonal division of a single cell. Multi-level selection theory has proposed that the most complex multicellular organisms likely arose through clonal development rather than by aggregation of genetically diverse cells, since intra-organismal

**eLife digest** When living things made from many cells evolved from single-celled ancestors, it was a breakthrough in the history of life—and one that has occurred more than once. In fact, multicellular life has evolved independently at least 25 times, in groups as diverse as animals, fungi, plants, slime molds and seaweeds. There are broadly two ways to become multicellular. The most complex multicellular species, such as animals, will replicate a single cell, over and over, without separating the resultant cells. However, in species that are only occasionally multicellular, free-living cells tend instead to join together in one mass of many cells.

Evolution is constrained by its raw materials; so looking at the living relatives of a given species, or group, can lead to a better understanding of its evolution because its relatives contain clues about its ancestors. To gain insights into how animal multicellular life might have began; Sebé-Pedrós et al. studied the life cycle of the amoeboid organism *Capsaspora owczarzaki*. Found within the bodies of freshwater snails, this single-celled amoeba is a close relative of multicellular animals and could resemble one of their earliest ancestors.

At certain stages of the life cycle Sebé-Pedrós et al. noticed that individual amoebae gathered together to form a multicellular mass—something that had not been seen before in such a close relative of the animals. Moreover, the genes that ‘switched on’ when the amoebae began to aggregate are also found in animals; where, together with other genes, they control development and the formation of tissues. Sebé-Pedrós et al. suggest that the first multicellular animals could have recycled the genes that control the aggregation of single-celled species: in other words, genes that once controlled the changes that happen at different times in a life cycle, now control the changes that develop between different tissues at the same time.

Sebé-Pedrós et al. also observed that alternative splicing—a process that allows different proteins to be made from a single gene—occurs via two different mechanisms during the life cycle of *Capsaspora*. Most of the time *Capsaspora* employs a form of alternative splicing that is often seen in plants and fungi, and only rarely in animals; for the rest of the time it uses a form of alternative splicing similar to that used by animal cells.

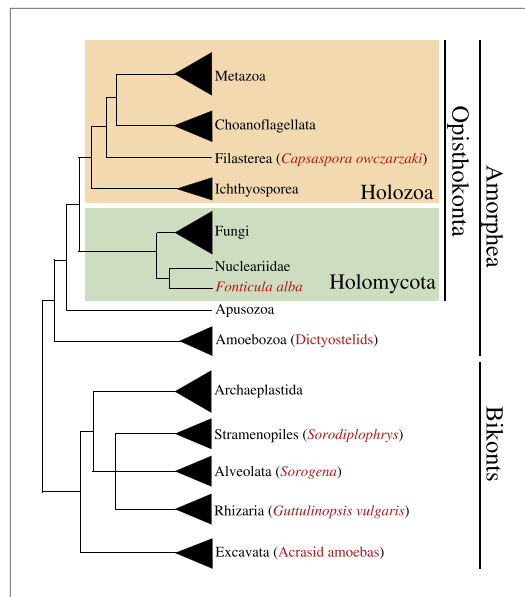
The evolution of complex alternative splicing mechanisms is a hallmark feature of multicellular animals. The exploitation of two major forms of alternative splicing in *Capsaspora* could thus reflect an important transition during evolution that resulted in an increased diversity of proteins and in more complex gene regulation. Such a transition may ultimately have paved the way for the increased specialization of cell types seen in animals.

This glimpse into the possible transitions in gene regulation that contributed to the birth of multicellular animals indicates that the single-celled ancestor of the animals was likely more complex than previously thought. Future analyses of the animals’ close relatives may further improve our understanding of how single-celled organisms became multicellular animals.

DOI: [10.7554/eLife.01287.002](https://doi.org/10.7554/eLife.01287.002)

competition in the latter might be expected to be evolutionarily unstable (**Grosberg and Strathmann, 2007; Michod, 2007; Newman, 2012**). Accordingly, eukaryotic lineages that attained the most complex multicellular lifestyles (i.e., plants and metazoans) arose through clonal cell division (**Grosberg and Strathmann, 2007**). In contrast to clonal multicellularity, aggregative cell behavior typically represents a transient life cycle stage. This type of multicellularity arose within several eukaryotic clades, including the dictyostelids (Amoebozoa) (**Schaap, 2011**), acrasid amoebas (Heterolobosea, Discicristata, Discoba) (**Brown et al., 2011**), *Guttulinopsis vulgaris* (Cercozoa, Rhizaria) (**Brown et al., 2012**), the genus *Sorogena* (Ciliata, Alveolata) (**Lasek-Nesselquist and Katz, 2001**), the holomycota *Fonticula alba* (Opisthokonta) (**Brown et al., 2009**) and the genus *Sorodiplophrys* (Labyrinthulomycetes, Heterokonta) (**Dykstra and Olive, 1975**) (**Figure 1**).

Within the opisthokont clade that comprises Metazoa, Fungi and their unicellular relatives (**Cavalier-Smith, 2003; Steenkamp et al., 2006; Ruiz-Trillo et al., 2008**), so far only a single taxon has been described to have aggregative behavior, which is *F. alba* (**Brown et al., 2009**), a close relative of Fungi. Moreover, among close unicellular relatives of Metazoa, clonal development is the only known multicellular behavior, as described in choanoflagellates and ichthyosporeans (**Dayel et al., 2011; Suga and Ruiz-Trillo, 2013**). Within metazoans, which have largely clonal development, some cells show



**Figure 1.** Phylogenetic position of *Capsaspora owczarzaki* within the eukaryotes. The Holozoa clade (in yellow) includes Metazoa and their closest unicellular relatives: Choanoflagellata, Filasterea, and Ichthyosporea. *C. owczarzaki* represents one of the two known filastereans taxa that form the sister-group of choanoflagellates and metazoans. Other major eukaryotic groups are shown. Within each group, species or clades with aggregative multicellularity (see text) are highlighted in red.

DOI: [10.7554/eLife.01287.003](https://doi.org/10.7554/eLife.01287.003)

aggregative behaviors; for example, mesenchymal (O'Shea, 1987) and germ line cells (Savage and Danilchik, 1993) during development, sponge cells after cell dissociation (Wilson, 1907) and arthropod blood cells through active amoeboid movement (Loeb, 1903, 1921).

To gain deeper insight into the possible transitions that arose during the emergence of metazoan multicellularity, we have performed a detailed examination of the life cycle and associated transcriptomic changes of *Capsaspora owczarzaki*, one of the closest known unicellular relatives of metazoans (Figure 1). Isolated decades ago as an endosymbiont of the fresh-water snail *Biomphalaria glabrata* (Owczarzak et al., 1980), *C. owczarzaki* belongs to the clade Filasterea, the sister-group of Metazoa and choanoflagellates (Torruella et al., 2012). Filasterea also includes a free-living marine unicellular species known as *Ministeria vibrans* (Shalchian-Tabrizi et al., 2008).

We analyzed the *C. owczarzaki* life cycle and its regulation using electron microscopy, flow cytometry and high-throughput RNA sequencing (RNA-Seq). Through these analyses, we show that *C. owczarzaki* life cycle is tightly regulated at the level of gene expression and alternative splicing (AS). Moreover, we demonstrate the existence of an aggregative multicellular stage in *C. owczarzaki* in which many orthologs of genes important for metazoan clonal multicellularity are upregulated.

## Results and discussion

Under initial culture conditions ('Materials and methods'), *C. owczarzaki* differentiates into an amoeba that crawls over substrate (Video 1), surveying its environment with its filopodia (Sebé-Pedrós et al., 2013). At this stage, active DNA replication occurs (with >10% of the cells in S-phase) and, within 48 hr, the cells enter an exponential growth phase (Figures 2 and 3). Subsequently, the cells start to detach from the surface and begin to retract their filopodia and encyst (Figure 3B,C). After 8 days, no attached amoebas remain and growth is stabilized (Figures 2 and 3). This cystic stage may represent a dispersal resistance form. Strikingly, we observe an alternative path to this process involving the active formation of cell aggregates (Videos 2 and 3). In these aggregates, the cells attach to each other and produce cohesive extracellular material (Figure 3D) until a compact cell aggregate, in which cells no longer bear filopodia, is formed (Figure 3E). Transmission electron microscopy demonstrates the presence of a thick, unstructured, extracellular material within the aggregates that appears to prevent direct contact between cells (Figure 3F). Clusters of cells appear to occur at random under normal culture conditions.

To investigate whether *C. owczarzaki* cell clusters are formed through aggregation or clonal division, we first mixed two differentially stained populations of cells ('Materials and methods') and induced aggregate formation, resulting in dual-colored cell clusters (Figure 4A). This indicates that cell clusters are not composed of daughter cells resulting from successive cell divisions, which would result in single-color cell clusters, but instead by aggregation of multiple cells. We also observed that aggregates could form efficiently even when cell division was blocked by two different inhibitors, hydroxyurea and aphidicolin (Figure 4B). Finally, by flow cytometry, we observed that the proliferation rate of aggregative cells (Figure 4—figure supplement 1) is extremely low (compared with the observations in Figure 2). Overall, these results show that *C. owczarzaki* cell clusters form by active cell aggregation, not by clonal cell division. This observation represents the first reported case of aggregative multicellularity in a close unicellular relative of Metazoa.



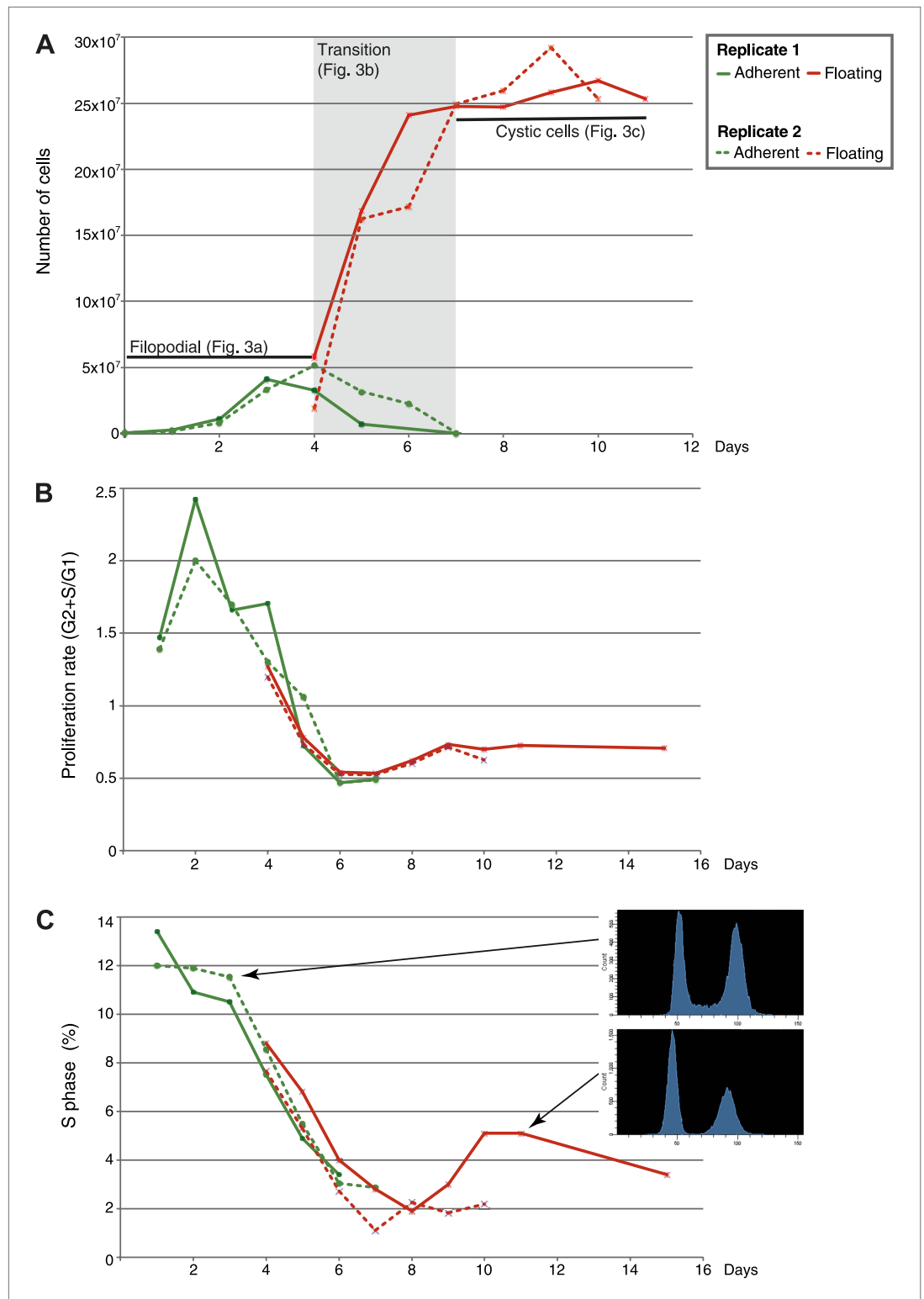
**Video 1.** *C. owczarzaki* filopodial amoeba stage cells crawling. Dark and refringent vesicles can be observed inside each cell. Up to nine different cells can be observed in the video. Also available on YouTube: [http://youtu.be/0Uyhor\\_nDts](http://youtu.be/0Uyhor_nDts). DOI: [10.7554/eLife.01287.004](https://doi.org/10.7554/eLife.01287.004)

animal multicellularity, ancestral premetazoan cell types may have been integrated into a single multicellular entity by means of controlling cell differentiation spatially, rather than temporally (Mikhailov et al., 2009). An alternative explanation is that some of the cell behaviors observed in extant unicellular relatives of Metazoa may have evolved independently in some particular unicellular holozoan lineages or species, and do not represent ancestral states. The limited taxon sampling in many of these poorly studied lineages makes it difficult to reliably assess whether these are derived or ancestral characters and this situation is especially dramatic in the case of filastereans, in which only two species have been described so far (*C. owczarzaki* and the free-living *M. vibrans*).

To investigate the molecular composition and regulation of the distinct life cycle stages of *C. owczarzaki*, we isolated filopodiated amoebae, aggregates, and cysts, and analyzed their transcriptomes using RNA-Seq. Of 8637 annotated genes, 4486 showed statistically significant differential regulation ('Materials and methods') in one or more pair-wise comparisons between life cycle stages, including 1354 changes between filopodial and aggregate stages, 3227 between filopodial and cystic stages, and 3096 between aggregate and cystic stages. Moreover, when performing one-versus-all comparisons, each cell stage had a specific transcriptomic profile (Figure 5A), indicating tight regulation at the level of gene expression. Using both pairwise and one-versus-all comparisons, we identified significantly enriched gene ontology (GO) categories (Figures 5 and 6) and Pfam protein domains (Figure 7) in each set of differentially expressed (both up and down-regulated) genes ( $p < 0.01$  for each significant category; Fisher's exact test). Genes upregulated in the filopodial stage were enriched in signalling functions, such as tyrosine kinase activity and G-protein-coupled receptor activity, as well as in transcription factors, especially of the Basic Leucine Zipper Domain (bZIP) superfamily (Figures 5 and 6). Genes involved in protein synthesis and DNA replication were also significantly upregulated, consistent with the rapid cell proliferation at this stage observed by flow cytometry (Figure 2), and further suggesting a high metabolic rate.

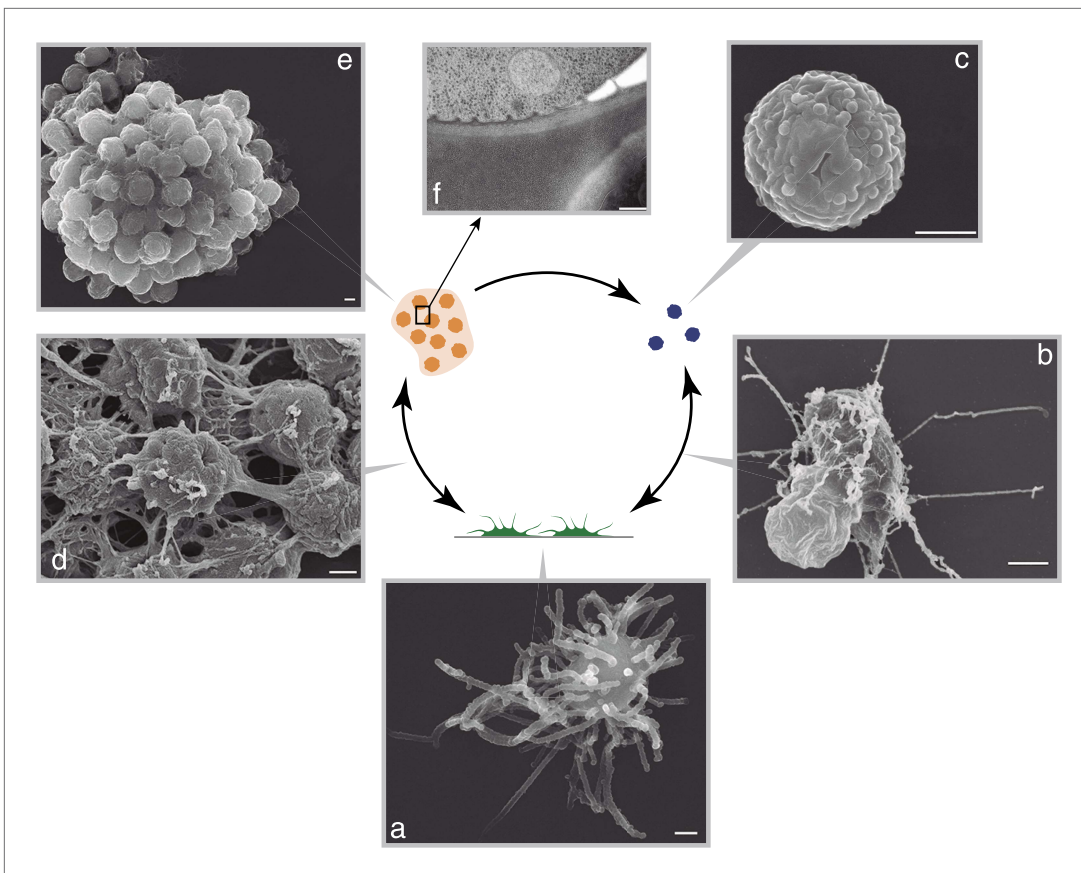
When compared to filopodial and aggregative cells, cystic cells showed significant downregulation of genes associated with myosin transport, translation, DNA replication and metabolic activities (especially mitochondrial energy production). However, genes involved in vesicle transport and autophagy were significantly upregulated at this stage (Figure 5). These differences may reflect recycling of intracellular components triggered by starvation or other adverse conditions, as has been observed under conditions of adaptive cell survival in other eukaryotes (Kiel, 2010). Protein domains involved in the ubiquitin pathway (e.g., UQ\_con, zf-RING2 and Cullin domains) and in synaptic cell-cell communication, such as SNARE, synaptobrevin and syntaxin, as well specific transcription factor families (e.g., bHLH transcription factors), were also significantly upregulated in the cystic cells (Figure 7). Altogether, these results suggest that major cytosolic rearrangement and protein turnover occur at the cystic stage.

The aggregative multicellularity observed in *C. owczarzaki* adds an additional cell behavior to those already known among extant close unicellular relatives of Metazoa (i.e., choanoflagellates, ichthyosporeans, and filastereans [Torruella et al., 2012]). We can then infer that multiple cell types and behaviors (including aggregative behavior, flagellar motility, amoeboid movement, clonal colony formation, etc) were most likely present among the unicellular ancestors of metazoans. This range of cell behaviors may have provided the basis for the evolution of the diverse cell types seen in animals (Arendt, 2008; Arendt et al., 2009). Interestingly, each one of the three known unicellular lineages closely related to Metazoa (choanoflagellates, ichthyosporeans and filastereans) has some kind of simple multicellularity. Moreover, the tight regulation observed in *C. owczarzaki* (see below) emphasizes that a regulated temporal cell differentiation was already in place among unicellular ancestor of animals. This reinforces the view that, during the transition to



**Figure 2.** Flow-cytometry analysis of *C. owzarzaki* cell cycle. **(A)** Total number of cells per day in each fraction (adherent and floating, see 'Materials and methods'). **(B)** Proliferation rate per day. **(C)** Percentage of cells in S-phase per day and two examples of DNA-content profiles obtained from days 3 and 11. Note the reduction in the number of G2/M cells (second peak) and the drastic reduction in S-phase cells (the area between the two peaks). Data from adherent cells ('Materials and methods') is shown in green and data from floating cells in red. Experimental replicate 1 results are shown with solid lines and replicate 2 results with dashed lines.

DOI: [10.7554/eLife.01287.005](https://doi.org/10.7554/eLife.01287.005)



**Figure 3.** *C. owzarzaki* life cycle. (A) Filopodial stage cells, amoebas with long filopodia. (B) Transition from filopodial to cystic stage: cells retract filopodia. (C) Cystic stage cells are rounded cysts without filopodia and slightly smaller than filopodial cells. (D) Transition from filopodial to aggregative stage: cells attach to each other and an extracellular matrix appears. (E) Mature aggregate. (F) Transmission EM showing adjacent cells in the aggregate separated by extracellular matrix. Arrows indicate the observed stage inter-conversions. Scale bars = 1  $\mu\text{m}$ , except in panel D = 200 nm.

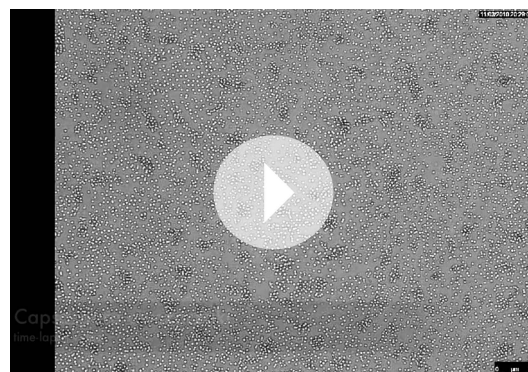
DOI: [10.7554/eLife.01287.006](https://doi.org/10.7554/eLife.01287.006)

Remarkably, the aggregative stage showed strong upregulation of the components of the integrin adhesome and associated signalling and cell-adhesion proteins (**Figures 5 and 8A,B**), such as the LamininG domain-containing protein CAOG\_07351 (which contains a N-terminal signal peptide sequence and therefore is likely to be secreted) (**Figure 8C**), the IPP complex (ILK-PINCH-Parvin) signalling module, G-protein  $\alpha$ -13 (**Gong et al., 2010**), several cytoplasmatic tyrosine kinases (**Hamazaki et al., 1998**) and two receptor tyrosine kinases (which possess extracellular DERM [Lewandowska et al., 1991] and fibronectin\_3 domains, known to interact with integrins [Figure 8C]). These observations strongly suggest that the integrin adhesome and the likely associated tyrosine kinase signalling genes play an important role in the formation of the *C. owzarzaki* aggregates. Furthermore, we also observed upregulation of genes involved in mitosis and in the tubulin cytoskeleton (e.g., kinesins) at the aggregative stage (**Figure 5**). These results indicate that a molecular repertoire associated with animal multicellularity, could function either in aggregative or in clonal multicellularity and in different phylogenetic contexts, in line with previous hypotheses (**Newman, 2012**).

A hallmark feature of the evolution of metazoan multicellularity and cell type diversity is the expansion of AS complexity and regulation through exon skipping, which has entailed the formation of cell type-specific networks of co-regulated exons belonging to functionally related or pathway-specific genes (**Irimia and Blencowe, 2012**). In contrast, differential intron retention is the most widespread form of AS found in non-metazoan eukaryotic species (**McGuire et al., 2008**). To assess the extent to which these forms of AS may contribute to gene regulation in *C. owzarzaki*, we systematically mapped reads from each life cycle stage to a comprehensive set of intron–exon and exon–exon junctions



**Video 2.** *C. owczarzaki* cells aggregation. Also available on YouTube: <http://youtu.be/83HB8srWQw4>. DOI: [10.7554/eLife.01287.007](https://doi.org/10.7554/eLife.01287.007)



**Video 3.** *C. owczarzaki* cells aggregation. Also available on YouTube: <http://youtu.be/Ovl6BvBucrc>. DOI: [10.7554/eLife.01287.008](https://doi.org/10.7554/eLife.01287.008)

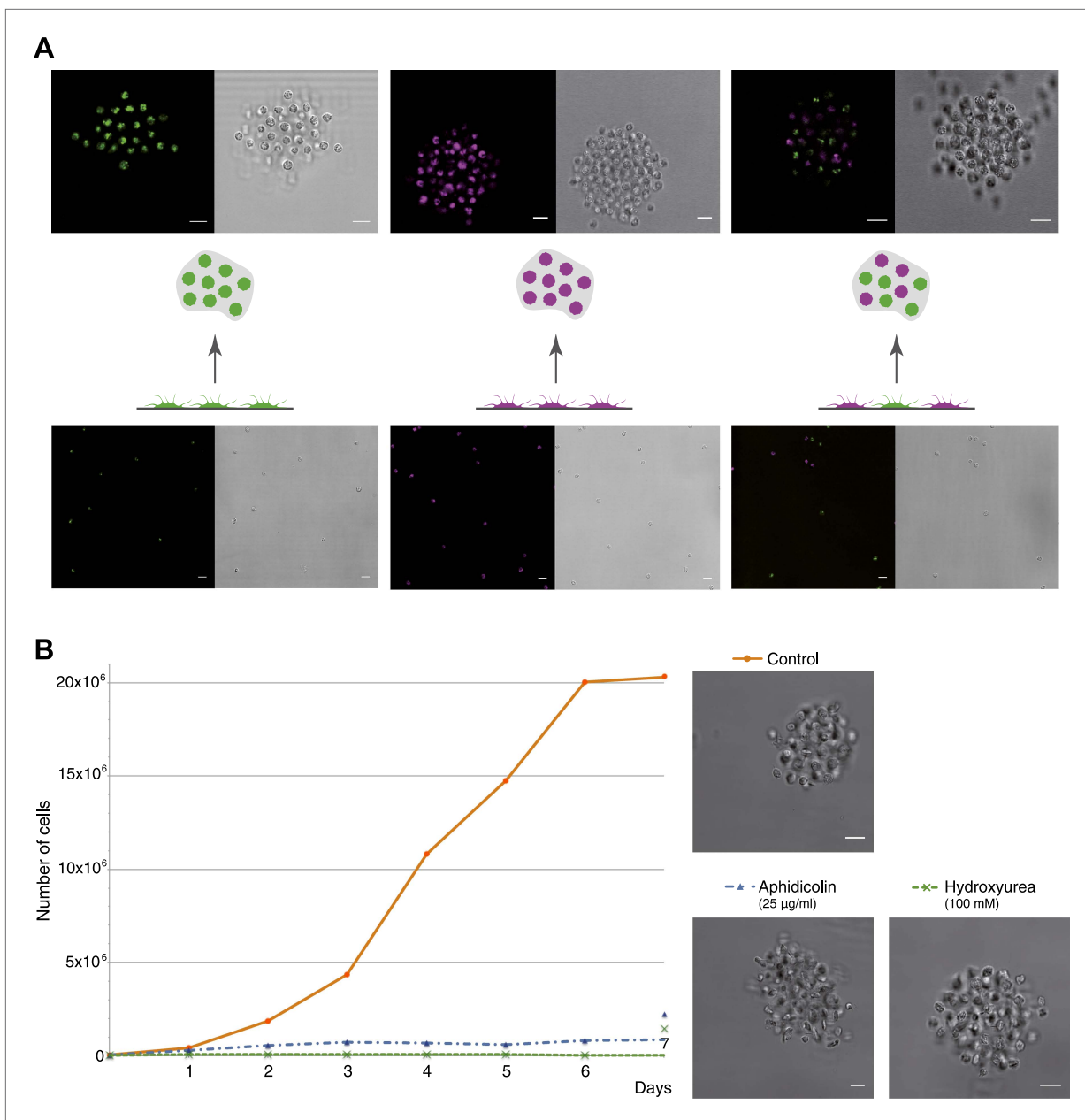
that most if not all of these retained introns act by reducing the levels of spliced mRNAs exported from the nucleus and translated into protein, as has been observed previously for regulated retained introns in metazoan species (Yap *et al.*, 2012). Moreover, we observe that multiple introns belonging to a gene can be retained in a stage-specific manner. For instance, >73% and >29% of multi-intronic genes with one differentially retained intron had at least one additional differentially retained intron at the filopodial or cystic-specific stages, respectively, and 22% and 5% of genes at these stages showed evidence of high retention for all introns in the same genes. Furthermore, RT-PCR analyses and mate information from paired-end read analyses suggested that multiple intron retention events often occur in a combinatorial manner (Figure 9—figure supplement 1), thereby increasing the potential impact of intron retentions on mRNA regulation. All of the above observations were highly consistent across three biological replicates (Figure 9B), and not observed for neighbouring genes, ruling out contamination of genomic DNA.

We analyzed different features of differentially retained introns that may account for their stage-specific regulation. First, we compared intron lengths. While filopodial-specific differentially retained introns have a similar length distribution to constitutive (PSI less than 2% across all stages) introns, cystic stage-specific introns were significantly longer ( $p=1.7e^{-14}$  Wilcoxon rank sum test) (Figure 9D). In line with this observation, the average level of intron retention increased steadily with intron length only in the cystic stage (Figure 9E). Furthermore, cystic stage-specific retained introns harbored significantly weaker canonical 5' and 3' splice site signals than other intron sets ( $p<0.0013$  Wilcoxon rank sum test for all comparisons). Collectively, these data suggest that differential intron retention in the cystic stage may be associated with suboptimal introns (i.e., long and with weak splice sites) that are more efficiently spliced in the other cell stages. In the case of the filopodial-specific differentially

(i.e., formed by exon/intron inclusion and skipping) to score their differential usage. Of 25,677 introns with sufficient RNA-Seq read coverage across the three life cycle stages, 2986 (11.6%) showed  $\geq 20\%$  PSI (Percent Spliced In, percent of transcripts from a given gene in which the intron sequence is present) in at least one stage, and approximately a third of genes had at least one such intron retention event. Moreover, we observed marked differences in the extent to which detected intron retention is differentially regulated between the different life cycle stages (Figure 9A). In particular, 797 retained introns (in 441 genes) and 259 retained introns (in 232 genes) display differential PSI (dPSI) values of 25% or more in the filopodial and cystic stages compared to the other stages, respectively (Figure 9B). In contrast, no retained introns were found to be differentially spliced at the aggregative stage. Most (12 out of 15, 80%) of the analyzed cases of differentially retained introns were validated by RT-PCR (Figure 9C and Figure 9—figure supplement 1).

GO enrichment analysis for the two sets of differentially retained introns showed distinct gene function enrichment (e.g., protein kinase activity and intracellular targeting in the filopodial stage, and histone modification and myosin complex in the cystic stage) (Figure 6), implying that regulated intron retention plays different roles at these stages. A low fraction of read-through introns (with length-multiples of three and no in-frame stop codons) were found among the two sets of differentially retained introns, suggesting





**Figure 4.** *C. owczarzaki* cell clusters form by active aggregation, not clonal cell division. Aggregation was induced in different stained cell populations ('Materials and methods'). (A) Left, population of cells stained with Lysotracker (green), uniformly green aggregates are formed. Center, population of cells stained with Chormeo Mitochondrial Staining (cyan), uniformly cyan aggregates are formed. Right, two independently stained populations of cells (green or cyan) are mixed and dual color aggregates are formed, indicating that cells from different origins aggregate to each other. (B) Total number of cells per day in control cells, aphidicolin-treated cells and hydroxyurea-treated cells. Note that cell division is blocked by both aphidicolin and hydroxyurea. Aggregate formation was evaluated under each condition. All cells, even those treated with aphidicolin or hydroxyurea, developed aggregates. A representative aggregate is shown for each condition. Scale bars= 10 µm.

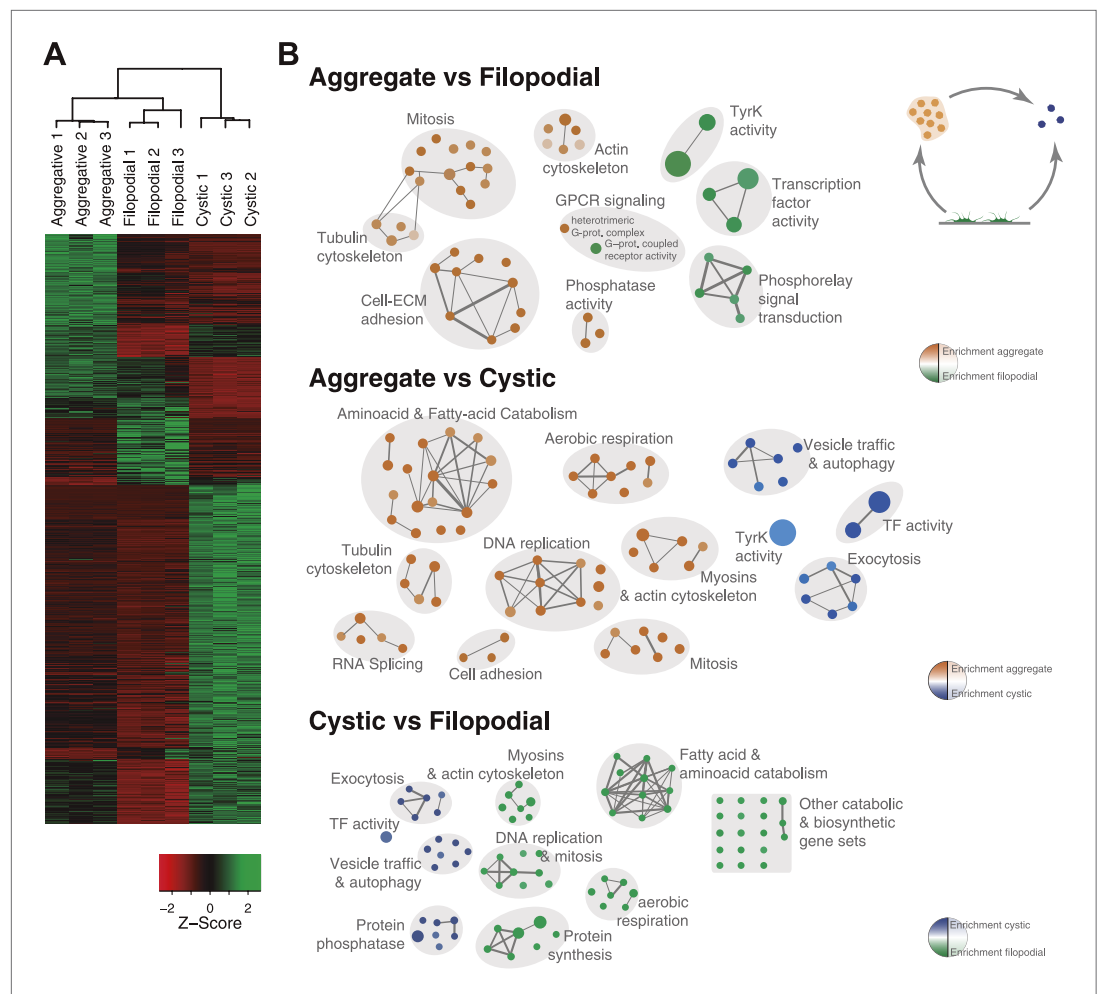
DOI: [10.7554/eLife.01287.009](https://doi.org/10.7554/eLife.01287.009)

The following figure supplements are available for figure 4:

**Figure supplement 1.** Proliferation rate per day of aggregative cells.

DOI: [10.7554/eLife.01287.010](https://doi.org/10.7554/eLife.01287.010)

retained introns, analyses of sequence motif enrichment with MEME show enrichment of a long T/G-rich motif that highly resembles a recently identified consensus binding site for Elav-like protein in mammals (Ince-Dunn et al., 2012) (Figure 10). Interestingly, the single ortholog for this gene in



**Figure 5.** Differential gene expression in *C. owczarzaki*. **(A)** Heatmap showing differential gene expression in the different biological replicates of each stage. Only genes with  $\text{cRPKM} \geq 5$  in at least one sample and with a 2-fold expression change in at least one pair-wise comparison are shown. **(B)** Gene set enrichment analysis (GSEA) for the different cell stages ('Materials and methods'). Orange represents enrichment in the aggregative stage, blue in the cystic stage, and green in the filopodial stage, with color intensity proportional to enrichment significance. The node size is proportional to the number of genes associated to the GO category, and the width of the edges is proportional to the number of genes shared between GO categories. Groups of functionally related GOs are manually circled and assigned a label.

DOI: [10.7554/eLife.01287.011](https://doi.org/10.7554/eLife.01287.011)

*C. owczarzaki* shows a highly-regulated expression pattern, with lowest expression in the filopodial stage (**Figure 10C**). Therefore, it is tempting to speculate that Elav-like protein may negatively regulate filopodial-specific intron retention of some introns. Experimental depletion of Elav-like protein in *C. owczarzaki* will require the development of RNAi or gene-targeting methods in this species before this hypothesis can be tested.

Next, we investigated differential exon splicing and identified 191 cassette exons with PSIs <95% in at least one life cycle stage, 39 of which display PSIs <85%. 29 of these exons showed a  $\geq 15\%$  PSI difference in pairwise comparisons between the cell stages, with lower PSIs typically associated with the filopodial stage (**Supplementary file 1**); RT-PCR assays confirmed skipping for 7 out of 8 tested cases (**Figure 11A**, and **Figure 11—figure supplement 1**). Most (~60%) of these exons maintain an open reading frame when skipped. In contrast to previous reports demonstrating that differentially-regulated exons are significantly under represented in modular, folded domains in metazoans (**Romero et al., 2006; Ellis et al., 2012**), two thirds of the differentially regulated exons in *C. owczarzaki* overlap annotated domains (**Supplementary file 1**). Furthermore, genes with differential exon

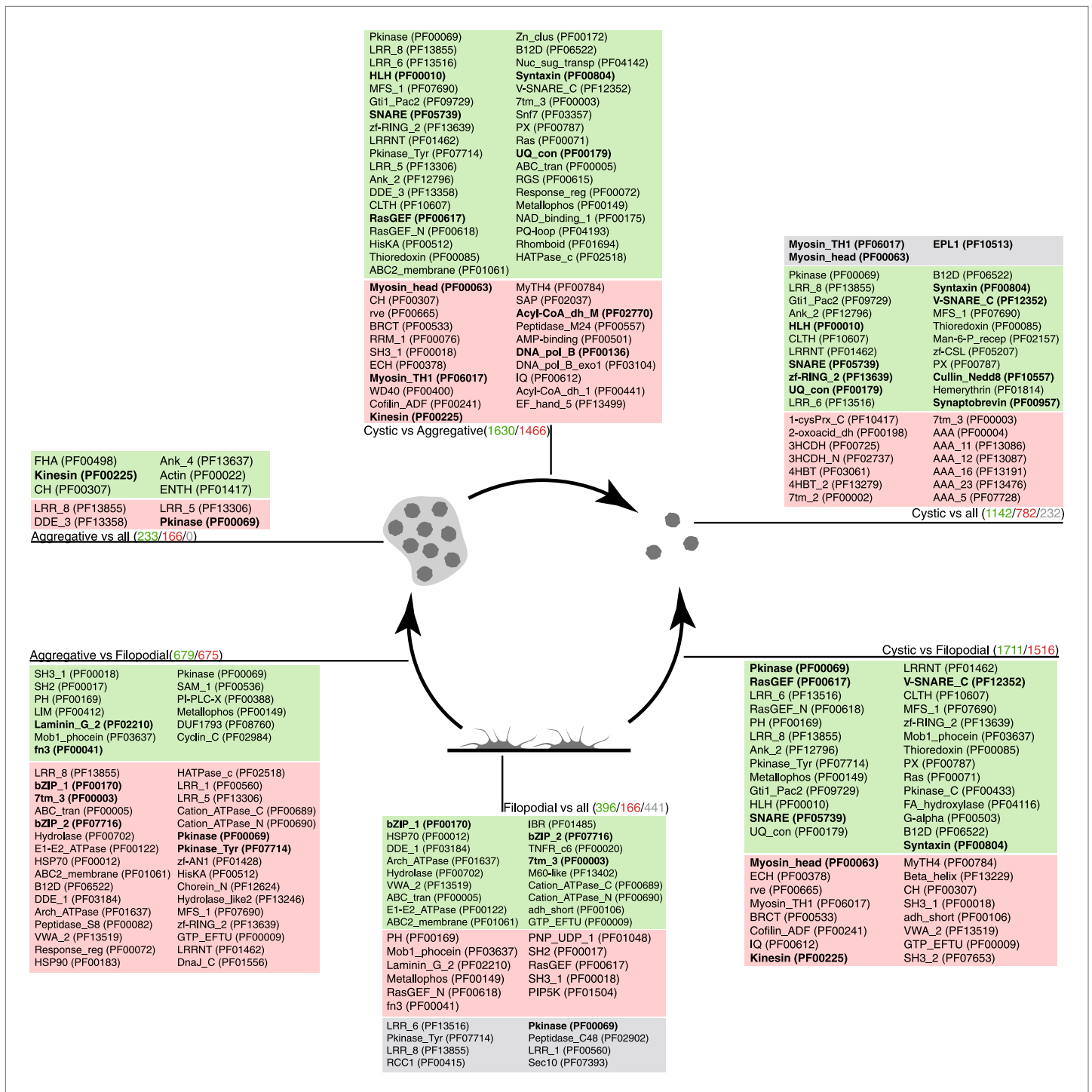


**Figure 6.** GO enrichment in sets of differentially expressed genes. Pairwise (Aggregative vs Filopodial, Cystic vs Aggregative and Cystic vs Filopodial) and one-versus-all comparisons are indicated. The significantly overrepresented GO categories ('Materials and methods') are shown for sets of overexpressed (green) and downregulated (red) genes and for genes with differential intron retention (gray). The number of genes included in each set is indicated with the same color code.

DOI: 10.7554/eLife.01287.012

skipping are statistically significantly enriched in protein kinase activity, impacting both tyrosine and serine/threonine kinases (**Figure 11B**). This observation strongly suggests a role for coordinated exon skipping in the modulation of cell signaling in *C. owczarzaki*. To our knowledge, this represents the first example of a regulated exon network linked to a specific biological function in a unicellular organism.

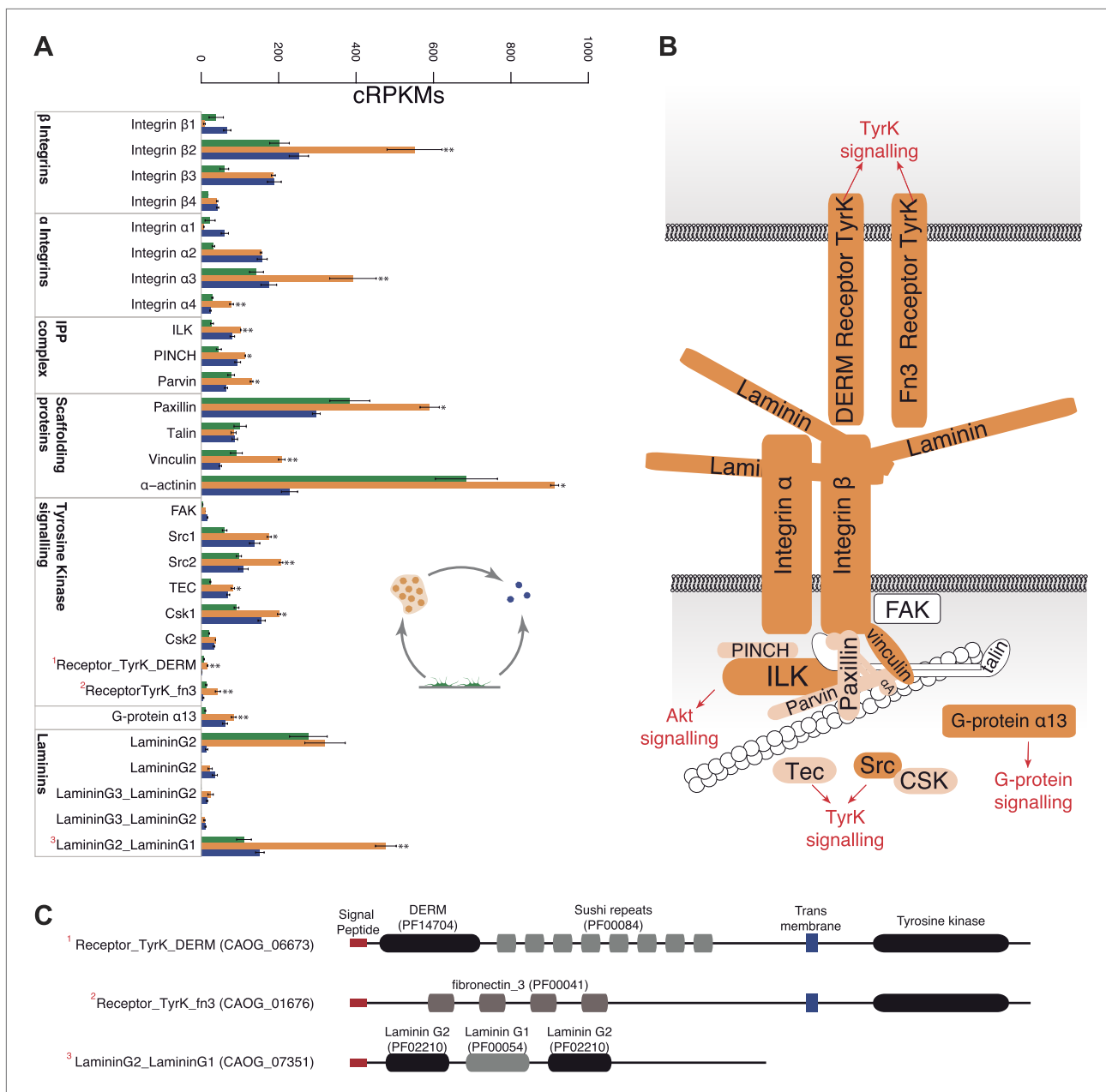
In summary, our results offer new insight into the origin of metazoan multicellularity. In particular, the observation of an aggregative multicellular stage in *C. owczarzaki* represents the first example of such cellular behavior in a close unicellular relative of metazoans. This observation therefore adds to the repertoire of reported complex cellular behaviors among extant unicellular relatives of metazoans—including clonal colony formation in choanoflagellates and sporangia formation by hypertrophic syncytial growth in ichthyosporeans (**Dayel et al., 2011; Suga and Ruiz-Trillo, 2013**), thus expanding



**Figure 7.** Pfam protein domain enrichment in sets of differentially expressed genes. Pairwise (Aggregative vs Filopodial, Cystic vs Aggregative and Cystic vs Filopodial) and one-versus-all comparisons are indicated. Significantly overrepresented Pfam domains ('Materials and methods') are shown for sets of overexpressed (green) and downregulated (red) genes and for genes with differential intron retention (gray). The number of genes included in each set is indicated with the same color code. Those Pfam domains mentioned in the text are shown in bold.

DOI: 10.7554/eLife.01287.013

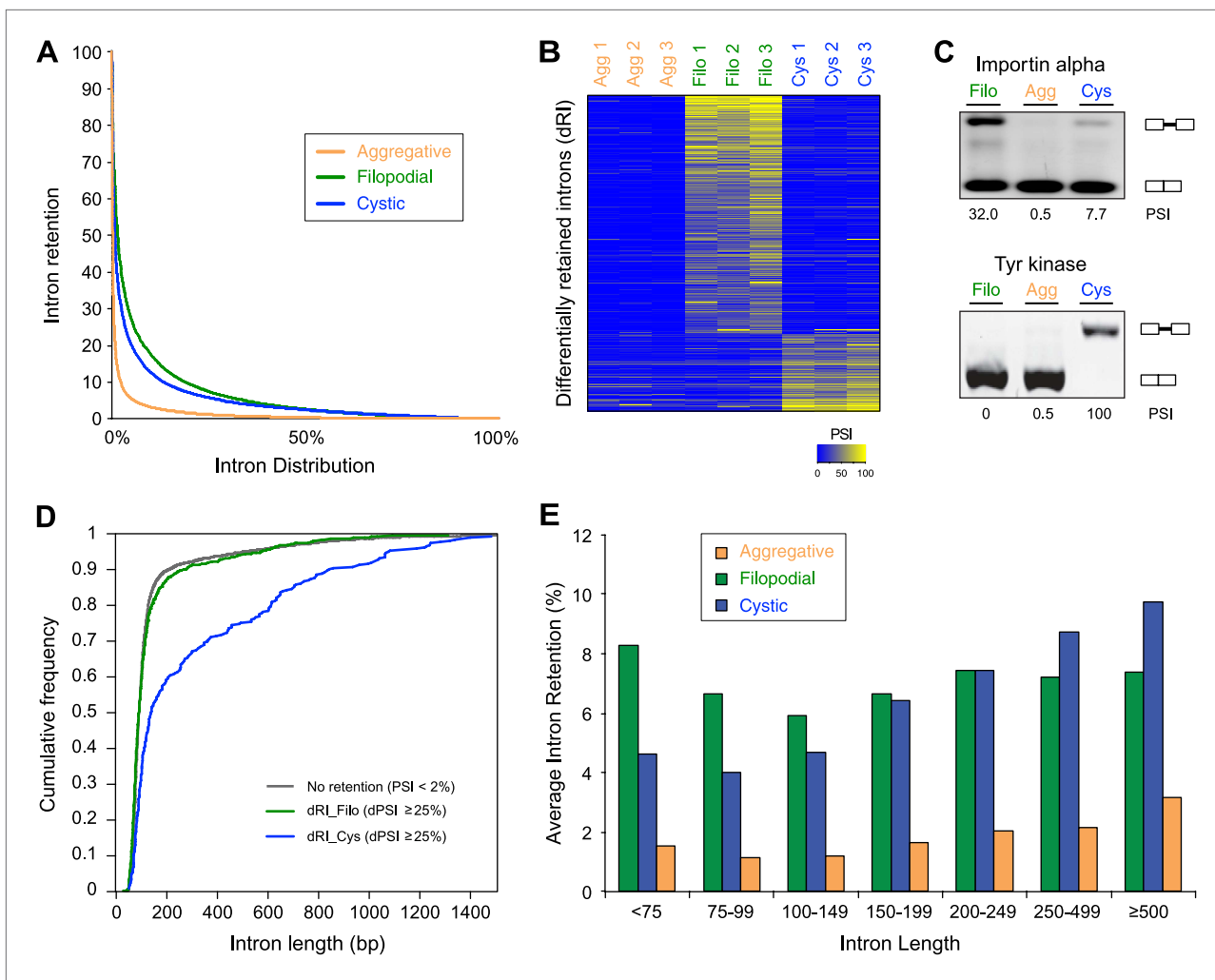
the potential starting 'raw material' available for the evolution of animal multicellularity. We note that the current evolutionary framework on the opisthokonts, based on phylogenomic analyses (Steenkamp et al., 2006; Ruiz-Trillo et al., 2008; Shalchian-Tabrizi et al., 2008; Torruella et al., 2012; Paps et al., 2013), discards the possibility that *C. owczarzaki* (or choanoflagellates) derives from a more complex



**Figure 8.** Expression of cell–ECM adhesion genes. **(A)** Barplot of the expression values of each gene in the different stages, showing overexpression of most components in the aggregate stage (orange). Asterisks indicate that the gene is significantly differentially expressed in both (two asterisks) or only one (one asterisk) pair-wise comparison (agg. vs fil. and agg. vs cyst.). Bars show standard error. **(B)** Schematic representation of the putative *C. owczarzaki* integrin adhesome and putative associated signalling proteins, colored according to overexpression in aggregates as shown in the barplot (dark orange, two asterisks; light orange, one asterisk; and white, no differences in expression). **(C)** Specific protein domain architectures for the fibronectin and DERM receptor tyrosine kinases (CAOG\_01676 and CAOG\_06673) and for the laminin protein (CAOG\_07351). DOI: 10.7554/eLife.01287.014

multicellular lineage. The sister-group of opisthokonts is the unicellular biflagellates Apusozoa, and complex multicellularity has not yet been observed in any of the non-metazoan holozoan lineages.

Furthermore, we show that the complex, metazoan-like genetic ‘toolkit’ of *C. owczarzaki* (Sebé-Pedrós et al., 2010, 2011; Suga et al., 2012) is dynamically deployed during its highly-regulated life cycle, with upregulation of integrin adhesome and signalling genes linked to multicellularity in



**Figure 9.** Regulated alternative splicing in *C. owczarzaki*. **(A)** Plot of percentage of intron inclusion by intron in rank order for the three studied cellular stages. Filopodial (green) and cystic (blue) stages show higher intron retention levels than the aggregative stage (orange) ( $p < 2.2 \times 10^{-16}$ , Wilcoxon Rank Sum test). **(B)** Heatmap of PSIs of filopodial- and cystic-specific differentially retained introns across three replicates for each cellular stage. **(C)** Examples of stage-specific intron retention. **(D)** Intron length distributions for differentially retained introns in cystic (blue), filopodial (green), and weakly retained introns (gray). **(E)** Relationship between intron length and retention. Percentage of average intron retention in each of the three cellular stages for different bins of intron size. In the cystic stage, the percentage of intron retention increased with intron length.

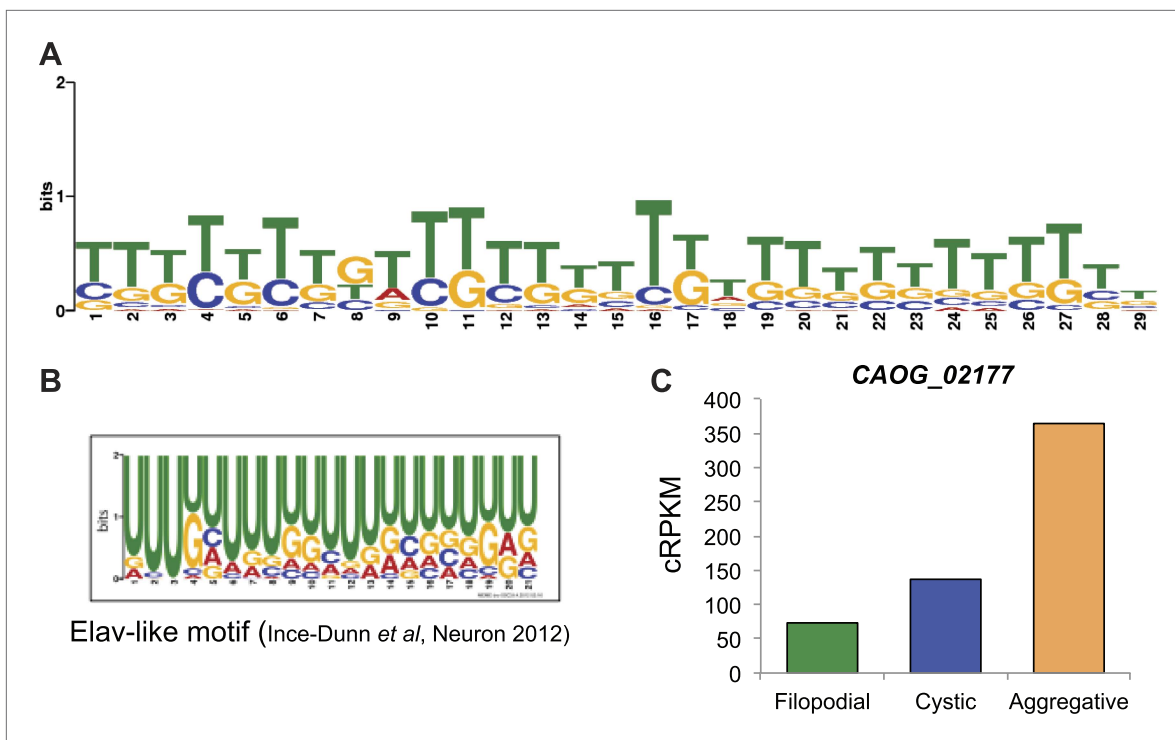
DOI: [10.7554/eLife.01287.015](https://doi.org/10.7554/eLife.01287.015)

The following figure supplements are available for figure 9:

**Figure supplement 1.** Intron retention validation (see 'Materials and methods').

DOI: [10.7554/eLife.01287.016](https://doi.org/10.7554/eLife.01287.016)

metazoans during the aggregative stage. Extensive differential AS between the *C. owczarzaki* life cycle stages likely further contributes to the dynamic gene regulation observed in this species, with differential intron retention likely acting as an important mechanism in the control of transcript levels between life cycle stages, probably through triggering non-sense mediated decay (NMD). Our discovery of an exon network associated with tyrosine kinase genes in *C. owczarzaki* further adds to the metazoan-like features of this species. Together with genes resembling those that function in metazoan multicellular processes, the emergence of an exon network that functions in conjunction with differentially-regulated intron retention may have provided a degree of proteomic and regulatory complexity that was key in the evolution of cell type complexity in metazoans (Nilsen and Graveley, 2010). Based on the collective results from our investigation of *C. owczarzaki*, it is intriguing to consider that the integration of regulatory innovations involving differential expression



**Figure 10.** Possible role for an Elav-like ortholog in the negative regulation of filopodial stage-specific dRIs. **(A)** Most significantly enriched motif in filopodial stage-specific dRIs, obtained by MEME. **(B)** Consensus motif obtained by CLIP-Seq data for an Elav-like member in mammals by *Ince-Dunn et al. (2012)* and that closely resembles the motif in **(A)**. T~U. **(C)** Expression (measured as cRPKMs) of CAOG\_02177, a Elav-like ortholog from *C. owczarzaki* that shows lower expression in filopodial stage.

DOI: [10.7554/eLife.01287.017](https://doi.org/10.7554/eLife.01287.017)

and splicing of metazoan-like genes set the stage for the evolution of cell specialization in the common ancestors of metazoans and *C. owczarzaki*.

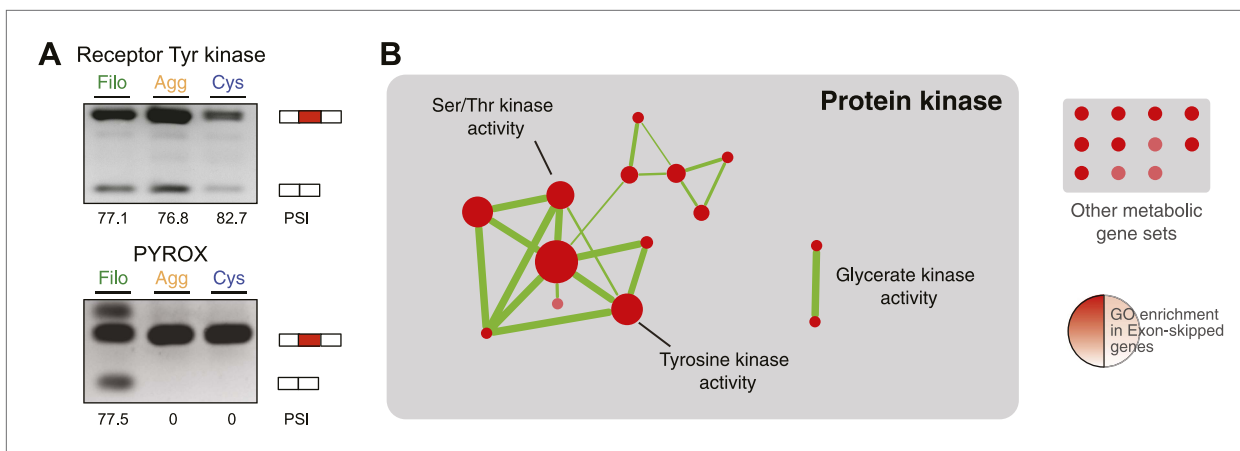
## Materials and methods

### Scanning electron microscopy

*C. owczarzaki* cells of the corresponding stage were fixed for 1 hr with 2.5% glutaraldehyde (Sigma-Aldrich, St. Louis, MO, USA), and for another hour with 1% osmium tetroxide (Sigma-Aldrich), followed by dehydration in a graded ethanol series (25%, 50%, 70%, 99%) for 15 min per step, followed by three 15-min rinses in 100% ethanol. Samples were critical-point dried in liquid CO<sub>2</sub> using a BAL-TEC CPD 030 critical-point drying apparatus. They were subsequently glued to SEM stubs with colloidal silver, sputter-coated with gold-palladium, and examined with a Hitachi S-3500N (Hitachi High-Technologies Europe GmbH, Krefeld, Germany).

### Transmission electron microscopy

Cell aggregates were loaded into the copper tubes and immediately cryoimmobilized using a Self-Pressurized Freezing System (EM SPF) (Leica-Microsystems, Vienna, Austria). The cells were then stored in liquid nitrogen until further use. Peeled copper tubes were freeze-substituted in anhydrous acetone containing 2% osmium tetroxide and 0.1% uranyl acetate at –90°C for 72 hr and warmed to room temperature, following a 2°C increase per hour in five consecutive steps (–60°C, –30°C, 0°C, 4°C, and room temperature) with a total of 8 hr at each temperature and using an EM AFS (Leica-Microsystems, Vienna). After several acetone rinses, samples were infiltrated with Epon resin during 7 days and embedded in resin and polymerised at 60°C during 48 hr. Ultrathin sections were obtained using a Leica Ultracut UC6 ultramicrotome (Leica-Microsystems) and mounting on Formvar-coated copper grids. The sections were stained with 2% uranyl acetate in water and lead citrate, and were observed in a Tecnai Spirit 120 kv electron microscope (FEI Company, Eindhoven, Netherlands) equipped with a Megaview III CCD camera.



**Figure 11.** Regulated exon-skipping in *C. owczarzaki*. **(A)** Examples of exon skipping. **(B)** Gene set enrichment analysis (GSEA) of the genes containing cassette exons that are differentially-regulated among cellular stages showing high enrichment for protein kinase-associated activities.

DOI: [10.7554/eLife.01287.018](https://doi.org/10.7554/eLife.01287.018)

The following figure supplements are available for figure 11:

**Figure supplement 1.** Exon Skipping validation by RT-PCR ('Materials and methods').

DOI: [10.7554/eLife.01287.019](https://doi.org/10.7554/eLife.01287.019)

## Cell culture conditions

*C. owczarzaki* cells were grown axenically in 5-ml flasks with ATCC medium 1034 (modified PYNFH medium) in a 23°C incubator. Three biological replicates (three independent cell lines) were generated by subculturing from a single-founding cell and grown for 2 months. Adherent filopodiated cells were obtained by initiating a new 1/100 sub-culture (from an approximately  $5 \times 10^6$  cells/ml initial culture) and, after 3–4 days, cells were scratched from the substrate. Aggregate formation was induced by initiating a new 1/250 sub-culture (from an approximately  $5 \times 10^6$  cells/ml initial culture) and by gentle agitation at 60 rpm during 4–5 days. Finally, floating cystic cells were obtained from a 14-day-old culture, starting from a new 1/100 sub-culture (from an approximately  $5 \times 10^6$  cells/ml initial culture).

## Aggregation experiments

Two different groups of cells (from two different starting cultures, 5 ml flasks with a cell density of  $10^6$  cells/ml, consisting exclusively of adherent filopodial cells) were stained either with 75 nM (in PBS1x) LysoTracker Green DND-26 (Life Technologies, Carlsbad, CA, USA) or with 1  $\mu$ M (in PBS1x) Chromeo Live Cell Mitochondrial Staining Kit (Active Motif Inc, Carlsbad, CA, USA). The cells were stained for 30 min at 23°C. After staining, 1/3 of the cells from the two differentially stained populations were mixed in a new culture flask and the remaining 2/3 of cells for each staining were kept as a control. All three cultures were grown for 2 hr and then the aggregate formation was induced (see above, 'Results and Discussion'). After 8 hr, aggregates were visualized in poly-L-lysine covered (Sigma-Aldrich, St. Louis, MO) glass-bottom plates in a Leica TCS SP5 confocal microscope (Leica-Microsystems).

*C. owczarzaki* cell division was blocked using 100 mM hydroxyurea (Sigma-Aldrich) or 25  $\mu$ g/ml aphidicolin (Sigma-Aldrich). The effect of both drugs was evaluated by following cultures treated with each drug during 7 days, using Neubauer chamber. The cells were cultured in 16-multiwell plates and with an initial density of  $5 \times 10^4$  cells/ml. Once these conditions were established, different cultures were treated with each drug for 1 day and, then, aggregate formation was induced (see above). 2 days later, the formation of aggregates was visualized in poly-L-lysine covered (Sigma-Aldrich) glass-bottom plates in a Leica TCS SP5 confocal microscope (Leica-Microsystems).

## RNA-Seq and analysis

*C. owczarzaki* cells were grown in 5-ml flasks with ATCC medium 1034 (modified PYNFH medium) in a 23°C incubator. Total RNA from each cell stage (and from three biological replicates from each stage) was extracted using Trizol reagent (Life Technologies). Nine libraries were sequenced over 2 lanes HiSeq 2000 instrument (Illumina, San Diego, CA, USA), generating a total of 197M 76-base paired



reads. Reads were aligned to the reference genome using Tophat (Trapnell et al., 2012) with default options and specifying that this was a strand-specific sequencing, rendering an average mapping of 90%. Significant differential expression was calculated by performing pairwise comparisons with DESeq (Anders and Huber, 2010) (threshold 1e-05), EdgeR (Robinson et al., 2010) (threshold 1e-05), CuffDiff (Trapnell et al., 2012) (threshold 1e-05) and NOISeq (Tarazona et al., 2012) (threshold 0.8) and only genes that appear to be significant at least in three out of the four methods were considered as differentially expressed. Quality control analyses of the data were performed using cummeRbund R package (Trapnell et al., 2012). These include count vs dispersion plot to estimate over-dispersion, density plot to assess the distributions of FPKM scores across samples and squared coefficient of variation plot to check for cross-replicate variability.

A gene ontology of *C. owczarzaki*'s 8637 genes was generated using Blast2GO (Conesa et al., 2005) and GO enrichments of the different lists of differentially expressed genes (see above) were analyzed using Ontologizer (Bauer et al., 2008) using the Topology-Weighted method. A p-value threshold of 0.01 was used. The results from Ontologizer were loaded into Enrichment map cytoscape plug-in (Merico et al., 2010) to generate a network visualization. Pfam domains of all genes were analyzed using Pfamscan v26 with default Gathering Threshold, and counts were generated using custom Perl scripts. Fisher's exact tests were performed using custom R scripts and a p-value threshold of 0.01 was used.

### Alternative splicing analysis

Exon skipping and intron retention were analyzed as previously described (Curtis et al., 2012; Han et al., 2013). In short, for exon-skipping analyses, multifasta libraries of exon–exon junctions were built by combining all forward annotated splicing donors and acceptors. A minimum of eight base pairs was required at each boundary to assure specificity. Next, the number of effective mappable positions was calculated for each exon–exon junction, as previously described (Labbé et al., 2012; Barbosa-Morais et al., 2012). Then, RNA-seq reads (previously trimmed to 50 nucleotides and combining each three replicates to increase read depth) were aligned to these sequences using Bowtie, with  $-m\ 1\ -v\ 2$  parameters (single mapping and two or fewer mismatches). Percentage of exon inclusion was calculated and a minimal read coverage was required, as previously described (Khare et al., 2012). For intron retention, a similar approach was taken for each contiguous intron–exon and exon–exon junction, and percentage of intron inclusion (PSI, Percent Spliced In, the percentage of transcript for a given gene that contain the intron) was calculated as previously described (Curtis et al., 2012). For comparisons among cellular stages, only events with enough read coverage in the three samples were considered (either (i)  $\geq 10$  reads in the exon–exon junction or (ii)  $\geq 10$  reads in one intron–exon junction and  $\geq 5$  in the other), and introns showing  $>95\%$  inclusion in the three samples were discarded. To assess whether differentially retained introns in the same genes were included in a coordinated or in a combinatorial manner, mate information of read pairs was used. If each end of a read mapped to two different intron retention events, each end may be providing support for retention of both introns or splicing of both introns (coordinated regulation), or retention of one and splicing of another (combinatorial intron retention). For the 555 pairs of retained introns that had read mate information, 196 (35.3%) showed evidence for combinatorial regulation. Finally, for sequence motif enrichment analyses, full intron sequences were compared using MEME (Bailey et al., 2009).

### RT-PCR

To validate AS analysis predictions, the three stages were induced (RNA-Seq and analysis sections) and RNA was extracted using Trizol reagent (Life Technologies). To eliminate genomic DNA, total RNA was treated with DNase I (Roche, Basel, Switzerland) and purified using RNeasy columns (Qiagen, Venlo, Netherlands). For each stage, cDNA was produced from 1  $\mu\text{g}$  of total RNA using SuperScript III reverse transcriptase (Life Technologies). Pairs of primers of similar melting temperature (60°C) and spanning the putative alternatively spliced segments were designed using Geneious software. PCR was performed using ExpandTaq polymerase (Roche).

### Flow cytometry

*C. owczarzaki* cells were grown for 10 to 15 days, sampling every day from both the supernatant (to obtain floating cells, which after day 7 are completely encysted) and the scratched flask (to obtain filopodial adherent cells). Thus, two samples were obtained daily, for floating and adherent cells. For DNA-content analysis, a sample was fixed using 70% ethanol and stored at  $-20^{\circ}\text{C}$  for one month. The

samples were subsequently fixed and stained with Propidium Iodide (as described in **Darzynkiewicz and Huang, 2004**) and DNA content estimated using FACScalibur flow cytometer (Becton Dickinson, Franklin Lakes, NJ, USA). For cell counting, 1 ml of fresh sample (one from the supernatant and one from the flask surface) was mixed in a BD Trucount Tube (Becton Dickinson), with a known number of beads, so absolute cell number counts could be calculated, using an LSR Fortessa flow cytometer (Becton Dickinson). Two replicate experiments (R1 and R2) were performed independently in order to confidently establish growth dynamics. Two measures were calculated from the DNA-content analysis. First, the proliferation rate, which indicates the proportion of number of cells in S and G2/M phases vs the number of cells in G0/G1. Second, the percentage of cells in S-phase.

## Acknowledgements

We thank Joshua Levin and Lin Fan for generating RNA-Seq libraries and the Broad Institute Genomics Platform for Illumina sequencing. The authors are grateful to Carmen López Iglesias of the Cryo-Electron Microscopy unit (CCiT-UB) for her help and advice on electron microscopy. We also thank Cristina Peligero, for her help in performing and interpreting the flow cytometry results, Ignacio Maeso for critical reading of the manuscript, and Professor Yves van de Peer, Stephane Rombauts and Brian Haas for their advice. ASP is supported by a pregraduate Formacion Profesorado Universitario grant from MICINN. MI is supported by a postdoctoral fellowship from the Human Frontiers Science Program Organization. BJB acknowledges grant funding from the Canadian Institutes of Health Research.

---

## Additional information

### Competing interests

BJB: Reviewing editor, *eLife*. The other authors declare that no competing interests exist.

### Funding

Funder	Grant reference number	Author
European Commission	ERC-2007-StG- 206883	Iñaki Ruiz-Trillo
Ministerio de Economía y Competitividad	BFU2011-23434	Iñaki Ruiz-Trillo
Canadian Institutes of Health Research	MOP-67011	Benjamin J Blencowe

The funders had no role in study design, data collection and interpretation, or the decision to submit the work for publication.

---

### Author contributions

AS-P, MI, Conception and design, Acquisition of data, Analysis and interpretation of data, Drafting or revising the article; JC, HP-A, Acquisition of data, Analysis and interpretation of data, Drafting or revising the article; CR, Acquisition of data, Drafting or revising the article; CN, Conception and design, Acquisition of data, Drafting or revising the article; BJB, IR-T, Conception and design, Analysis and interpretation of data, Drafting or revising the article

---

## Additional files

### Supplementary files

- Supplementary file 1. Exon-skipping events in *C. owczarzaki*. For each event, the table shows the affected gene, a reference of the skipping event, the coordinates of the skipped exon, its length and if it is multiple of three nucleotides, the full coordinates of the event (i.e., including the donor of the upstream constitutive exons and acceptor of the downstream constitutive exon), percentage of inclusion of the exon among all transcripts present at each stage, general protein domain structure and a brief description of the gene, and information on which region of the protein is affected by the exon-skipping event (in orange if impacting a structured domain). Protein kinases are highlighted in dark green and tyrosine kinases in light green.

DOI: [10.7554/eLife.01287.020](https://doi.org/10.7554/eLife.01287.020)

## Major dataset

The following dataset was generated:

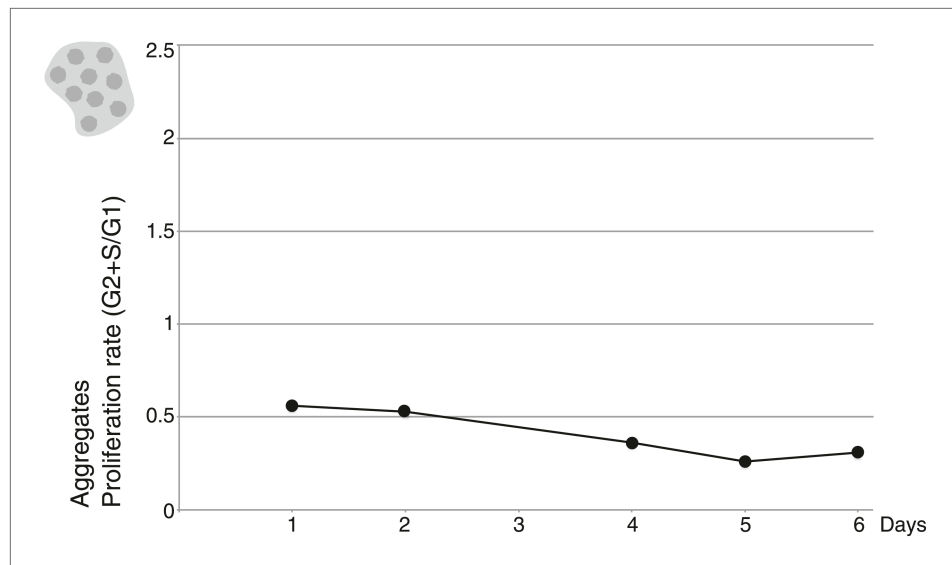
Author(s)	Year	Dataset title	Dataset ID and/or URL	Database, license, and accessibility information
Russ C, Nusbaum C, Sebé-Pedrós A	2013	RNAseq data from <i>Capsaspora owczarzaki</i>	<a href="http://www.ncbi.nlm.nih.gov/biosample/?term=txid595528">http://www.ncbi.nlm.nih.gov/biosample/?term=txid595528</a> [Organism: noexp]	Publicly available at NCBI ( <a href="http://www.ncbi.nlm.nih.gov/">http://www.ncbi.nlm.nih.gov/</a> )

## References

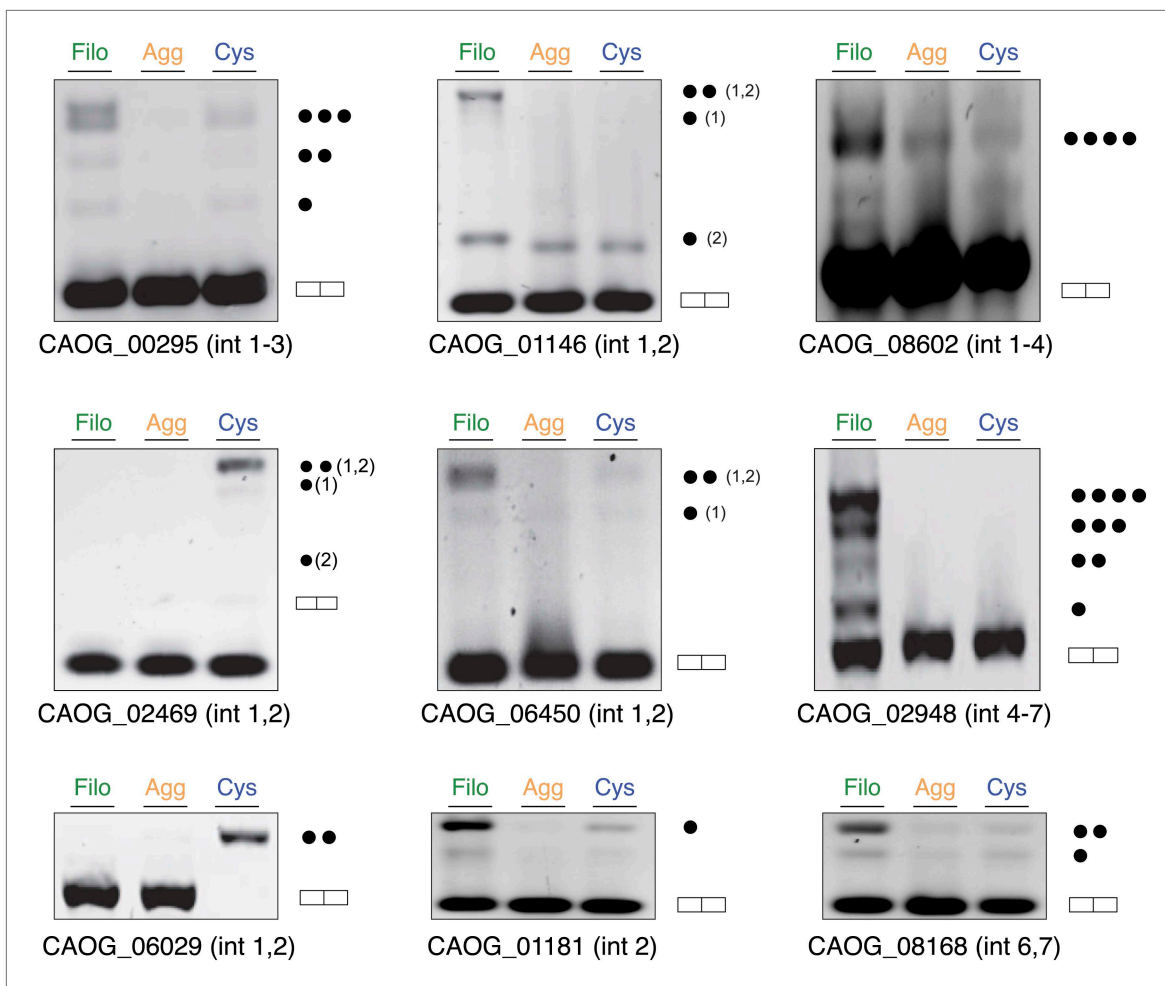
- Abedin M**, King N. 2008. The premetazoan ancestry of cadherins. *Science* **319**:946–948. doi: [10.1126/science.1151084](https://doi.org/10.1126/science.1151084).
- Anders S**, Huber W. 2010. Differential expression analysis for sequence count data. *Genome Biology* **11**:R106. doi: [10.1186/gb-2010-11-10-r106](https://doi.org/10.1186/gb-2010-11-10-r106).
- Arendt D**. 2008. The evolution of cell types in animals: emerging principles from molecular studies. *Nature Reviews Genetics* **9**:868–882. doi: [10.1038/nrg2416](https://doi.org/10.1038/nrg2416).
- Arendt D**, Hausen H, Purschke G. 2009. The “division of labour” model of eye evolution. *Philosophical Transactions of the Royal Society of London B Biological Sciences* **364**:2809. doi: [10.1098/rstb.2009.0104](https://doi.org/10.1098/rstb.2009.0104).
- Bailey TL**, Boden M, Buske FA, Frith M, Grant CE, Clementi L, Ren J, Li WW, Noble WS. 2009. MEME Suite: tools for motif discovery and searching. *Nucleic Acids Research* **37**:W202–W208. doi: [10.1093/nar/gkp335](https://doi.org/10.1093/nar/gkp335).
- Barbosa-Morais NL**, Irimia M, Pan Q, Xiong HY, Gueroussov S, Lee LJ, Slobodenic V, Kutter C, Watt S, Colak R, Kim T, Misquitta-Ali CM, Wilson MD, Kim PM, Odom DT, Frey BJ, Blencowe BJ. 2012. The evolutionary landscape of alternative splicing in vertebrate species. *Science* **338**:1587–1593. doi: [10.1126/science.1230612](https://doi.org/10.1126/science.1230612).
- Bauer S**, Grossmann S, Vingron M, Robinson PN. 2008. Ontologizer 2.0—a multifunctional tool for GO term enrichment analysis and data exploration. *Bioinformatics* **24**:1650–1651. doi: [10.1093/bioinformatics/btn250](https://doi.org/10.1093/bioinformatics/btn250).
- Brown MW**, Silberman JD, Spiegel FW. 2011. A contemporary evaluation of the acrasids (Acrasidae, Heterolobosea, Excavata). *European Journal of Protistology* **48**:103–123. doi: [10.1016/j.ejop.2011.10.001](https://doi.org/10.1016/j.ejop.2011.10.001).
- Brown MW**, Spiegel FW, Silberman JD. 2009. Phylogeny of the “forgotten” cellular slime mold, *Fonticula alba*, reveals a key evolutionary branch within Opisthokonta. *Molecular Biology and Evolution* **26**:2699–2709. doi: [10.1093/molbev/msp185](https://doi.org/10.1093/molbev/msp185).
- Brown MW**, Kolisko M, Silberman JD, Roger AJ. 2012. Aggregative multicellularity evolved independently in the eukaryotic supergroup Rhizaria. *Current Biology* **22**:1–5. doi: [10.1016/j.cub.2012.04.021](https://doi.org/10.1016/j.cub.2012.04.021).
- Burt DB**. 2001. Evolutionary stasis, constraint and other terminology describing evolutionary patterns. *Biol J Linn Soc Lond* **72**:509–517. doi: [10.1111/j.1095-8312.2001.tb01334.x](https://doi.org/10.1111/j.1095-8312.2001.tb01334.x).
- Cavalier-Smith T**. 2003. Phylogeny of choanozoa, apusozoa, and other protozoa and early eukaryote megaevolution. *Journal of Molecular Evolution* **56**:540–563. doi: [10.1007/s00239-002-2424-z](https://doi.org/10.1007/s00239-002-2424-z).
- Conesa A**, Götz S, García-Gómez JM, Terol J, Talón M, Robles M. 2005. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* **21**:3674–3676. doi: [10.1093/bioinformatics/bti610](https://doi.org/10.1093/bioinformatics/bti610).
- Curtis BA**, Tanifuji G, Burki F, Gruber A, Irimia M, Maruyama S, Arias MC, Ball SG, Gile GH, Hirakawa Y, Hopkins JF, Kuo A, Rensing SA, Schmutz J, Symeonidi A, Elias M, Eveleigh RJ, Herman EK, Klute MJ, Nakayama T, Obornik M, Reyes-Prieto A, Armbrust EV, Aves SJ, Beiko RG, Coutinho P, Dacks JB, Durnford DG, Fast NM, Green BR, Gridale CJ, Hempel F, Henrissat B, Höpner MP, Ishida K, Kim E, Kořený L, Kroth PG, Liu Y, Malik SB, Maier UG, McRose D, Mock T, Neilson JA, Onodera NT, Poole AM, Pritham EJ, Richards TA, Rocap G, Roy SW, Sarai C, Schaack S, Shirato S, Slamovits CH, Spencer DF, Suzuki S, Worden AZ, Zauner S, Barry K, Bell C, Bharti AK, Crow JA, Grimwood J, Kramer R, Lindquist E, Lucas S, Salamov A, McFadden GI, Lane CE, Keeling PJ, Gray MW, Grigoriev IV, Archibald JM. 2012. Algal genomes reveal evolutionary mosaicism and the fate of nucleomorphs. *Nature* **492**:59–65. doi: [10.1038/nature11681](https://doi.org/10.1038/nature11681).
- Darzynkiewicz Z**, Huang X. 2004. Analysis of cellular DNA content by flow cytometry. *Current Protocols in Immunology* **60**:5.7.1–5.7.18. doi: [10.1002/0471142735.im0507s60](https://doi.org/10.1002/0471142735.im0507s60).
- Dayel MJ**, Alegado RA, Fairclough SR, Levin TC, Nichols SA, McDonald K, King N. 2011. Cell differentiation and morphogenesis in the colony-forming choanoflagellate *Salpingoeca rosetta*. *Developmental Biology* **357**:73–82. doi: [10.1016/j.ydbio.2011.06.003](https://doi.org/10.1016/j.ydbio.2011.06.003).
- Dykstra MJ**, Olive LS. 1975. Sorodiplophrys: an unusual Sorocarp-producing protist. *Mycologia* **67**:873–879. doi: [10.2307/3758346](https://doi.org/10.2307/3758346).
- Ellis JD**, Barrios-Rodiles M, Colak R, Irimia M, Kim T, Calarco JA, Wang X, Pan Q, O’Hanlon D, Kim PM, Wrana JL, Blencowe BJ. 2012. Tissue-specific alternative splicing remodels protein-protein interaction networks. *Molecular Cell* **46**:884–892. doi: [10.1016/j.molcel.2012.05.037](https://doi.org/10.1016/j.molcel.2012.05.037).
- Gong H**, Shen B, Flevaris P, Chow C, Lam S, Voyno-Yasenetskaya TA, Kozasa T, Du X. 2010. G protein subunit Galpha13 binds to integrin alpha11bBeta3 and mediates integrin “Outside-In” signaling. *Science* **327**:1–4. doi: [10.1126/science.1174779](https://doi.org/10.1126/science.1174779).
- Grosberg RK**, Strathmann RR. 2007. The evolution of multicellularity: a Minor major transition? *Annual Review of Ecology, Evolution, and Systematics* **38**:621–654. doi: [10.1146/annurev.ecolsys.36.102403.114735](https://doi.org/10.1146/annurev.ecolsys.36.102403.114735).

- Hamazaki Y**, Kojima H, Mano H, Nagata Y, Todokoro K, Abe T, Nagasawa T. 1998. Tec is involved in G protein-coupled receptor- and integrin-mediated signalings in human blood platelets. *Oncogene* **16**:2773–2779. doi: [10.1038/sj.onc.1201799](https://doi.org/10.1038/sj.onc.1201799).
- Han H**, Irimia M, Ross PJ, Sung H-K, Alipanahi B, David L, Golipour A, Gabut M, Michael IP, Nachman EN, Wang E, Trcka D, Thompson T, O’Hanlon D, Slobodeniuc V, Barbosa-Morais NL, Burge CB, Moffat J, Frey BJ, Nagy A, Ellis J, Wrana JL, Blencowe BJ. 2013. MBNL proteins repress ES-cell-specific alternative splicing and reprogramming. *Nature* **498**:241–245. doi: [10.1038/nature12270](https://doi.org/10.1038/nature12270).
- Ince-Dunn G**, Okano H, Jensen KB, Park WY, Zhong R, Ule J, Mele A, Fak JJ, Yang C, Zhang C, Yoo J, Herre M, Okano H, Noebels JL, Darnell RB. 2012. Neuronal Elav-like (Hu) proteins regulate RNA splicing and abundance to control glutamate levels and neuronal excitability. *Neuron* **75**:1067–1080. doi: [10.1016/j.neuron.2012.07.009](https://doi.org/10.1016/j.neuron.2012.07.009).
- Irimia M**, Blencowe BJ. 2012. Alternative splicing: decoding an expansive regulatory layer. *Current Opinion In Cell Biology* **24**:323–332. doi: [10.1016/j.ceb.2012.03.005](https://doi.org/10.1016/j.ceb.2012.03.005).
- Jacob F**. 1977. Evolution and tinkering. *Science* **196**:1161–1166. doi: [10.1126/science.860134](https://doi.org/10.1126/science.860134).
- Khare T**, Pai S, Koncevicius K, Pal M, Kriukiene E, Liutkeviciute Z, Irimia M, Jia P, Ptak C, Xia M, Tice R, Tochigi M, Moréra S, Nazarians A, Belsham D, Wong AH, Blencowe BJ, Wang SC, Kapranov P, Kustra R, Labrie V, Klimasauskas S, Petronis A. 2012. 5-hmC in the brain is abundant in synaptic genes and shows differences at the exon-intron boundary. *Nature structural & molecular biology* **19**:1037–1043. doi: [10.1038/nsmb.2372](https://doi.org/10.1038/nsmb.2372).
- Kiel JA**. 2010. Autophagy in unicellular eukaryotes. *Philosophical Transactions of the Royal Society of London B Biological Sciences* **365**:819–830. doi: [10.1098/rstb.2009.0237](https://doi.org/10.1098/rstb.2009.0237).
- King N**, Westbrook MJ, Young SL, Kuo A, Abedin M, Chapman J, Fairclough S, Hellsten U, Isogai Y, Letunic I, Marr M, Pincus D, Putnam N, Rokas A, Wright KJ, Zuzov R, Dirks W, Good M, Goodstein D, Lemons D, Li W, Lyons JB, Morris A, Nichols S, Richter DJ, Salamov A, Sequencing JG, Bork P, Lim WA, Manning G, Miller WT, McGinnis W, Shapiro H, Tjian R, Grigoriev IV, Rokhsar D. 2008. The genome of the choanoflagellate *Monosiga brevicollis* and the origin of metazoans. *Nature* **451**:783–788. doi: [10.1038/nature06617](https://doi.org/10.1038/nature06617).
- King N**. 2004. The unicellular ancestry of animal development. *Developmental Cell* **7**:313–325. doi: [10.1016/j.devcel.2004.08.010](https://doi.org/10.1016/j.devcel.2004.08.010).
- Knoll AH**. 2011. The multiple origins of complex multicellularity. *Annual Review of Earth and Planetary Science* **39**:217–239. doi: [10.1146/annurev.earth.031208.100209](https://doi.org/10.1146/annurev.earth.031208.100209).
- Labbé RM**, Irimia M, Currie KW, Lin A, Zhu SJ, Brown DD, Ross EJ, Voisin V, Bader GD, Blencowe BJ, Pearson BJ. 2012. A comparative transcriptomic analysis reveals conserved features of stem cell pluripotency in planarians and mammals. *Stem Cells* **30**:1734–1745. doi: [10.1002/stem.1144](https://doi.org/10.1002/stem.1144).
- Lasek-Nesselquist E**, Katz L. 2001. Phylogenetic position of *Sorogena stonianovitchae* and relationships within the class Colpodea (Ciliophora) based on SSU rDNA sequences. *The Journal of Eukaryotic Microbiology* **48**:604–607. doi: [10.1111/j.1550-7408.2001.tb00197.x](https://doi.org/10.1111/j.1550-7408.2001.tb00197.x).
- Lewandowska K**, Choi HU, Rosenberg LC, Sasse J, Neame PJ, Culp LA. 1991. Extracellular matrix adhesion-promoting activities of a dermatan sulfate proteoglycan-associated protein (22K) from bovine fetal skin. *Journal of Cell Science* **99**:657–668.
- Loeb L**. 1903. On the coagulation of the blood of some Arthropods and on the influence of pressure and traction on the protoplasm of the blood cells of Arthropods. *The Biological Bulletin* **4**:301–318. doi: [10.2307/1535851](https://doi.org/10.2307/1535851).
- Loeb L**. 1921. Amœboid movement, tissue formation and consistency of protoplasm. *Science* **53**:261–262. doi: [10.1126/science.53.1368.261](https://doi.org/10.1126/science.53.1368.261).
- McGuire AM**, Pearson MD, Neafsey DE, Galagan JE. 2008. Cross-kingdom patterns of alternative splicing and splice recognition. *Genome Biology* **9**:R50. doi: [10.1186/gb-2008-9-3-r50](https://doi.org/10.1186/gb-2008-9-3-r50).
- Merico D**, Isserlin R, Stueker O, Emili A, Bader GD. 2010. Enrichment map: a network-based method for gene-set enrichment visualization and interpretation. *PLOS ONE* **5**:e13984. doi: [10.1371/journal.pone.0013984](https://doi.org/10.1371/journal.pone.0013984).
- Michod RE**. 2007. Evolution of individuality during the transition from unicellular to multicellular life. *Proceedings of the National Academy of Sciences of the United States of America* **104**:8613–8618. doi: [10.1073/pnas.0701489104](https://doi.org/10.1073/pnas.0701489104).
- Mikhailov K**, Konstantinova A, Nikitin M. 2009. The origin of Metazoa: a transition from temporal to spatial cell differentiation. *Bioessays* **31**:758–768. doi: [10.1002/bies.200800214](https://doi.org/10.1002/bies.200800214).
- Newman SA**. 2012. Physico-genetic determinants in the evolution of development. *Science* **338**:217–219. doi: [10.1126/science.1222003](https://doi.org/10.1126/science.1222003).
- Nilsen TW**, Graveley BR. 2010. Expansion of the eukaryotic proteome by alternative splicing. *Nature* **463**:457–463. doi: [10.1038/nature08909](https://doi.org/10.1038/nature08909).
- O’Shea KS**. 1987. Differential deposition of basement membrane components during formation of the caudal neural tube in the mouse embryo. *Development* **99**:509–519.
- Owczarzak A**, Stibbs HH, Bayne CJ. 1980. The destruction of *Schistosoma mansoni* mother sporocysts in vitro by amoebae isolated from *Biomphalaria glabrata*: an ultrastructural study. *Journal of Invertebrate Pathology* **35**:26–33. doi: [10.1016/0022-2011\(80\)90079-8](https://doi.org/10.1016/0022-2011(80)90079-8).
- Paps J**, Medina-Chacón LA, Marshall W, Suga H, Ruiz-Trillo I. 2013. Molecular phylogeny of unikonts: new insights into the position of apusomonads and ancyromonads and the internal relationships of opisthokonts. *Protist* **164**:2–12. doi: [10.1016/j.protis.2012.09.002](https://doi.org/10.1016/j.protis.2012.09.002).
- Parfrey LW**, Lahr DJG. 2013. Multicellularity arose several times in the evolution of eukaryotes. *Bioessays* **33**:339–347. doi: [10.1002/bies.201200143](https://doi.org/10.1002/bies.201200143).
- Robinson MD**, McCarthy DJ, Smyth GK. 2010. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**:139–140. doi: [10.1093/bioinformatics/btp616](https://doi.org/10.1093/bioinformatics/btp616).

- Rokas A. 2008. The molecular origins of multicellular transitions. *Current Opinion In Genetics and Development* **18**:472–478. doi: [10.1016/j.gde.2008.09.004](https://doi.org/10.1016/j.gde.2008.09.004).
- Romero PR, Zaidi S, Fang YY, Uversky VN, Radivojac P, Oldfield CJ, Cortese MS, Sickmeier M, LeGall T, Obradovic Z, Dunker AK. 2006. Alternative splicing in concert with protein intrinsic disorder enables increased functional diversity in multicellular organisms. *Proceedings of the National Academy of Sciences of the United States of America* **103**:8390–8395. doi: [10.1073/pnas.0507916103](https://doi.org/10.1073/pnas.0507916103).
- Ruiz-Trillo I, Burger G, Holland PWH, King N, Lang BF, et al. 2007. The origins of multicellularity: a multi-taxon genome initiative. *Trends in Genetics* **23**:113–118. doi: [10.1016/j.tig.2007.01.005](https://doi.org/10.1016/j.tig.2007.01.005).
- Ruiz-Trillo I, Roger AJ, Burger G, Gray MW, Lang BF. 2008. A phylogenomic investigation into the origin of Metazoa. *Molecular Biology and Evolution* **25**:664–672. doi: [10.1093/molbev/msn006](https://doi.org/10.1093/molbev/msn006).
- Savage RM, Danilchik MV. 1993. Dynamics of germ plasm localization and its inhibition by ultraviolet irradiation in early cleavage *Xenopus* embryos. *Developmental Biology* **157**:371–382. doi: [10.1006/dbio.1993.1142](https://doi.org/10.1006/dbio.1993.1142).
- Schaap P. 2011. Evolutionary crossroads in developmental biology: *Dictyostelium discoideum*. *Development* **138**:387–396. doi: [10.1242/dev.048934](https://doi.org/10.1242/dev.048934).
- Sebé-Pedrós A, Burkhardt P, Sánchez-Pons N, Fairclough SR, Lang BF, King N, Ruiz-Trillo I. 2013. Insights into the origin of metazoan filopodia and microvilli. *Molecular Biology and Evolution* **30**:2013–2023. doi: [10.1093/molbev/mst110](https://doi.org/10.1093/molbev/mst110).
- Sebé-Pedrós A, de Mendoza A, Lang BF, Degnan BM, Ruiz-Trillo I. 2011. Unexpected repertoire of metazoan transcription factors in the unicellular holozoan *Capsaspora owczarzaki*. *Molecular Biology and Evolution* **28**:1241–1254. doi: [10.1093/molbev/msq309](https://doi.org/10.1093/molbev/msq309).
- Sebé-Pedrós A, Roger A, Lang FB, King N, Ruiz-Trillo I. 2010. Ancient origin of the integrin-mediated adhesion and signaling machinery. *Proceedings of the National Academy of Sciences of the United States of America* **107**:10142–10147. doi: [10.1073/pnas.1002257107](https://doi.org/10.1073/pnas.1002257107).
- Shalchian-Tabrizi K, Minge MA, Espelund M, Orr R, Ruden T, Jakobsen KS, Cavalier-Smith T, Aramayo R. 2008. Multigene phylogeny of choanozoa and the origin of animals. *PLOS ONE* **3**:e2098. doi: [10.1371/journal.pone.0002098](https://doi.org/10.1371/journal.pone.0002098).
- Steenkamp ET, Wright J, Baldauf SL. 2006. The protistan origins of animals and fungi. *Molecular Biology and Evolution* **23**:93–106. doi: [10.1093/molbev/msj011](https://doi.org/10.1093/molbev/msj011).
- Suga H, Chen Z, de Mendoza A, Sebé-Pedrós A, Brown MW, Kramer E, Carr M, Kerner P, Vervoort M, Sánchez-Pons N, Torruella G, Derelle R, Manning G, Lang BF, Russ C, Haas BJ, Roger AJ, Nusbaum C, Ruiz-Trillo I. 2013. The *Capsaspora* genome reveals a complex unicellular prehistory of animals. *Nature Communications* **4**:1–9. doi: [10.1038/ncomms3325](https://doi.org/10.1038/ncomms3325).
- Suga H, Dacre M, de Mendoza A, Shalchian-Tabrizi K, Manning G, Ruiz-Trillo I. 2012. Genomic survey of premetazoans shows deep conservation of cytoplasmic tyrosine kinases and multiple radiations of receptor tyrosine kinases. *Science Signaling* **5**:ra35–ra35. doi: [10.1126/scisignal.2002733](https://doi.org/10.1126/scisignal.2002733).
- Suga H, Ruiz-Trillo I. 2013. Development of ichthyosporeans sheds light on the origin of metazoan multicellularity. *Developmental Biology* **377**:284–292. doi: [10.1016/j.ydbio.2013.01.009](https://doi.org/10.1016/j.ydbio.2013.01.009).
- Tarazona S, Furio-Tari P, Ferrer A, Conesa A. 2012. NOISeq: exploratory analysis and differential expression for RNA-seq data. *R package version 1.0.0*.
- Torruella G, Derelle R, Paps J, Lang BF, Roger AJ, Shalchian-Tabrizi K, Ruiz-Trillo I. 2012. Phylogenetic relationships within the Opisthokonta based on phylogenomic analyses of conserved single-copy protein domains. *Molecular Biology and Evolution* **29**:531–544. doi: [10.1093/molbev/msr185](https://doi.org/10.1093/molbev/msr185).
- Trapnell C, Hendrickson DG, Sauvageau M, Goff L, Rinn JL, Pachter L. 2012. Differential analysis of gene regulation at transcript resolution with RNA-seq. *Nature Biotechnology* **31**:46–53. doi: [10.1038/nbt.2450](https://doi.org/10.1038/nbt.2450).
- Trapnell C, Roberts A, Goff L, Pertea G, Kim D, Kelley DR, Pimentel H, Salzberg SL, Rinn JL, Pachter L. 2012. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nature Protocols* **7**:562–578. doi: [10.1038/nprot.2012.016](https://doi.org/10.1038/nprot.2012.016).
- Wilson HV. 1907. On some phenomena of coalescence and regeneration in sponges. *Journal of Experimental Zoology* **5**:245–258. doi: [10.1002/jez.1400050204](https://doi.org/10.1002/jez.1400050204).
- Yap K, Lim ZQ, Khandelia P, Friedman B, Makeyev EV. 2012. Coordinated regulation of neuronal mRNA steady-state levels through developmentally controlled intron retention. *Genes & Development* **26**:1209–1223. doi: [10.1101/gad.188037.112](https://doi.org/10.1101/gad.188037.112).

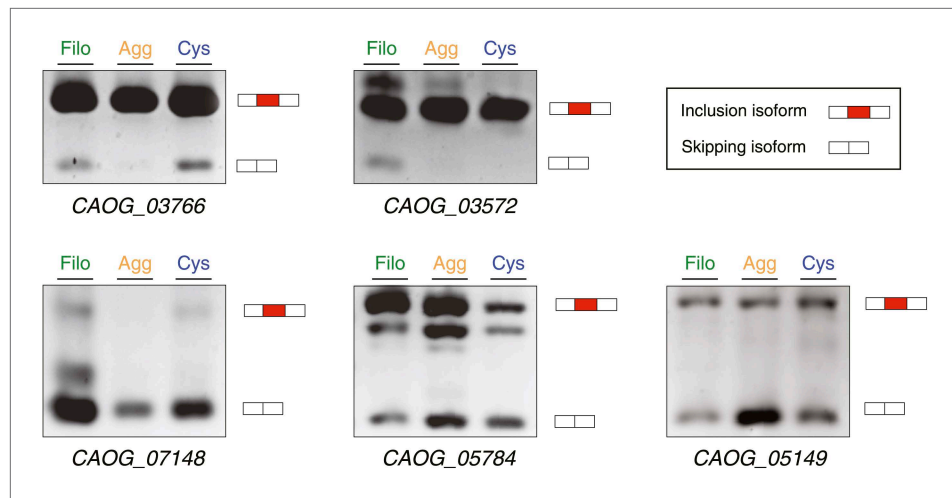


**Figure 4—figure supplement 1.** Proliferation rate per day of aggregative cells.  
DOI: [10.7554/eLife.01287.010](https://doi.org/10.7554/eLife.01287.010)



**Figure 9—figure supplement 1.** Intron retention validation (see 'Materials and methods'). Genes with confirmed intron retention events are indicated using gene IDs. Each dot corresponds to the inclusion of one intron (introns in *C. owczarzaki* usually have very similar lengths, and thus different combinations of the same size introns cannot be differentiated).

DOI: [10.7554/eLife.01287.016](https://doi.org/10.7554/eLife.01287.016)



**Figure 11—figure supplement 1.** Exon Skipping validation by RT-PCR ("Materials and methods"). Note that DNA heteroduplexes between the two isoforms may be formed and appear as extra bands. Bars in the key correspond to the exons amplified: the alternative exon in red, and the adjoining constitutive exons in white.

DOI: [10.7554/eLife.01287.019](https://doi.org/10.7554/eLife.01287.019)





## **R2: The Dynamic Regulatory Genome of *Capsaspora* and the Origin of Animal Multicellularity**

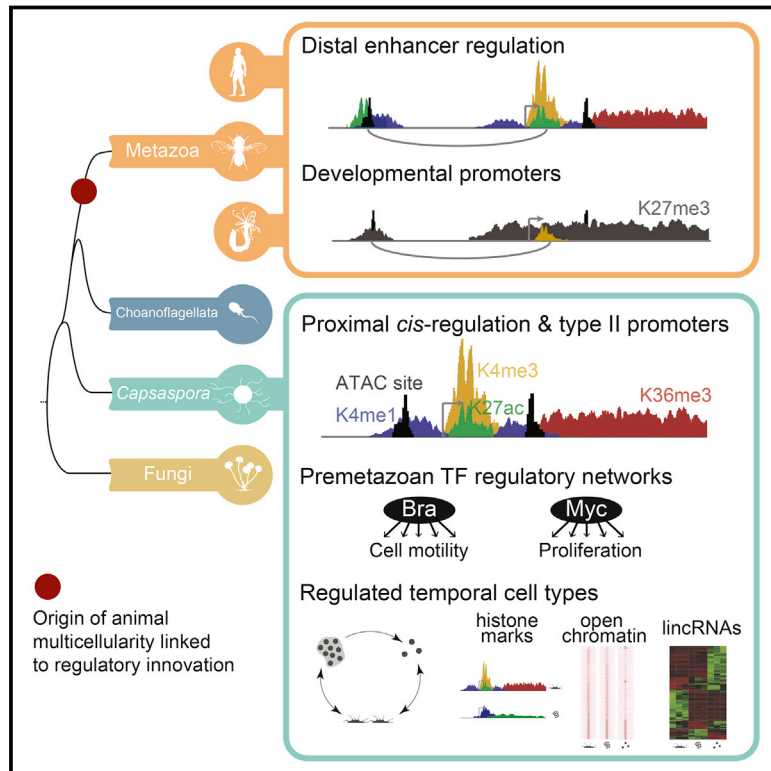
### *Abstract*

The unicellular ancestor of animals had a complex repertoire of genes linked to multicellular processes. This suggests that changes in the regulatory genome, rather than in gene innovation, were key to the origin of animals. Here, we carry out multiple functional genomic assays in *Capsaspora owczarzaki*, the unicellular relative of animals with the largest known gene repertoire for transcriptional regulation. We show that changing chromatin states, differential lincRNA expression, and dynamic cis-regulatory sites are associated with life cycle transitions in *Cap-saspora*. Moreover, we demonstrate conservation of animal developmental transcription-factor networks and extensive network interconnection in this preme-tazoan organism. In contrast, however, *Capsaspora* lacks animal promoter types, and its regulatory sites are small, proximal, and lack signatures of animal enhancers. Overall, our results indicate that the emergence of animal multicellularity was linked to a major shift in genome cis-regulatory complexity, most notably the appearance of distal enhancer regulation.



# The Dynamic Regulatory Genome of *Capsaspora* and the Origin of Animal Multicellularity

## Graphical Abstract



## Authors

Arnau Sebé-Pedrós, Cecilia Ballaré, Helena Parra-Acero, ..., José Luis Gómez-Skarmeta, Luciano Di Croce, Iñaki Ruiz-Trillo

## Correspondence

arnau.sebe-pedros@weizmann.ac.il (A.S.-P.), inaki.ruiz@ibe.upf-csic.es (I.R.-T.)

## In Brief

Analysis of the regulatory genome in one of our closest unicellular relatives suggests that the appearance of developmental promoters and distal enhancer elements, rather than of gene innovations, may have been the critical events underlying the origin of multicellular organisms.

## Highlights

- Dynamic chromatin states and *cis*-regulatory sites in a unicellular context
- Elaborate lincRNA regulation associated with a unicellular life cycle
- Premetazoan origin of core metazoan developmental transcription-factor networks
- Distal enhancer elements are a metazoan innovation

## Accession Numbers

PXD002342  
GSE71131



# The Dynamic Regulatory Genome of *Capsaspora* and the Origin of Animal Multicellularity

Arnau Sebé-Pedrós,<sup>1,7,\*</sup> Cecilia Ballaré,<sup>2,3</sup> Helena Parra-Acero,<sup>1</sup> Cristina Chiva,<sup>2,3</sup> Juan J. Tena,<sup>4</sup> Eduard Sabidó,<sup>2,3</sup> José Luis Gómez-Skarmeta,<sup>4</sup> Luciano Di Croce,<sup>2,3,5</sup> and Iñaki Ruiz-Trillo<sup>1,5,6,\*</sup>

<sup>1</sup>Institut de Biologia Evolutiva (CSIC-Universitat Pompeu Fabra), Passeig Marítim de la Barceloneta 37-49, 08003 Barcelona, Spain

<sup>2</sup>Center for Genomic Regulation, Doctor Aiguader 88, 08003 Barcelona, Spain

<sup>3</sup>Universitat Pompeu Fabra (UPF), Doctor Aiguader 88, 08003 Barcelona, Spain

<sup>4</sup>Centro Andaluz de Biología del Desarrollo (CABD), CSIC-Universidad Pablo de Olavide-Junta de Andalucía, Carretera de Utrera Km1, 41013 Sevilla, Spain

<sup>5</sup>Institució Catalana de Recerca i Estudis Avançats, Pg Lluís Companys 23, 08010 Barcelona, Spain

<sup>6</sup>Departament de Genètica, Universitat de Barcelona, 08028 Barcelona, Spain

<sup>7</sup>Present address: Department of Computer Science and Applied Mathematics, Weizmann Institute of Science, Rehovot 76100, Israel

\*Correspondence: [arnau.sebe-pedros@weizmann.ac.il](mailto:arnau.sebe-pedros@weizmann.ac.il) (A.S.-P.), [inaki.ruiz@ibe.upf-csic.es](mailto:inaki.ruiz@ibe.upf-csic.es) (I.R.-T.)

<http://dx.doi.org/10.1016/j.cell.2016.03.034>

## SUMMARY

The unicellular ancestor of animals had a complex repertoire of genes linked to multicellular processes. This suggests that changes in the regulatory genome, rather than in gene innovation, were key to the origin of animals. Here, we carry out multiple functional genomic assays in *Capsaspora owczarzaki*, the unicellular relative of animals with the largest known gene repertoire for transcriptional regulation. We show that changing chromatin states, differential lincRNA expression, and dynamic *cis*-regulatory sites are associated with life cycle transitions in *Capsaspora*. Moreover, we demonstrate conservation of animal developmental transcription-factor networks and extensive network interconnection in this premetazoan organism. In contrast, however, *Capsaspora* lacks animal promoter types, and its regulatory sites are small, proximal, and lack signatures of animal enhancers. Overall, our results indicate that the emergence of animal multicellularity was linked to a major shift in genome *cis*-regulatory complexity, most notably the appearance of distal enhancer regulation.

## INTRODUCTION

A defining feature of multicellular animals is their capacity to generate multiple specialized cell types through temporally and spatially regulated developmental programs. These programs of individual cell differentiation involve the generation of cell-specific transcriptional profiles. Recent genomic analyses, however, have shown that the unicellular ancestor of Metazoa already had a complex gene repertoire involved in multicellular functions, including specific differentiation programs (Fairclough et al.,

2013; King et al., 2008; de Mendoza et al., 2015; Sebé-Pedrós et al., 2013b; Srivastava et al., 2010; Suga et al., 2013).

Since the origin of animals was not solely dependent on the appearance of new genes, it is likely that animal evolution involved a shift in the genome regulatory capabilities required to generate cell-type-specific transcriptional profiles during animal development. In animals, these profiles are established and maintained by a complex combination of chromatin regulatory dynamics, distal *cis*-regulatory elements, and transcription factor networks (Bernstein et al., 2007; Buecker and Wysocka, 2012; Ho et al., 2014; de Laat and Duboule, 2013; Levine, 2010; Levine and Tjian, 2003). Interestingly, a recent analysis of an early branching and morphologically simple animal, the cnidarian *Nematostella vectensis*, has shown that cnidarians and bilaterians share a conserved gene regulatory landscape (Schwaiger et al., 2014). However, it is unclear whether these ancient genome regulatory features are animal innovations or whether they were already present in the unicellular ancestor of Metazoa.

To determine the timing and importance of regulatory changes in the origin of Metazoa, we need to unravel the genomic regulation of the extant animal relatives. Among the closest extant unicellular relatives of Metazoa, the amoeboid filasterean *Capsaspora owczarzaki* (herein *Capsaspora*), has the richest repertoire of transcription factors described to date (Sebé-Pedrós et al., 2011). These include genes, such as *Brachyury*, *Myc*, and *Runx*, that are essential for animal development. Moreover, *Capsaspora* is known to differentiate into three temporal life stages that are transcriptionally tightly regulated (Sebé-Pedrós et al., 2013b). These temporal cell types include (1) a filopodiated amoeba, which corresponds to the proliferative trophic stage, (2) an aggregative multicellular stage, in which the cells produce an extracellular matrix, and (3) a cystic resistance form without filopodia (see an schematic representation of the life cycle in Figure 3). Its key phylogenetic position as the sister group of animals and choanoflagellates, its rich gene repertoire, and the observed regulatory capabilities of *Capsaspora*, therefore, make it an ideal candidate to explore the origin of animal genome regulation.

The advent of functional genomics assays based on next-generation sequencing (NGS) has revolutionized the study of the regulatory genome. These techniques have shown that different chromatin biochemical signatures and accessibility are associated with *cis*-regulatory elements (Creyghton et al., 2010; Rada-Iglesias et al., 2011; Thurman et al., 2012), promoter types (Lenhard et al., 2012), ncRNAs (Marques et al., 2013), and gene transcriptional states (Dunham et al., 2012; Schwaiger et al., 2014). To date, however, this new paradigm has only been systematically applied to a handful of model species (Ho et al., 2014), and our understanding of most eukaryotic genomes remains limited to primary sequence. These techniques hold the potential to go beyond genome content description and systematically explore genome regulation in non-model systems like *Capsaspora*. Here, we apply these principles to study the dynamic *Capsaspora* regulatory genome in a comparative evolutionary framework and demonstrate that a major change in genome regulation was linked to the origin and the subsequent diversification of animal body plans.

## RESULTS

### Histone Modifications in *Capsaspora*

Posttranslational modifications of histone tails (hPTMs) are important components of the regulatory genomic landscape in eukaryotes. hPTMs play a crucial role in maintaining and transmitting on-off transcriptional signals (Zhou et al., 2011) by modifying the chromatin structure, and they are associated with specific regulatory elements in animals (Creyghton et al., 2010; Rada-Iglesias et al., 2011). To determine whether hPTMs are conserved between animals and their closest relatives or across all eukaryotes, we first analyzed the hPTMs of *Capsaspora* by chemical derivatization coupled to mass spectrometry and compared those with eukaryotes for which hPTMs are known (Figures 1 and S1). We found that H3 and H4 modifications are largely conserved across the eukaryotes analyzed. In contrast, we identified several novel *Capsaspora*-specific modifications in H2B and H2AZ and a *Capsaspora*-specific H2A variant, indicating that H2AZ and H2B histones and histone variants are the fastest evolving components of the histone code. Additionally, there was a correspondence between hPTMs and histone-modifying enzymes in the genome of *Capsaspora* (Figure 1). An example is the lack of H3K9me3 and H3K27me3, the two best-characterized animal repressive marks, co-occurring with the absence of the enzymes responsible for writing and erasing them (Suv3/9, G9a, and SETD1B for H3K9me3 and EZH2 (PRC2 complex) for H3K27me3). Despite some lineage-specific changes, H3 and H4 hPTMs are mostly conserved across eukaryotes, and thus, informative comparative analyses can be performed across distant taxa.

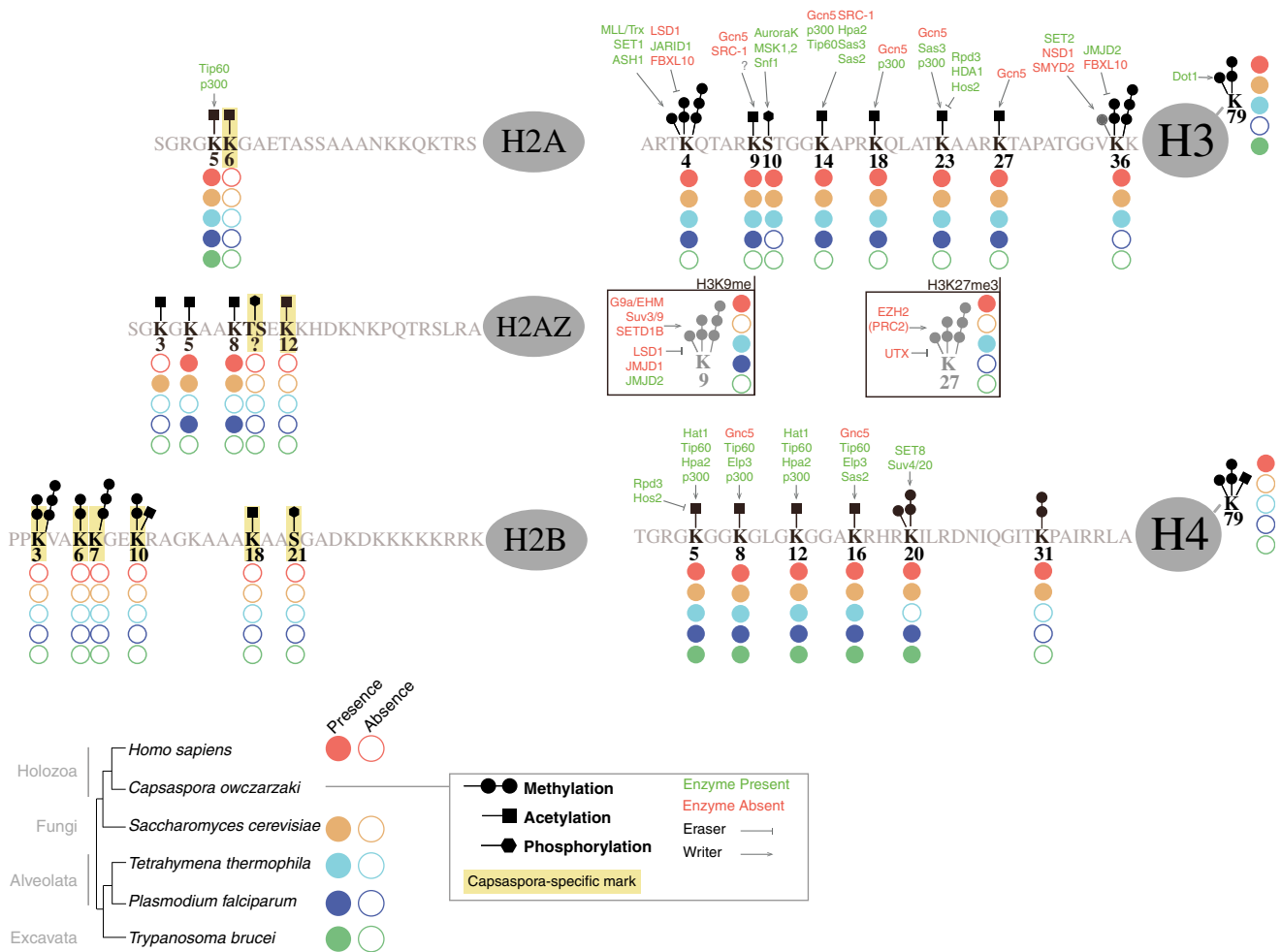
### Dynamic Chromatin States in *Capsaspora*

To investigate the genome-wide distribution of *Capsaspora* hPTMs across temporally segregated cell types, we selected those marks that have been widely used in animals to characterize chromatin states (Ho et al., 2014). Chromatin immunoprecipitation sequencing (ChIP-seq) was carried out for H3 lysine 4 trimethylation and monomethylation (H3K4me3 and H3K4me1),

H3 lysine 27 acetylation (H3K27ac), and H3 lysine 36 trimethylation (H3K36me3). Deep-sequencing reads were mapped in the *Capsaspora* genome, and their correlation with different genomic features and gene expression was analyzed (Figures 2, S2, and S3). Additionally, we undertook transposase-accessible chromatin sequencing (ATAC-seq) (Buenrostro et al., 2013) in each cell stage in order to interrogate nucleosome positioning and accessible chromatin as a proxy for active *cis*-regulatory elements. Normalized ChIP-seq read coverage around the transcription start site (TSS) reveals a unimodal H3K4me3 peak well positioned after the TSS of active genes that strongly colocalizes with H3K27ac (Figure 2A). In contrast, two sharp H3K4me1 peaks appear flanking H3K4me3/H3K27ac peaks, both before and after the TSS. Finally, H3K36me3 spreads through the gene bodies of active genes. All these marks correlate with the level of expression of active genes (Figure 2A), in a pattern similar to that observed in human cells (van Galen et al., 2016). It must be noted, though, that histone modifications might also be related to other regulatory processes; e.g., H3K36me3 has been linked to splicing (Kolasinska-Zwierz et al., 2009). Nucleosomes appear in highly ordered positions after the TSS of expressed genes, while, in contrast, nucleosomal fuzziness (which measures the deviation of each nucleosome position in the cell population) increases in weakly expressed and silent genes (Figures 2B and 2C). ATAC nucleosome-free reads are preferentially distributed in the surroundings of the TSS (Figure 2B). Finally, we also analyzed the distribution of RNAPolIII in *Capsaspora* genes (Figure S2), showing a strong peak around the TSS. In contrast, C-terminal domain (CTD) S2 phosphorylated RNA polymerase II (RNA Pol II) is distributed along the gene body, consistent with the known association of this S2 phosphorylated RNA Pol II form with transcriptional elongation (Egloff et al., 2012; Eick and Geyer, 2013; Schwer and Shuman, 2011). RNA Pol II coverage is associated with increased gene expression (Figure S2B) and changes dynamically between life stages (Figure S2C).

Next, we integrated these hPTM maps and ATAC nucleosome-free reads in order to predict chromatin states and their genome-wide distribution in *Capsaspora*, using a hidden Markov model (ChromHMM) (Ernst and Kellis, 2012) (Figure 2D). Overall, we defined seven different chromatin states that preferentially associated with specific genomic features (Figure 2E). For example, state one (defined by H3K36me3) is the most abundant and associates with coding regions and non-first introns (Figure 2E), consistent with the function of H3K36me3 as a transcriptional elongation mark (Dunham et al., 2012). In contrast, state seven corresponds to ATAC nucleosome-free signal, together with H3K4me1, and is strongly enriched around TSS (Figure 2E), corresponding to potential regulatory sites.

Given the absence of known repressive marks in *Capsaspora* (see Figure 1), we asked whether strongly repressed genes show any particular biochemical signature. Thus, we compared lowly expressed genes (<2 FKPMs) with active genes (Figures 2E and 2F) and observed a particular profile in which H3K4me1 shifts from two flanking peaks to a single post-TSS peak, H3K27ac is spread across the gene body, and both H3K4me3 and H3K36me3 are absent (Figures 2F, S2, and S3). Similarly, we observe a strong enrichment of state four across the gene



**Figure 1. Histone Modifications in *Capsaspora***

Histone N-terminal tail sequences of *Capsaspora* with the identified posttranslational modifications are shown. Below: filled or empty circles indicate whether the particular histone mark is present or absent, respectively, in the different eukaryotic species represented in the phylogenetic tree (left). Above: the presence (green) or absence (red) of specific histone modifiers in the *Capsaspora* genome is shown; both enzymes that add the mark (writers) and enzymes that remove it (erasers) are indicated. *Capsaspora*-specific marks are highlighted in yellow. The repressive marks H3K9me and H3K27me3 are absent in *Capsaspora* and indicated separately in a box below the corresponding position. See also Figure S1.

body and of state three around TSS. If we specifically select genes with H3K27ac across the gene body (>800 bp from TSS) and post-TSS H3K4me1 peaks (TSS+800 bp), we recover the population of repressed genes (Figure 2G). This signature of repression has never been described in any other organism and might represent a *Capsaspora*-specific mechanism.

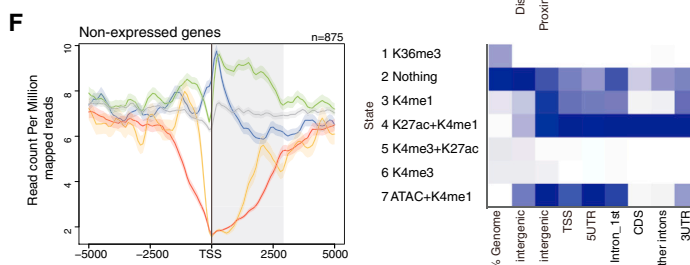
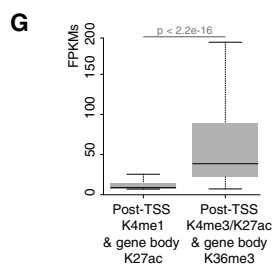
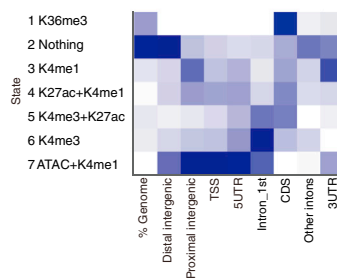
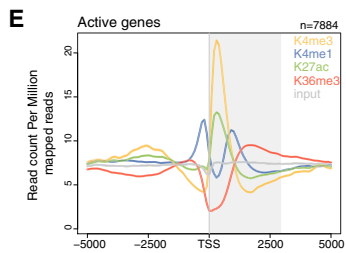
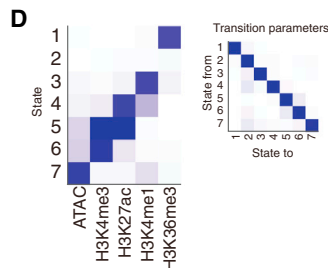
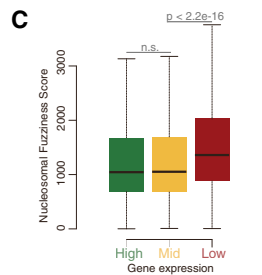
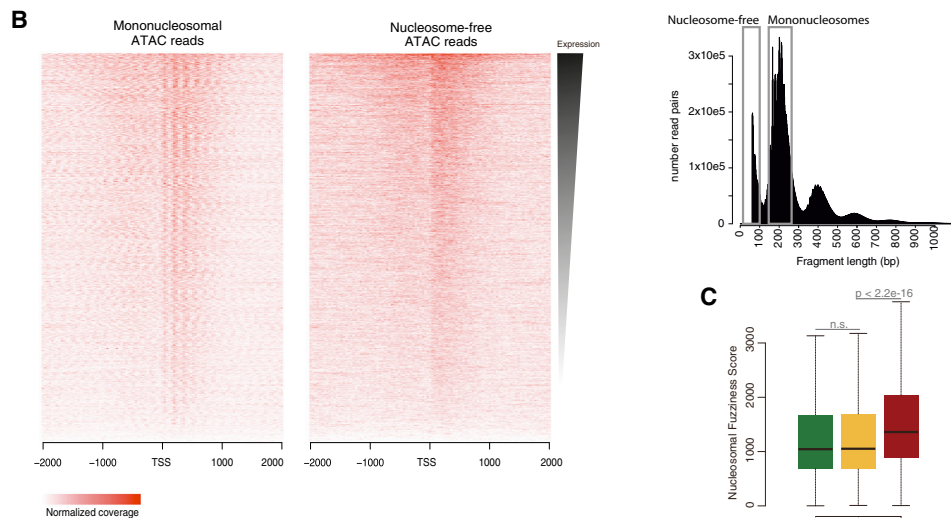
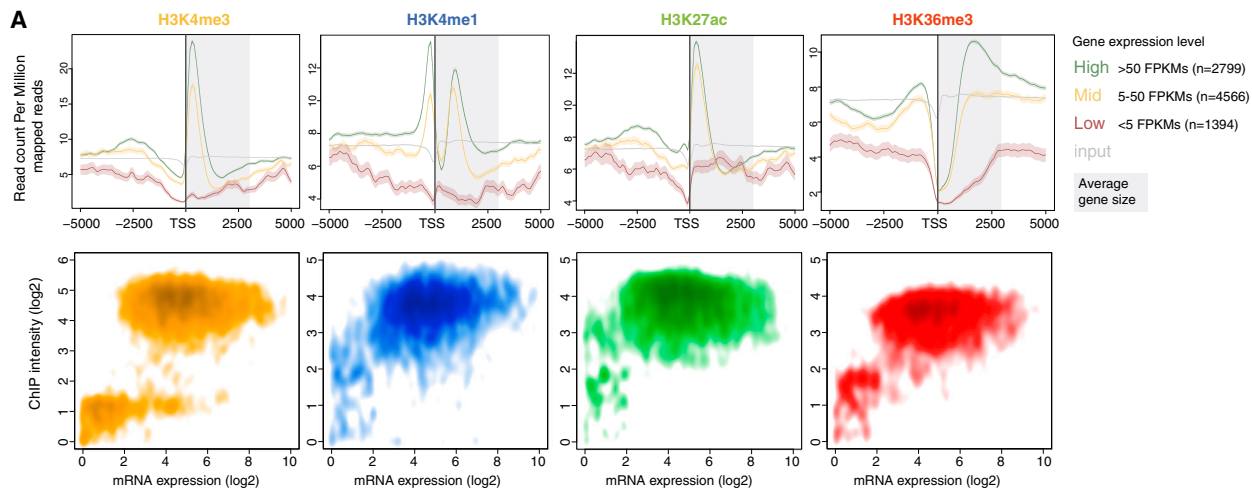
Finally, we evaluated how changes in chromatin features correlate with life stage transitions in *Capsaspora*. First, we observed that chromatin marks change between life stages, correlating with changes in genes expression (Figures 3A, 3B, and S3). Second, we treated *Capsaspora* cells with Trichostatin A (TSA), a widely used histone deacetylase (HDAC) inhibitor (Simola et al., 2016), in order to study the role of histone acetylation in the life cycle of *Capsaspora*. Treatment with 3 μM TSA blocked life cycle transitions, e.g., from cystic to filopodial stage (Figure 3C). As expected when blocking HDACs, TSA induced

an increase in histone acetylation levels (Figure 3D). Using RNA sequencing (RNA-seq), we also observed that TSA caused a generalized activation of gene expression (Figure 3E). These observations directly link histone modifications with life cycle transitions and gene expression in *Capsaspora*.

Overall, we obtained high-coverage linear maps of multiple epigenomic features, which show consistent patterns of association with expression states, specific genomic regions and temporal cell-type transitions. These maps allowed us to further systematically dissect functional elements in *Capsaspora* genome.

**The Origin of Animal Promoter Types**

To understand the evolution of proximal promoter chromatin regulatory signatures, we compared TSS profiles of *Capsaspora* with different metazoan taxa and *Saccharomyces cerevisiae* using publicly available ChIP-seq datasets (Figure 4). All species



(legend on next page)



show well-positioned post-TSS H3K4me3/H3K27ac peaks correlated with active gene expression. *Homo sapiens* show a strong bimodal peak, whereas a *Drosophila melanogaster* and *Caenorhabditis elegans* show weak bimodality. In comparison, the cnidarian *Nematostella vectensis*, as well as *Capsaspora* and *Saccharomyces*, present sharp unimodal post-TSS H3K4me3/H3K27ac peaks. This difference between bilaterians and others could be related to the presence (bimodality) or absence (unimodality) of anti-sense transcript production from some TSS (Ho et al., 2014). Moreover, H3K36me3 is present in the gene bodies of active genes in all species. Interestingly, H3K4me1 is enriched on both sides of the TSS in animals and *Capsaspora*. The signal is weaker and less sharp in animals, whereas it is sharp and complementary to the H3K4me3/H3K27ac peak in *Capsaspora*. In contrast, *Saccharomyces* has only one post-TSS H3K4me1 peak after the H3K4me3/H3K27ac peak.

The distribution of histone modifications around TSS has been used to define three different promoter types in metazoans (Lenhard et al., 2012). Type I promoters are associated with tissue-specific expression in terminal-differentiated cell types, and they are characterized by fuzzy nucleosomes, strongly positioned H3K4me3 and H3K27ac peaks, and no H3K4me1 and H3K27me3 marks. Type II promoters are found in ubiquitously expressed genes and show strongly positioned nucleosomes and flanking H3K4me1 marks (in addition to post-TSS H3K4me3 and K27ac). Finally, type III promoters, also called bivalent promoters, are associated with developmentally regulated genes and present both activation (H3K4me3) and repression (H3K27me3) marks (Lenhard et al., 2012). Thus, the different configurations observed here are likely to reflect different promoter specification modes. Interestingly, *Capsaspora* TSS signatures strongly resemble those of animal type II promoters (also called “ubiquitous”), including highly ordered nucleosome positioning (Figure 2). In contrast, no type I (without flanking H3K4me1 and fuzzy nucleosomes) or type III (H3K27me3-regulated developmental promoters) promoters could be identified in *Capsaspora*. This indicates that type I and type III promoters are animal innovations and related to the emergence of cell-type-specific (type I) and developmental regulation of gene expression (type III).

### Premetazoan Long Intergenic Non-coding RNAs Regulation

Long intergenic non-coding RNAs (lincRNAs) are an important component of animal genome regulation (Marques and Ponting, 2014; Ulitsky and Bartel, 2013). lincRNAs exert multiple developmental and cell-type-specific regulatory functions, and their number is greatly expanded in multicellular animals and plants (Gaiti et al., 2015; Kapusta and Feschotte, 2014; Ulitsky and Bartel, 2013). In order to understand the evolution of lincRNAs in the lineage leading to metazoan multicellularity, we used deep strand-specific polyA-enriched RNA-sequencing data to annotate lincRNAs in *Capsaspora*. After applying multiple filters, we predicted 632 lincRNAs and validated 17 of them by RT-PCR (Figures S4 and S5). This is less than those identified in multicellular animals, but more than those found in yeast (Kapusta and Feschotte, 2014). *Capsaspora* lincRNAs show dynamic expression (Figure S4A) and have multiple features that differentiate them from coding genes (Figure S4B). Interestingly, we found that predicted *Capsaspora* lincRNAs can be separated into two populations based on their association with H3K4me1 and H3K4me3 (Figures S4C and S4E), resembling those found in mouse lincRNAs (Marques et al., 2013). Moreover, similar to mouse, these two lincRNA populations show only slight differences in length, expression level, and expression variation (Figure S4D), so the functional significance of these two populations remains unclear. Thus, our data reveal that elaborate lincRNA genome regulation was already present in unicellular premetazoans.

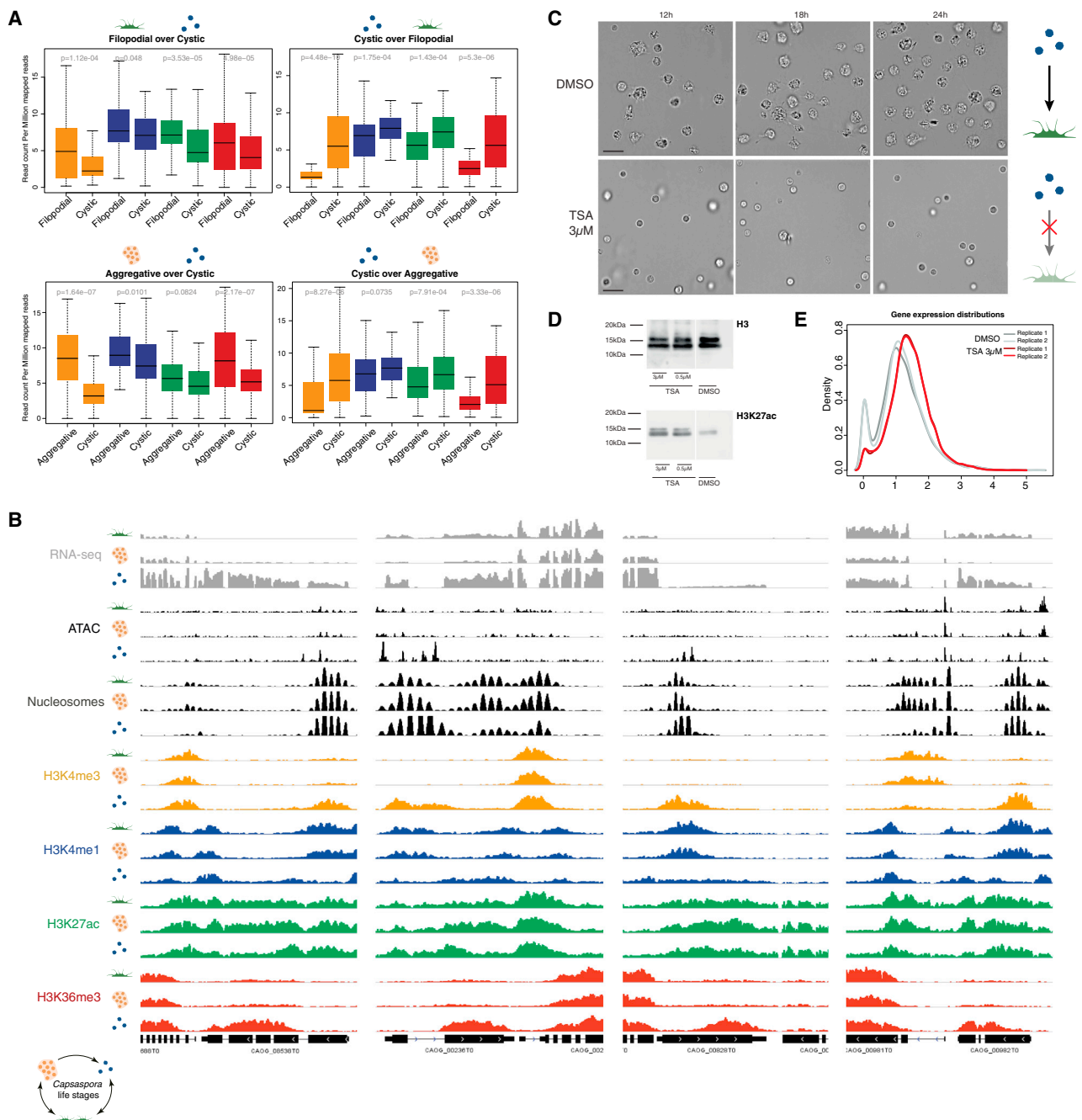
### Accessible Chromatin Landscape of *Capsaspora*

Transcription factors and other regulatory proteins bind to discrete DNA sequences, creating nucleosome-depleted areas of high-nuclease/transposase accessibility. We therefore used high-coverage nucleosome-free ATAC reads to identify all active regulatory sites in the *Capsaspora* genome and study their distribution and cell-type dynamics (Figures 5 and S6). In *Capsaspora*, 95% of the genome lies within 6.4 kb of one of the 11,927 discrete regulatory sites identified, and 63% of genes are associated with at least one site (Figure 5A). As an estimation of the number of regulatory inputs, we calculated the number of

#### Figure 2. Genome-Wide Chromatin Annotation in *Capsaspora*

(A) Top: TSS-centered average normalized read coverage plots of hPTMs in the filopodial stage for genes with high (green), intermediate (yellow), and low (red) expression levels. The x axis spans  $-5$  to  $+5$  kb around the TSS. The shaded gray area represents the average size of *Capsaspora* genes. Bottom: scatterplots of hPTMs coverage ( $\log_2$  normalized reads) compared to mRNA expression levels ( $\log_2$  fragments per kilobase of transcript per million mapped reads [FPKMs]). (B) Heatmaps of ATAC mononucleosome-associated (left) and nucleosome-free (right) reads centered around the TSS of genes sorted by level of expression in the filopodial stage. Right: histogram showing an example of the distribution of ATAC-seq fragment sizes obtained. (C) Boxplot representing the mean fuzziness score of the first four post-TSS nucleosomes of genes grouped by the level of expression in the filopodial stage. The p value is indicated for the Wilcoxon rank-sum test. (D) Heatmaps representing the emission (left) and transition (right) parameters of a seven-state hidden Markov model. In the left heatmap, the white-blue (0–1) scale represents the frequency with which a given mark is found at genomic positions corresponding to the chromatin state. In the right heatmap, the white-blue (0–1) scale represents the frequency with which a given state changes into another state at the neighboring location. (E) Chromatin signatures in active genes ( $>2$  FPKMs) in the filopodial stage. The plot (left) represents the average normalized read coverage of histone modifications around the TSS of these active genes, and the heatmap (right) indicates the relative percentage of the genome represented by each chromatin state (first column) and relative fold enrichment for different genome features (other columns). (F) Chromatin signatures in silent genes in the filopodial stage (heatmap and plot as in C). (G) Boxplot representing the expression levels in the filopodial stage of genes (left) selected for having a significant peak of H3K27ac in the gene body (more than 800 bp from the TSS) and a significant peak of H3K4me1 after the TSS (within 800 bp), and vice versa (right). The p value is indicated for Wilcoxon the rank-sum test.

See also Figures S2 and S4 and Data S1.



### Figure 3. Dynamic Chromatin Modifications

(A) Boxplots showing hPTMs coverage levels in differentially expressed genes between h stages, as indicated above each boxplot. The p value is indicated for the Wilcoxon signed-rank test.

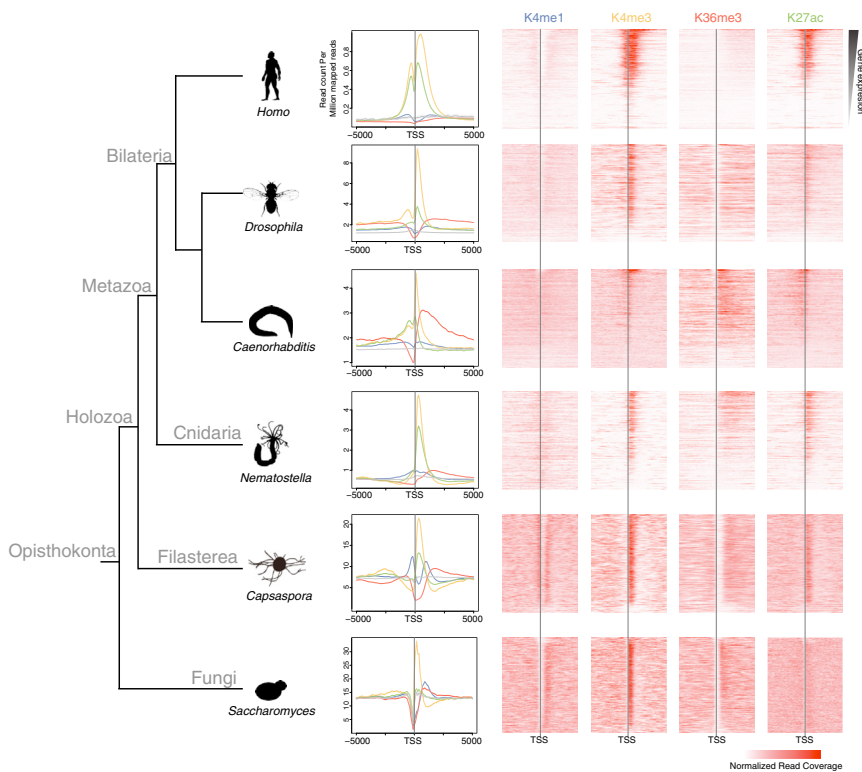
(B) Illustrative examples of dynamic chromatin modifications in *Capsaspora*. Different genomic windows show normalized coverage for different chromatin features and their dynamic association with gene expression. For each feature, the top track corresponds to the filopodial stage, the middle track to the aggregative stage, and the bottom track to the cystic stage.

(C) Histone deacetylase inhibition experiments. Pictures of *Capsaspora* cells at different time points of incubation with DMSO (negative control) and TSA 3  $\mu$ M. Transition from cystic to filopodial stage is blocked in the TSA-treated cells. Scale bar, 10  $\mu$ m.

(D) Western blot against total H3 and H3K27ac on histone extracts from control cells (DMSO) and cells treated with 0.5 and 3  $\mu$ M TSA. White line indicates a lane was removed.

(E) Gene expression distributions from biological replicates of control (DMSO, gray colors) and TSA-treated (red colors) cells. Notice the decrease in the fraction of non-expressed genes and the general shift in the distribution of TSA-treated cells.

See also Figures S2 and S3.



**Figure 4. Comparative Proximal Distribution of Chromatin Marks across Opisthokonta Species**

For each species, a plot shows the average normalized read coverage of four different histone modifications around the TSS ( $\pm 5$  kb), and heat-maps represent the same coverage for all genes sorted by level of expression. ChIP-seq data were obtained from publicly available datasets: *Homo sapiens*, *Drosophila melanogaster*, *Caenorhabditis elegans*, *Nematostella vectensis*, and *Saccharomyces cerevisiae*.

ATAC-defined regulatory sites per gene. Comparison of the number of associated sites across different types of genes revealed that transcription factors (TFs) are significantly enriched in regulatory sites (Figure 5B). In particular, T-box, bHLH, and bZIP TFs have the largest number of regulatory sites (Figure 5C). In contrast to previous predictions (Sebé-Pedrós and De Mendoza, 2015), this indicates the presence of intricate TF networks in *Capsaspora*. In addition, these regulatory sites were strongly enriched around TSS, in particular at proximal intergenic regions, first introns, and 5' UTRs (Figure 5D), and depleted at gene bodies and distal intergenic regions. Interestingly, many of these regulatory sites show dynamic changes in ATAC-seq signal across life stages in *Capsaspora* (Figure 5E). In particular, 36% are stage specific and only 22% are constitutive in all three stages. Therefore, this specific and primarily proximal regulatory lexicon supports temporal cell-type transitions in *Capsaspora* and very likely also in the unicellular ancestors of animals.

### Ancient Transcription Factor Networks

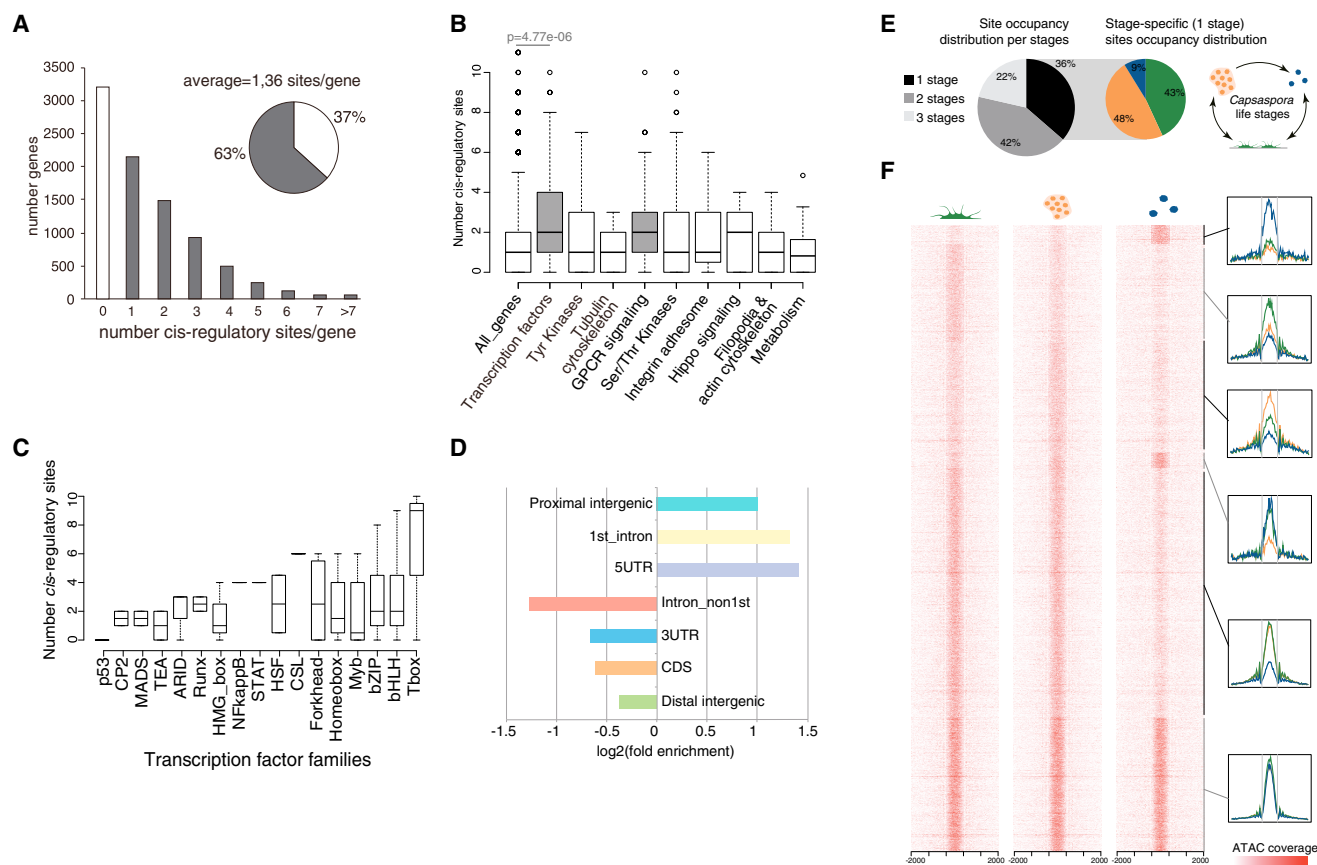
*Capsaspora* has a rich repertoire of metazoan-like TFs that are enriched in regulatory sites; however, it is unclear which specific genes are regulated by these TFs. To gain insights into premetazoan TF networks, we used motif analysis of the ATAC-defined regulatory sites. First, we looked for sites potentially bound by *Capsaspora-Brachyury*, an essential gene for animal gastrulation and mesoderm differentiation and the only TF whose binding site has been experimentally validated in *Capsaspora* (Sebé-Pedrós et al., 2013a). We found approximately 900 instances of this motif in the regulatory sites, all of them consistently displaying a similar tag density profile (Figure 6A). When compared with

the whole population of *cis*-regulatory elements, these inferred *Bra* sites are preferentially located at the first intron and 5' UTR and are predominantly associated with the filopodial amoeba and aggregative stages (Figure 6B). Accordingly, these *Capsaspora-Bra* sites are also more strongly correlated with the activating marks H3K4me3 and H3K27ac in these two stages (Figure 6C) compared with the cystic stage, and they are also enriched in these active histone marks compared with random *Bra* motifs found outside ATAC-defined regions (Figure 6D).

In order to validate some of our *Bra* downstream target predictions, we developed an anti-*Capsaspora-Bra* (CoBra) antibody (Figures 6E and 6F) and performed ChIP-qPCR experiments. We selected 20 ATAC-defined regulatory sites with *Bra* motifs (e.g., Figure 6H), including several with our lower limit selection threshold (0.80 Matscan cutoff [Blanco et al., 2006]), and compared them with ten random regions in the genome with strongly conserved *Bra* motifs (>0.90 Matscan cutoff). The ATAC-defined *Bra* regulatory sites were strongly enriched in CoBra compared with random motifs (Figure 6G), validating our *Bra* target prediction approach.

The *Capsaspora-Bra* downstream target network includes genes involved in establishment of cell polarity, phagocytosis, metabolism, transcription factors, and GPCR signaling genes (Figure 6I). Moreover, we identified 63 shared orthologs between inferred *Capsaspora-Bra* targets and those known for mouse *Brachyury* (Lolas et al., 2014). Interestingly, those shared orthologs are enriched in actin cytoskeleton and amoeboid cell-motility functions (Figure 6J). This suggests that there was a conserved *Brachyury* downstream target network already present in premetazoan lineages and involved in cell migration, an essential cellular function later used in animal gastrulation.

Next, we performed a blind motif-enrichment analysis of all ATAC-defined sites in order to gain additional information on other TFs. Among the 29 significantly enriched nucleotide motifs, three of them strongly resemble (~90% similarity) known motifs for animal *Runx*, *NFAT/NFkappaB*, and *Myc* TFs. *Capsaspora* has clear orthologs of these three TFs (Sebé-Pedrós et al., 2011). Assuming that the motifs represent the consensus motifs for these *Capsaspora* orthologs, this provides evidence of



**Figure 5. The Genomic Landscape of *cis*-Regulatory Elements in *Capsaspora***

- (A) Distribution of the number of regulatory sites per gene.  
 (B) Number of *cis*-regulatory elements associated with different gene categories. Highlighted in gray are those with a significant enrichment (Wilcoxon rank-sum test  $p$  value  $< 0.01$ ) compared with all genes.  
 (C) *Capsaspora* transcription factor families sorted by the number of *cis*-regulatory elements associated per gene.  
 (D) Preferential distribution of *cis*-regulatory sites across genomic features.  
 (E) Pie charts showing the distribution of the number of stages in which each site is occupied (left) and the stage distribution of the stage-specific fraction of regulatory sites (right).  
 (F) Heatmaps of clustered *cis*-regulatory elements ( $\pm 2$  kb) showing dynamic normalized ATAC nucleosome-free read coverage between stages. Plots show the associated average coverage profiles of each cluster.
- See also [Figure S6](#).

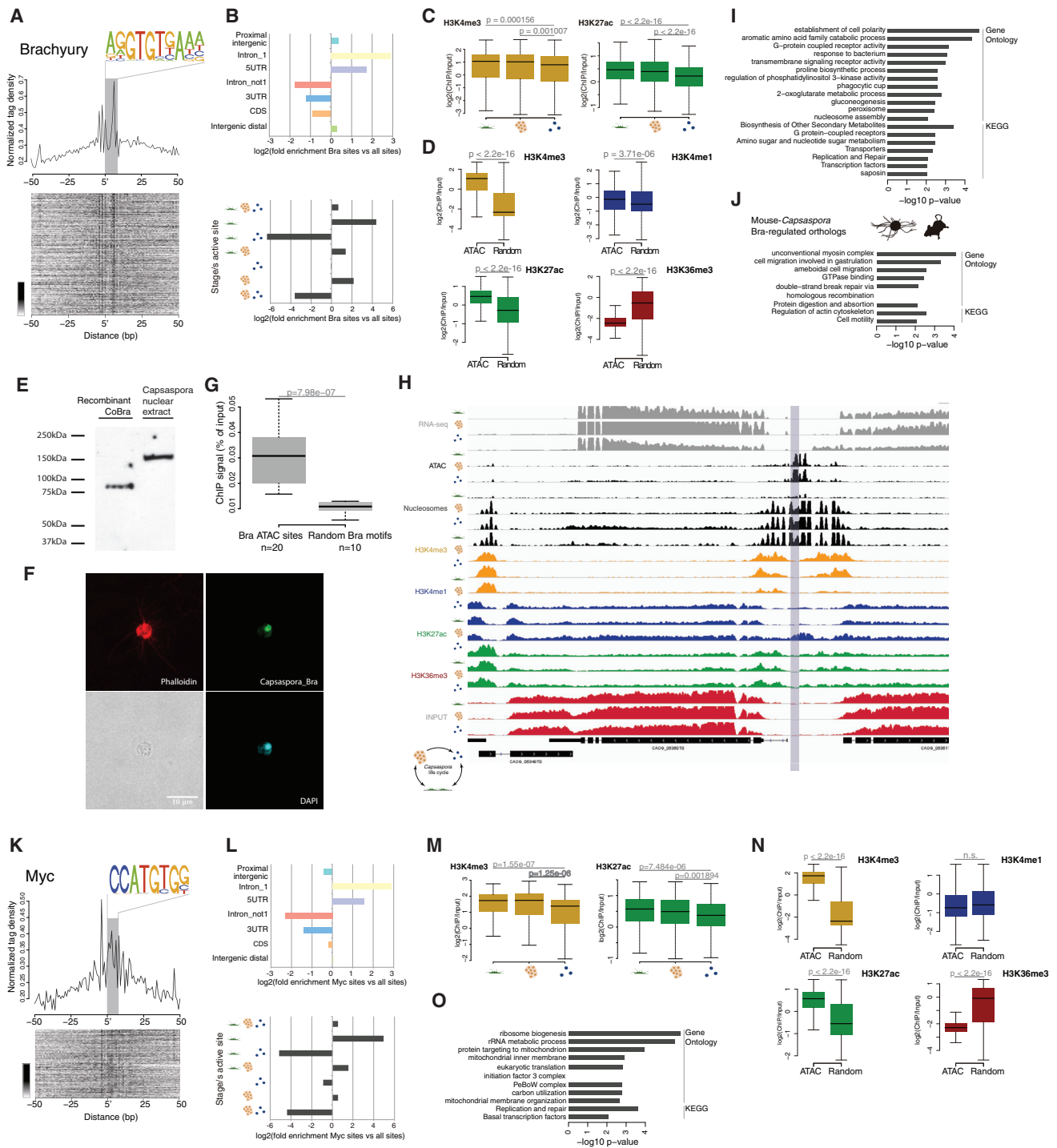
associations with genomic features and hPTMs (Figures 6 and S7). In particular, *Capsaspora-Myc*, a well-studied proto-oncogene in animals, appears to be strongly associated with regulatory sites that show higher ATAC-seq signal in the filopodial stage (Figure 6L), the proliferative stage in *Capsaspora* (Sebé-Pedrós et al., 2013b). These *Capsaspora-Myc* sites are more strongly correlated with the activating marks H3K4me3 and H3K27ac in filopodial and aggregative stages (Figure 6M) compared with the cystic stage, and they are also enriched in these active histone marks compared with random *Myc* motifs found outside ATAC-defined regions (Figure 6N). Moreover, *Myc* regulates genes mainly involved in ribosome biogenesis and translation (Figure 6O), similar to what is known for animal *Myc* networks (van Riggelen et al., 2010).

Interestingly, all TFs analyzed here show an enrichment of other TFs in their inferred downstream networks, reinforcing

the idea of relatively complex TF-TF regulatory interactions in *Capsaspora*. The expansion of the TF repertoire at the stem of Metazoa (Sebé-Pedrós and De Mendoza, 2015), both in the total number of genes and of TF families, was probably associated with an increase in complexity of these TF networks. Remarkably, however, the inferred *Capsaspora* TF downstream targets suggest that at least some TF downstream regulatory networks were already conserved in the unicellular ancestor of metazoans and then subsequently remodeled within the animal lineage.

#### Distal Enhancers Are Animal Specific

To address whether there are potential distal enhancer elements in the genome of *Capsaspora*, we compared the regulatory sites defined by ATAC between *Capsaspora* and animals. Regulatory sites in *Capsaspora* are significantly smaller and more uniformly distributed than are sites in *Drosophila* and *Homo sapiens*



**Figure 6. Capsaspora Brachyury and Myc Regulation**

(A) Plot of ATAC-seq nucleosome-free reads average density around *Bra* motifs (top) and heatmap of the signal around the individual sites (bottom).  
 (B) Differential distribution of regulatory sites containing *Bra* motif compared with all sites according to genomic feature (top) and stage/s in which the site is active (bottom).  
 (C) Enrichment of different histone modifications (ChIP versus input) at *Bra* sites across stages.  
 (D) Enrichment of different histone modifications (ChIP versus input) at *Bra* motifs in ATAC-defined sites compared with motifs occurring randomly in the genome.  
 (E) Western blot of recombinant *Capsaspora-Brachyury* protein and *Capsaspora* nuclear protein extract, using *Capsaspora-Brachyury* affinity-purified antibody from guinea pig.

(legend continued on next page)

(Figure 7A). This means that regulatory sites in *Capsaspora* are bound by small numbers of proteins, whereas in animals large assemblies of regulatory factors bind active sites, allowing more complex combinatorial regulation. Comparison of the distribution of regulatory sites across genomic features revealed that distal sites, located at non-first introns and at intergenic regions, are extremely abundant in *Homo* and *Drosophila* but rare in *Capsaspora* (Figure 7B), and even rarer in *Saccharomyces cerevisiae* (Figure S7I) (Bulger and Groudine, 2011). Distal regulatory sites in animals, called enhancer elements, have unique chromatin signatures (Creyghton et al., 2010; Rada-Iglesias et al., 2011), including the presence of H3K4me1 constitutively and H3K27ac when activated. Using ATAC-defined proximal and distal intergenic *cis*-regulatory sites, we analyzed the read coverage for different histone marks for those sites in *Capsaspora*. At distal intergenic sites, *Capsaspora* shows no enrichment of H3K4me1 and H3K27ac compared with H3K4me3, whereas this enrichment is evident in *Homo sapiens* and *Drosophila* (Figure 7C). At proximal intergenic sites (800 bp upstream of TSS), H3K4me3 is significantly enriched over H3K4me1, although in *Capsaspora* this is less marked due to the proximal intergenic H3K4me1 enrichment described above. In all examined sites, H3K36me3 is depleted, as is expected outside gene bodies. Thus, regulatory sites in *Capsaspora* are mostly proximal, in contrast with the distal regulation observed in animals. Moreover, none of the regulatory sites in *Capsaspora* have biochemical signatures typical of animal enhancers. This indicates that distal regulation by enhancer elements is an animal evolutionary innovation and, probably, the most important difference in the genomic regulatory capabilities between premetazoans and metazoans.

## DISCUSSION

In order to understand the evolution of the metazoan regulatory genome, we have here performed the first integrative analysis of the genome regulatory biology of a close unicellular relative of metazoans, the amoeboid filasterean *Capsaspora owczarzaki*. Indeed, this is the first such analysis of temporal chromatin dynamics in any non-model eukaryote.

We show that histone postranslational modifications, particularly those in H3 and H4, are highly conserved between *Capsaspora* and animals and also in other eukaryotes. Furthermore, good correspondence exists between these modifications and the presence/absence of known histone-modifying enzymes in the *Capsaspora* genome. For example, *Capsaspora* lacks H3K27me3 Polycomb repression marks, and it also lacks the PRC2 complex proteins, including EZH2, the central

methyltransferase of the complex (Margueron and Reinberg, 2011).

We observe that 91.7% of the compact *Capsaspora* genome (28 Mb) includes regions producing transcripts (protein coding or lincRNAs) and/or regions with particular chromatin signatures and *cis*-regulatory sites. These signatures and regulatory sites are dynamically associated with life cycle progression and gene expression in *Capsaspora* and some, particularly active chromatin states, are shared with metazoans.

Our results indicate that *Capsaspora* has more numerous potential TF-TF regulatory connections than was previously thought, suggesting complex regulatory networks exist. Transcription factor networks tend to be quickly rewired during evolution (Li and Johnson, 2010; Sorrells and Johnson, 2015), and despite this, we find a remarkable degree of conservation between *Capsaspora* and animals in the downstream networks of orthologous TF that are key to animal multicellularity and development, such as *Brachyury* and *Myc*. These findings suggest that core downstream target networks of some developmental TF evolved long before the advent of animal multicellularity (Davidson and Erwin, 2006), controlling behaviors, such as proliferation and cell motility, in the first animal cells. These core conserved TF networks were subsequently integrated into complex developmental programs during animal evolution (Peter and Davidson, 2011).

*Capsaspora* also has a large repertoire of polyadenylated and, in some cases, alternatively spliced lincRNAs. These lincRNAs have temporal, cell-type-specific expression patterns, and they are associated with chromatin signatures similar to those found in metazoans (Marques et al., 2013). These *Capsaspora* lincRNAs show no homology with any known metazoan lincRNA, due to the fast evolution of lincRNA genes (Hezroni et al., 2015; Kapusta and Feschotte, 2014), and their functions are currently unknown. Despite this, our results indicate that elaborate genome regulation by long non-coding RNAs is not exclusive to multicellular organisms and was likely present in the protistan ancestors of Metazoa.

In contrast, the most important difference observed between *Capsaspora* and animal genome regulation is the marginal presence of distal *cis*-regulatory sites in *Capsaspora*, together with the absence of particular chromatin signatures associated with animal enhancers. This is in line with what is known in yeast, where regulation is proximal to the TSS (Bulger and Groudine, 2011) and no distal regulatory loops have been identified in genome 3D structure studies (Duan et al., 2010; Tanizawa et al., 2010). This result strongly indicates that distal enhancer elements are a major animal evolutionary innovation and constitute the basis of the sophisticated and highly evolvable gene

(F) *Capsaspora* filopodial stage cell stained with phalloidin (red, actin cytoskeleton), DAPI (blue, nucleus), and *Capsaspora-Brachyury* antibody (green). Notice *Bra* localization in the nucleus.

(G) Boxplot showing the *Capsaspora-Brachyury* ChIP-qPCR signal for predicted *Bra* regulatory sites versus random *Bra* motifs in the genome.

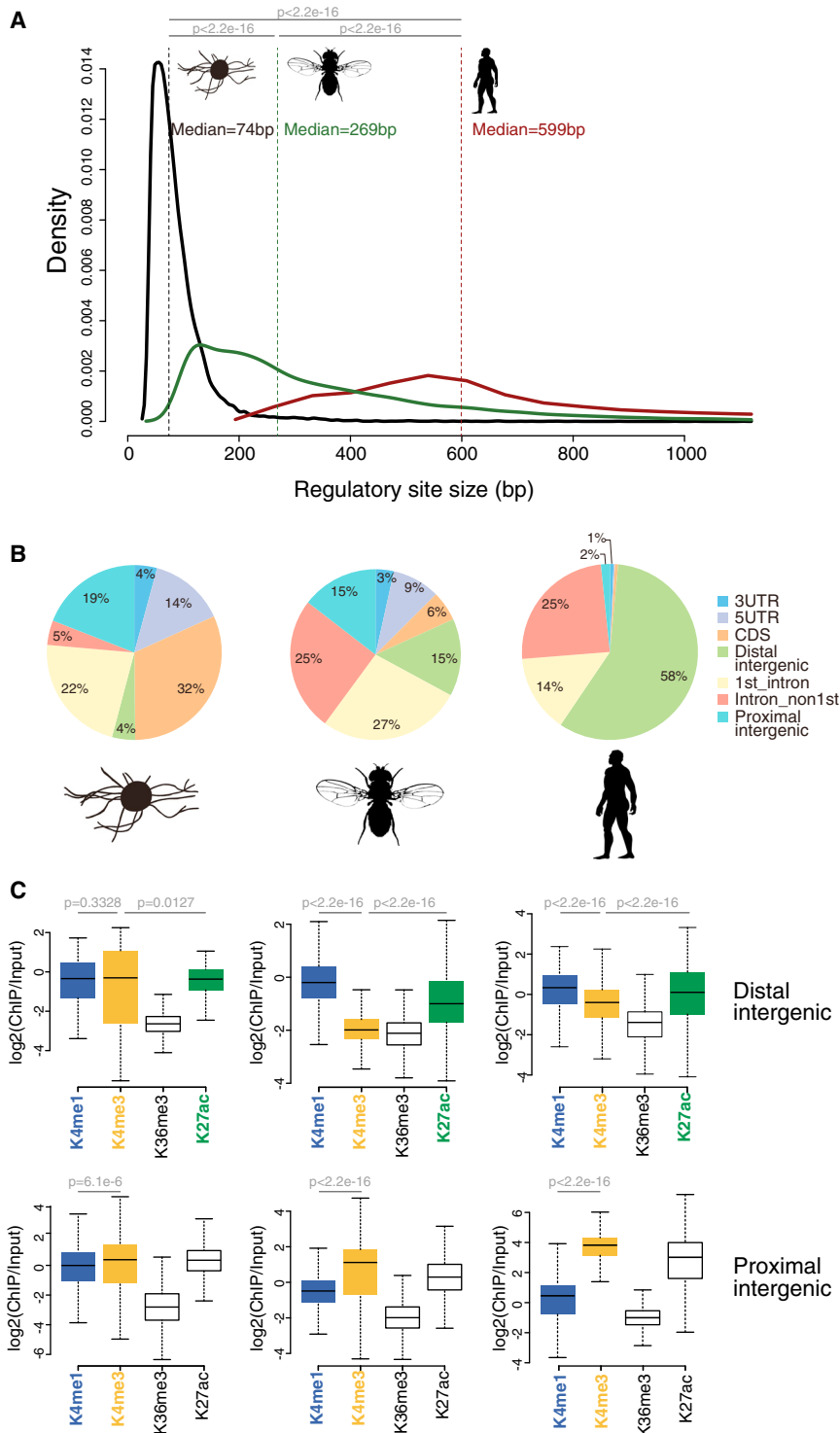
(H) Illustrative case example of a predicted *Bra* regulatory site (highlighted in blue). For each feature, the top track corresponds to the filopodial stage, the middle track to the aggregative stage, and the bottom track to the cystic stage. Notice the decreased ATAC signal in the putative *Bra*-regulatory site in the cystic stage.

(I) Enriched gene ontology (GO) terms and KEGG pathways among genes associated with *Bra* regulatory sites.

(J) Enriched GO terms and KEGG pathways among genes associated with *Bra* regulatory sites with shared orthologs regulated by *Bra* in mouse.

(K–O) Same as (A–D) and (I) for *Capsaspora Myc*.

See also Figure S7.



**Figure 7. Comparative Analysis of Regulatory Sites between *Capsaspora* and Animals**

(A) Distribution of ATAC-defined regulatory site sizes (bp) in *Capsaspora*, *Drosophila* and *Homo*. (B) Genomic feature distribution of regulatory sites in *Capsaspora*, *Drosophila* and *Homo*. (C) Enrichment of different histone modifications (ChIP versus input) at regulatory sites in distal (top) and proximal (bottom) intergenic regions in *Capsaspora*, *Drosophila* and *Homo*. In each boxplot, p values are indicated for Wilcoxon signed rank tests between H3K4me3 and H3K4me1 and between H3K4me3 and K27ac (only for distal intergenic). See also Figure S7.

blocks in animal unicellular relatives like *Capsaspora* (Irimia et al., 2012). Moreover, the observation that *cis*-regulatory sites in *Capsaspora* are much smaller than those of animals further indicates that complex combinatorial TF binding appeared after the divergence of animals. Thus, not only did TF numbers expand (de Mendoza et al., 2013) and TF interaction networks became more intricate (Reinke et al., 2013; Seb e-Pedr os et al., 2013a) at the stem of Metazoa, but also the combinatorial binding of multiple TFs increased the potential number of developmental regulatory states in animals (Erwin, 2009; Erwin and Davidson, 2009; Peter and Davidson, 2011). Finally, specific promoter types for cell-type-specificity and developmental regulation, defined by chromatin signatures, appear also to be an animal innovation, since *Capsaspora* only has type II promoters. The emergence of additional promoter architectures in animals allowed distinct groups of genes to be controlled in different ways (Lenhard et al., 2012).

Overall, we reconstruct an evolutionary scenario in which the emergence of specific enhancer and promoter features at the onset of Metazoa, together with the expansion and remodeling of TF networks and non-coding RNA systems, allowed for fine-tuned spatiotemporal control of gene expression. Thus, the increase in regulatory genome complexity was probably a crucial step for the integration of cell types associated with the emergence of animal multicellularity. The precise mo-

regulatory landscapes observed in animals (Andersson et al., 2014; Schwaiger et al., 2014; Villar et al., 2014). The emergence of these long-range *cis*-regulatory elements could also explain the pervasiveness of conserved syntenic regulatory blocks in animal genomes (Irimia et al., 2013) and the absence of these

lecular basis for this regulatory change remains to be determined. However, we hypothesize that it could be associated with the emergence of new chromatin modifying and remodeling enzymes and/or linked to the evolution of mechanisms for long-range genomic interaction and compartmentalization (Tanay and

Cavalli, 2013). Future analyses in other unicellular holozoans and in early branching animals, together with the study of the three-dimensional genome architecture of these taxa, will be crucial to further delineate the early evolution of the animal regulatory genome.

## EXPERIMENTAL PROCEDURES

### Capsaspora Cultures

*Capsaspora* strain ATCC30864 cells were grown axenically in ATCC medium 1034 at 23°C and differentiated as described in the [Supplemental Experimental Procedures](#).

### Histone Mass Spectrometry

*Capsaspora* histones were isolated by acid extraction, derivatized with propionic anhydride, and digested as described in [Garcia et al. \(2007\)](#). Tryptic peptides were analyzed via liquid chromatography–tandem mass spectrometry on an LTQ-Orbitrap Velos Pro mass spectrometer. Peptides were identified using the Mascot search engine.

### Chromatin Immunoprecipitation

ChIP-seq and ChIP-qPCR were performed at three different life stages using antibodies against H3K4me3, H3K4me1, H3K27ac, H3K36me3, RNAPolII, and CoBra as detailed in the [Supplemental Experimental Procedures](#). 50 bp single-end Illumina sequencing reads were aligned to the *Capsaspora* genome (v.2) using Bowtie ([Langmead et al., 2009](#)), and regions of enrichment were determined using MACS2 ([Zhang et al., 2008](#)), correcting for genome mappability. Chromatin state definition and genomic feature enrichment was performed using ChromHMM ([Ernst and Kellis, 2012](#)). *Capsaspora* genome was reannotated as described in the [Supplemental Experimental Procedures](#).

### HDAC Inhibition Experiments

*Capsaspora* cystic stage cells were transferred to fresh medium and treated with 3 μM TSA and DMSO (negative control), and stage transition to the filopodial stage was monitored every 6 hr. Histones were isolated from *Capsaspora* cells incubated with DMSO or TSA by acid extraction, and the levels of histone acetylation were measured by western blot. Total RNA from treated cells was also extracted for RNA-seq. Further details are provided in the [Supplemental Experimental Procedures](#).

### ATAC-Seq

ATAC-seq was performed as originally described in [Buenrostro et al. \(2013\)](#), using 500,000 cells per cell stage. 50 bp paired-end sequencing reads were aligned to the *Capsaspora* genome (v.2) using Bowtie. Nucleosomal-free reads were used to define *cis*-regulatory sites using MACS2. The blind TF motif enrichment analysis was performed in these sites using HOMER ([Heinz et al., 2010](#)). Mononucleosomal reads were used to define nucleosome positions and fuzziness using Danpos2 ([Chen et al., 2013](#)).

### lincRNA Annotation

High-coverage RNA-seq data were used for de novo annotation *Capsaspora* lincRNAs as detailed in the [Supplemental Experimental Procedures](#).

## ACCESSION NUMBERS

The accession number for the mass spectrometry proteomics data reported in this paper has been uploaded to PRIDE repository: PXD002342. The accession number for the ChIP-seq and ATAC-seq data reported in this paper has been uploaded to GEO: GSE71131.

## SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures, seven figures, one table, and three data files and can be found with this article online at <http://dx.doi.org/10.1016/j.cell.2016.03.034>.

## AUTHOR CONTRIBUTIONS

A.S.-P., C.B., and H.P.-A. performed the experiments. A.S.-P., C.C., and E.S. planned, acquired, and analyzed the proteomics data. J.J.T. was involved in ATAC-seq data analysis. A.S.-P., J.L.G.-S., I.R.-T., and L.D. were involved in the study design. A.S.-P. analyzed the data and generated the figures. A.S.-P. and I.R.-T. wrote the paper. All authors discussed the results and commented on the manuscript.

## ACKNOWLEDGMENTS

We thank Alex de Mendoza, Xavier Grau-Bové, Ignacio Maeso, and Manuel Irimia for comments on the manuscript and figures. This work was supported by an Institutió Catalana de Recerca i Estudis Avançats contract, a European Research Council Consolidator Grant (ERC-2012-Co-616960), and a grant from Ministerio de Economía y Competitividad (MINECO) (BFU-2011-23434) (to I.R.-T.). We also acknowledge financial support from Secretaria d'Universitats i Recerca del Departament d'Economia i Coneixement de la Generalitat de Catalunya (project 2014 SGR 619). The work in L.D.'s laboratory was supported by grants from the Spanish "Ministerio de Educación y Ciencia" (SAF2013-48926-P), AGAUR, and the European Commission's 7th Framework Program 4DCellFate (277899). A.S.-P. is supported by an EMBO Long-Term Fellowship (ALTF 841-2014). J.L.G.-S. was funded by grants from Ministerio de Economía y Competitividad (BFU2013-41322-P) and the Andalusian Government (BIO-396). J.J.T. has a postdoctoral grant from the University Pablo de Olavide. The CRG/UPF Proteomics Unit is part of the "Plataforma de Recursos Biomoleculares y Bioinformáticos (ProteoRed)" supported by a grant from Instituto de Salud Carlos III (ISCIII) (PT13/0001). We thank Guadalupe Espadas for her support with the histone derivatization protocol and Núria Ros and Meritxell Antó for technical support. Finally, we thank the CRG Genomics Unit for helping with ChIP-seq and RNA-seq sequencing.

Received: July 21, 2015

Revised: February 3, 2016

Accepted: March 18, 2016

Published: April 21, 2016

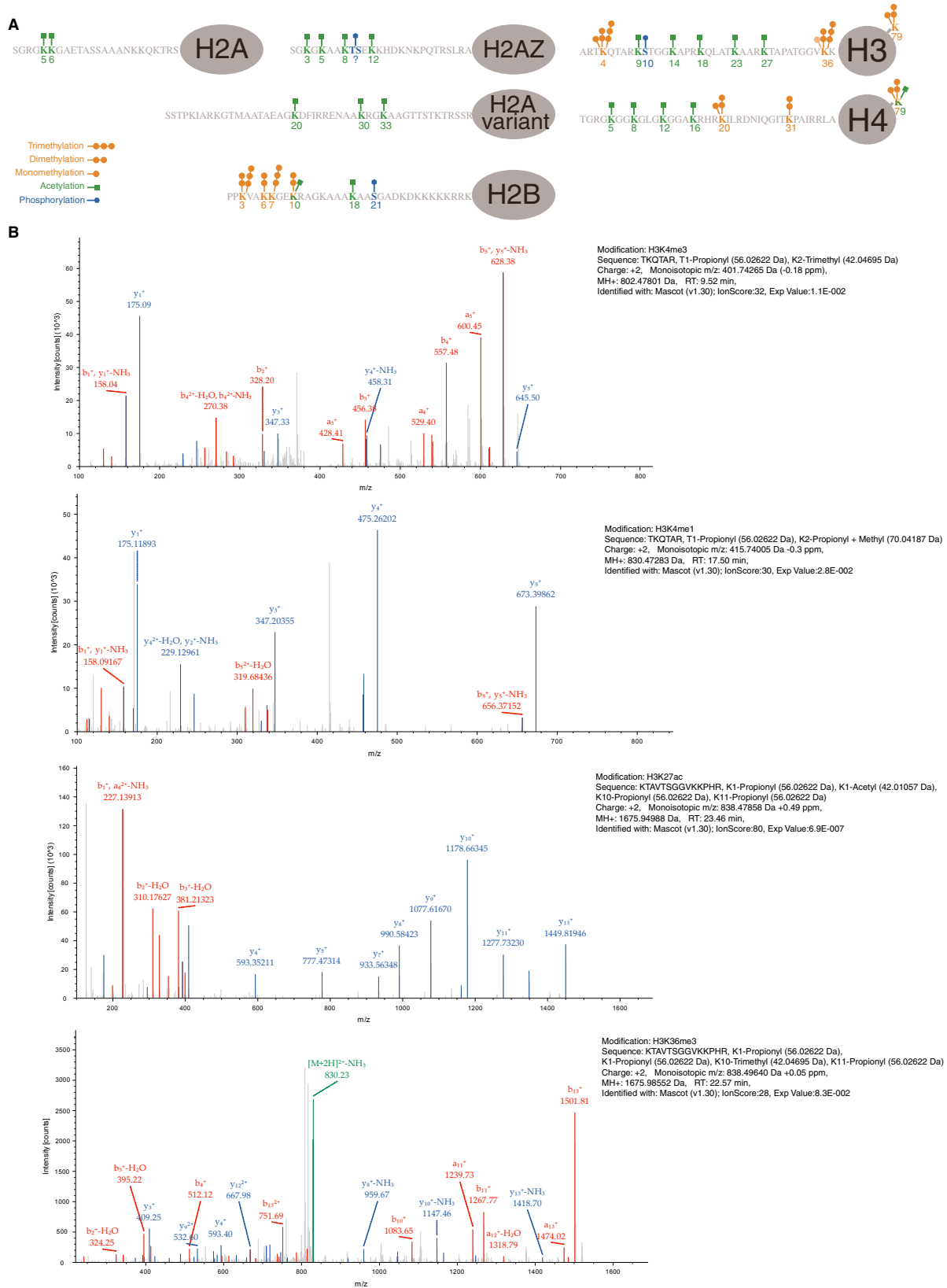
## REFERENCES

- Andersson, R., Gebhard, C., Miguel-Escalada, I., Hoof, I., Bornholdt, J., Boyd, M., Chen, Y., Zhao, X., Schmidl, C., Suzuki, T., et al.; FANTOM Consortium (2014). An atlas of active enhancers across human cell types and tissues. *Nature* **507**, 455–461.
- Bernstein, B.E., Meissner, A., and Lander, E.S. (2007). The mammalian epigenome. *Cell* **128**, 669–681.
- Blanco, E., Messegueur, X., Smith, T.F., and Guigó, R. (2006). Transcription factor map alignment of promoter regions. *PLoS Comput. Biol.* **2**, e49.
- Buecker, C., and Wysocka, J. (2012). Enhancers as information integration hubs in development: lessons from genomics. *Trends Genet.* **28**, 276–284.
- Buenrostro, J.D., Giresi, P.G., Zaba, L.C., Chang, H.Y., and Greenleaf, W.J. (2013). Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat. Methods* **10**, 1213–1218.
- Bulger, M., and Groudine, M. (2011). Functional and mechanistic diversity of distal transcription enhancers. *Cell* **144**, 327–339.
- Chen, K., Xi, Y., Pan, X., Li, Z., Kaestner, K., Tyler, J., Dent, S., He, X., and Li, W. (2013). DANPOS: dynamic analysis of nucleosome position and occupancy by sequencing. *Genome Res.* **23**, 341–351.
- Creyghton, M.P., Cheng, A.W., Welstead, G.G., Kooistra, T., Carey, B.W., Steine, E.J., Hanna, J., Lodato, M.A., Frampton, G.M., Sharp, P.A., et al. (2010). Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proc. Natl. Acad. Sci. USA* **107**, 21931–21936.
- Davidson, E.H., and Erwin, D.H. (2006). Gene regulatory networks and the evolution of animal body plans. *Science* **311**, 796–800.



- de Laat, W., and Duboule, D. (2013). Topology of mammalian developmental enhancers and their regulatory landscapes. *Nature* 502, 499–506.
- de Mendoza, A., Sebé-Pedrós, A., Šestak, M.S., Matejčić, M., Torruella, G., Domazet-Lošo, T., and Ruiz-Trillo, I. (2013). Transcription factor evolution in eukaryotes and the assembly of the regulatory toolkit in multicellular lineages. *Proc. Natl. Acad. Sci. USA* 110, E4858–E4866.
- de Mendoza, A., Suga, H., Permanyer, J., Irimia, M., and Ruiz-Trillo, I. (2015). Complex transcriptional regulation and independent evolution of fungal-like traits in a relative of animals. *eLife* 4, e08904.
- Duan, Z., Andronescu, M., Schutz, K., McIlwain, S., Kim, Y.J., Lee, C., Shendure, J., Fields, S., Blau, C.A., and Noble, W.S. (2010). A three-dimensional model of the yeast genome. *Nature* 465, 363–367.
- Dunham, I., Kundaje, A., Aldred, S.F., Collins, P.J., Davis, C., Doyle, F., Epstein, C.B., Fritze, S., Harrow, J., Kaul, R., et al.; ENCODE Project Consortium (2012). An integrated encyclopedia of DNA elements in the human genome. *Nature* 489, 57–74.
- Egloff, S., Dienstbier, M., and Murphy, S. (2012). Updating the RNA polymerase CTD code: adding gene-specific layers. *Trends Genet.* 28, 333–341.
- Eick, D., and Geyer, M. (2013). The RNA polymerase II carboxy-terminal domain (CTD) code. *Chem. Rev.* 113, 8456–8490.
- Ernst, J., and Kellis, M. (2012). ChromHMM: automating chromatin-state discovery and characterization. *Nat. Methods* 9, 215–216.
- Erwin, D.H. (2009). Early origin of the bilaterian developmental toolkit. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 364, 2253–2261.
- Erwin, D.H., and Davidson, E.H. (2009). The evolution of hierarchical gene regulatory networks. *Nat. Rev. Genet.* 10, 141–148.
- Fairclough, S.R., Chen, Z., Kramer, E., Zeng, Q., Young, S., Robertson, H.M., Begovic, E., Richter, D.J., Russ, C., Westbrook, M.J., et al. (2013). Premetazoan genome evolution and the regulation of cell differentiation in the choanoflagellate *Salpingoeca rosetta*. *Genome Biol.* 14, R15.
- Gaiti, F., Fernandez-Valverde, S.L., Nakanishi, N., Calcino, A.D., Yanai, I., Taurdzic, M., and Degnan, B.M. (2015). Dynamic and widespread lncRNA expression in a sponge and the origin of animal complexity. *Mol. Biol. Evol.* 32, 2367–2382.
- Garcia, B.A., Mollah, S., Ueberheide, B.M., Busby, S.A., Muratore, T.L., Shabanowitz, J., and Hunt, D.F. (2007). Chemical derivatization of histones for facilitated analysis by mass spectrometry. *Nat. Protoc.* 2, 933–938.
- Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y.C., Laslo, P., Cheng, J.X., Murre, C., Singh, H., and Glass, C.K. (2010). Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol. Cell* 38, 576–589.
- Hezroni, H., Koppstein, D., Schwartz, M.G., Avrutin, A., Bartel, D.P., and Ulitsky, I. (2015). Principles of long noncoding RNA evolution derived from direct comparison of transcriptomes in 17 species. *Cell Rep.* 11, 1110–1122.
- Ho, J.W.K., Jung, Y.L., Liu, T., Alver, B.H., Lee, S., Ikegami, K., Sohn, K.-A., Minoda, A., Tolstorukov, M.Y., Appert, A., et al. (2014). Comparative analysis of metazoan chromatin organization. *Nature* 512, 449–452.
- Irimia, M., Tena, J.J., Alexis, M.S., Fernandez-Miñan, A., Maeso, I., Bogdanovic, O., de la Calle-Mustienes, E., Roy, S.W., Gómez-Skarmeta, J.L., and Fraser, H.B. (2012). Extensive conservation of ancient microsynteny across metazoans due to cis-regulatory constraints. *Genome Res.* 22, 2356–2367.
- Irimia, M., Maeso, I., Roy, S.W., and Fraser, H.B. (2013). Ancient cis-regulatory constraints and the evolution of genome architecture. *Trends Genet.* 29, 521–528.
- Kapusta, A., and Feschotte, C. (2014). Volatile evolution of long noncoding RNA repertoires: mechanisms and biological implications. *Trends Genet.* 30, 439–452.
- King, N., Westbrook, M.J., Young, S.L., Kuo, A., Abedin, M., Chapman, J., Fairclough, S., Hellsten, U., Isogai, Y., Letunic, I., et al. (2008). The genome of the choanoflagellate *Monosiga brevicollis* and the origin of metazoans. *Nature* 451, 783–788.
- Kolasinska-Zwierz, P., Down, T., Latorre, I., Liu, T., Liu, X.S., and Ahinger, J. (2009). Differential chromatin marking of introns and expressed exons by H3K36me3. *Nat. Genet.* 41, 376–381.
- Langmead, B., Trapnell, C., Pop, M., and Salzberg, S.L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* 10, R25.
- Lenhard, B., Sandelin, A., and Carninci, P. (2012). Metazoan promoters: emerging characteristics and insights into transcriptional regulation. *Nat. Rev. Genet.* 13, 233–245.
- Levine, M. (2010). Transcriptional enhancers in animal development and evolution. *Curr. Biol.* 20, R754–R763.
- Levine, M., and Tjian, R. (2003). Transcription regulation and animal diversity. *Nature* 424, 147–151.
- Li, H., and Johnson, A.D. (2010). Evolution of transcription networks—lessons from yeasts. *Curr. Biol.* 20, R746–R753.
- Lolas, M., Valenzuela, P.D.T., Tjian, R., and Liu, Z. (2014). Charting Brachyury-mediated developmental pathways during early mouse embryogenesis. *Proc. Natl. Acad. Sci. USA* 111, 4478–4483.
- Margueron, R., and Reinberg, D. (2011). The Polycomb complex PRC2 and its mark in life. *Nature* 469, 343–349.
- Marques, A.C., and Ponting, C.P. (2014). Intergenic lncRNAs and the evolution of gene expression. *Curr. Opin. Genet. Dev.* 27, 48–53.
- Marques, A.C., Hughes, J., Graham, B., Kowalczyk, M.S., Higgs, D.R., and Ponting, C.P. (2013). Chromatin signatures at transcriptional start sites separate two equally populated yet distinct classes of intergenic long noncoding RNAs. *Genome Biol.* 14, R131.
- Peter, I.S., and Davidson, E.H. (2011). Evolution of gene regulatory networks controlling body plan development. *Cell* 144, 970–985.
- Rada-Iglesias, A., Bajpai, R., Swigut, T., Brugmann, S.A., Flynn, R.A., and Wysocka, J. (2011). A unique chromatin signature uncovers early developmental enhancers in humans. *Nature* 470, 279–283.
- Reinke, A.W., Baek, J., Ashenberg, O., and Keating, A.E. (2013). Networks of bZIP protein-protein interactions diversified over a billion years of evolution. *Science* 340, 730–734.
- Schwaiger, M., Schönauer, A., Rendeiro, A.F., Pribitzer, C., Schauer, A., Gilles, A.F., Schinko, J.B., Renfer, E., Fredman, D., and Technau, U. (2014). Evolutionary conservation of the eumetazoan gene regulatory landscape. *Genome Res.* 24, 639–650.
- Schwer, B., and Shuman, S. (2011). Deciphering the RNA polymerase II CTD code in fission yeast. *Mol. Cell* 43, 311–318.
- Sebé-Pedrós, A., and de Mendoza, A. (2015). Transcription factors and the origin of animal multicellularity. In *Evolutionary Transitions to Multicellular Life*, I. Ruiz-Trillo and A.M. Nedelcu, eds. (Springer), pp. 379–394.
- Sebé-Pedrós, A., de Mendoza, A., Lang, B.F., Degnan, B.M., and Ruiz-Trillo, I. (2011). Unexpected repertoire of metazoan transcription factors in the unicellular holozoan *Capsaspora owczarzaki*. *Mol. Biol. Evol.* 28, 1241–1254.
- Sebé-Pedrós, A., Ariza-Cosano, A., Weirauch, M.T., Leininger, S., Yang, A., Torruella, G., Adamski, M., Adamska, M., Hughes, T.R., Gómez-Skarmeta, J.L., and Ruiz-Trillo, I. (2013a). Early evolution of the T-box transcription factor family. *Proc. Natl. Acad. Sci. USA* 110, 16050–16055.
- Sebé-Pedrós, A., Irimia, M., Del Campo, J., Parra-Acero, H., Russ, C., Nusbaum, C., Blencowe, B.J., and Ruiz-Trillo, I. (2013b). Regulated aggregative multicellularity in a close unicellular relative of metazoa. *eLife* 2, e01287–e01287.
- Simola, D.F., Graham, R.J., Brady, C.M., Enzmann, B.L., Desplan, C., Ray, A., Zwiebel, L.J., Bonasio, R., Reinberg, D., Liebig, J., and Berger, S.L. (2016). Epigenetic (re)programming of caste-specific behavior in the ant *Camponotus floridanus*. *Science* 351, aac6633.
- Sorrells, T.R., and Johnson, A.D. (2015). Making sense of transcription networks. *Cell* 161, 714–723.
- Srivastava, M., Simakov, O., Chapman, J., Fahey, B., Gauthier, M.E., Mitros, T., Richards, G.S., Conaco, C., Dacre, M., Hellsten, U., et al. (2010). The

- Amphimedon queenslandica genome and the evolution of animal complexity. *Nature* 466, 720–726.
- Suga, H., Chen, Z., de Mendoza, A., Sebé-Pedrós, A., Brown, M.W., Kramer, E., Carr, M., Kerner, P., Vervoort, M., Sánchez-Pons, N., et al. (2013). The *Capsaspora* genome reveals a complex unicellular prehistory of animals. *Nat. Commun.* 4, 2325.
- Tanay, A., and Cavalli, G. (2013). Chromosomal domains: epigenetic contexts and functional implications of genomic compartmentalization. *Curr. Opin. Genet. Dev.* 23, 197–203.
- Tanizawa, H., Iwasaki, O., Tanaka, A., Capizzi, J.R., Wickramasinghe, P., Lee, M., Fu, Z., and Noma, K. (2010). Mapping of long-range associations throughout the fission yeast genome reveals global genome organization linked to transcriptional regulation. *Nucleic Acids Res.* 38, 8164–8177.
- Thurman, R.E., Rynes, E., Humbert, R., Vierstra, J., Maurano, M.T., Haugen, E., Sheffield, N.C., Stergachis, A.B., Wang, H., Vernot, B., et al. (2012). The accessible chromatin landscape of the human genome. *Nature* 489, 75–82.
- Ulitsky, I., and Bartel, D.P. (2013). lincRNAs: genomics, evolution, and mechanisms. *Cell* 154, 26–46.
- van Galen, P., Viny, A.D., Ram, O., Ryan, R.J., Cotton, M.J., Donohue, L., Sievers, C., Drier, Y., Liao, B.B., Gillespie, S.M., et al. (2016). A multiplexed system for quantitative comparisons of chromatin landscapes. *Mol. Cell* 61, 170–180.
- van Riggelen, J., Yetil, A., and Felsher, D.W. (2010). MYC as a regulator of ribosome biogenesis and protein synthesis. *Nat. Rev. Cancer* 10, 301–309.
- Villar, D., Flicek, P., and Odom, D.T. (2014). Evolution of transcription factor binding in metazoans - mechanisms and functional implications. *Nat. Rev. Genet.* 15, 221–233.
- Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nusbaum, C., Myers, R.M., Brown, M., Li, W., and Liu, X.S. (2008). Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* 9, R137.
- Zhou, V.W., Goren, A., and Bernstein, B.E. (2011). Charting histone modifications and the functional organization of mammalian genomes. *Nat. Rev. Genet.* 12, 7–18.



(legend on next page)

---

**Figure S1. Identification of Histone Modifications in *Capsaspora*, Related to Figure 1**

(A) Histone N-terminal tail sequences of *Capsaspora* with all identified post-translational modifications and their location. A quotation mark indicates the impossibility of reliably assigning a modification to one or another of a pair of neighboring residues.

(B) Representative MSMS analysis of modified peptides from H3 (from top to bottom): K4me3 (TK(me3)QTAR); K4me1 (TK(me)QTAR); K27ac (K(ac)TAVTSGGVKKPHR); K36me3 (KTAVTSGGVK(me3)KPHR). The b- and y-ion series are represented in red and blue, respectively. Non-fragmented precursor peptides are shown in green.



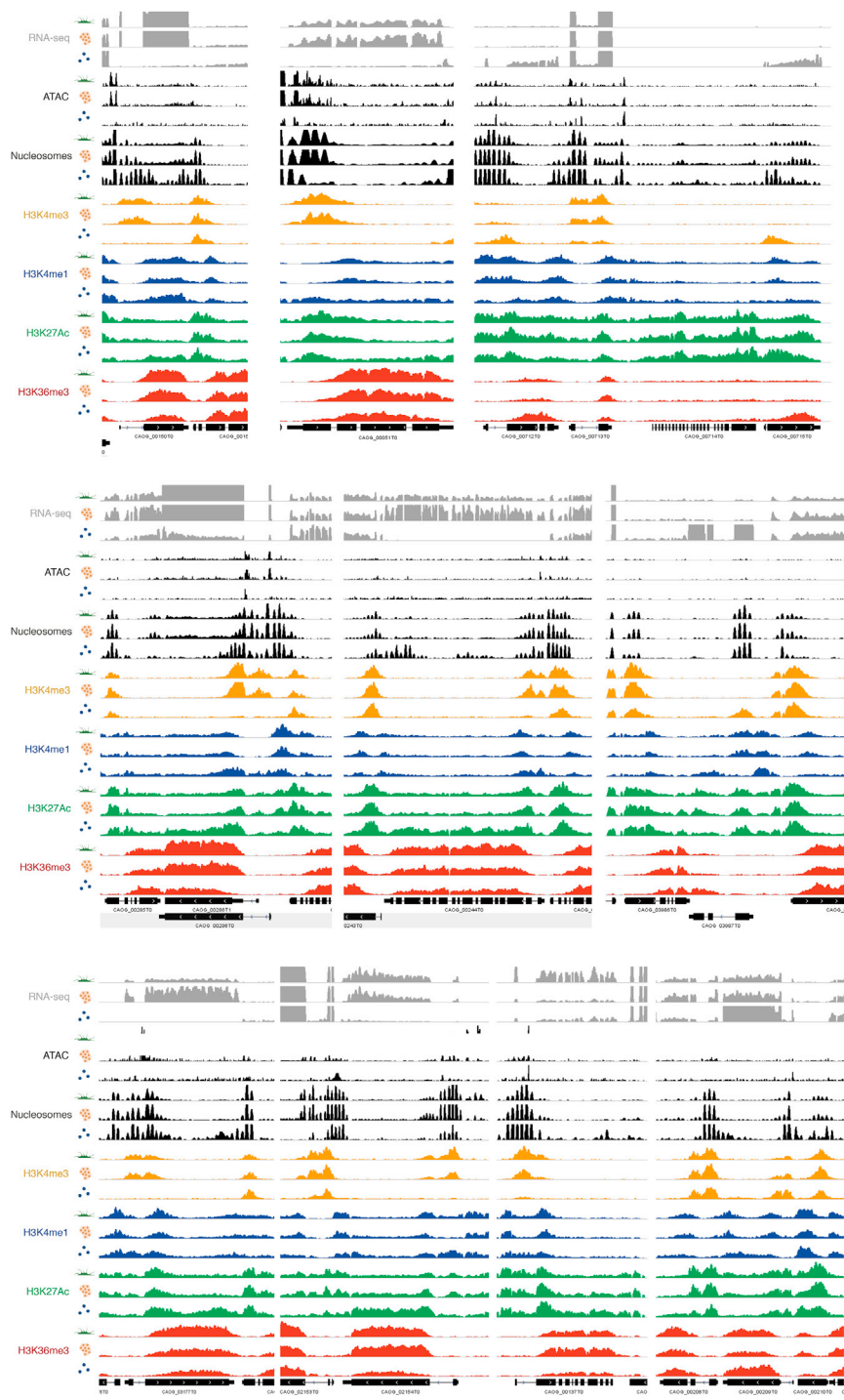
---

**Figure S2. RNAPolIII ChIP-Seq Experiments, Related to Figure 2**

(A) TSS-centered average normalized read coverage plots for RNAPolIII in the filopodial stage, using three different antibodies: 8WG16 (which preferentially recognizes unphosphorylated RNAPolIII), CTD4H8 (which recognizes both phospho- and unphosphorylated RNAPolIII) and S2P (which recognizes S2P-CTD phosphorylated RNAPolIII, the form associated to transcriptional elongation). The x axis spans  $-5\text{Kb}$  to  $+5\text{Kb}$  around the TSS. Shaded gray area represents the average size of *Capsaspora* genes.

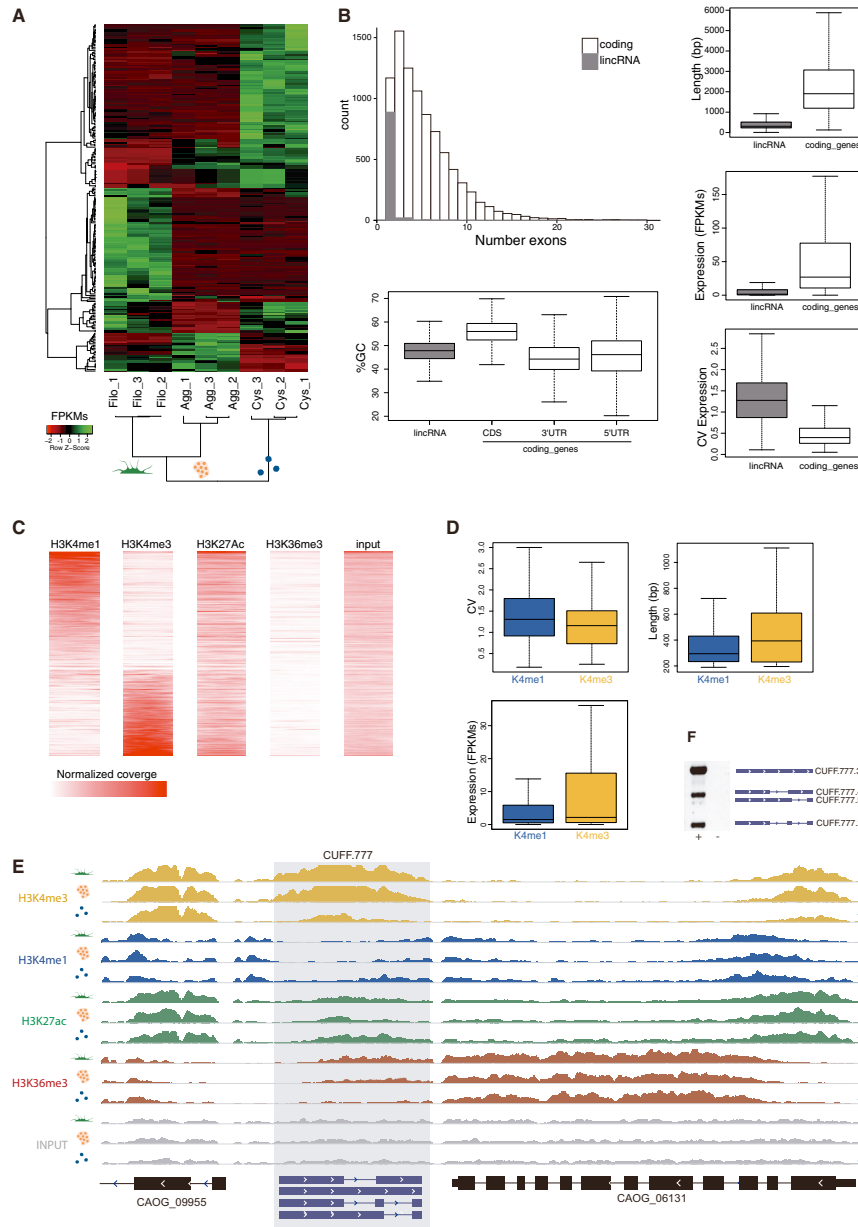
(B) Scatterplots of RNAPolIII coverage ( $\log_2$  normalized reads) compared to mRNA expression levels ( $\log_2$  FPKMs) in the filopodial stage.

(C) Illustrative examples of RNAPolIII dynamic changes. Different genomic windows showing normalized coverage for different chromatin features. For each feature, the top track corresponds to filopodial stage, middle track to aggregative stage and bottom track to cystic stage.



**Figure S3. Additional Illustrative Examples of Dynamic Chromatin Modifications in *Capsaspora*, Related to Figure 3**

Different genomic windows showing normalized coverage for different chromatin features and their dynamic association with gene expression. For each feature, the top track corresponds to filopodial stage, middle track to aggregative stage and bottom track to cystic stage.



**Figure S4. *Capsaspora* lincRNA Populations Defined by Chromatin Marks, Related to Figure 2**

(A) Heatmap showing clustered lincRNA expression (RPKM) across replicates of each stage. Only significantly differentially expressed lincRNAs (DESeq FDR < 0.05) are represented.

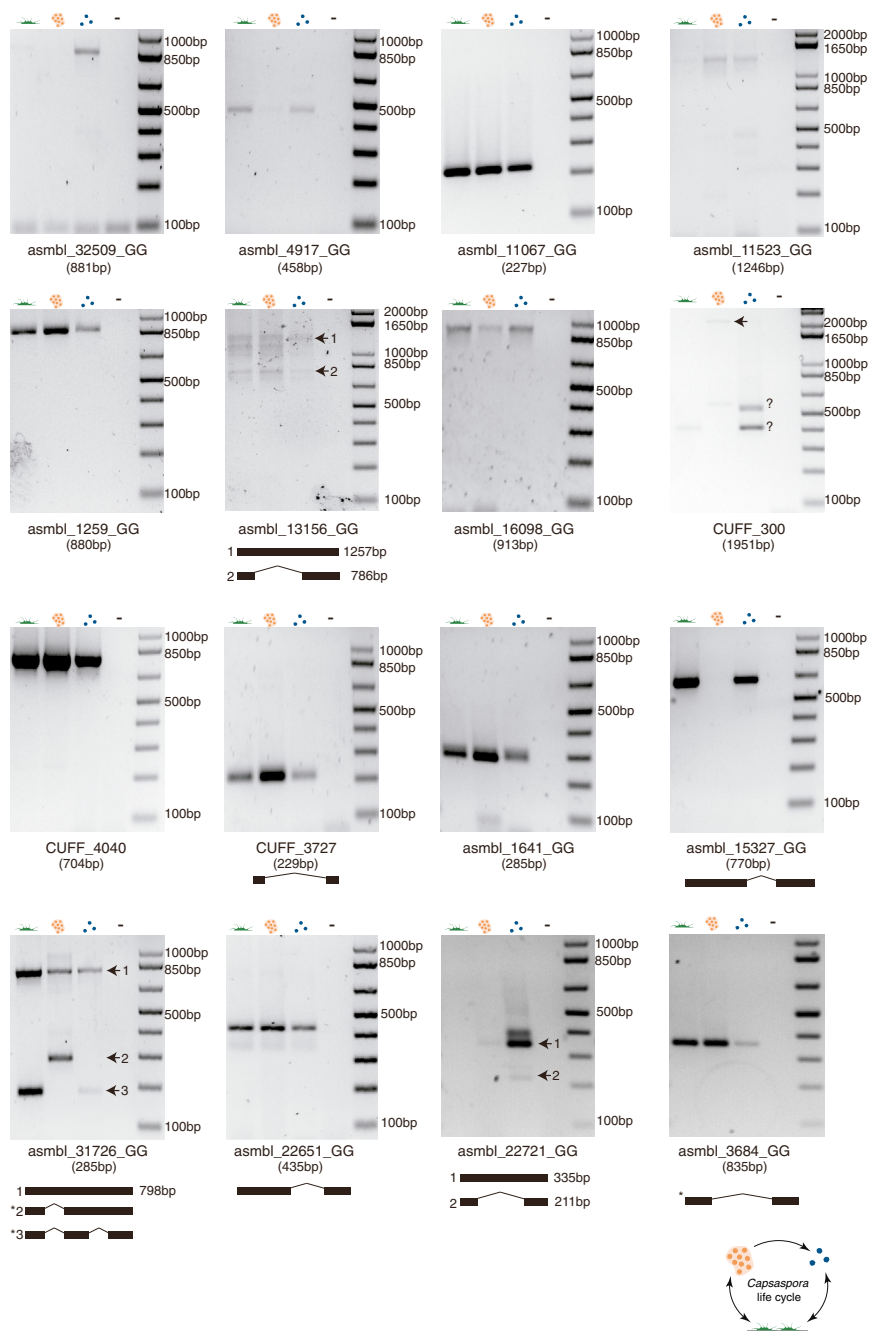
(B) Characteristics of lincRNA loci compared with coding protein genes, including exon number distribution (top left), GC content (bottom left), length (top right), level of expression (middle right) and coefficient of variation in expression between stages and replicates (bottom right).

(C) Heatmaps showing average read normalized coverage of four different histone modifications along lincRNA loci.

(D) Characteristics of H3K4me1 versus H3K4me3 marked lincRNA loci.

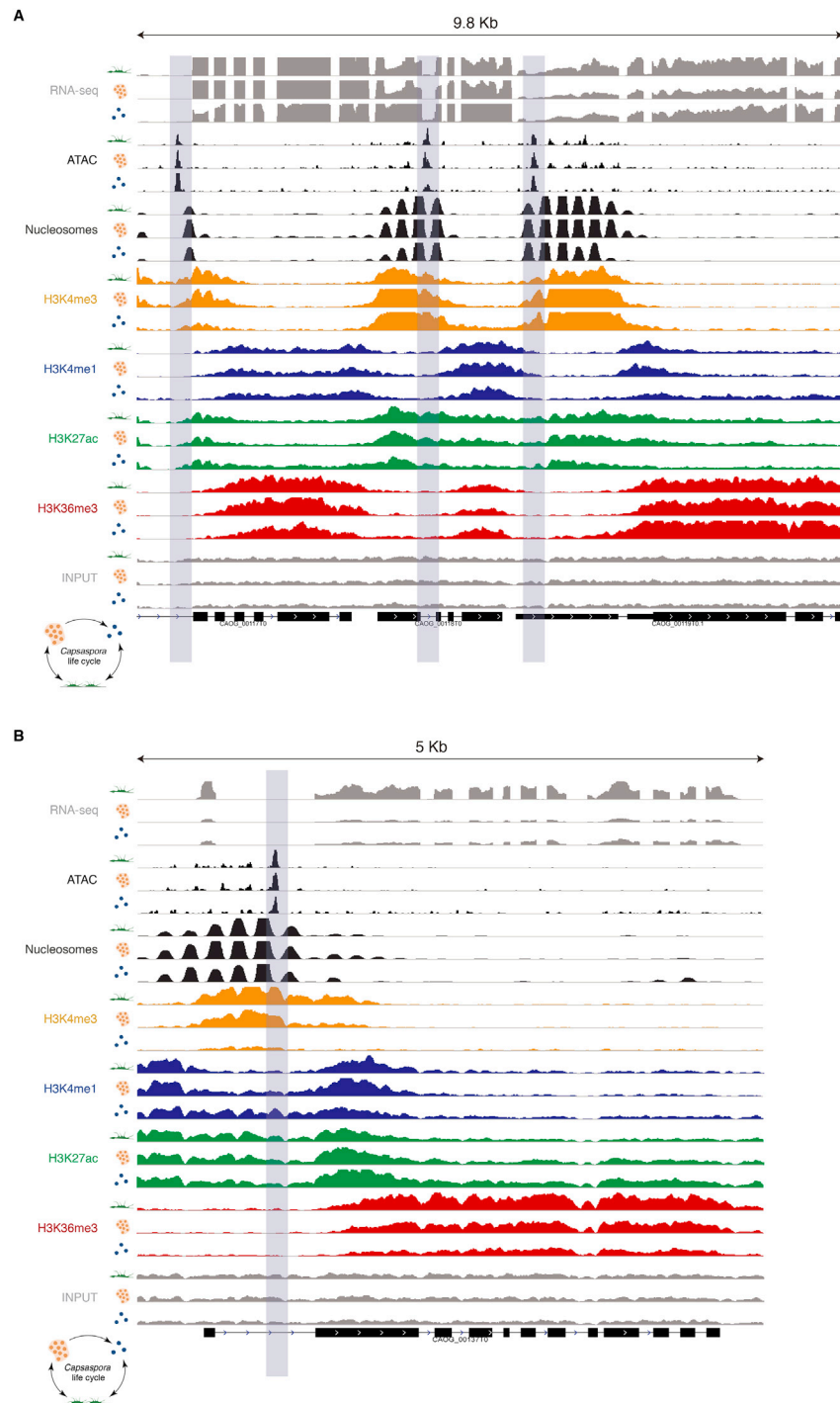
(E) Illustration of the genomic location of a lincRNA locus and normalized read coverage of histone modifications ChIP-seq. (F) RT-PCR validation of CUFF.777 lincRNA, revealing the existence of 3 isoforms. The minus sign indicates the negative control performed using RNA without reverse transcription to check for genomic DNA contamination.





**Figure S5. RT-PCR Validation of lincRNAs, Related to Figure S6**

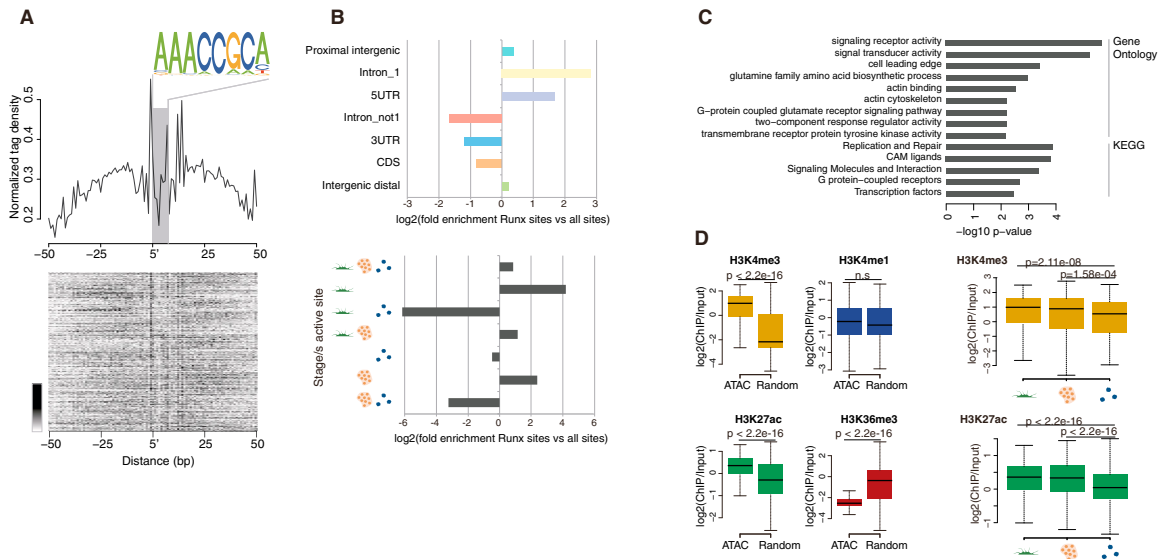
Each panel shows the result of a PCR over poly-A-selected cDNA for each stage, as well as a control sample for genomic contamination (a cross-stages pool of the original non-retrotranscribed RNAs). The name of the lincRNA locus and the predicted size is indicated below each image. Cases of splicing-event validation are indicated schematically. An asterisk indicates that the observed sizes are smaller than predicted.



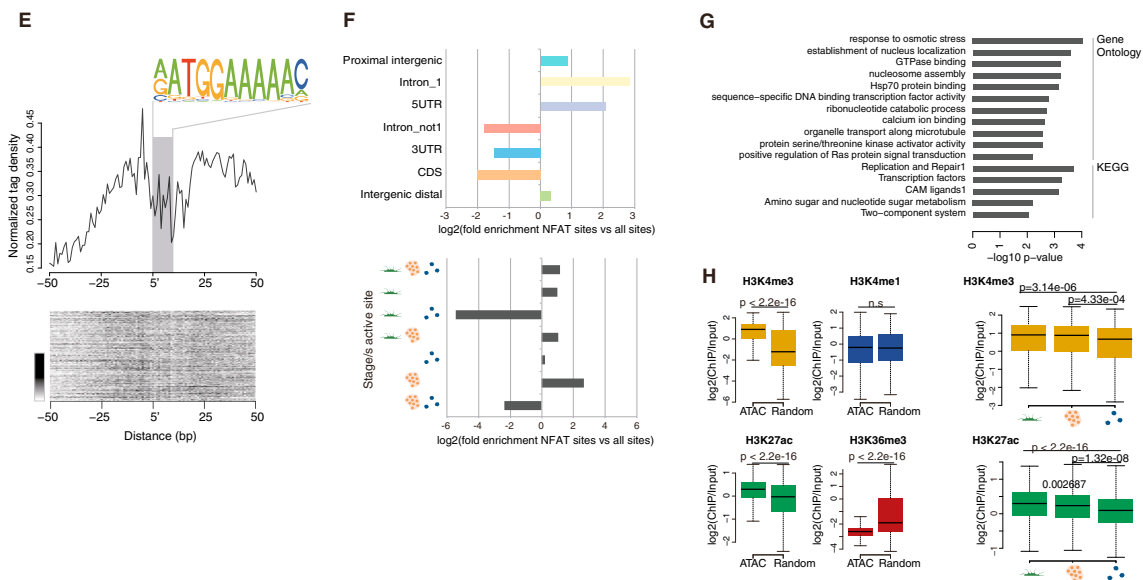
**Figure S6. Examples of ATAC Profiling of Regulatory Sites, Related to Figure 5**

(A and B) Different genomic windows (size indicated above) showing normalized coverage for different chromatin features and their dynamic association with gene expression. For each feature, the top track corresponds to filopodial stage, middle track to aggregative stage and bottom track to cystic stage. Significant peaks of ATAC nucleosome-free reads are highlighted.

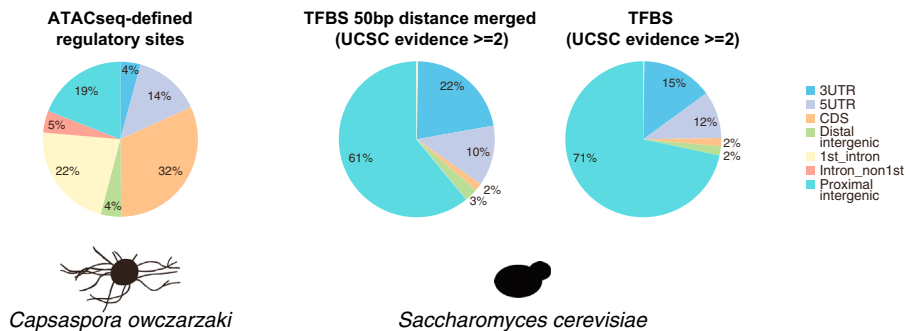
### Runx



### NFkappaB/NFAT



I



(legend on next page)

---

**Figure S7. *Capsaspora Runx* and *NFAT/NFkappaB* Regulatory Networks, Related to Figures 6 and 7**

- (A) Plot of ATAC-seq nucleosome-free average signal density around *Runx* motifs (top) and heatmap of signal around individual sites (bottom).
- (B) Differential distribution of regulatory sites containing *Runx* motif compared with all sites according to genomic feature (top) and stage/s where the site is active (bottom).
- (C) Enriched GO terms and KEGG pathways among genes associated with *Runx* regulatory sites.
- (D) Enrichment of different histone modifications (ChIP versus input) at *Runx* motifs in ATAC-defined sites compared with motifs occurring randomly in the genome (left) and at *Runx* sites across stages (right).
- (E) Plot of ATAC-seq nucleosome-free average signal density around *NFAT/NFkappaB* motifs (top) and heatmap of signal around individual sites (bottom).
- (F) Differential distribution of regulatory sites containing *NFAT/NFkappaB* motif compared with all sites according to genomic feature (top) and stage/s where the site is active (bottom).
- (G) Enriched GO terms and KEGG pathways among genes associated with *NFAT/NFkappaB* regulatory sites.
- (H) Enrichment of different histone modifications (ChIP versus input) at *NFAT/NFkappaB* motifs in ATAC-defined sites compare with motifs occurring randomly in the genome (left) and at *NFAT/NFkappaB* sites across stages (right).
- (I) Regulatory Site Distribution in *Capsaspora* and yeast. Genomic feature distribution of ATAC-defined regulatory sites in *Capsaspora* compared with the distribution of transcription factor binding sites (TFBS) extracted from UCSC (with evidence support > 2). In the left pie chart, TFBS at < 50bp of distance were merged.

Cell, Volume 165

## Supplemental Information

### The Dynamic Regulatory Genome of *Capsaspora* and the Origin of Animal Multicellularity

Arnau Sebé-Pedrós, Cecilia Ballaré, Helena Parra-Acero, Cristina Chiva, Juan J. Tena, Eduard Sabidó, José Luis Gómez-Skarmeta, Luciano Di Croce, and Iñaki Ruiz-Trillo

## Supplemental Experimental Procedures

### Cell Culture and Differentiation Conditions

*Capsaspora* strain ATCC30864 cells were grown axenically in 5 ml flasks with ATCC medium 1034 (modified PYNFH medium) in an incubator at 23°C (Sebé-Pedrós et al., 2013a). Filopodial stage cells were obtained by initiating a new 1/100 sub-culture (from an approximately  $5 \times 10^6$  cells/ml initial culture) and, after 3–4 days, cells were harvested. Aggregative stage was induced by initiating a new 1/250 sub-culture (from an approximately  $5 \times 10^6$  cells/ml initial culture) and by gentle agitation at 60 rpm during 4–5 days. Finally, cystic stage cells were obtained from a 14-day-old culture, starting from a new 1/100 sub-culture (from an approximately  $5 \times 10^6$  cells/ml initial culture).

### Histone Extraction and Analysis of Histone Modifications.

*Capsaspora owczarzaki* cells were centrifuged at 8000g for 10 min and washed once with cold PBS. The pellet was resuspended in 1 ml of lysis buffer (10 mM TrisHCl pH 6.5, 50 mM Natrium-Bisulfite, 1% Triton X-100, 10 mM MgCl<sub>2</sub>, 8.6% Sucrose, 10 mM Na-Butyrate, plus protease and phosphatase inhibitors and 0.5 mM DTT). The cell lysate was centrifuged at 14000g for 15 seconds and the supernatant was discarded. This process was repeated twice. Next, the pellet was washed once in 1 ml wash buffer (10mM TrisHCl pH 7.4, 13mM EDTA, 10mM Na-Butyrate, plus protease and phosphatase inhibitors), resuspended in 150µl of H<sub>2</sub>SO<sub>4</sub> 0.4N and incubated 1h at 4°C. After centrifugation at 14000g for 5 min, the supernatant was recovered and histones precipitated ON at -20°C, adding 1350 µl of cold acetone. The mixture was centrifuged at 14000g for 10 min and the histone pellet air dried for 10 min, before suspending it in 50 µl of water.

Histones were quantified by the BCA (Bicinchoninic acid assay) method and 10mg of each sample were derivatized with propionic anhydride, digested with trypsin and derivatized again with propionic anhydride as described before (Garcia et al., 2007b). Briefly, samples were dissolved in 30 mL of 100 mM ammonium bicarbonate and 0.5 mL of ammonium hydroxide was added to bring the pH between 7 and 9. The propionic anhydride was prepared by adding 25 mL of propionic anhydride to 75 mL of anhydrous isopropanol. 15 ml of propionic anhydride was added to the samples and immediately 8 ml of ammonium hydroxide were added to maintain the pH at around 8 and samples were incubated at 37°C for 15 minutes. Samples were vacuum dried and the propionylation procedure repeated. Dried samples were dissolved in 100 mL of 100 mM ammonium bicarbonate and digested 6h at 37°C with 0.5 mg of trypsin. The digestion was quenched adding 10 mL of glacial acetic acid, vacuum centrifuged and pH adjusted again between 7 and 9. The propionylation procedure was repeated twice. Samples

were vacuum dried and desalted with C18 ultramicrospin columns (The Nest Group Inc, Southborough, MA).

A 2- $\mu$ g aliquot of the peptide mixture was analyzed using a LTQ-Orbitrap Velos Pro mass spectrometer (Thermo Fisher Scientific, San Jose, CA) coupled to a nano-LC (Proxeon, Odense, Denmark) equipped with a reversed-phase chromatography 12 cm column with an inner diameter of 75  $\mu$ m, packed with 3  $\mu$ m C18 particles (Nikkyo Technos, Japan) with both collision induced dissociation (CID) and high energy collision dissociation (HCD) fragmentation. Chromatographic gradients started at 3% buffer B with a flow rate of 300 nL/min and gradually increased to 10% buffer B in 1 min and to 35% buffer B in 30 min. After each analysis, the column was washed for 10 min with 90% buffer B (Buffer A: 0.1% formic acid in water. Buffer B: 0.1% formic acid in acetonitrile). The mass spectrometer was operated in positive ionization mode with nanospray voltage set at 2.2 kV and source temperature at 300 °C. Ultramark 1621 for the FT mass analyzer was used for external calibration prior the analyses. The background polysiloxane ion signal at m/z 445.1200 was used as lock mass. The instrument was operated in data-dependent acquisition (DDA) mode, and full MS scans with 1 microscan at resolution of 60 000 were used over a mass range of m/z 350–2000 with detection in the Orbitrap. Auto gain control (AGC) was set to 1e6, dynamic exclusion was set at 30 s, and the charge-state filter disqualifying singly charged peptides for fragmentation was activated. Following each survey scan, the 20 (CID) or the 10 (HCD) most intense ions with multiple charged ions above a threshold ion count of 5000 (CID) or 10000 (HCD) were selected for fragmentation at normalized collision energy of 35%. Fragment ion spectra produced via CID were acquired in the linear ion trap and the produced via HCD were acquired in the Orbitrap, AGC was set to 1e4 (CID) or 4e4 (HCD) and isolation window of 2.0 m/z, activation time of 10 ms (CID) or 0.1ms (HCD), and maximum injection time of 100 ms were used. All data were acquired with Xcalibur software v2.2. Data Analysis. Acquired data were analyzed using the Proteome Discoverer software suite (v1.3.0.339, Thermo Fisher Scientific), and the Mascot search engine (v2.3, Matrix Science) (Perkins et al., 1999) was used for peptide identification. Data were searched against a *Capsaspora* protein database derived from the Broad Institute plus the most common contaminants (total of 9407 sequences). A precursor ion mass tolerance of 7 ppm at the MS1 level was used, and up to three missed cleavages for trypsin were allowed. The fragment ion mass tolerance was set to 0.5 Da (CID) or 20 mmu (HCD). Dimethyl lysine, trimethyl lysine, propionyl lysine and propionyl + methyl lysine were defined as variable modification. Propionylation on N-terminal was set as a fix modification. The identified peptides were filtered by mascot ion score higher than 20. The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository with the dataset identifier PXD002342.

Identified histone modifications were compared with known modifications (identified by similar methods) in homologous residues in other organisms, including *Homo sapiens* (Garcia et al., 2007a; Robin et al., 2007), *Saccharomyces cerevisiae* (Garcia et al., 2007a; Krebs, 2007; Millar et al., 2006), *Tetrahymena thermophila* (Bonenfant et al., 2006; Garcia et al., 2007a; Medzihradzky et al., 2004; Wei et al., 1998), *Plasmodium falciparum* (Nardelli et al., 2013; Trelle et al., 2009) and *Trypanosoma brucei* (Cross, 2008; Mandava et al., 2008).

### **Reannotation of *Capsaspora owczarzaki* Genome.**

Nine different RNA-seq experiments (Sebé-Pedros et al., 2013a), obtained over 2 lanes HiSeq 2000 instrument (Illumina, San Diego, CA, USA), were pooled, representing 197M 76-base strand-specific paired reads and a high depth of coverage (>1000x) (<http://www.ncbi.nlm.nih.gov/biosample/?term=txid595528%5BOrganism:noexp%5D>).

Genome re-annotation was performed using the PASA pipeline (Haas et al., 2003) as described here: <http://pasapipeline.github.io/>. In brief, a genome-guided *de novo* transcriptome assembly was generated using Trinity (Haas et al., 2013), with the Jaccard\_clip option. This assembly was used in the PASA pipeline, with default options, to perform an incremental annotation over *Capsaspora* v3 annotation ([https://www.broadinstitute.org/annotation/genome/multicellularity\\_project/download/?sp=EATranscriptsGtf&sp=SC\\_owczarzaki\\_V2&sp=S.zip](https://www.broadinstitute.org/annotation/genome/multicellularity_project/download/?sp=EATranscriptsGtf&sp=SC_owczarzaki_V2&sp=S.zip)). UTR annotation was significantly improved, both in terms of length and also number of genes with annotated UTRs (from 40% to 83.3% in the case of 5'UTRs and from 31.2% to 77.6% in the case of 3'UTRs). Consequently, this resulted in an increment of the accuracy of Transcription Start Site (TSS) and intergenic regions delimitation (Supplementary File 1).

### **Histone Deacetylase Inhibition Experiments**

For the life stage transition assay, 300 $\mu$ l of a 1/100 dilution of a *Capsaspora* culture in cystic stage were plated in 400 $\mu$ l of fresh ATCC medium 1034 in glass-bottom dish. Immediately, 3 $\mu$ M Trichostatin A (TSA, #T8552, Sigma-Aldrich, St. Louis, MO) and the equivalent volume of DMSO (negative control) were added to the cells. Cells were observed at 12h, 18h, and 24h in an inverted microscope with a 63X objective (Zeiss Axio Observer Z.1, Zeiss, Oberkochen, Germany).

For the histone acetylation assay, histones were extracted as described above (acid extraction method) from TSA (0.5  $\mu$ M and 3  $\mu$ M) and DMSO treated cells. 5  $\mu$ g of histones per lane were separated by SDS-PAGE and transferred to nitrocellulose membranes. Histones were probed with antibodies against total H3 (1:2000, #ab1791, Abcam, Cambridge, UK) and H3K27ac (1:1000, #07-360, Millipore, Darmstadt, Germany). Proteins were detected with HRP-



conjugated goat anti-rabbit IgG antibody (1:20000, #12-348, Millipore) and visualized with Supersignal WestPico chemiluminescent substrate (#34078, ThermoScientific, Rockford, IL). For the RNA-seq experiments, *Capsaspora* filopodial stage cells were cultured as described above and incubated with DMSO (negative control) and 3  $\mu$ M TSA during 24h in a 23°C incubator. Two replicate experiments were performed per each condition. Total RNA from each condition (and from two replicates of each condition) was extracted using Trizol reagent (#15596018, ThermoScientific). Four strand-specific libraries (one per sample) were sequenced over 1 lane of an Illumina HiSeq2000 instrument in the Genomics Unit at the Centre for Genomic Regulation (CRG). We obtained around 35M paired-end 50bp reads per sample. Reads were aligned to the reference genome using Tophat (Trapnell et al., 2012a) with default options. Transcript abundances were quantified using kallisto (Bray et al., 2015).

### **Capsaspora-Brachyury Antibody Production**

An epitope near the C-terminal region of *Capsaspora*-Brachyury was used as antigen for antibody production (QQPVSLVQSMQPGQPQSMQPIQQQPIQQQQPIQQQQQQQQQLGQYAAQTNVLPYGQP QMVDRRVFYEQQQPQLQQQQQLQLQPLQQQQQLQPLQR). The polypeptide was produced by the Biomolecular Screening and Protein Technologies Unit of the CRG (cloned in pETM44 with the His-MBP-tag, expressed in *E.coli*, cut the tag with PreScission Protease, purified by exclusion chromatography). The polyclonal antibody was produced in guinea-pig (one animal) by Rockland-TebuBio (conjugated with KLH and 3mg of antigen injected). The serum corresponding to the terminal bleed was affinity-purified against the antigen by the Biomolecular Screening and Protein Technologies Unit of the CRG. The antibody was validated by western blot and immunostaining.

For Western blot, we used *Capsaspora* nuclear protein extracts and recombinant *Capsaspora*-Brachyury protein. Nuclear proteins were extract from 5x10e9 cells, collected by centrifugation at 6,000g 10 min. This pellet was resuspended in 12ml Nuclear Extraction Buffer (10mM Hepes-KOH pH7.9, 1.5mM MgCl<sub>2</sub>, 10mM KCl, 0.2% NP-40, 0.5mM DTT, and protease and phosphatase inhibitors), incubated 10min on ice, and centrifuged at 5000g 5 min at 4°C. The pellet of nuclei was lysed for 10min on ice in 2ml Lysis Buffer (50mM Tris-HCl pH8.8, 300mM NaCl, 0.1% SDS, 5mM EDTA, 1mM EGTA, 1% NP-40, 1mM MgCl<sub>2</sub>, 1mM CaCl<sub>2</sub>, 1mM DTT, and protease and phosphatase inhibitors), then sonicated (3 pulses of 15 sec, 10% amplitude) using a Branson Digital Sonifier, and centrifuged 20,000g 20min at 4°C. *Capsaspora*-Brachyury protein was expressed with a His-tag in *E.coli* and purified by affinity chromatography (Nickel resin) in denaturing conditions by the Biomolecular Screening and Protein Technologies Unit of the CRG. 30  $\mu$ g of nuclear proteins and 3.5  $\mu$ g of *Capsaspora*-

Brachyury recombinant protein were separated by SDS-PAGE and transferred to nitrocellulose membranes. Proteins were probed with the anti-*Capsaspora*-Brachyury antibody (1 $\mu$ g/ml) and detected with HRP-conjugated goat anti-guinea pig IgG antibody (1:5000, #PA128679, ThermoScientific) and visualized with Supersignal WestPico chemiluminescent substrate (#34078, ThermoScientific).

For the immunostaining validation of the antibody, filopodial stage cells were grown on coverslips and fixed for 5 min with 6% acetone and for 5 min with 4% formaldehyde. The coverslips were washed gently four times with PBS1x, incubated for 30 min in blocking solution (1% BSA, 0.3% Triton X-100 in PBS1x), incubated overnight in primary antibody solution (1 $\mu$ g in 100 $\mu$  of blocking solution), and washed four times in blocking solution. After that, samples were incubated 1h in the dark with alexa-488 goat anti-guinea pig (1:1000, #A11073, ThermoScientific). and washed again four times (now with PBS1x). To visualize F-actin, samples were incubated for 15 min in the dark with with Phalloidin Texas Red (1:100, #T7471, ThermoScientific), washed twice with PBS, and incubated for 20 min with DAPI (1:100), to visualize the nucleus. After two final washes with PBS1x, coverslips were mounted onto slides with Fluorescent Mounting Media (4  $\mu$ L; Prolong Gold Antifade, #P36930, ThermoScientific). Images were taken with a 63x oil immersion objective on a Leica TCS SP5 confocal microscope (Leica-Microsystems, Vienna, Austria).

### **Chromatin Immunoprecipitation (ChIP) Assays**

Cells were crosslinked in 1% formaldehyde for 10 min at room temperature (RT). Crosslinking was quenched with 0.125 M glycine for 5 min RT. Pelleted cells were lysed in Lysis buffer I (10 mM HEPES.KOH pH 7.9, 1.5 mM MgCl<sub>2</sub>, 10 mM KCl, 0.2 % NP40, plus protease and phosphatase inhibitors and 0.5 mM DTT), incubated on ice for 10 minutes and centrifuged at 8000g 10 min to pellet the nuclei. Nuclei were resuspended in Lysis buffer II (1%SDS, 10 mM EDTA, 50 mM tris ClH pH 8.1 plus protease and phosphatase inhibitors), incubated on ice for 10 min and sonicated for 15 min (15 cycles, each one 30sec “on”, 30 sec “off”) in a Bioruptor (Diagenode, Seraing, Belgium) in order to generate 200bp fragments. Optimal sonication conditions were previously set up by testing a range of sonication cycles (from 3 to 24), determining that 15 cycles was the optimal.

An amount of chromatin equivalent to 40  $\mu$ g of DNA was used per ChIP. Antibodies used were: anti-H3K4me<sub>3</sub> (#pAb-003-050, Diagenode), anti-H3K4me<sub>1</sub> (#ab8895, Abcam, Cambridge, UK), anti-H3K36me<sub>3</sub> (#ab9050, Abcam), and anti-H3K27Ac (#07-360, Millipore, Darmstadt, Germany). Immunocomplexes were recovered with Protein A-Agarose Beads (Diagenode). Immunoprecipitated material was washed once with low salt washing buffer (0.1% SDS, 1%Triton x100, 2 mM EDTA, 20 mM Tris HCl pH8, 150 mM NaCl) and twice with high salt

buffer (0.1% SDS, 1% Triton x100, 2 mM EDTA, 20 mM Tris HCl pH8 , 500 mM NaCl). DNA complexes were eluted 30 min at 65°C (Elution buffer: 1% SDS, 0.1 M NaHCO<sub>3</sub>), decrosslinked ON at 65°C, treated with proteinase K and purified using MinElute PCR Purification Kit (Qiagen, Hilden, Germany). Libraries of immunoprecipitated and input DNA were prepared using the NEBNext DNA sample prep reagent Set 1 kit (New England Biolabs, Ipswich, MA, USA) according to the manufacturer's protocol. These libraries were sequenced over 1 lane of an Illumina HiSeq2000 instrument in the Genomics Unit at the CRG. We obtained between 10 and 20 M single-end 50bp reads per sample (Supplemental Table S1).

Bra-ChIP was performed as described above, with the only modifications of higher input chromatin (60  $\mu$ g) and High Salt Washing Buffer containing 250 mM NaCl instead of 500 mM. The immunoprecipitated DNA fragments were analyzed by quantitative PCR (ChIP-qPCR) using SYBR Green I PCR Master Mix (Roche) and the Roche LightCycler 480. Primer sequences are available upon request. The fold enrichment of the target sequence in the immunoprecipitated material compared with the input (% of input) was calculated using the comparative Ct method (Frank et al., 2001).

For RNAPolII ChIP-seq experiments, cells were crosslinked in 1% formaldehyde for 10 min at room temperature (RT). Crosslinking was quenched with 0.125 M glycine for 5 min RT. Pelleted cells were resuspended in Lysis/Sonication buffer (50 mM HEPES pH 7.9, 140 mM NaCl, 1mM EDTA, 1% Triton X-100, 0.1% Na-deoxycholate, 0.1% SDS plus protease and phosphatase inhibitors), incubated on ice for 10 minutes and sonicated for 14 min (14 cycles, each one 30sec “on”, 30 sec “off”) in a Bioruptor (Diagenode). ChIPs assays were performed as previously described (Stock et al., 2007) with some modifications. An amount of chromatin equivalent to 60  $\mu$ g of DNA was used per ChIP. Antibodies used were: anti-RNAPII [8WG16] (Abcam, ab817), anti-RNAPII [CTD4H8] (Biolegend/Covance, 904001), anti-RNAPII Phospho-S2 (Abcam, ab5095). Immunocomplexes were recovered with Dynabeads Protein G or A (Novex) for mouse or rabbit IgG antibodies respectively. Beads were then washed once with Lysis/Sonication buffer, once with Wash Buffer A (50 mM HEPES pH 7.9, 500 mM NaCl, 1mM EDTA, 1% Triton X-100, 0.1% Na-deoxycholate, 0.1% SDS), once with Wash Buffer B (20 mM Tris pH 8.0, 1 mM EDTA, 250 mM LiCl, 0.5% NP-40, 0.5% Na-deoxycholate) and twice with TE (10 mM Tris pH 8.0, 1 mM EDTA). DNA complexes were eluted 30 min at 65°C (Elution buffer: 1% SDS, 0.1 M NaHCO<sub>3</sub>), decrosslinked ON at 65°C, treated with proteinase K and purified using QIAgen PCR Purification Kit. The immunoprecipitated DNA was processed in the Ultrasequencing Unit at the CRG and subjected to deep sequencing using the Illumina HiSeq2000 sequencer.

## ATAC-seq

ATAC-seq was performed as originally described by Buenrostro et al. (2013) (Buenrostro et al., 2013). In brief, for each stage, 500,000 cells were collected and nuclei were obtained as described above. Nuclei were resuspended in 22.5  $\mu$ l of water, 25  $\mu$ l of 2x TD Buffer and 2.5  $\mu$ l of Tn5 Transposase from the Nextera DNA Library Prep Kit (Illumina, San Diego, CA) and incubated for 30 min at 37°C. Transposed DNA was purified using MinElute PCR Purification Kit (Qiagen) and immediately we performed 12 cycles of PCR amplification, using the following primers:

Forward: AATGATACGGCGACCACCGAGATCTACACTCGTCGGCAGCGTCAGATGTG

and

Reverse:

CAAGCAGAAGACGGCATAACGAGATNNNNNNGTCTCGTGGGCTCGGAGATGT (N

indicates barcode nucleotides). Amplified libraries were purified using MinElute PCR Purification Kit (Qiagen) and quantified using Qubit fluorimeter (Life Technologies, San Francisco, CA). The quality and profile of the libraries was analysed using Agilent 2100 Bioanalyzer (Agilent, Santa Clara, CA), in order to select only those with nucleosomal periodicity signal as an indicative of successful transposition reaction. We sequenced 2 replicates for the filopodial stage, 2 for the aggregative stage and 1 for the cystic stage. We obtained in total 140M 50bp paired-end reads over 2 lanes of an Illumina HiSeq2000 instrument in the Genomics Unit at the CRG.

## ChIP-seq and ATAC-seq Data Analysis

ChIP-seq and input reads were mapped into the *Capsaspora* reference genome using Bowtie v1.1.1 (Langmead et al., 2009) with -v 1 -m 1 parameters (single mapping and 1 or 0 mismatches). Duplicates reads were removed using samtools v1.1 (Li et al., 2009). Peak calling was performed using MACS2 (Zhang et al., 2008) with --nomodel, --shiftsize 100, -n 24500000 (genome mappability) and a q-value threshold of 0.01; except for H3K36me3 samples, for which additionally used --broad parameter and a q-value threshold of 0.05. Genome mappability was estimated using the gem-mappability function from the GEM library ([http://algorithms.cnag.cat/wiki/The\\_GEM\\_library](http://algorithms.cnag.cat/wiki/The_GEM_library)).

ATAC-seq reads were mapped into the *Capsaspora* reference genome using Bowtie (Langmead et al., 2009) with -v 1 -m 1 and -X2000 (only fragments up to 2Kb are aligned) parameters. Replicate samples of the same cell stage (for the filopodial stage, and aggregative stage) were pooled as high correlation ( $R > 0.98$ ) between them was observed. Duplicates reads were removed using samtools v1.1 (Li et al., 2009). Read start sites were corrected to account for the 9 bp insert between adaptors introduced by Tn5 transposases (Buenrostro et al., 2013): all reads aligning to the plus strand were offset by +4 bp, and all reads aligning to the minus strand were

offset  $-5$  bp. Finally, reads were classified into nucleosomal free reads (paired-read distance  $<100$ bp) and mononucleosomal reads (pair-read distance between 150 and 240bp), representing single nucleosomes.

Nucleosomal free reads were used to identify *cis*-regulatory sites. Peak calling was performed using MACS2 (Zhang et al., 2008) with the following parameters: `-g 24500000 -q 0.01 --extsize 40 --call-summits --nomodel`. Peaks from different samples were merged using Bedtools (Quinlan and Hall, 2010) to generate the final dataset of 11927 peaks (Supplemental Table 1). The *Capsaspora-Brachyury* binding motif, determined using a Protein Binding Microarray (Sebé-Pedrós et al., 2013b), was scanned in these peaks using Matscan (Blanco et al., 2006), with a cut-off of 0.80. *De novo* motif enrichment analysis of the ATAC-defined regulatory sites was performed using HOMER software (Heinz et al., 2010), with default parameters, except `-size` given. The predicted binding sites for *Capsaspora* Myc, Runx and NFAT/NFkappaB ( $>0.85$  similarity with metazoan orthologs binding site) were scanned in the ATAC-defined regulatory sites using the HOMER-defined cut-off. Gene ontology (GO) functional enrichment analyses were calculated using Ontologizer (Bauer et al., 2008) using the Topology-Weighted method and a p-value cut-off of 0.01. The gene ontology of 8,637 *Capsaspora* genes generated in Sebé-Pedrós et al. (2013) (Sebé-Pedrós et al., 2013a) was used. Additionally, a KEGG pathway annotation of all *Capsaspora* genes was generated using the WebMGA (Wu et al., 2011) and KEGG pathway enrichments were calculated using Hoesa (<http://hoesa.sourceforge.net/>).

Mononucleosomal reads were used to define nucleosome positions and fuzziness using Danpos2 'Dpos' function (Chen et al., 2013), with `-a 1 -p 1 -m 1` parameters.

Chromatin states across the genome were defined using ChromHMM (Ernst and Kellis, 2012) with default parameters, except for binary size of 10bp. We analysed four chromatin marks plus the nucleosomal-free fraction of ATAC-seq. We tested ChromHMM with different *a priori* defined states (from 4 to 20) and chose seven states as the best number to maximize informative features (i.e. different chromatin marks and their unique combinations) with minimal redundancy. The seven ChromHMM defined states were used to calculate the percentage of the genome occupied by any mark. State enrichment in different genomic features was calculated dividing the percentage of nucleotides occupied by a particular state in a particular genomic feature by the percentage of nucleotides that this genomic feature represents in all genome.

bedGraph coverage files were obtained from normalized alignment files, using Bedtools genomeCoverageBed tool (Quinlan and Hall, 2010) with `-bg` option (and `-split` option in the case of RNA-seq data), and loaded into Integrative Genomics Viewer (IGV) (Robinson et al., 2011) for visualization.

Transcription Start Site (TSS) read coverage profiles and heatmaps, as well as transcription factor motif-centered tag density profiles and heatmaps, were calculated using ngs.plot (Shen et

al., 2014). We restricted our TSS profile analyses to tail-to-head oriented genes, to avoid the potential impact of gene orientation in the analysis of chromatin states, given the compact genome of *Capsaspora*. Bedtools intersectBed tool (Quinlan and Hall, 2010) was used to calculate overlaps between peaks and chromatin states with the different genomic features, as well as to assign each *cis*-regulatory site to a particular gene.

### **Cross-species Comparison**

Available ChIP-seq datasets for different species and for H3K4me1, H3K4me3, H3K27Ac, H3K36me3 and the corresponding input files, were used to compare with the *Capsaspora* data obtained in this study. In the case of *Homo sapiens*, ENCODE experiments (<https://www.encodeproject.org/experiments/>) corresponding to cell line GM12878 and the hg19 reference genome were used. modENCODE experiments (<http://data.modencode.org/>) corresponding to larval stage L3 and the ce6 reference genome were used for *Caenorhabditis elegans*. modENCODE experiments (<http://data.modencode.org/>) corresponding to mixed adults and the dm3 reference genome were used for *Drosophila melanogaster*. Data from Schwaiger et al. (2014) (Schwaiger et al., 2014) (GSE46488, GEO accession number) corresponding to gastrula stage and the Nemvec1 reference genome were used for *Nematostella vectensis*. Finally, data from Weiner et al. (2015) (Weiner et al., 2015) (GSE61888, GEO accession number) corresponding to mid-log phase strain BY4741 yeast cells and the SacCer3 reference genome were used for *Saccharomyces cerevisiae*. For each species, RNA-seq data from the same sample type was used to classify genes according to expression and the TSS of the longest isoform for each gene were used for TSS profile analyses. Finally, ATAC-based *cis*-regulatory sites defined in *Homo sapiens* by Buenrostro et al. (2013) (Buenrostro et al., 2013) and those defined in *Drosophila melanogaster* by Davie et al. (2015) (Davie et al., 2015) were used to compare with *Capsaspora*.

### **lincRNA Annotation, Validation and Analysis**

Two complementary approaches were used to identify candidate lincRNAs. The first one was based on the Tophat-Cufflinks pipeline (Trapnell et al., 2012a). We used our pooled 197M strand-specific paired reads RNA-seq dataset (see above) and aligned it to the *Capsaspora* reference genome using Tophat2 with default parameters (except --min-intron-length 30). Aligned reads were assembled into transcripts using Cufflinks2.1.1 with -u, --min-intron-length 30, --max-intron-length 2000 and --intron-overhang-tolerance 30 parameters and with the improved *Capsaspora* gene annotation (see above) as reference (-g). Only newly assembled transcripts were considered and TBLASTX against the predicted *Capsaspora* proteome (e-value < e-3) was used to verify these were non-previously annotated transcripts.

The second approach was based on *de novo* genome-guided Trinity assembly and the PASA pipeline (see above) (Haas et al., 2003, 2013). We used both programs with default parameters, except `--MAX_INTRON_LENGTH 2000` (in order to, like in Cufflinks, minimize over fusion of transcripts). TBLASTX against the predicted *Capsaspora* proteome (e-value < e-3) was used to filter out transcripts representing previously known genes.

Both population (from Cufflinks and from TrinityGG+PASA) were pooled and a series of filters were applied:

1. First, we filtered out transcripts intersecting any annotated *Capsaspora* gene using Bedtools intersectBed tool (Quinlan and Hall, 2010).
2. We selected only transcripts above 200bp in length.
3. We performed TBLASTX against NCBI non-redundant database and against the transcriptomes and genomes of several closely related species (including 2 choanoflagellates, *Salpingoeca rosetta* and *Monosiga brevicollis*; one other filasterean, *Ministeria vibrans*; and 6 ichthyosporeans; *Sphaeroforma arctica*, *Ichthyophonus hoferi*, *Pirum gemmata*, *Amoebidium parasiticum*, *Abeoforma whisleri*, *Creolimax fragrantissima* and *Corallochytrium limacisporum*) and selected only those transcripts that did not retrieve any significant hit (e-value < e-3).
4. We analyzed the remaining transcripts with RfamScan\_2, using Rfam 11 database (Burge et al., 2013), in order to filter out (threshold < e-5) those transcripts corresponding to other types of known ncRNAs (like tRNAs or ribosomal RNAs).
5. Next, we used Coding Potential Calculator (CPC) (Kong et al., 2007) to discard transcripts suspected to have coding potential (coding potential score < -0.5).
6. Finally, we collapsed the transcripts into single loci using Cuffcompare (Trapnell et al., 2012b) and further discarded those potential lincRNAs with an expression level below 1 RPKMs (See below).

These resulted in 632 predicted lincRNA loci (Supplementary File 2 and 3). To validate lincRNA predictions, the three stages were induced (see above) and RNA was extracted using Trizol reagent (Life Technologies). To eliminate genomic DNA, total RNA was treated with DNase I (Roche, Basel, Switzerland) and purified using RNeasy columns (Qiagen, Venlo, Netherlands). Polyadenylated RNA was enriched from total RNA using Poly(A)Purist MAG kit (Life Technologies). For each stage, cDNA was produced from 50ng of polyA-enriched RNA using SuperScript III reverse transcriptase (Life Technologies). PCR was performed using ExpandTaq polymerase (Roche), using 45 cycles and a melting temperature of 65°C.

The expression levels (RPKMs) of each lincRNA were calculated using bamutils 'count' function, from the NGSUtils suite (Breese and Liu, 2013). Differentially expressed lincRNAs (included in Figure S4A) were identified using DESeq2 (Love et al., 2014) with a p-adjusted cut-off of 0.01.

## Supplemental References

- Bauer, S., Grossmann, S., Vingron, M., and Robinson, P.N. (2008). Ontologizer 2.0—a multifunctional tool for GO term enrichment analysis and data exploration. *Bioinformatics* 24, 1650–1651.
- Bonenfant, D., Coulot, M., Towbin, H., Schindler, P., and van Oostrum, J. (2006). Characterization of histone H2A and H2B variants and their post-translational modifications by mass spectrometry. *Mol. Cell. Proteomics* 5, 541–552.
- Bray, N.L., Pimentel, H., Melsted, P., and Pachter, L. (2015). Near-optimal RNA-Seq quantification. *aRxiv*.
- Breese, M.R., and Liu, Y. (2013). NGSUtils: a software suite for analyzing and manipulating next-generation sequencing datasets. *Bioinformatics* 29, 494–496.
- Burge, S.W., Daub, J., Eberhardt, R., Tate, J., Barquist, L., Nawrocki, E.P., Eddy, S.R., Gardner, P.P., and Bateman, A. (2013). Rfam 11.0: 10 years of RNA families. *Nucleic Acids Res.* 41, D226–D232.
- Cross, G.A.M. (2008). Histone modifications in *Trypanosoma brucei*. *Mol. Biochem. Parasitol.* 156, 41–50.
- Davie, K., Jacobs, J., Atkins, M., Potier, D., Christiaens, V., Halder, G., and Aerts, S. (2015). 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.
- Frank, S.R., Schroeder, M., Fernandez, P., Taubert, S., and Amati, B. (2001). Binding of c-Myc to chromatin mediates mitogen-induced acetylation of histone H4 and gene activation. *Genes Dev.* 15, 2069–2082.
- Garcia, B. a, Hake, S.B., Diaz, R.L., Kauer, M., Morris, S. a, Recht, J., Shabanowitz, J., Mishra, N., Strahl, B.D., Allis, C.D., et al. (2007a). Organismal differences in post-translational modifications in histones H3 and H4. *J. Biol. Chem.* 282, 7641–7655.
- Haas, B.J., Delcher, A.L., Mount, S.M., Wortman, J.R., Smith Jr, R.K., Hannick, L.I., Maiti, R., Ronning, C.M., Rusch, D.B., Town, C.D., et al. (2003). Improving the Arabidopsis genome annotation using maximal transcript alignment assemblies. *Nucleic Acids Res.* 31, 5654–5666.
- Haas, B.J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P.D., Bowden, J., Couger, M.B., Eccles, D., Li, B., Lieber, M., et al. (2013). De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat. Protoc.* 8, 1494–1512.
- Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y.C., Laslo, P., Cheng, J.X., Murre, C., Singh, H., and Glass, C.K. (2010). Simple Combinations of Lineage-Determining Transcription Factors Prime cis-Regulatory Elements Required for Macrophage and B Cell Identities. *Mol. Cell* 38, 576–589.



- Kong, L., Zhang, Y., Ye, Z.-Q., Liu, X.-Q., Zhao, S.-Q., Wei, L., and Gao, G. (2007). CPC: assess the protein-coding potential of transcripts using sequence features and support vector machine. *Nucleic Acids Res.* *35*, W345–W349.
- Krebs, J.E. (2007). Moving marks: dynamic histone modifications in yeast. *Mol. Biosyst.* *3*, 590–597.
- Langmead, B., Trapnell, C., Pop, M., and Salzberg, S.L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* *10*, R25.
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., and Subgroup, 1000 Genome Project Data Processing (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics* *25* 2078–2079.
- Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-Seq data with DESeq2. *Genome Biol.* *15*, 1–21.
- Mandava, V., Fernandez, J.P., Deng, H., Janzen, C.J., Hake, S.B., and Cross, G.A.M. (2008). Histone modifications in *Trypanosoma brucei*. *Mol. Biochem. Parasitol.* *156*, 41–50.
- Medzihradzky, K.F., Zhang, X., Chalkley, R.J., Guan, S., McFarland, M. a, Chalmers, M.J., Marshall, a G., Diaz, R.L., Allis, C.D., and Burlingame, a L. (2004). Characterization of *Tetrahymena* histone H2B variants and posttranslational populations by electron capture dissociation (ECD) Fourier transform ion cyclotron mass spectrometry (FT-ICR MS). *Mol. Cell. Proteomics* *3*, 872–886.
- Millar, C.B., Xu, F., Zhang, K., and Grunstein, M. (2006). Acetylation of H2AZ Lys 14 is associated with genome-wide gene activity in yeast. *Genes Dev.* *20*, 711–722.
- Nardelli, S.C., Che, F., and De, N.C.S. (2013). The Histone Code of *Toxoplasma gondii* Comprises Conserved and Unique Posttranslational Modification. *MBio.* *4*, e00922-13.
- Perkins, D.N., Pappin, D.J.C., Creasy, D.M., and Cottrell, J.S. (1999). Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis* *20*, 3551–3567.
- Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* *26*, 841–842.
- Robin, P., Fritsch, L., Philipot, O., Svinarchuk, F., and Ait-Si-Ali, S. (2007). Post-translational modifications of histones H3 and H4 associated with the histone methyltransferases Suv39h1 and G9a. *Genome Biol.* *8*, R270.
- Robinson, J.T., Thorvaldsdottir, H., Winckler, W., Guttman, M., Lander, E.S., Getz, G., and Mesirov, J.P. (2011). Integrative genomics viewer. *Nat. Biotechnol.* *29*, 24–26.
- Schwaiger, M., Schonauer, A., Rendeiro, a. F., Pribitzer, C., Schauer, A., Gilles, a. F., Schinko, J.B., Renfer, E., Fredman, D., and Technau, U. (2014). Evolutionary conservation of the eumetazoan gene regulatory landscape. *Genome Res.* *24*, 639–650.
- Shen, L., Shao, N., Liu, X., and Nestler, E. (2014). ngs.plot: Quick mining and visualization of next-generation sequencing data by integrating genomic databases. *BMC Genomics* *15*, 284.

- Stock, J.K., Giadrossi, S., Casanova, M., Brookes, E., Vidal, M., Koseki, H., Brockdorff, N., Fisher, A.G., and Pombo, A. (2007). Ring1-mediated ubiquitination of H2A restrains poised RNA polymerase II at bivalent genes in mouse ES cells. *Nat Cell Biol* 9, 1428–1435.
- Trapnell, C., Roberts, A., Goff, L., Pertea, G., Kim, D., Kelley, D.R., Pimentel, H., Salzberg, S.L., Rinn, J.L., and Pachter, L. (2012a). Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat. Protoc.* 7, 562–578.
- Trapnell, C., Hendrickson, D.G., Sauvageau, M., Goff, L., Rinn, J.L., and Pachter, L. (2012b). Differential analysis of gene regulation at transcript resolution with RNA-seq. *Nat. Biotechnol.* 31, 46–53.
- Trelle, M.B., Salcedo-Amaya, A.M., Cohen, A.M., Stunnenberg, H.G., and Jensen, O.N. (2009). Global histone analysis by mass spectrometry reveals a high content of acetylated lysine residues in the malaria parasite *Plasmodium falciparum*. *J. Proteome Res.* 8, 3439–3450.
- Wei, Y., Mizzen, C. a, Cook, R.G., Gorovsky, M. a, and Allis, C.D. (1998). Phosphorylation of histone H3 at serine 10 is correlated with chromosome condensation during mitosis and meiosis in *Tetrahymena*. *Proc. Natl. Acad. Sci. U. S. A.* 95, 7480–7484.
- Weiner, A., Hsieh, T.S., Rando, O.J., Friedman, N., Weiner, A., Hsieh, T.S., Appleboim, A., Chen, H. V, Rahat, A., and Amit, I. (2015). High-Resolution Chromatin Dynamics during a Yeast Stress Response. *Mol. Cell* 58, 1–16.
- Wu, S., Zhu, Z., Fu, L., Niu, B., and Li, W. (2011). WebMGA: a customizable web server for fast metagenomic sequence analysis. *BMC Genomics* 12, 444.
- Zhang, Y., Liu, T., Meyer, C.A., Eeckhoutte, J., Johnson, D.S., Bernstein, B.E., Nusbaum, C., Myers, R.M., Brown, M., Li, W., et al. (2008). Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* 9, R137.



### **R3: Transfection of *Capsaspora owczarzaki*, a close unicellular relative of animals**

#### *Abstract*

How animals emerged from their unicellular ancestor remains a major evolutionary question. New genome data from the closest unicellular relatives of animals have provided important insights into the evolution of animal multicellularity. We know that the unicellular ancestor of animals had an unexpectedly complex genetic repertoire, including many genes that are key to animal development and multicellularity. Thus, assessing the function of these genes among unicellular relatives of animals is key to understanding how they were co-opted at the onset of the Metazoa. However, such analyses have been hampered by the lack of genetic tools. Progress has been made in choanoflagellates and teretosporeans, two of the three lineages closely related to animals, whereas no tools are yet available for functional analysis in the third lineage: the filastereans. Importantly, filastereans have a striking repertoire of genes involved in transcriptional regulation and other developmental processes. Here, we describe a reliable transfection method for the filasterean *Capsaspora owczarzaki*. We also provide a set of constructs for visualising subcellular structures in live cells. These tools convert *Capsaspora* into a unique experimentally tractable organism to use to investigate the origin and evolution of animal multicellularity.



# Transfection of *Capsaspora owczarzaki*, a close unicellular relative of animals

Helena Parra-Acero<sup>1,\*</sup>, Núria Ros-Rocher<sup>1,\*</sup>, Alberto Perez-Posada<sup>1</sup>, Aleksandra Kozyczkowska<sup>1</sup>, Núria Sánchez-Pons<sup>1</sup>, Azusa Nakata<sup>2</sup>, Hiroshi Suga<sup>2</sup>, Sebastián R. Najle<sup>1,3</sup> and Iñaki Ruiz-Trillo<sup>1,4,5,‡</sup>

## ABSTRACT

How animals emerged from their unicellular ancestor remains a major evolutionary question. New genome data from the closest unicellular relatives of animals have provided important insights into the evolution of animal multicellularity. We know that the unicellular ancestor of animals had an unexpectedly complex genetic repertoire, including many genes that are key to animal development and multicellularity. Thus, assessing the function of these genes among unicellular relatives of animals is key to understanding how they were co-opted at the onset of the Metazoa. However, such analyses have been hampered by the lack of genetic tools. Progress has been made in choanoflagellates and teretosporeans, two of the three lineages closely related to animals, whereas no tools are yet available for functional analysis in the third lineage: the filastereans. Importantly, filastereans have a striking repertoire of genes involved in transcriptional regulation and other developmental processes. Here, we describe a reliable transfection method for the filasterean *Capsaspora owczarzaki*. We also provide a set of constructs for visualising subcellular structures in live cells. These tools convert *Capsaspora* into a unique experimentally tractable organism to use to investigate the origin and evolution of animal multicellularity.

**KEY WORDS:** Transfection, *Capsaspora owczarzaki*, Holozoa, Multicellularity, Origin of Metazoa

## INTRODUCTION

The transition to animal multicellularity from a single-celled ancestor is one of the most intriguing events in the history of life (King, 2004; Ruiz-Trillo et al., 2007; Rokas, 2008; Knoll, 2011; Richter and King, 2013; Cavalier-Smith, 2017; Sebé-Pedrós et al., 2017). Analysis of the genomes of extant unicellular relatives of animals, hereafter unicellular holozoans, recently showed that the unicellular ancestor of animals was genetically more complex than

previously thought (Sebé-Pedrós et al., 2017). Strikingly, genes thought to be animal specific are now known to be present in unicellular holozoans. These include genes involved in cell adhesion, such as those encoding integrins and cadherins, cell-to-cell communication, such as those encoding tyrosine kinases, and transcriptional regulation, such as the developmental transcription factor *Brachyury* (Sebé-Pedrós et al., 2010, 2011; Sebé-Pedrós et al., 2013a; Nichols et al., 2012; Suga et al., 2012). These findings imply that the co-option of ancestral genes into new functions was an important mechanism for the transition to animal multicellularity. However, understanding how these genes were co-opted will only be possible through functional analyses in extant unicellular relatives of animals.

There are three known lineages of unicellular holozoans: choanoflagellates, teretosporeans (ichthyosporeans and corallochytreaans) and filastereans (Torruella et al., 2015; Grau-Bové et al., 2017). These three lineages show very different developmental modes, such as the clonal colony formation of choanoflagellates (Fairclough et al., 2013), the coenocytic growth of teretosporeans (Marshall et al., 2008; Suga and Ruiz-Trillo, 2013) and the aggregative behaviour present in filastereans (Sebé-Pedrós et al., 2013b). To develop a comprehensive view of the transition to multicellularity, we need to understand these three different modes of development. So far, a forward genetics approach has been developed in choanoflagellates, leading to the discovery of *rosetteless*, a gene related to colony formation in *Salpingoeca rosetta* (Levin et al., 2014). Efforts are also underway to develop transfection in choanoflagellates. Transfection has already been developed in the ichthyosporean *Creolimax fragrantissima*, where it allowed the description of synchronous nuclear division during coenocytic development (Suga and Ruiz-Trillo, 2013). To date, however, there are still no genetic tools reported in filastereans.

Recent analysis of the genome, transcriptome, proteome and phosphoproteome of the filasterean amoeba *Capsaspora owczarzaki* (Fig. S1), hereafter *Capsaspora*, provided important insights into the origins of animal multicellularity and the nature of their unicellular ancestor (Suga et al., 2013; Sebé-Pedrós et al., 2016a,b). The *Capsaspora* genome encodes an unexpected set of transcription factors known to be involved in animal development that were previously thought to be metazoan specific, such as NFκB, Runx and T-box (Sebé-Pedrós et al., 2011; Suga et al., 2013; de Mendoza et al., 2013). Similar to animals, these transcription factors are differentially regulated at the transcriptional level and are also differentially phosphorylated during the *Capsaspora* life cycle (Sebé-Pedrós et al., 2016a,b). *Capsaspora* also contains the most complete set of proteins linked to cell–extracellular matrix adhesion (the Integrin adhesome) among unicellular holozoans (Sebé-Pedrós et al., 2010; Suga et al., 2013). This highlights *Capsaspora* as the closest relative of animals in which such genes can be studied.

<sup>1</sup>Institut de Biologia Evolutiva (CSIC-Universitat Pompeu Fabra), Passeig Marítim de la Barceloneta 37-49, 08003 Barcelona, Catalonia, Spain. <sup>2</sup>Faculty of Life and Environmental Sciences, Prefectural University of Hiroshima, Nanatsuka 5562, Shobara, Hiroshima 727-0023, Japan. <sup>3</sup>Instituto de Biología Molecular y Celular de Rosario (IBR-CONICET) and Facultad de Ciencias Bioquímicas y Farmacéuticas, Universidad Nacional de Rosario, Ocampo y Esmeralda s/n, Rosario S2000FHQ, Argentina. <sup>4</sup>Departament de Genètica, Microbiologia i Estadística, Universitat de Barcelona, Av. Diagonal, 645, 08028 Barcelona, Catalonia, Spain. <sup>5</sup>ICREA, Passeig Lluís Companys 23, 08010, Barcelona, Catalonia, Spain. \*These authors are co-first authors

‡Author for correspondence (inaki.ruiz@ibe.upf-csic.es)

© N.R., 0000-0003-0897-0186; I.R., 0000-0001-6547-5304

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/3.0>), which permits unrestricted use, distribution and reproduction in any medium provided that the original work is properly attributed.

Here, we present the first protocol for transfecting the filasterean *Capsaspora* with plasmid DNA. The protocol is based on the classical calcium phosphate precipitation method (Graham and van der Eb, 1973), which we coupled with a glycerol shock to increase transfection efficiency. We also constructed a set of expression vectors containing an endogenous promoter and fluorescent reporters that allow labelling of multiple subcellular structures *in vivo*. Altogether, this work provides the key step necessary to perform functional assays requiring foreign nucleic acid delivery, including overexpression, RNA interference and genome editing, rendering *Capsaspora* experimentally tractable towards addressing the transition to animal multicellularity.

## RESULTS AND DISCUSSION

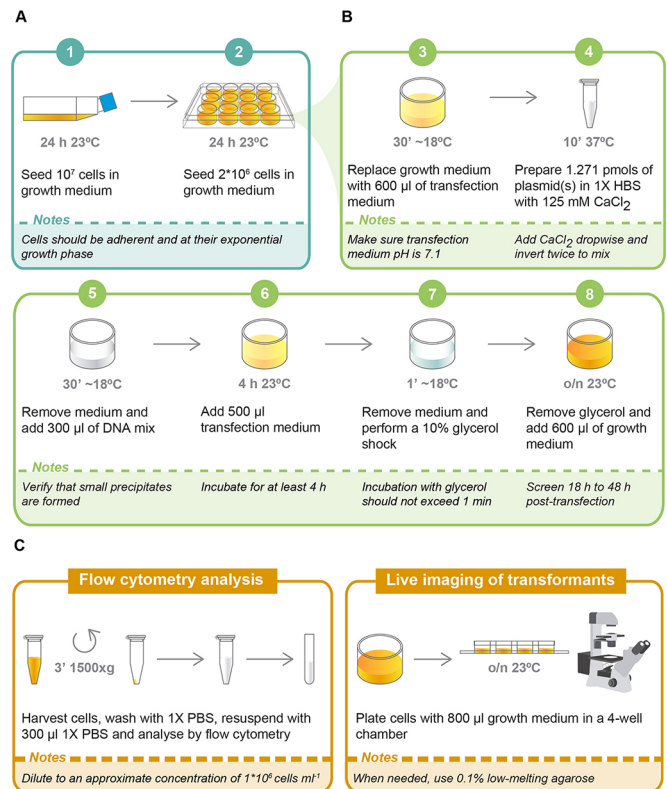
### *Capsaspora* transfection using calcium phosphate precipitation

There are several protocols available for the transient transfection of plasmid DNA for eukaryotic cells. For *Capsaspora*, we initially tried electroporation because it has been successfully used to transiently transfect the ichthyosporian *C. fragrantissima* (Suga and Ruiz-Trillo, 2013). However, we obtained no more than 20 positive cells out of thousands of cells. We additionally tried lipid-based transfection (Felgner et al., 1987) and magnetofection (Buerli et al., 2007; Ensenaer et al., 2011), which have been reported to work in eukaryotic cells that are difficult to transfect. Nevertheless, both approaches resulted in few, if any, positive cells. Finally, we tested the classical calcium phosphate precipitation-based transfection method (Graham and van der Eb, 1973), which has been reported to successfully transfect *Dictyostelium discoideum* (Nellen et al., 1984; Gaudet et al., 2007), an amoebozoan without a cell wall. Given that we initially obtained ~100 times more cells that were positive cells with the calcium phosphate precipitation protocol than with the other methods, we focused on this protocol to further improve its efficiency.

As a first step to increase the efficiency of transfection, we investigated which life stage to transfect. Under culture conditions, *Capsaspora* presents three distinct life stages: adherent, cystic and aggregative (Sebé-Pedrós et al., 2013b). We tried using cells in the adherent stage because the culture is at its exponential growth phase at this stage (Fig. 1A) (Sebé-Pedrós et al., 2013b). We observed that transfecting *Capsaspora* adherent cells at 90–95% confluence from a fresh culture resulted in higher transfection efficiency.

Next, we addressed crystal formation during DNA precipitation. We sought the smallest size of crystals possible, because smaller crystals have been associated with higher transfection efficiency (Jordan et al., 1996; Jordan and Wurm, 2004). We achieved this by keeping the same ratio of DNA, calcium and phosphate as previously described for *D. discoideum* (Gaudet et al., 2007), and setting an incubation time of 10 min at 37°C (Fig. 1B-4). Additionally, we used a transfection medium containing minimal growth components but lacking phosphate (Fig. 1B-3,6), to maintain the optimal concentration of calcium phosphate for DNA precipitation. This medium also contains buffering agents at pH 7.1 to avoid pH fluctuations that might affect the solubility of any precipitates.

Finally, to increase the number of transfected cells, we coupled the protocol with a glycerol shock, because the latter has shown to increase the transfection efficiency in other systems (Grosjean et al., 2006; Gaudet et al., 2007; Guo et al., 2017). We performed the shock using 10% glycerol in 1× HBS for 1 min (Fig. 1B-7) and immediately added growth medium to avoid compromising cell viability (Fig. 1B-8, see Supplementary Materials and Methods for further details).

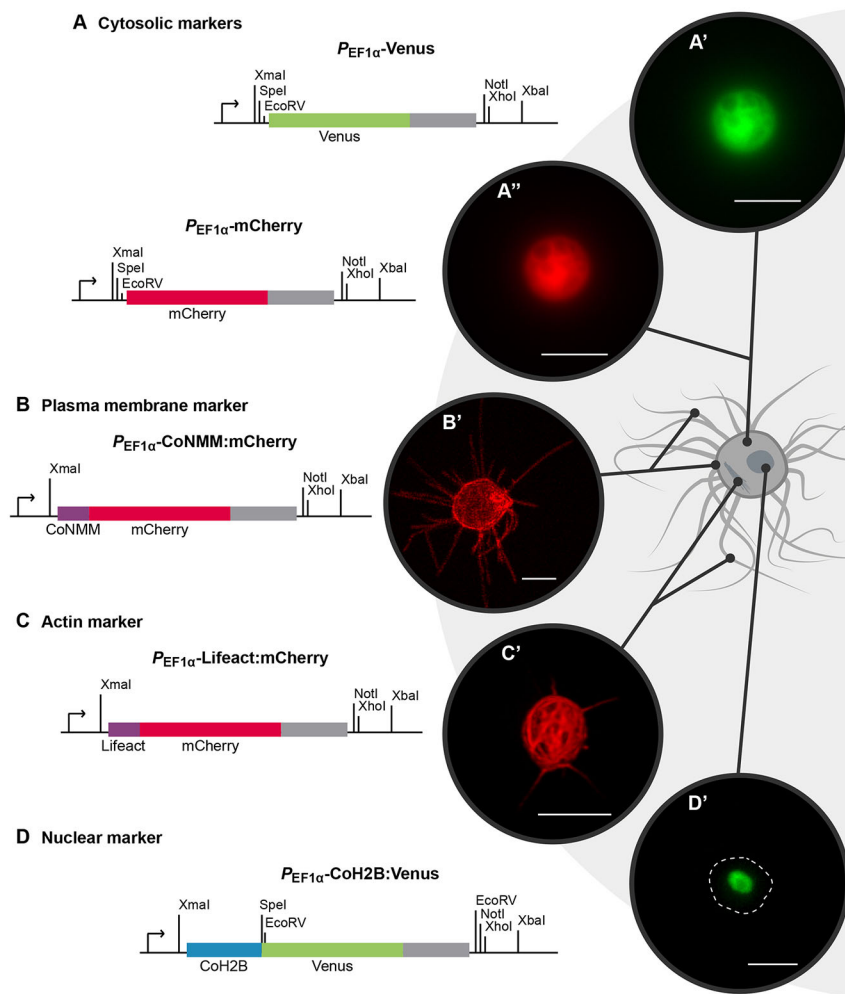


**Fig. 1. Protocol for transfection of plasmid DNA in *Capsaspora*.** (A) Preparation of cells (1-2). (B) Calcium phosphate precipitation (3-8). (C) Screening of transformants. o/n, overnight.

### Analysis of transfected *Capsaspora* cells by flow cytometry

To evaluate DNA incorporation by *Capsaspora*, we constructed two expression vectors containing either Venus (pONSY-Venus) or mCherry (pONSY-mCherry) fluorescent proteins as cytosolic markers (Fig. 2A). These vectors contain the promoter and terminator regions of the endogenous *elongation factor 1-α* (*EF1-α*) gene (CAOG\_07807) from *Capsaspora* (see Materials and Methods). We confirmed the successful expression of both fluorescent proteins by fluorescence microscopy (Fig. 2A'-A'') and flow cytometry. We performed an immunofluorescence assay on sorted cells and confirmed that the fluorescent population identified was expressing Venus (Fig. S2).

Next, we analysed the transfection efficiency by quantifying the number of positive cells by flow cytometry (Fig. 3A,B). We performed single transfection experiments using either pONSY-Venus or pONSY-mCherry in seven independent experiments (each experiment performed with a different batch of cells) with at least six technical replicates each. In both cases, the positive populations were defined using a negative control (Fig. S3). *Capsaspora* transfection efficiency was 1.132%±0.529 (mean±s.d.) with a 95% confidence interval of (0.983-1.281%) from a total of 4.9 million cells (Fig. 3C). In these experiments, individual transfection efficiencies ranged from 0.347%±0.193 to 2.083%±0.248 (Table S1). Importantly, these transfection efficiencies are sufficient to screen for positive cells and perform further manipulations, because they correspond to thousands of positive cells per well. Additionally, we compared transfection rates between *Capsaspora* cells transfected with either pONSY-Venus or pONSY-mCherry (Fig. 3D, experiments 7a and 7b, respectively, in Table S1), but no significant difference was observed ( $P=0.5625$ , Wilcoxon Signed Rank Test).



**Fig. 2. *Capsaspora* expression cassettes and live imaging of transfected cells.** (A) Cytosolic marker cassettes expressing mCherry (A') or Venus (A'') fluorescent proteins. (B) Plasma membrane marker cassette containing the *Capsaspora Src2* NMM fused to mCherry (B'). (C) Actin marker cassette containing *Lifeact* fused to mCherry (C'). (D) Nuclear marker cassette containing *Capsaspora* histone *H2B* (*CoH2B*) fused to Venus (D'). *EF-1α* promoter (arrows) and terminator (grey boxes) and single-cut restriction enzymes are shown. Cells in (A) and (D) were imaged using wide-field fluorescence microscopy. The cell in B was imaged using a Spinning Disk confocal microscope and the cell in C was imaged using a confocal laser scanning microscope. Dashed line indicates the cell body. Scale bars: 5  $\mu$ m.

Given that transfection is transient, it is of interest to know how long the expression of the reporter gene persists for. Thus, we performed three independent experiments transfecting pONSY-Venus. We analysed the percentage of positive cells by flow cytometry every 24 h for 10 days (Fig. S4 and Table S2). We observed an exponential decrease in the number of positive cells. Although there was a significant reduction in the number of positive cells after 48 h (~39% of the total of positive cells), positive cells (~3%) could still be detected by Day 10. At this point, most of the cells are expected to be in the cystic stage (Seb e-Pedr os et al., 2013b), indicating that transient expression of a gene of interest can be analysed during each of the life stages of *Capsaspora*.

We also tested whether the two different constructs could express a protein simultaneously by co-transfecting both pONSY-mCherry and pONSY-Venus at equimolar concentrations in seven independent experiments (Fig. 3E). The red (Q2) and green (Q4) positive populations were defined using their corresponding negative controls (Fig. S5). The mean relative percentage of cells showing red and green fluorescence simultaneously (Q3 in Fig. 3F) from the total number of positive cells (sum of Q2, Q3 and Q4 in Fig. 3F) was  $72.909\% \pm 5.468$ , ranging from ~65% to ~83% (Fig. 3G, Table S3). Thus, it is possible to co-transfect two different vectors with a high probability of incorporating both of them in the same cell. This result is similar to those observed in other unicellular eukaryotes [40-80% in *Volvox carteri* (Schiedlmeier et al., 1994), 84% in *Pandorina morum* (Lerche and Hallmann, 2014) and

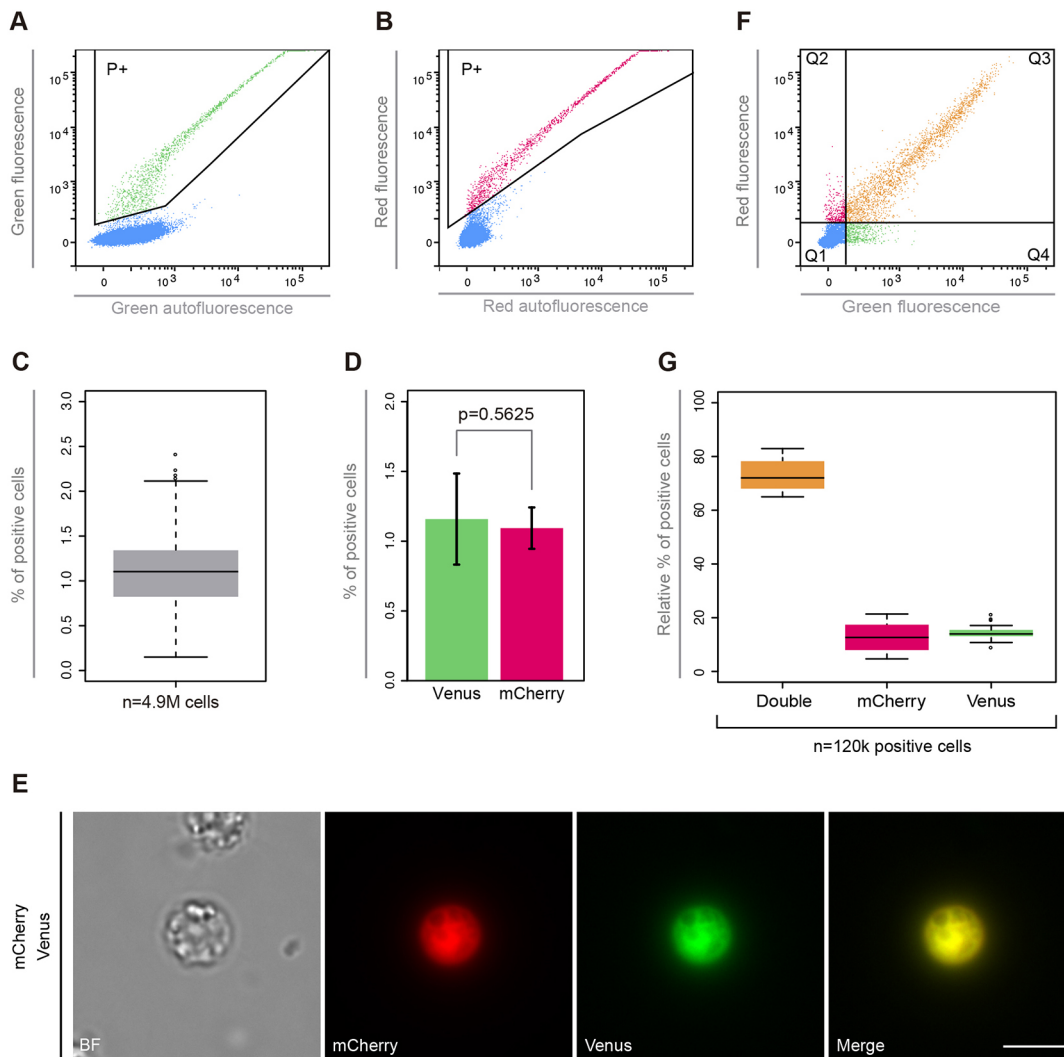
50-100% in *Eudorina elegans* (Lerche and Hallmann, 2013)]. Co-transfection is a useful strategy when more than one cassette is needed, such as when labelling two different cellular structures simultaneously, delivering resistance cassettes against an antibiotic with a reporter gene, or delivering different elements required for CRISPR/Cas9 assays.

### Live imaging of *Capsaspora* by labelling endogenous proteins

To understand the biological role of certain genes in *Capsaspora*, it is important to subcellularly localise the protein of interest in the cell. Thus, as a means of labelling the cellular structures in *Capsaspora*, we designed three additional vectors that allowed live imaging of the plasma membrane, the actin cytoskeleton and the nucleus.

To label the plasma membrane, we built a construct expressing an endogenous membrane-binding motif fused to mCherry (pONSY-CoNMM:mCherry, Fig. 2B). We used the *N*-myristoylation motif (NMM), a well-known membrane-binding motif, present in the *Src* tyrosine kinases (Sigal et al., 1994). *Capsaspora* has two homologs of *Src*, *CoSrc1* (CAOG\_02182) and *CoSrc2* (CAOG\_06360), the localisation of which has been reported in filopodia (Schultheiss et al., 2012). We used *CoSrc2* NMM to create the CoNMM:mCherry fusion, which successfully localised at the plasma membrane, including filopodia (Fig. 2B'). To label the actin cytoskeleton in *Capsaspora*, we built a construct containing *Lifeact* fused to mCherry (pONSY-*Lifeact*:mCherry, Fig. 2C). *Lifeact* is a





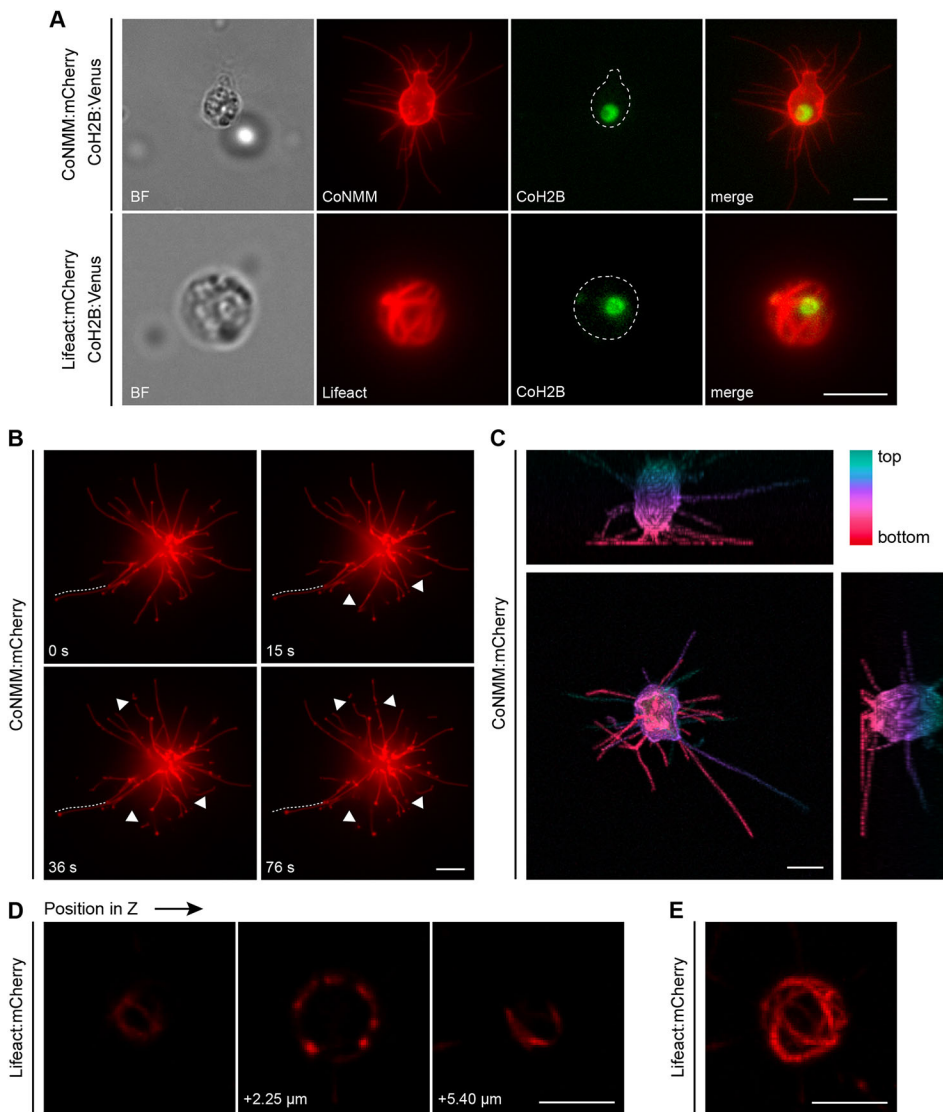
**Fig. 3. Transfection efficiency analysis of *Capsaspora*.** (A) Flow cytometry distribution of pONSY-Venus transfected cells. Area selected (P+) represents the Venus-positive population. (B) Flow cytometry distribution of pONSY-mCherry transfected cells. Area selected (P+) represents the mCherry-positive population. (C) Percentage of positive cells in single transfection. The box plot represents the transfection efficiency distribution over seven independent experiments with at least six technical replicates each ( $n=4.9$  M cells). (D) Percentage of positive cells in a paired experiment with six technical replicates, transfecting either pONSY-Venus or pONSY-mCherry. Error bars represent s.d. ( $P=0.5625$ , Wilcoxon Signed Rank Test). (E) Wide-field fluorescence microscopy of a live cell co-transfected with pONSY-Venus and pONSY-mCherry. (F) Flow cytometry distribution of pONSY-Venus and pONSY-mCherry co-transfected cells. Cell population was divided into quartiles: negative cells (Q1), fluorescent cells expressing mCherry only (Q2), co-transfected cells expressing both fluorescent proteins (Q3), and fluorescent cells expressing Venus only (Q4). (G) Relative percentage of positive cells co-transfected with pONSY-Venus and pONSY-mCherry; expressing both fluorescent proteins (double), mCherry only or Venus only, calculated from the total number of positive cells in seven independent experiments with six replicates each ( $n=120,000$  cells). Scale bar:  $5\ \mu\text{m}$ .

17-amino acid peptide from the N-terminal region of yeast Abp140 (Riedl et al., 2008) that works as a marker of filamentous actin. The Lifeact:mCherry fusion successfully labelled the actin cytoskeleton (Fig. 2C'). This construct also labels actin in filopodia (Fig. S7), although the signal is much weaker than that observed with the membrane marker. Finally, to label the nucleus, we built a construct containing the coding sequence of *Capsaspora* histone *H2B* (CAOG\_01818) fused to Venus (pONSY-CoH2B:Venus, Fig. 2D). We confirmed nuclear localisation by staining transfected cells with DAPI (Fig. 2D', Fig. S6).

To better understand the subcellular structures of *Capsaspora* cells, we combined the nuclear, plasma membrane and actin markers. We co-transfected pONSY-CoH2B:Venus with either pONSY-CoNMM:mCherry or pONSY-Lifeact:mCherry (Fig. 4A). Furthermore, we performed live imaging in cells transfected with

either the membrane marker or the actin marker. The use of the membrane marker allowed us to observe the dynamic behaviour of filopodia on the substrate with unprecedented detail. We observed the retraction of filopodia, filopodia breakage and foci of membrane accumulation (Fig. 4B and Movie 1). In particular, we observed that filopodia are distributed around the cell body. More importantly, the projections constructed from the z-stack clearly demonstrated that the *Capsaspora* cell body is not in direct contact with the substrate, with the numerous filopodia instead holding the cell up (Fig. 4C). Moreover, we tracked a cell transfected with the actin cytoskeleton marker and observed the organisation of actin bundles around the cell body (Fig. 4D,E and Movie 2).

The accumulated knowledge on its well-annotated genome, transcriptome, proteome and phosphoproteome and histone modifications, and its key phylogenetic position as a close



**Fig. 4. Live imaging of transfected *Capsaspora* cells.** (A) Wide-field fluorescence microscopy of live cells co-transfected with pONSY-CoNMM:mCherry and pONSY-CoH2B:Venus, and live cells co-transfected with pONSY-Lifeact:mCherry and pONSY-CoH2B:Venus. CoNMM:mCherry labelling is presented as a maximum projection of the cell. Dashed lines indicate cell bodies. (B) Time-points on a *Capsaspora* cell transfected with pONSY-CoNMM:mCherry imaged using wide-field fluorescence microscopy. Filopodia attached to the substrate are in focus. A retracting filopodia can be observed (dotted line), whereas four filopodia are broken (arrowheads). (C) Maximum-intensity projections in each axis of a cell transfected with pONSY-CoNMM:mCherry. Colour scale represents depth through the projection. Imaging was performed using confocal microscopy. (D) z-stack on a *Capsaspora* cell transfected with pONSY-Lifeact:mCherry imaged using confocal microscopy. Actin bundles can be observed shaping the cell in a basket-like structure that is hollow in the middle. (E) Full z-stack maximum intensity projection of cell in (D). Cells in (C-E) were imaged using a Spinning Disk confocal microscope. Scale bars: 5  $\mu$ m.

unicellular relative to animals render *Capsaspora* as a powerful system to understand the emergence of multicellular animals. The reliable transfection protocol for *Capsaspora* presented here will allow us to study the function of genes that were key to the evolution of multicellularity, opening new avenues of functional research to better understand the transition to animal multicellularity.

## MATERIALS AND METHODS

### Cell strain and growth conditions

*C. owczarzaki* cell cultures (strain ATCC<sup>®</sup>30864) were grown axenically in 25 cm<sup>2</sup> culture flasks (Falcon<sup>®</sup> VWR, #734-0044) with 5 ml ATCC medium 1034 (modified PYNFH medium), hereafter growth medium, in a 23°C incubator (see supplementary Materials and Methods).

### Construction of *Capsaspora* expression vectors

DNA from *Capsaspora* cells was extracted as in Suga et al. (2013). RNA was extracted using a Trizol reagent (Invitrogen/Thermo Fisher Scientific, #15596026). cDNA was obtained by RT-PCR using SuperScript<sup>®</sup> III Reverse Transcriptase (Invitrogen, #18080044) following the manufacturer's instructions.

*Capsaspora* expression vectors, named pONSY, bear the promoter and terminator regions from the endogenous *EF-1 $\alpha$*  gene (CAOG\_07807). To build the pONSY-Venus vector (5.849 kb), the *EF-1 $\alpha$*  promoter (906 bp upstream from methionine) and terminator (320 bp downstream from the

stop codon) were amplified from gDNA with primers 1 and 2 and primers 4 and 5, respectively (Table S4). *Venus* was amplified from a plasmid available in H.P.-A.'s lab using primers 7 and 8, which contain overlap regions with the promoter and terminator regions, respectively. The three amplicons were fused together by overlapping PCR using primers 1 and 5. The resulting *P*<sub>EF1 $\alpha$</sub> -Venus-terminator cassette was digested using the *KpnI* restriction enzyme and cloned into the pCR2.1 vector (Life Technologies, #K203001) linearised at the *KpnI* restriction site.

To build pONSY-mCherry (5.828 kb), we followed the same strategy as described above for the promoter, and used primers 4 and 6 to amplify the terminator region. This modification was introduced to eliminate an extra *EcoRV* site that affects further cloning. *mCherry* was amplified from a plasmid available in H.P.-A.'s lab using primers 9 and 10, which contain overlap regions with both the promoter and terminator regions. The three amplicons were fused together by overlapping PCR using primers 1 and 6. The resulting *P*<sub>EF1 $\alpha$</sub> -mCherry-terminator cassette was digested with *KpnI* and *KspI* enzymes and cloned into the respective restriction sites of the pCR2.1 vector.

A pONSY (empty) vector (5.127 kb) was created by releasing a mCherry-terminator fragment from pONSY-mCherry using *SpeI* and *KspI* restriction enzymes and inserting the terminator in this backbone by Gibson Assembly<sup>®</sup> (New England Biolabs, E2611L) using primers 7 and 8.

pONSY-CoH2B:Venus (6.230 kb) was created by fusing *Capsaspora* histone *H2B* (CAOG\_01818) to Venus. *CoH2B* was PCR amplified from

cDNA using primers 13 and 14 and cloned into the pONSY-Venus multicloning site using *SmaI* and *SpeI* restriction enzymes.

pONSY-CoNMM:mCherry (5.904 kb) was created by fusing an NMM to mCherry. NMM was predicted in the *Capsaspora* Src homolog CoSrc2 (CAOG\_06360) using 'NMT - The MYR Predictor' online software (<http://mendel.imp.ac.at/myristate/SUPLpredictor.htm>), which is based on an in-depth study of *N*-myristoyltransferase substrate proteins (Maurer-Stroh et al., 2002). The NMM predicted sequence GCSNSKPHDPSDFKVSP plus seven extra amino acids (SGVASNS) and an *mCherry* overlap region were included in primer 11. Primers 11 and 12 were used to build a CoNMM-mCherry cassette by PCR using the pONSY-mCherry vector as a template. This cassette was then cloned into pONSY (empty) using *XmaI* and *EcoRV* restriction enzymes.

pONSY-Lifeact:mCherry (5.882 kb) was created by fusing the Lifeact peptide MGVADLIKKFESISKEE(GDPP) (linker in parentheses) to mCherry using primers 15 and 16. The codons were optimised according to their usages in *C. owczarzewski* and *C. fragrantissima*. The Lifeact DNA fragment was first cloned into a pTAC-2 vector (BioDynamics Laboratory) by TA cloning, and the *XmaI*- and *XbaI*-excised fragment was cloned into the pONSY:mCherry vector.

All plasmids DNA were obtained using the plasmid GenElute™ Plasmid Midiprep Kit (Sigma, #NA0200-UKT), lyophilised and resuspended at an approximate concentration of 1 µg/µl in distilled water.

### Transfection of *Capsaspora owczarzewski*

*Capsaspora* cells were transfected using a calcium-phosphate DNA precipitation protocol coupled with a glycerol shock. At Day 0,  $2 \times 10^6$  cells were seeded in a 12-well plate (Nunc/Corning/Costar, #55428) containing growth medium and grown at 23°C overnight.

At Day 1, growth medium was replaced by transfection medium (see Supplementary Materials and Methods), and incubated for 30 min at room temperature (~18°C). During incubation, the DNA mix was prepared with 1.271 pmol of plasmid DNA for single transfection experiments or 0.636 pmol of each plasmid DNA for co-transfection experiments in 1× HBS Buffer. CaCl<sub>2</sub> was added dropwise to a final concentration of 125 mM. The DNA mix was inverted immediately twice and incubated for 10 min at 37°C. After incubation, the transfection medium was removed gradually and the DNA mix was added dropwise to the centre of each well. The cell:DNA mixes were incubated for 30 min at room temperature, after which the transfection medium was added to each well. Cells were incubated for 4 h at 23°C. After incubation, medium was removed and an osmotic shock was performed using 10% glycerol in 1× HBS buffer, for 1 min at room temperature. After the osmotic shock, the glycerol solution was replaced by growth medium and cells were incubated at 23°C overnight. Screening of positive cells was performed 18 h post transfection. More details about the transfection protocol and preparation of transfection reagents are listed in the Supplementary Materials and Methods.

### Flow cytometry and FACS

Transfection efficiency was analysed 18 h post transfection. pONSY (empty) transfected cells, mock-transfected cells and nontransfected cells were used as controls for all transfection experiments to discriminate autofluorescence and to distinguish the positive population. For co-transfection experiments, pONSY-mCherry and pONSY-Venus single-transfected controls were used to correctly identify double-positive cells.

Cells were scraped and harvested by centrifugation at 1500 ×g for 3 min at 18°C, washed once with 500 µl 1× PBS (Sigma, #P5368-10 PAK) and diluted to a final concentration of  $1 \times 10^6$  cells ml<sup>-1</sup> in a minimum volume of 300 µl 1× PBS. Samples were analysed by flow cytometry using a BD LSRFortessa analyser (Becton Dickinson).

To evaluate plasmid persistence over time, transfected cells from 12 wells per experiment were pooled to homogenise the sample, were then split again into 12 new wells and grown for 10 days. Samples were scraped and harvested by centrifugation at 1500 ×g for 3 min at 18°C. Samples were fixed with 4% formaldehyde (Sigma-Aldrich, #F8775-4X25ML) in 1× PBS for 10 min at 18°C and washed once with 500 µl 1× PBS (Sigma, #P5368-10 PAK). Finally, cells were re-suspended in 400 µl 1× PBS and kept at 4°C until analysed.

SSC-A and FSC-A parameters were used to detect populations of cells (P1). Single cells were gated by FSC-H and FSC-A (P2). Around 100,000

events were recorded from P2, whenever possible. Venus-positive cells (P+ or Q1) were detected using a 488 nm laser with a 530/28 bandpass filter (green fluorescence) and differentiated from autofluorescent cells with a 670/50 bandpass filter.

mCherry-positive cells (P+ or Q4) were detected using a 561 nm laser with a 610/20 bandpass filter (red fluorescence) and differentiated from autofluorescent cells with a 780/60 bandpass filter. Around 2000 events in the population expressing both Venus and mCherry (Q2) were recorded.

For immunofluorescence validation, pONSY-Venus transfected cells were harvested as before and diluted to a concentration of  $1 \times 10^7$  cells ml<sup>-1</sup> in a minimum volume of 500 µl 1× PBS. Cells from nine replicates were pooled. Then, 40,000 Venus-positive cells (P+) and 1 million Venus-negative cells (P-) were sorted using a BD FACSAria II SORP flow cytometer cell sorter (Becton Dickinson) equipped with a 100 µm nozzle. The cell population (P2) was gated as before. P+ and P- were detected using a 488 nm laser with a 525/50 bandpass filter (green fluorescence) and differentiated from autofluorescent cells with a 605/40 bandpass filter. Flow cytometry data were visualised and analysed using FlowJo software (FlowJo LLC, version 9.9.3).

### Immunostaining

Sorted cells were collected in 200 µl of 1× PBS and seeded in a Nunc glass-bottom dish (Thermo Fisher Scientific, #150680) previously treated with 200 µl of 20 µg ml<sup>-1</sup> fibronectin (Sigma-Aldrich, #F1141-2MG) overnight at 4°C. Cells were incubated for 3 h at 23°C, then 1× PBS was substituted with 200 µl growth medium and grown overnight at 23°C.

Cells were fixed for 5 min at room temperature with 4% formaldehyde in 1× PBS and washed once with 200 µl 1× PBS. Cells were then blocked for 30 min at room temperature in blocking solution [1% bovine serum albumin (Sigma-Aldrich, #A3294-10G), 0.1% Triton-X100 (Sigma-Aldrich, X100) in 1× PBS] and incubated for 1.5 h at room temperature with 1:100 anti-green fluorescent protein (GFP) primary antibody (Abcam, ab5450, Lot GR277059-1) in blocking solution [Venus is an improved version of GFP (Nagai et al., 2002)]. Cells were washed twice for 10 min in blocking solution and incubated for 1.5 h in the dark at room temperature with 1:1000 Alexa Fluor 568 goat anti-rat IgG (Life Technologies, A11077, Lot 1512105) in the blocking solution. After three washes of 10 min in 1× PBS, the preparation was overlaid with fluorescence mounting media (DAKO/Agilent Technologies, #S3023), covered with a coverslip and sealed with nail polish.

### Imaging of transfected cells

Immunostained samples were imaged using a Leica TCS SP5 II inverted confocal microscope with a 63× immersion oil objective. Acquisition settings were adjusted using Venus-positive cells without primary antibody and Venus-negative cells.

For live imaging, all samples were plated in a µ-Slide 4-well glass-bottom dish (Ibidi, #80427) and grown overnight at 23°C. In the case of pONSY-H2B:Venus transfected cell samples, plated cells were washed once with 200 µl 1× PBS, fixed for 5 min at room temperature with 4% formaldehyde in 1× PBS and washed again as before. Cells were covered using Vectashield with DAPI (Vector, #H-1200). For live imaging of the cytoskeleton, cells were plated in a µ-Slide 4-well Ph+ glass-bottom dish (Ibidi, #80447) in 800 µl of growth medium containing 0.1% low-melting agarose (Sudelab #8085). Wide-field microscopy was performed using a Zeiss Axio Observer Z.1 epifluorescence inverted microscope equipped with LED illumination and a AxioCam 503 mono camera.

Time-lapse videos were recorded using the same microscope. For Movie 1, acquisition was performed at 1 frame/s (fps) and video export was performed at 10 fps. For Movie 2, images were taken every 10 min and video export was performed at 2 fps. A maximum intensity projection was used of two slices from a z-stack.

When indicated, membrane and cytoskeleton labelling were additionally imaged using confocal microscopy with an Andor Revolution XD Spinning Disk microscope equipped with an Andor Ixon 897E Dual Mode EM-CCD camera. These images were deconvolved using The Huygens System 17.10-64 Multi-Processing edition software.

Confocal microscopy was performed with a 63× immersion oil objective using either a confocal laser scanning Leica TCS SP5 II microscope or an Andor Revolution XD Spinning Disk microscope equipped with an Andor Ixon 897E Dual Mode EM-CCD camera. These images were deconvolved using The Huygens System 17.10-64 Multi-Processing edition software.

All images were edited using Fiji Imaging Software version 2.0.0-rc-44/1.50e (Schindelin et al., 2012).

### Statistical analysis

Results are shown as mean±standard deviation (s.d.) per experiment. The 95% confidence intervals were calculated using the Student's *t*-test. The significance of differences in the percentage of positive cells from single-transfection experiments were tested using the non-parametric Wilcoxon Signed Rank Test for paired samples. All statistical analyses were performed using the R Stats Package version 3.3.1 (R Core Team, 2016).

### Acknowledgements

We thank UPF Flow Cytometry Core Facility staff for helpful discussions. We thank the Advanced Light Microscopy Unit of the CRG for support on confocal images acquisition and edition. We thank Arnau Sebé-Pedrós and Alex de Mendoza for help in the project design and the initial steps. We thank Omayya Dudin for fruitful discussions on the manuscript. We also thank the many other people who have worked in our lab and who have provided support and feedback over the years.

### Competing interests

The authors declare no competing or financial interests.

### Author contributions

Conceptualisation: H.P.-A., N.R.-R., S.R.N., I.R.-T.; Methodology: H.P.-A., N.R.-R., A.P.-P., A.K., N.S.-P., S.R.N.; Formal analysis: H.P.-A., N.R.-R., A.P.-P., A.K., N.S.-P., S.R.N.; Resources: A.N., H.S.; Writing - original draft: H.P.-A., N.R.-R.; Writing - review & editing: H.P.-A., N.R.-R., A.P.-P., S.R.N., I.R.-T.; Supervision: I.R.-T.; Funding acquisition: I.R.-T.

### Funding

This work was supported by European Research Council Starting (ERC-2007-StG-206883) and Consolidator (ERC-2012-Co-616960) grants, and a grant (BFU2014-57779-P) from Ministerio de Economía y Competitividad (MINECO) to I.R.-T., which was co-funded by the European Regional Development Fund (fondos FEDER). We also acknowledge financial support from Secretaria d'Universitats i Recerca del Departament d'Economia i Coneixement de la Generalitat de Catalunya (Project 2014 SGR 619) and 'Formación del Profesorado Universitario (FPU13/01840)' from Ministerio de Educación, Cultura y Deporte (MECD) PhD fellowship to N.R.-R. and 'la Caixa' Grant for Doctoral Studies from 'la Caixa' Banking Foundation (Spain) to A.P.-P. We also acknowledge financial support for H.S. from Japan Society for the Promotion of Science (KAKENHI 16K07468) and research grants from the Novartis foundation for the Promotion of Science, ITOH Science Foundation, and Naito Foundation.

### Data availability

Capsaspora expression vectors have been deposited in Addgene ([www.addgene.org](http://www.addgene.org)) under the following accession numbers: pONSY (empty) (111873), pONSY-mCherry (111874), pONSY-Venus (111875), pONSY-Lifeact-mCherry (111876), pONSY-CoH2B:Venus (111877) and pONSY-CoNMM:mCherry (111878). The Capsaspora transfection protocol has been uploaded to protocols.io under [dx.doi.org/10.17504/protocols.io.p4daqse](https://doi.org/10.17504/protocols.io.p4daqse)

### Supplementary information

Supplementary information available online at <http://dev.biologists.org/lookup/doi/10.1242/dev.162107.supplemental>

### References

Buerli, T., Pellegrino, C., Baer, K., Lardi-Studler, B., Chudotvorova, I., Fritschy, J.-M., Medina, I. and Fuhrer, C. (2007). Efficient transfection of DNA or shRNA vectors into neurons using magnetofection. *Nat. Protoc.* **2**, 3090-3101.

Cavalier-Smith, T. (2017). Origin of animal multicellularity: precursors, causes, consequences—the choanoflagellate/sponge transition, neurogenesis and the Cambrian explosion. *Philos. Trans. R. Soc. B Biol. Sci.* **372**, 20150476.

de Mendoza, A., Sebé-Pedrós, A., Šesták, M. S., Matejčić, M., Torruella, G., Domazet-Lošo, T. and Ruiz-Trillo, I. (2013). Transcription factor evolution in eukaryotes and the assembly of the regulatory toolkit in multicellular lineages. *Proc. Natl. Acad. Sci. USA* **110**, E4858-E4866.

Ensenauer, R., Hartl, D., Vockley, J., Roscher, A. A. and Fuchs, U. (2011). Efficient and gentle siRNA delivery by magnetofection. *Biotech. Histochem.* **86**, 226-231.

Fairclough, S. R., Chen, Z., Kramer, E., Zeng, Q., Young, S., Robertson, H. M., Begovic, E., Richter, D. J., Russ, C., Westbrook, M. J. et al. (2013). Premetazoan genome evolution and the regulation of cell differentiation in the choanoflagellate *Salpingoeca rosetta*. *Genome Biol.* **14**, R15.

Felgner, P. L., Gadek, T. R., Holm, M., Roman, R., Chan, H. W., Wenz, M., Northrop, J. P., Ringold, G. M. and Danielsen, M. (1987). Lipofection: a highly efficient, lipid-mediated DNA-transfection procedure. *Proc. Natl. Acad. Sci. USA* **84**, 7413-7417.

Gaudet, P., Pilcher, K. E., Fey, P. and Chisholm, R. L. (2007). Transformation of *Dictyostelium discoideum* with plasmid DNA. *Nat. Protoc.* **2**, 1317-1324.

Graham, F. L. and van der Eb, A. J. (1973). A new technique for the assay of infectivity of human adenovirus 5 DNA. *Virology* **52**, 456-467.

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

Grosjean, F., Bertschinger, M., Hacker, D. L. and Wurm, F. M. (2006). Multiple glycerol shocks increase the calcium phosphate transfection of non-synchronized CHO cells. *Biotechnol. Lett.* **28**, 1827-1833.

Guo, L., Wang, L., Yang, R., Feng, R., Li, Z., Zhou, X., Dong, Z., Gharthey-Kwansah, G., Xu, M. M., Nishi, M. et al. (2017). Optimizing conditions for calcium phosphate mediated transient transfection. *Saudi J. Biol. Sci.* **24**, 622-629.

Jordan, M. and Wurm, F. (2004). Transfection of adherent and suspended cells by calcium phosphate. *Methods* **33**, 136-143.

Jordan, M., Schallhorn, A. and Wurm, F. M. (1996). Transfecting mammalian cells: optimization of critical parameters affecting calcium-phosphate precipitate formation. *Nucleic Acids Res.* **24**, 596-601.

King, N. (2004). The unicellular ancestry of animal development. *Dev. Cell* **7**, 313-325.

Knoll, A. H. (2011). The multiple origins of complex multicellularity. *Annu. Rev. Earth Planet. Sci.* **39**, 217-239.

Lerche, K. and Hallmann, A. (2013). Stable nuclear transformation of *Eudorina elegans*. *BMC Biotechnol.* **13**, 1-19.

Lerche, K. and Hallmann, A. (2014). Stable nuclear transformation of *Pandorina morum*. *BMC Biotechnol.* **14**, 1-16.

Levin, T. C., Greaney, A. J., Wetzel, L. and King, N. (2014). The rosetteless gene controls development in the choanoflagellate *S. rosetta*. *eLife* **3**, e04070.

Marshall, W. L., Celio, G., McLaughlin, D. J. and Berbee, M. L. (2008). Multiple isolations of a culturable, motile Ichthyosporean (Mesomycetozoa, Opisthokonta), *Creolimax fragrantissima* n. gen., n. sp., from marine invertebrate digestive tracts. *Protist* **159**, 415-433.

Maurer-Stroh, S., Eisenhaber, B. and Eisenhaber, F. (2002). N-terminal N-myristoylation of proteins: refinement of the sequence motif and its taxon-specific differences. *J. Mol. Biol.* **317**, 523-540.

Nagai, T., Ibata, K., Park, E. S., Kubota, M., Mikoshiba, K. and Miyawaki, A. (2002). A variant of yellow fluorescent protein with fast and efficient maturation for cell-biological applications. *Nat. Biotechnol.* **20**, 87-90.

Nellen, W., Silan, C. and Firtel, R. A. (1984). DNA-mediated transformation in *Dictyostelium discoideum*: regulated expression of an actin gene fusion. *Mol. Cell. Biol.* **4**, 2890-2898.

Nichols, S. A., Roberts, B. W., Richter, D. J., Fairclough, S. R. and King, N. (2012). Origin of metazoan cadherin diversity and the antiquity of the classical cadherin/β-catenin complex. *Proc. Natl. Acad. Sci. USA* **109**, 13046-13051.

R Core Team (2016). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. URL: <https://www.R-project.org/>.

Richter, D. J. and King, N. (2013). The genomic and cellular foundations of animal origins. *Annu. Rev. Genet.* **47**, 509-537.

Riedl, J., Crevenna, A. H., Kessenbrock, K., Yu, J. H., Neukirchen, D., Bista, M., Bradke, F., Jenne, D., Holak, T. A., Werb, Z. et al. (2008). Lifeact: a versatile marker to visualize F-actin. *Nat. Methods* **5**, 605-607.

Rokas, A. (2008). The origins of multicellularity and the early history of the genetic toolkit for animal development. *Annu. Rev. Genet.* **42**, 235-251.

Ruiz-Trillo, I., Burger, G., Holland, P. W. H., King, N., Lang, B. F., Roger, A. J. and Gray, M. W. (2007). The origins of multicellularity: a multi-taxon genome initiative. *Trends Genet.* **23**, 113-118.

Schiedmeier, B., Schmitt, R., Müller, W., Kirk, M. M., Gruber, H., Mages, W. and Kirk, D. L. (1994). Nuclear transformation of *Volvox carteri*. *Proc. Natl. Acad. Sci. USA* **91**, 5080-5084.

Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., Schmid, B. et al. (2012). Fiji: an open-source platform for biological-image analysis. *Nat. Methods* **9**, 676-682.

Schultheiss, K. P., Suga, H., Ruiz-Trillo, I. and Miller, W. T. (2012). Lack of Csk-mediated negative regulation in a unicellular Src Kinase. *Biochemistry* **51**, 8267-8277.

Sebé-Pedrós, A., Roger, A. J., Lang, F. B., King, N. and Ruiz-trillo, I. (2010). Ancient origin of the integrin-mediated adhesion and signaling machinery. *Proc. Natl. Acad. Sci. USA* **107**, 10142-10147.

- Sebé-Pedros, A., de Mendoza, A., Lang, B. F., Degnan, B. M. and Ruiz-Trillo, I.** (2011). Unexpected repertoire of metazoan transcription factors in the unicellular holozoan *Capsaspora owczarzaki*. *Mol. Biol. Evol.* **28**, 1241-1254.
- Sebé-Pedros, A., Ariza-Cosano, A., Weirauch, M. T., Leininger, S., Yang, A., Torruella, G., Adamski, M., Adamska, M., Hughes, T. R., Gómez-Skarmeta, J. L. et al.** (2013a). Early evolution of the T-box transcription factor family. *Proc. Natl. Acad. Sci. USA* **110**, 16050-16055.
- Sebé-Pedros, A., Irimia, M., del Campo, J., Parra-Acero, H., Russ, C., Nusbaum, C., Blencowe, B. J. and Ruiz-Trillo, I.** (2013b). Regulated aggregative multicellularity in a close unicellular relative of metazoa. *eLife* **2**, e01287.
- Sebé-Pedros, A., Ballaré, C., Parra-Acero, H., Chiva, C., Tena, J. J., Sabidó, E., Gómez-Skarmeta, J. L., Di Croce, L. and Ruiz-Trillo, I.** (2016a). The dynamic regulatory genome of *Capsaspora* and the origin of animal multicellularity. *Cell* **165**, 1224-1237.
- Sebé-Pedros, A., Peña, M. I., Capella-Gutiérrez, S., Antó, M., Gabaldón, T., Ruiz-Trillo, I. and Sabidó, E.** (2016b). High-Throughput proteomics reveals the unicellular roots of animal phosphosignaling and cell differentiation. *Dev. Cell* **39**, 186-197.
- Sebé-Pedros, A., Degnan, B. M. and Ruiz-Trillo, I.** (2017). The origin of metazoa: a unicellular perspective. *Nat. Rev. Genet.* **18**, 498-512.
- Sigal, C. T., Zhou, W., Buser, C. A., McLaughlin, S. and Resh, M. D.** (1994). Amino-terminal basic residues of Src mediate membrane binding through electrostatic interaction with acidic phospholipids. *Proc. Natl. Acad. Sci. USA* **91**, 12253-12257.
- Suga, H. and Ruiz-Trillo, I.** (2013). Development of ichthyosporeans sheds light on the origin of metazoan multicellularity. *Dev. Biol.* **377**, 284-292.
- Suga, H., Dacre, M., de Mendoza, A., Shalchian-Tabrizi, K., Manning, G. and Ruiz-Trillo, I.** (2012). Genomic survey of premetazoans shows deep conservation of cytoplasmic Tyrosine Kinases and multiple radiations of receptor Tyrosine Kinases. *Sci. Signal.* **5**, ra35.
- Suga, H., Chen, Z., de Mendoza, A., Sebé-Pedros, A., Brown, M. W., Kramer, E., Carr, M., Kerner, P., Vervoort, M., Sánchez-Pons, N. et al.** (2013). The *Capsaspora* genome reveals a complex unicellular prehistory of animals. *Nat. Commun.* **4**, 2325.
- Torruella, G., de Mendoza, A., Grau-Bové, X., Antó, M., Chaplin, M. A., del Campo, J., Eme, L., Pérez-Cordón, G., Whipps, C. M., Nichols, K. M. et al.** (2015). Phylogenomics reveals convergent evolution of lifestyles in close relatives of animals and fungi. *Curr. Biol.* **25**, 2404-2410.

## Supplementary Materials and Methods

### Transfection of *Capsaspora* using Calcium-phosphate precipitation

Adherent stage cells at the exponential growth phase were obtained after a two-day culture passage as follows. Two days before transfection,  $1 \times 10^7$  cells were seeded in a 25 cm<sup>2</sup> culture flask containing 5 mL growth medium and grown overnight at 23°C (Fig. 1A-1). Hereafter, all amounts are indicated per well.

At day 0,  $2 \times 10^6$  cells were seeded from this previous confluent culture to attain 90-95% cell confluence at the time of transfection. Cells were seeded in a 12-well plate (Nunc/DDBioLab #55428) containing 600  $\mu$ L growth medium and grown overnight at 23°C (Fig. 1A-2). Cell concentration was determined using a Neubauer Chamber Hemocytometer (DDBiolab #900505).

**▲ CRITICAL STEP: Adherent stage cells in confluency.** *Cultures should be fresh to maximize transfection efficiency. Ideally, they should be maintained weekly, and used for transfection at their exponential growth phase. Do not let cultures reach higher cell densities ( $< 5 \times 10^7$  cells mL<sup>-1</sup>).*

At day 1, growth medium was replaced by 600  $\mu$ L of transfection medium (see Reagent preparation), and incubated for 30 min at room temperature ( $\sim 18^\circ\text{C}$ ) (Fig. 1B-3). During incubation, 1.271 pmols of plasmid DNA for single transfection experiments or 0.636 pmols of each plasmid DNA for co-transfection experiments were diluted in sterile distilled water up to 120  $\mu$ L plus an additional volume of 150  $\mu$ L of 2X HBS Buffer. Next, 30  $\mu$ L of 1.25 M CaCl<sub>2</sub> were added dropwise while flickering the tube carefully, reaching a final DNA mix volume of 300  $\mu$ L. DNA mix was inverted immediately two times to ensure proper mixing of reagents and incubated 10 min at 37°C (Fig. 1B-4). After incubation, transfection medium was removed and the DNA mix was added dropwise in the centre of the wells. Cells:DNA mix were incubated for 30 min at 18°C (Fig. 1B-5).

**▲ CRITICAL STEP: DNA-Calcium-phosphate precipitates formation.** *Check the cultures periodically under the microscope to check crystal size. Big cloudy precipitates may compromise transfection efficiency. Instead, verify that small grains of refractant material are spread homogeneously in the plate.*

After this period, an additional volume of 500  $\mu$ L of transfection medium was added and cells were incubated for a minimum of 4 h at 23°C (Fig. 1B-6).

**▲ NOTE: Transfection medium incubation.** An incubation of less than 4 h yields lower transfection efficiency. This incubation time can be extended to 6 h.

After incubation, the medium was removed and an osmotic shock using 110  $\mu$ L 10% (v/v) glycerol in 1X HBS Buffer was performed, pouring the solution dropwise all over the well for one min at  $\sim$ 18°C (Fig. 1B-7).

**▲ CRITICAL STEP: Glycerol shock.** *Incubation with glycerol at this concentration should not exceed 1 min, counting from the first droplet, to avoid excessive cell death.*

After the osmotic shock, glycerol solution was removed and cells were grown at 23°C overnight with 700  $\mu$ L of growth medium (Fig. 1B-8). Screening of positive cells was performed 18 h post-transfection using fluorescence microscopy and flow cytometry analysis (Fig. 1C).

**▲ NOTE: Controls.** pONSY (empty) transfected cells, mock-transfected cells and non-transfected cells were used as controls.

## Transfection Reagents preparation

**Growth medium (for 1 L):** 10 g Peptone (BD, #211677), 10 g Yeast Extract (BD, #212750), 1 g Yeast nucleic acid (Ribonucleic Acid, Type VI from Torula Yeast) (Sigma, #R-6625), 15 mg Folic acid (Sigma, #F8758) in 880 mL distilled water. Autoclave for 15 min at 121°C. Cool down and aseptically add 0.4 mL of Hemin stock solution\* (Sigma, #H9039), 20 mL Buffer solution\*\* and 100 mL of heat-inactivated Fetal Bovine Serum (Sigma, #F9665-100ml). Filter-sterilise through 0.22 µm and store at 4°C.

**\*Hemin stock solution (for 200 mL):** 400 mg NaOH in 200 mL dH<sub>2</sub>O. Add 500 mg of Hemin and autoclave 20 min at 121°C. Store at 4°C protected from the light.

**\*\*Buffer solution (for 1 L):** 18.1 g KH<sub>2</sub>PO<sub>4</sub> (Sigma, #P5655), 25 g Na<sub>2</sub>HPO<sub>4</sub> (Sigma, #S5136) in 1 L distilled water. Adjust final pH to 6.5 with HCl 37% and filter-sterilise through 0.22 µm. Store at 4°C.

**Transfection medium (for 1 L):** 10 g Peptone, 15 mg Folic Acid in 990 mL distilled water. Autoclave for 20 min at 121°C. Aseptically add 10 mL HEPES 1 M (Sigma, #H4034) to a final concentration of 10 mM and 2.1 g Bis-Tris methane (Sigma, #B9754) final concentration 0.21% w/w. adjust pH to 7.1 with NaOH, filter-sterilise through 0.22 µm and store at 4°C.

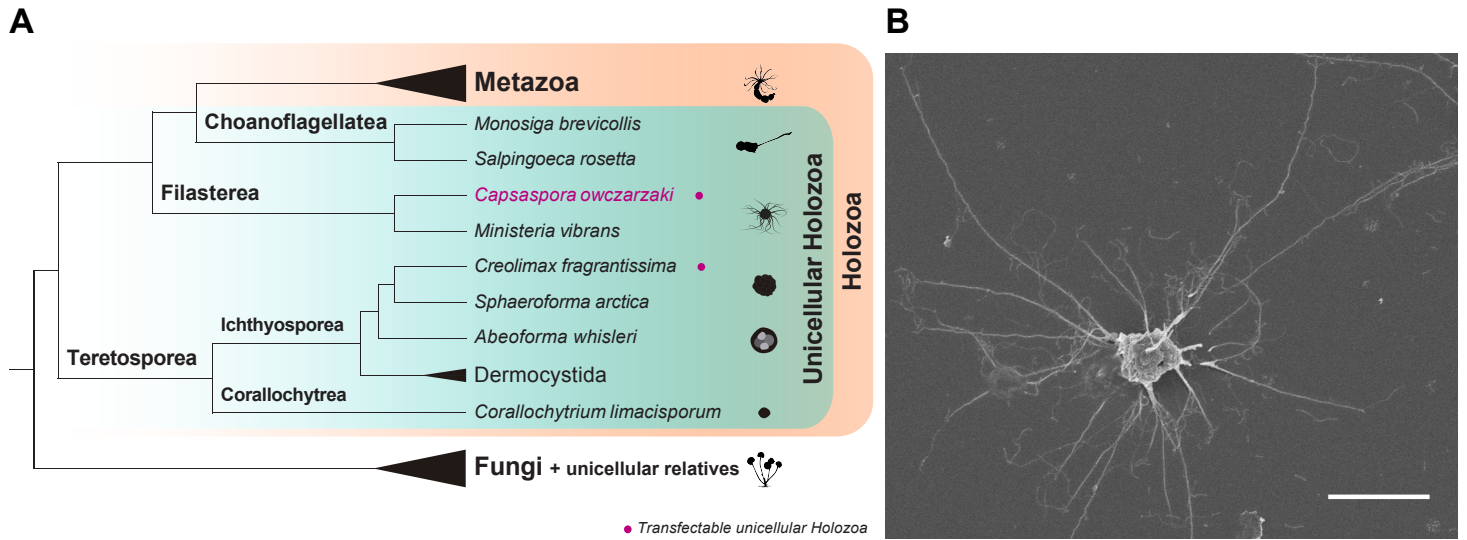
**2X HBS (for 250 mL):** Dissolve 4 g NaCl (Sigma, #S3014), 0.18 g KCl (Sigma, #P9541), 0.05 g Na<sub>2</sub>HPO<sub>4</sub> (Sigma, #S5136), 2.5 g HEPES and 0.5 g D-glucose (Sigma, #G8270) in autoclaved distilled water. Adjust pH to 7.1 with NaOH. Filter-sterilise through 0.22 µm, flash-freeze with liquid Nitrogen and store at -80°C.

**1.25M CaCl<sub>2</sub> (for 10 mL):** 1.84 g CaCl<sub>2</sub> (Sigma, #C1016) in 10 mL autoclaved distilled water. Filter-sterilise through 0.22 µm, flash-freeze with liquid Nitrogen and store at -80°C.

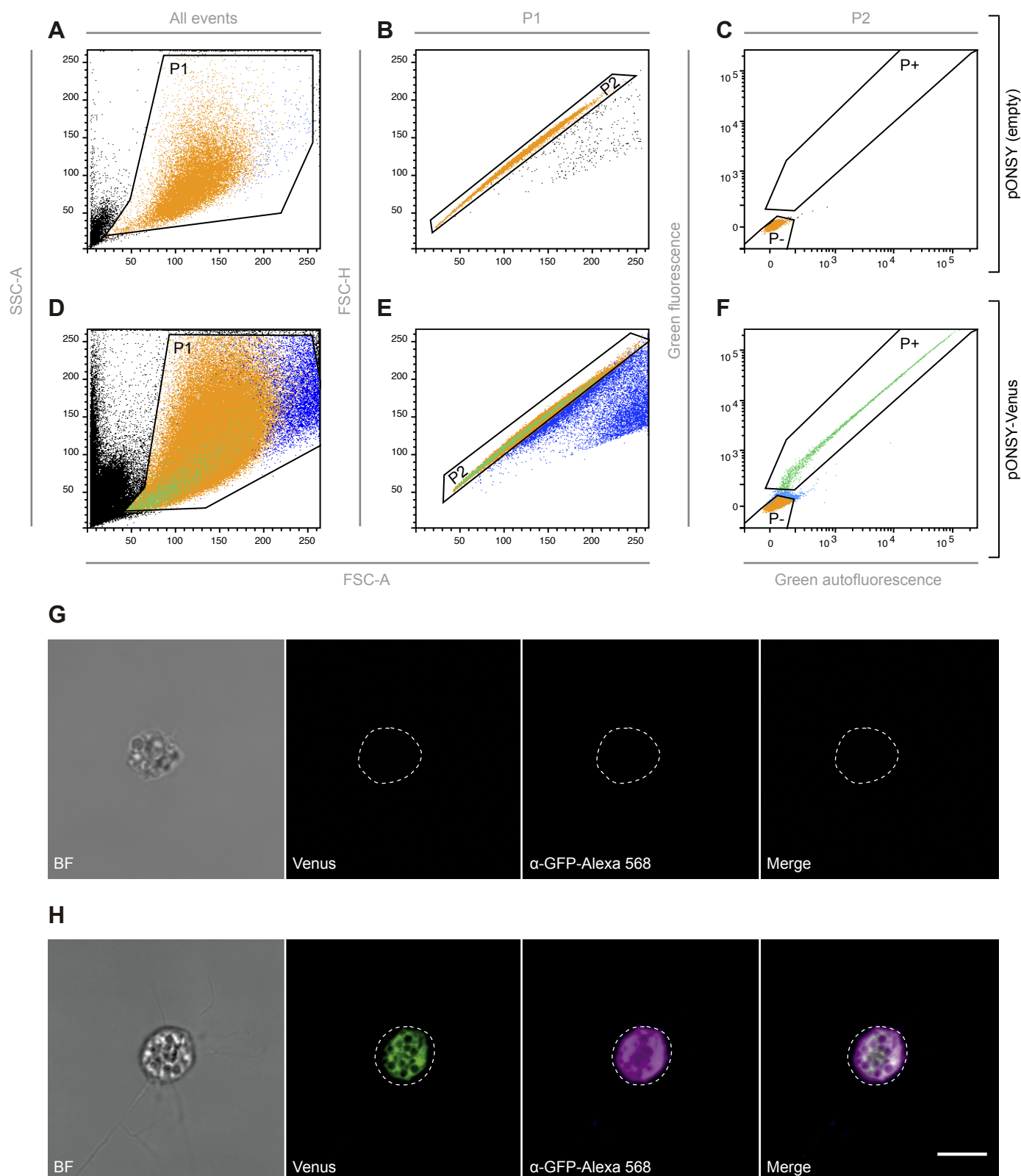
**10% glycerol (for 4 mL):** 0.8 mL of filter-sterilised 50% (v/v) glycerol (Sigma, #G7757) in 1.2 mL autoclaved distilled water and 2 mL 2X HBS. Filter-sterilise through 0.22 µm, flash-freeze with liquid Nitrogen and store at -80°C.



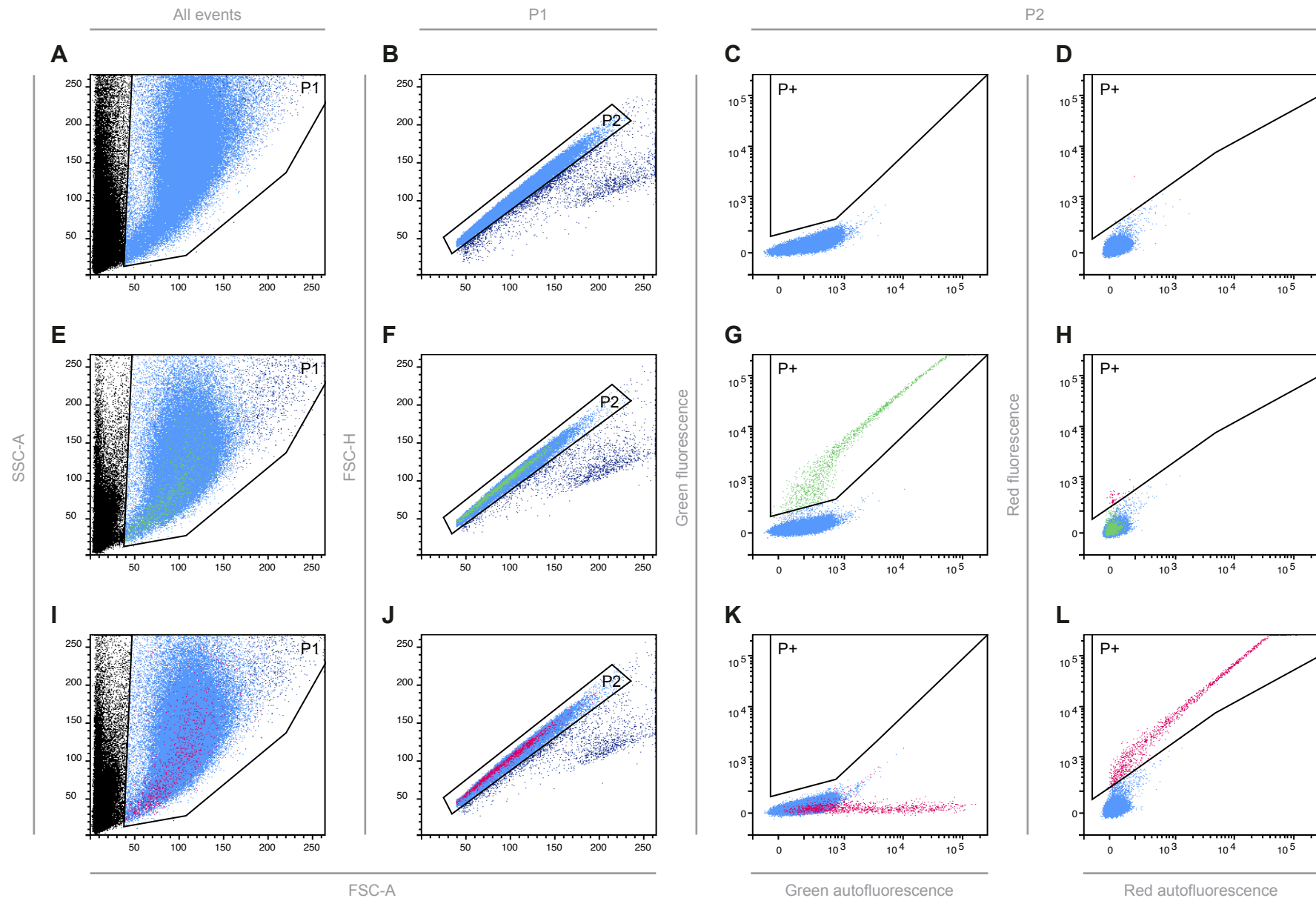
## Supplementary Figures



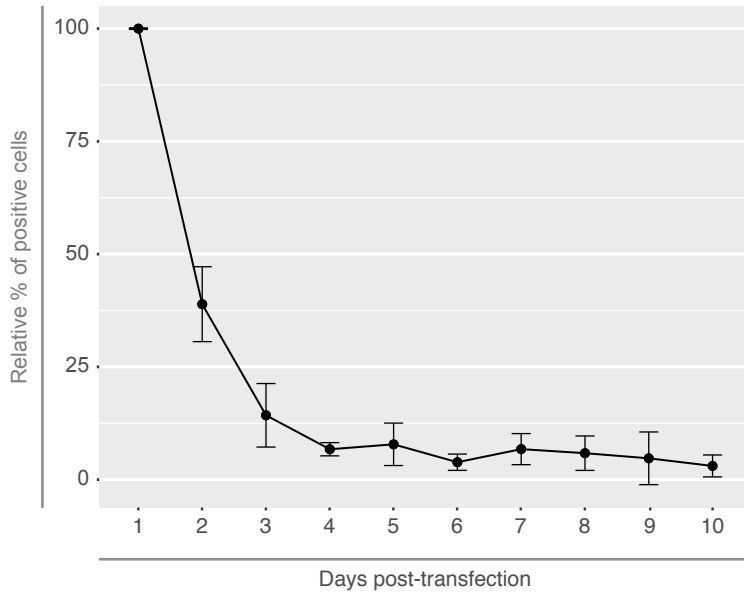
**Fig. S1. Transfectable unicellular Holozoa and *Capsaspora owczarzaki*.** (A) Metazoa and their unicellular relatives; Choanoflagellata, Filasterea and Teretosporea, comprise the Holozoa clade. Transfectable unicellular Holozoa to date are *C. fragrantissima* and *C. owczarzaki*. (B) SEM image of a *Capsaspora* cell. Scale bar represents 5  $\mu\text{m}$ .



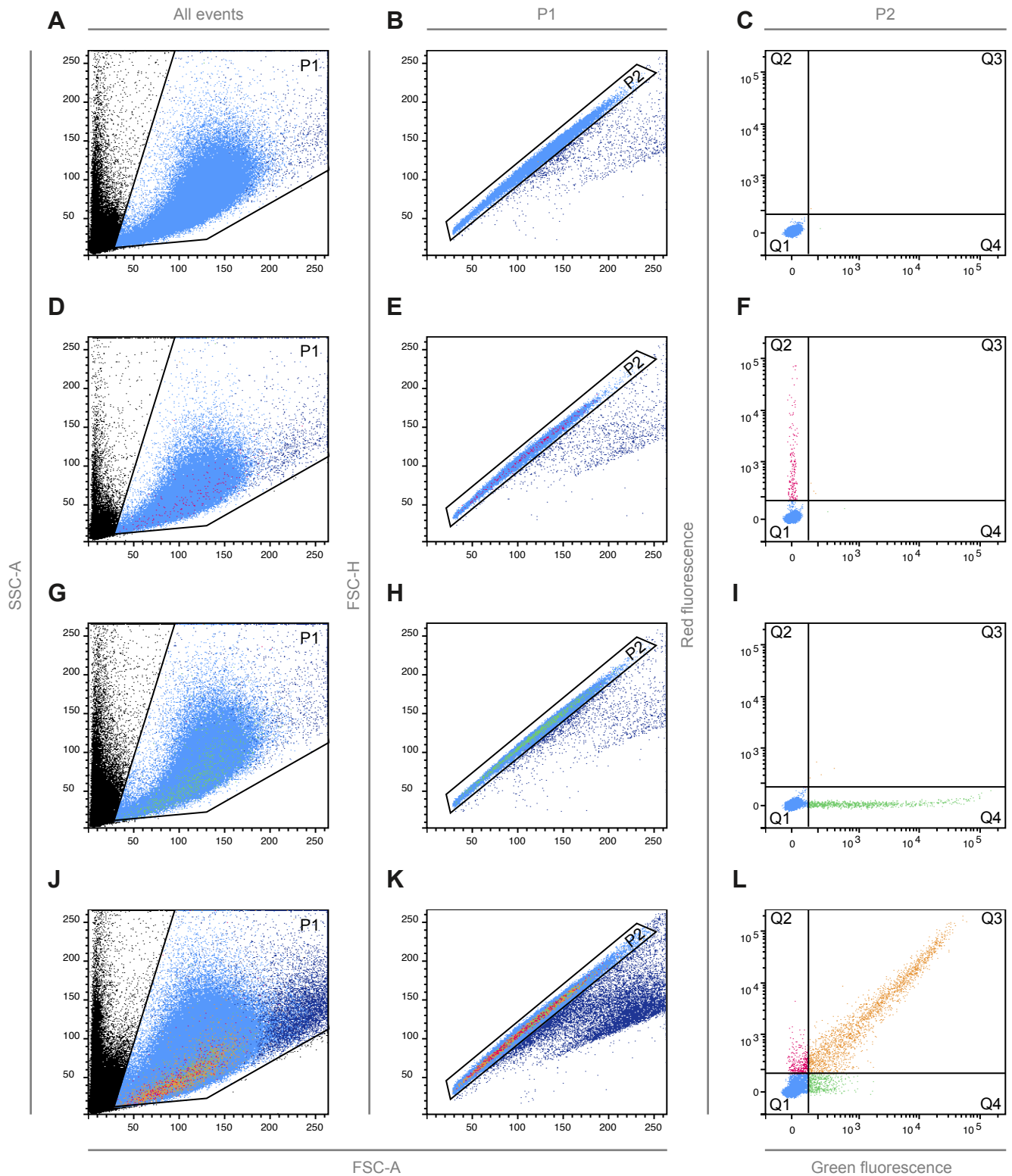
**Fig. S2. FACS of *Capsaspora* transfected cells and immunofluorescence validation.** (A-C) Cells transfected with pONSY (empty) as control to gate positive and negative populations. (D-F) Cells transfected with pONSY-Venus. Areas selected in (A) and (D) define total population of cells (P1). Areas selected in (B) and (E) define single cells (P2). Areas in (C) and (F) define sorted Venus positive cells (P+) and sorted Venus negative cells (P-), respectively. (G-H) Immunofluorescence validation of Venus expression of P- (G) and P+ (H) sorted populations from (F) using an anti-GFP antibody. Dashed line indicates cell body. Scale bar represents 5  $\mu$ m.



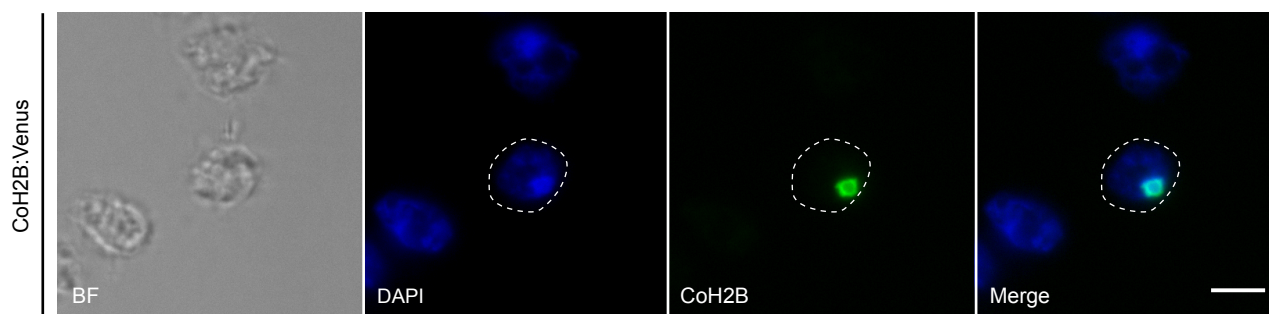
**Fig. S3. Flow cytometry analysis of *Capsaspora* transfected cells.** (A-D) Cells transfected with pONSY (empty) as control to gate positive and negative populations. (E-H) Cells transfected with pONSY-Venus. (I-L) Cells transfected with pONSY-mCherry. Areas selected in A, E and I define total population of cells (P1). Areas selected in B, F and J define single cells (P2). P+ in C, G and K defines positive cells in the green channel (Venus). P+ in D, H and L defines positive cells in the red channel (mCherry). Figure associated to Fig. 3A-B.



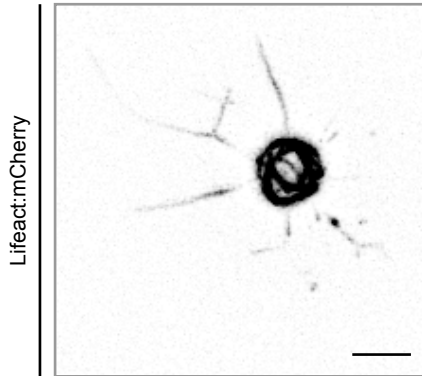
**Fig. S4: Persistence of positive cells along 10 days after transfection.** Percentage of positive cells transfected with pONSY-Venus, measured every 24h by flow cytometry (number of positive cells at day 1 was considered as 100%). Error bars represent s.d. Figure associated to Table S2.



**Fig. S5: *Capsaspora* co-transfected with both pONSY-mCherry and pONSY-Venus.** (A-C) Cells transfected with pONSY (empty) as control. (D-F) Cells transfected with pONSY-mCherry only. (G-H) Cells transfected with pONSY-Venus only. (J-L) Cells co-transfected with both pONSY-mCherry and pONSY-Venus. Areas selected in panels A,D,G and J define total population of cells (P1). Areas selected in B, E, H, and K define single cells (P2). Quartiles define negative cells (Q1), red fluorescent cells expressing mCherry only (Q2), cells expressing both fluorescent proteins (Q3) and green fluorescent cells expressing Venus only (Q4). Figure associated to Fig. 3F.

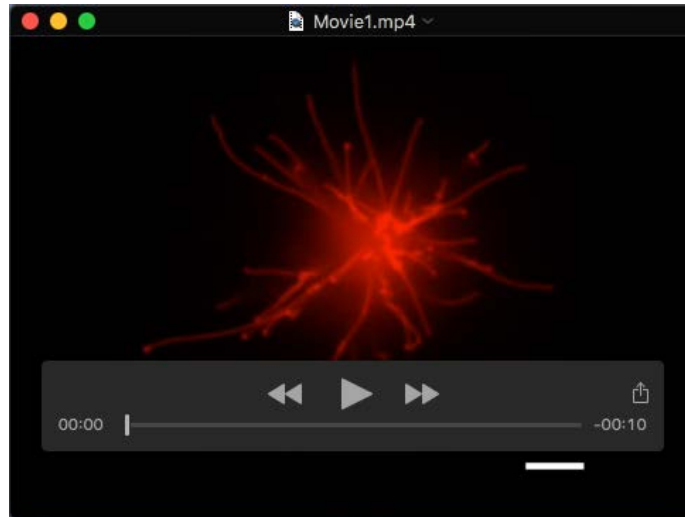


**Fig. S6. Localisation of nuclear marker in *Capsaspora* transfected cells.** Transfected cells with pONSY-CoH2B:Venus stained with DAPI. Dashed line indicates cell body. Scale bar represents 5  $\mu$ m.

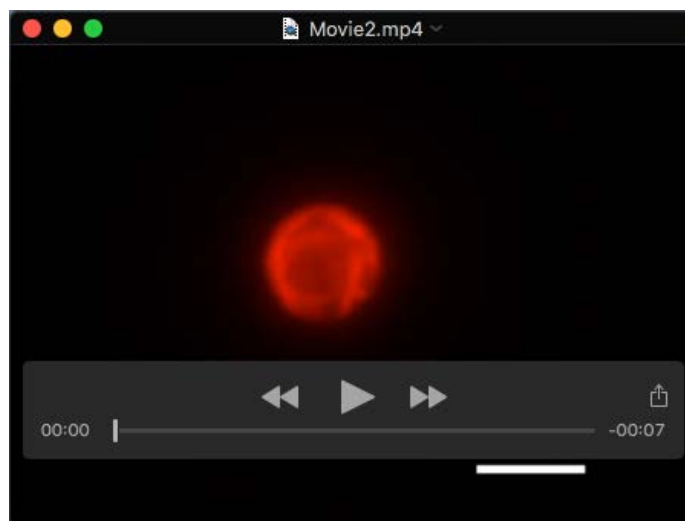


**Fig. S7. Labelling the actin cytoskeleton and filopodia in *Capsaspora*.** Transfected cell with pONSY-Lifeact:mCherry from Fig. 2C'. Image saturated and inverted to improve visualization of filopodia. Scale bar represents 5  $\mu$ m.

## Supplementary movies



**Movie 1. *Capsaspora* filopodia dynamics *in vivo*.** Time-lapse of a cell transfected with pONSY:CoNMM-mCherry. Images were taken every second during 100 seconds. Scale bar represents 5  $\mu$ m.



**Movie 2. *Capsaspora* actin cytoskeleton *in vivo*.** Time-lapse of a cell transfected with pONSY:Lifeact-mCherry. Images were taken every 10 minutes during 130 minutes. Scale bar represents 5  $\mu$ m.



## Supplementary Tables

### Tables S1-S4

**Table S1. Flow Cytometry analysis of *Capsaspora* cells transfected with a single vector.** Flow cytometry analysis of *Capsaspora* cells transfected with pONSY-Venus (1-7a) or pONSY-mCherry (7b) expression vectors. Results from 7 independent experiments with at least 6 replicates each (n=51) are shown. Transfection efficiency is calculated as the ratio of total number of positive cells (P+) from total number of cells (P2) and represented as mean $\pm$ s.d per experiment. Table associated to Fig. 3A-D and Fig. S3.

Experiment		Number of cells		Transfection efficiency	
Number	Sample	Total (P2)	Positive (P+)	(P+/P2)%	mean $\pm$ s.d.
1	Empty vector	100083	0	0.000	0.347 $\pm$ 0.193
	Replicate 1	100152	370	0.369	
	Replicate 2	100036	512	0.512	
	Replicate 3	100147	633	0.632	
	Replicate 4	100302	219	0.218	
	Replicate 5	99930	150	0.150	
	Replicate 6	100055	200	0.200	
2	Empty vector	100000	0	0.000	2.083 $\pm$ 0.248
	Replicate 1	47180	1139	2.414	
	Replicate 2	52604	1178	2.239	
	Replicate 3	91753	1632	1.779	
	Replicate 4	100000	2114	2.114	
	Replicate 5	100000	2146	2.146	
	Replicate 6	100000	1807	1.807	

3	Empty vector	100000	0	0.000	-
	Replicate 1	100000	814	0.814	
	Replicate 2	100000	1332	1.332	
	Replicate 3	100000	673	0.673	
	Replicate 4	100000	820	0.820	
	vReplicate 5	100000	950	0.950	0.859±0.227
	Replicate 6	100000	827	0.827	
	Replicate 7	100000	669	0.669	
	Replicate 8	100000	1051	1.051	
	Replicate 9	100000	596	0.596	
4	Empty vector	100003	6	0.006	-
	Replicate 1	100000	1250	1.250	
	Replicate 2	100000	1091	1.091	
	Replicate 3	100000	1103	1.103	
	Replicate 4	100000	1049	1.049	1.047±0.140
	Replicate 5	100000	849	0.849	
	Replicate 6	100000	938	0.938	

5	Empty vector	100229	9	0.009	-
	Replicate 1	100206	1048	1.046	
	Replicate 2	100062	1352	1.351	
	Replicate 3	100070	1160	1.159	1.204±0.128
	Replicate 4	100075	1368	1.367	
	Replicate 5	100067	1123	1.122	
	Replicate 6	100055	1182	1.181	
6	Empty vector	100000	0	0.000	-
	Replicate 1	100000	944	0.944	
	Replicate 2	100000	472	0.472	
	Replicate 3	100000	1681	1.681	1.399±0.621
	Replicate 4	100000	1802	1.802	
	Replicate 5	100000	2182	2.182	
	Replicate 6	100000	1315	1.315	
7a	Empty vector	12243	0	0.000	-
	Replicate 1	86084	1134	1.317	
	Replicate 2	100000	559	0.559	
	Replicate 3	100000	1469	1.469	1.159±0.326
	Replicate 4	100000	1376	1.376	
	Replicate 5	100000	1129	1.129	
	Replicate 6	100000	1101	1.101	
7b	Empty vector	12243	0	0.000	-
	Replicate 1	100000	980	0.980	
	Replicate 2	100000	1151	1.151	
	Replicate 3	100000	1284	1.284	1.094±0.148
	Replicate 4	100000	1118	1.118	
	Replicate 5	100000	1163	1.163	
	Replicate 6	100000	865	0.865	

**Table S2. Flow Cytometry analysis of *Capsaspora* transfected cells during 10 days.** Cells transfected with pONSY-Venus expression vector analysed every 24 hours during 10 days after transfection. Results from 3 independent experiments are shown. Transfection efficiency was calculated as the ratio of total number of positive cells (P+) from total number of cells (P2). Ratio of positive cells was calculated as the percentage of positive cells in a particular day relative to the percentage of positive cells at day 1 and represented as mean $\pm$ s.d per day. Table associated to Fig. S4.

Experiment		Number of cells		Transfection efficiency		
Days post-transfection	Sample	Total (P2)	Positive (P+)	(P+/P2)%	Ratio %	mean $\pm$ s.d.
1	Empty vector	21427	0	0.000	-	-
	Replicate 1	295873	313	0.106	100.000	
	Replicate 2	99871	338	0.338	100.000	100.000 $\pm$ 0.000
	Replicate 3	99968	79	0.079	100.000	
2	Empty vector	99944	0	0.000	-	-
	Replicate 1	99962	32	0.032	30.260	
	Replicate 2	99955	134	0.134	39.612	38.898 $\pm$ 8.303
	Replicate 3	100000	37	0.037	46.820	
3	Empty vector	99962	1	0.001	-	-
	Replicate 1	100000	22	0.022	20.796	
	Replicate 2	99953	23	0.023	6.799	14.260 $\pm$ 7.044
	Replicate 3	100000	12	0.012	15.185	
4	Empty vector	100000	0	0.000	-	-
	Replicate 1	100000	8	0.008	7.562	
	Replicate 2	99705	17	0.017	5.038	6.731 $\pm$ 1.466
	Replicate 3	100000	6	0.006	7.593	
5	Empty vector	100000	0	0.000	-	-
	Replicate 1	100000	8	0.008	7.562	
	Replicate 2	99942	11	0.011	3.252	7.823 $\pm$ 4.706
	Replicate 3	100000	10	0.010	12.654	

6	Empty vector	100000	0	0.000	-	-
	Replicate 1	100000	5	0.005	4.726	
	Replicate 2	99946	6	0.006	1.744	3.854±1.809
	Replicate 3	100000	4	0.004	5.062	
7	Empty vector	100000	0	0.000	-	-
	Replicate 1	100000	11	0.011	10.398	
	Replicate 2	99973	12	0.012	3.547	6.757±3.446
	Replicate 3	100000	5	0.005	6.327	
8	Empty vector	100000	0	0.000	-	-
	Replicate 1	100000	9	0.009	8.508	
	Replicate 2	99947	5	0.005	1.478	5.859±3.822
	Replicate 3	100000	6	0.006	7.593	
9	Empty vector	235982	0	0.000	-	-
	Replicate 1	100000	12	0.012	11.343	
	Replicate 2	99983	1	0.001	0.296	4.723±5.841
	Replicate 3	100000	2	0.002	2.531	
10	Empty vector	186504	1	0.001	-	-
	Replicate 1	100000	6	0.006	5.672	
	Replicate 2	99709	3	0.003	0.889	3.031±2.430
	Replicate 3	100000	2	0.002	2.531	

**Table S3. Flow Cytometry analysis of *Capsaspora* cells co-transfected with pONSY-Venus and pONSY-mCherry.** Results from 7 independent experiments with 6 replicates each (n=42) are shown. Transfection efficiency is calculated by total number of positive cells (Q2+Q3+Q4) from total number of cells (P2) and represented as mean±s.d per experiment. Relative percentages of Double, Venus and mCherry expression were calculated as number of double positive cells (Q2) or number of Venus positive cells (Q4) or number of mCherry positive cells (Q3) from total number of positive cells (Q2+Q3+Q4), respectively, and represented as mean±s.d per experiment. Table associated to Fig. 3F and G and Fig. S5.

Experiment		Number of cells						Transfection efficiency		Relative % over total number of positive cells		
Number	Sample	Total (P2)	Negative (Q1)	mCherry (Q2)	Double (Q3)	Venus (Q4)	Total positive (Q2+Q3+Q4)	% Total positive	% Double (Q3/P2)%	% Double	% Venus	% mCherry
1	Empty vector	100000	99996	1	2	1	4	0.004				
	Control Venus	100000	98177	1	2	1820	1823	1.823				
	Control mCherry	100000	99689	310	1	0	311	0.311				
	Replicate 1	520488	516878	706	2346	558	3610	0.694	0.451	64.986	15.457	19.557
	Replicate 2	358753	355200	731	2354	468	3553	0.990	0.656	66.254	13.172	20.574
	Replicate 3	322368	319598	501	1818	451	2770	0.859	0.564	65.632	16.282	18.087
	Replicate 4	408411	404855	620	2339	597	3556	0.871	0.573	65.776	16.789	17.435
	Replicate 5	426129	422602	592	2357	578	3527	0.828	0.553	66.827	16.388	16.785
	Replicate 6	380788	377503	501	2305	479	3285	0.863	0.605	70.167	14.581	15.251
		Mean±s.d.							0.851±0.095	0.567±0.068	66.607±1.850	15.445±1.364
2	Empty vector	100000	99988	2	5	5	12	0.012				
	Control Venus	100000	98648	7	2	1343	1352	1.352				
	Control mCherry	100000	98879	1116	2	3	1121	1.121				
	Replicate 1	209708	206792	191	2247	478	2916	1.391	1.071	77.058	16.392	6.550
	Replicate 2	249962	247044	230	2217	471	2918	1.167	0.887	75.977	16.141	7.882
	Replicate 3	231155	228463	180	2154	358	2692	1.165	0.932	80.015	13.299	6.686
	Replicate 4	342982	340145	222	2220	395	2837	0.827	0.647	78.252	13.923	7.825
	Replicate 5	231457	228666	222	2203	366	2791	1.206	0.952	78.932	13.114	7.954
	Replicate 6	348075	345308	233	2167	367	2767	0.795	0.623	78.316	13.263	8.421
		Mean±s.d.							1.092±0.233	0.852±0.179	78.091±1.416	14.355±1.508
3	Empty vector	100677	100662	4	9	2	15	0.015				
	Control Venus	100712	99079	5	8	1620	1633	1.621				
	Control mCherry	100673	99677	990	5	1	996	0.989				
	Replicate 1	374093	371723	186	1881	303	2370	0.634	0.503	79.367	12.785	7.848
	Replicate 2	380307	377928	145	1940	294	2379	0.626	0.510	81.547	12.358	6.095
	Replicate 3	292266	289888	132	1972	274	2378	0.814	0.675	82.927	11.522	5.551
	Replicate 4	218046	215653	112	1910	371	2393	1.097	0.876	79.816	15.504	4.680
	Replicate 5	279028	276622	138	1937	331	2406	0.862	0.694	80.507	13.757	5.736
	Replicate 6	179012	176633	140	1892	347	2379	1.329	1.057	79.529	14.586	5.885
		Mean±s.d.							0.894±0.275	0.719±0.215	80.616±1.386	13.419±1.480

4	Empty vector	100000	100000	0	0	0	0	0				
	Control Venus	100000	99020	5	3	972	980	0.980				
	Control mCherry	100000	99794	205	1	0	206	0.206				
	Replicate 1	459409	456258	383	2359	409	3151	0.686	0.513	74.865	12.980	12.155
	Replicate 2	412692	409512	375	2388	417	3180	0.771	0.579	75.094	13.113	11.792
	Replicate 3	837030	833626	480	2436	488	3404	0.407	0.291	71.563	14.336	14.101
	Replicate 4	402481	399267	481	2387	346	3214	0.799	0.593	74.269	10.765	14.966
	Replicate 5	332644	329608	357	2352	327	3036	0.913	0.707	77.470	10.771	11.759
	Replicate 6	582324	579008	502	2398	416	3316	0.569	0.412	72.316	12.545	15.139
Mean±s.d.								0.691±0.180	0.516±0.147	74.263±2.118	12.418±1.410	13.319±1.597
5	Empty vector	100000	99986	8	5	1	14	0.014				
	Control Venus	100000	99573	4	2	421	427	0.427				
	Control mCherry	100000	99983	16	1	0	17	0.017				
	Replicate 1	464014	461024	294	2300	396	2990	0.644	0.496	76.923	13.244	9.833
	Replicate 2	425876	423064	259	2300	253	2812	0.660	0.540	81.792	8.997	9.211
	Replicate 3	616927	613484	411	2300	732	3443	0.558	0.373	66.802	21.261	11.937
	Replicate 4	598473	595286	342	2300	545	3187	0.533	0.384	72.168	17.101	10.731
	Replicate 5	491626	488340	355	2300	631	3286	0.668	0.468	69.994	19.203	10.803
	Replicate 6	801753	798352	434	2300	667	3401	0.424	0.287	67.627	19.612	12.761
Mean±s.d.								0.581±0.095	0.425±0.093	72.551±5.805	16.57±4.623	10.879±1.309
6	Empty vector	100262	100248	9	4	1	14	0.014				
	Control Venus	100194	99427	3	0	764	767	0.766				
	Control mCherry	100000	99676	324	0	0	324	0.324				
	Replicate 1	1172129	1168353	697	2517	562	3776	0.322	0.215	66.658	14.883	18.459
	Replicate 2	780326	776622	645	2520	539	3704	0.475	0.323	68.035	14.552	17.414
	Replicate 3	1036442	1032768	645	2513	516	3674	0.354	0.242	68.400	14.045	17.556
	Replicate 4	725427	721735	692	2511	489	3692	0.509	0.346	68.012	13.245	18.743
	Replicate 5	757172	753425	709	2531	507	3747	0.495	0.334	67.547	13.531	18.922
	Replicate 6	697333	693820	643	2460	410	3513	0.504	0.353	70.026	11.671	18.303
Mean±s.d.								0.443±0.083	0.302±0.059	68.113±1.113	13.654±1.148	18.233±0.620
7	Empty vector	12243	12230	7	5	1	13	0.106				
	Control Venus	100000	98740	15	9	1236	1260	1.260				
	Control mCherry	100000	98809	1185	4	2	1191	1.191				
	Replicate 1	161940	160433	322	981	204	1507	0.931	0.606	65.096	13.537	21.367
	Replicate 2	155693	154319	172	1000	202	1374	0.883	0.642	72.780	14.702	12.518
	Replicate 3	123025	121624	186	1000	215	1401	1.139	0.813	71.378	15.346	13.276
	Replicate 4	163206	162117	174	783	132	1089	0.667	0.480	71.901	12.121	15.978
	Replicate 5	258248	256791	233	1000	224	1457	0.564	0.387	68.634	15.374	15.992
	Replicate 6	122506	121096	202	1000	208	1410	1.151	0.816	70.922	14.752	14.326
Mean±s.d.								0.889±0.240	0.624±0.173	70.119±2.826	14.305±1.260	15.576±3.165

**Table S4. List of primers used to build *Capsaspora* expression vectors with reporter genes.** Restriction enzymes sites are underlined. CoNMM sequence plus 7 extra aminoacids is highlighted in red.

Region/Gene	<i>Capsaspora</i> gene ID	Primer name	Sequence 5'-3'
<i>CoEF1<math>\alpha</math></i> promoter	CAOG_07807	1	CTGGTACCAAATGCACAGTTAGCAACGACC
		2	<u>GATATCACTAGTCCC</u> GGGATCCTGTGAAGGTTGTTCTG
		3	AAATGCACAGTTAGCAACGACC
<i>CoEF1<math>\alpha</math></i> terminator	CAOG_07807	4	GAGCTGTACAAGTAAATTTTGTGTTTGCCAAG
		5	CATTGCTAGTGCTGTTCTCACC
		6	GACCGCGGTGAGAACAGCACTAGCAATG
		7	<u>CCCGGGACTAGTGATATCTGAATTTTGTGTTTGCCAAGACAC</u>
		8	CGCCAGTGTGATGGATTGAAAGCTTCCGCGGTGA
<i>mCherry/Venus</i>	-	9	<u>CCCGGGACTAGTGATATCATGGT</u> GAGCAAGGGCG
		10	CTTGGCAAACACAAAATTTACTTGTACAGCTC
<i>CoSrc2</i> NMM	CAOG_06360	11	TATAC <u>CCGGGATGGGCTGCTCCA</u> ACTCTAAACCGCACGACCCGTCGGATTCAAGGTTTCCCTTCTGGCGTTGCGTCCAACAGCATGGTGAGCAAGGGCGAGGAG
		12	TTACTTGTACAGCTCGTCCATG
<i>CoH2B</i>	CAOG_01818	13	TAC <u>CCGGGATGCCGCCGA</u> AGGTC
		14	TAACTAGTCTTGGCGCCGGAGGT
Lifeact	-	15	<u>CCCGGGACCATGGGTGTGGCAGACCTGATTAAGAAGTTCGAGAGCATT</u>
		16	<u>TCTAGATGGTGGGTCA</u> CCCTCCTCCTTGCTAATGCTCTCGAACTTCTT





## **R4: A first approach into the function of the integrin adhesome of *Capsaspora owczarzaki***

### *Abstract*

Adhesion of animal cells to their extracellular matrix is primarily mediated by integrins. Integrins connect the ECM to the cytoskeleton by interacting with numerous cytoplasmic proteins (adaptors), that jointly form an integrin adhesion complex (IAC). Homologs of the proteins that form the core of an IAC have been found in unicellular relatives of animals, indicating that they were present in the unicellular ancestor of animals. *Capsaspora owczarzaki* is one of the closest unicellular relatives of animals that contains homologs of IAC core proteins and it can develop multicellular aggregative life-forms. Additionally, aggregates of *C. owczarzaki* show a significant increase in gene expression of the IAC homologs. In this study we localised some proteins of the integrin adhesome in *Capsaspora* by immunostaining and performed adhesion assays to analyse adhesive properties over different substrates. The results suggest that the integrin adhesome proteins could have a role in adhesion mediated by filopodia. Understanding the role that the integrin adhesome in *Capsaspora* will help to infer the function in the unicellular ancestor of animals and thus help to decipher its role in emergence of animal multicellularity.



## **A first approach into the function of the integrin adhesome of *Capsaspora owczarzaki***

Parra-Acero, H., Dudin, O., Harcet M., Ruiz-Trillo, I.

### ***Abstract***

Adhesion of animal cells to their extracellular matrix is primarily mediated by integrins. Integrins connect the ECM to the cytoskeleton by interacting with numerous cytoplasmic proteins (adaptors), that jointly form an integrin adhesion complex (IAC). Homologs of the proteins that form the core of an IAC have been found in unicellular relatives of animals, indicating that they were present in the unicellular ancestor of animals. *Capsaspora owczarzaki* is one of the closest unicellular relatives of animals that contains homologs of IAC core proteins and it can develop multicellular aggregative life-forms. Additionally, aggregates of *C. owczarzaki* show a significant increase in gene expression of the IAC homologs. In this study we localised some proteins of the integrin adhesome in *Capsaspora* by immunostaining and performed adhesion assays to analyse adhesive properties over different substrates. The results suggest that the integrin adhesome proteins could have a role in adhesion mediated by filopodia. Understanding the role that the integrin adhesome in *Capsaspora* will help to infer the function in the unicellular ancestor of animals and thus help to decipher its role in emergence of animal multicellularity.

### **Introduction**

The major mechanism that animal cells use to establish cell-matrix adhesions are mediated by heterodimeric transmembrane proteins called integrins (Hynes 2002). Integrins and adaptor proteins form an integrin-adhesion complex (IAC) that links to the cytoskeleton (Zaidel-Bar et al. 2007, Klapholz & Brown 2017). The set of proteins that have been identified in integrin adhesions is called the integrin adhesome (Zaidel-Bar et al. 2007, Horton et al. 2016). Integrins can bind different ligands and its specificity depends on the alpha and beta subunit combination (Barczyk et al. 2010; Maartens & Brown 2015). The most studied ligand is the RGD motif firstly found in fibronectin (Pierschbacher & Ruoslahti 1984), but other motifs in fibronectin have also been identified (Pankov & Yamada 2002). There are other components of the ECM that animal integrins can recognise, among them, laminin and collagen (Humphries et al. 2006).

The essential elements that are part of integrin-adhesion complexes existed already before animals emerged, as they have been found in different unicellular Opisthokonta lineages and Apusomonadida (Sebé-Pedrós et al. 2010). Integrin alpha and beta subunits, talin and some more adaptor proteins were found for the first time outside the animal lineage in the filasterean *Capsaspora owczarzaki* (from

herein *Capsaspora*) and the apusomonad *Thecamonas trahens* (previously known as *Amastigomonas sp.*). The analyses also showed that integrins and some adaptors seem to have been secondarily lost from Fungi and choanoflagellates (Sebé-Pedrós et al. 2010). The finding that the overall structure of *Capsaspora* and *Thecamonas* integrins is so well conserved (including the functional domains) and that the adaptor proteins are present, suggested that they could work in a similar way as the animal integrins (Sebé-Pedrós et al. 2010). The same elements were later found in the ichthyosporean *Creolimax fragrantissima* (de Mendoza et al. 2015), and recently in the other filastereans *Pigoraptor vietnamica* and *P. chiliana* (Hehenberger et al. 2017). In any case, given its phylogenetic position, *Capsaspora* is an ideal organism in which to analyse the function of integrins from an evolutionary point of view.

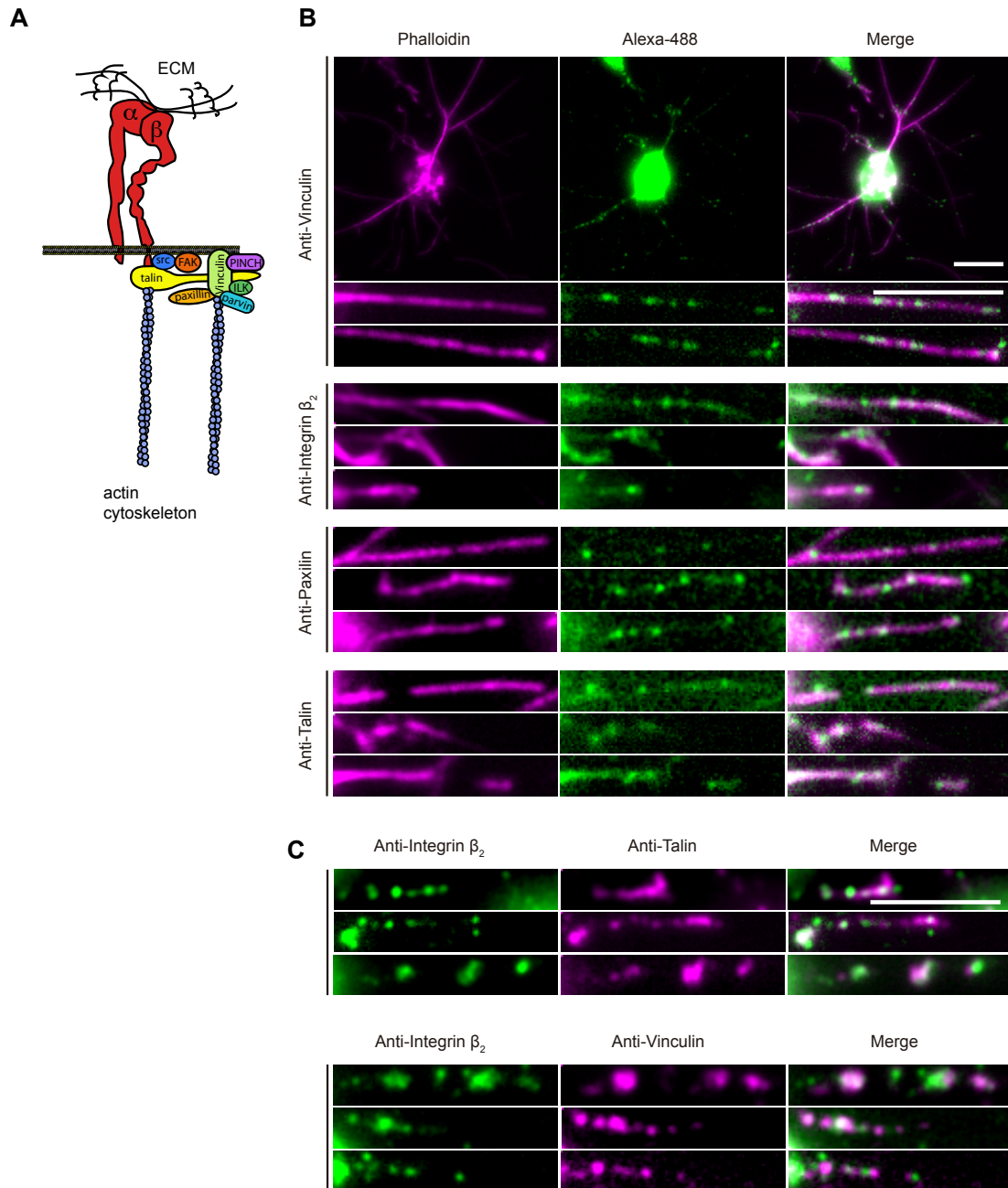
Moreover, *Capsaspora*, under culture conditions, forms aggregates at a specific stage of its lifecycle. Aggregates are essentially not clonal, but formed by different cells that actively join each other and stick together (Sebé-Pedrós, Irimia, et al. 2013). Despite not presenting true homologs for animal ECM proteins, some fibronectin and laminin domains have been found in *Capsaspora* proteins (Suga et al. 2013), which are surprisingly more expressed in the aggregative stage (Sebé-Pedrós, Irimia, et al. 2013). The expression levels of most of the integrin adhesome components were also found to be higher in the aggregates than in the other stages (i.e. adherent and cystic) (Sebé-Pedrós, Irimia, et al. 2013), a finding that accentuates the idea of the putative adhesive function of the integrin adhesome components in *Capsaspora*. However, the actual function of the adhesome components in *Capsaspora* remains unclear. Knowing this function will provide clues into the potential function of the pre-metazoan integrin adhesome.

To address this question, we performed experiments to try to decipher the function of the integrin adhesome in *Capsaspora*. We focused on adherent cells, since they are easier to handle. We performed localisation assays and set up a protocol to measure adhesion capacity of *Capsaspora* adherent cells.

## **Results**

### **Components of the integrin adhesome localise in the filopodia of *Capsaspora***

To address the role of integrin adhesome components in *Capsaspora*, we decided to localise them in *Capsaspora* cells in the first place. To this, we raised specific polyclonal antibodies against multiple components of the conserved integrin adhesome. We successfully obtained functional antibodies against integrin  $\beta 2$  (CAOG\_5058), talin (CAOG\_09384), paxillin (CAOG\_06505) and vinculin (CAOG\_05123) (Fig. S1, Table 1). Using these antibodies, we then performed immunostaining of *Capsaspora* in the adherent stage.



**Figure 1. Localisation of *Capsaspora* integrin adhesome proteins.**

(A) Schematic integrin adhesion complex in animals. (B) Immunostaining on adherent *Capsaspora* cells labelled with phalloidin (magenta) and antibodies for vinculin, integrinB, paxillin and talin. (C) Co-immunostaining to localise talin (magenta) or vinculin (magenta) with integrin B (green). Scale bars=5  $\mu$ m.

Interestingly, all four antibodies localise in the cell body and as specific dots in the filopodia (Fig. 1B). Due to the high fluorescence intensity of the antibodies in the cell body, we chose to show only filopodia to avoid contrast issues. We observed that localisation of all four components of the integrin adhesome co-localised with actin (phalloidin staining) present in *Capsaspora* filopodia. This localisation showed a particular pattern of multiple distant patches all over the length of the filopodia (Fig. 1B).

In animals, the different components of the integrin adhesome co-localise together at cell-matrix adhesion sites (Zamir & Geiger 2001). We thus assessed if components of the integrin adhesome in *Capsaspora* also localise together. For this, we performed immunostaining of *Capsaspora* adherent cells using combinations of these antibodies. We observed that co-localisation occurs at some points (Fig. 1C). Altogether, these results show that the integrin adhesome proteins in *Capsaspora* have a specific localisation in the filopodia.

### **Filopodia mediate cell-substrate adhesion in *Capsaspora***

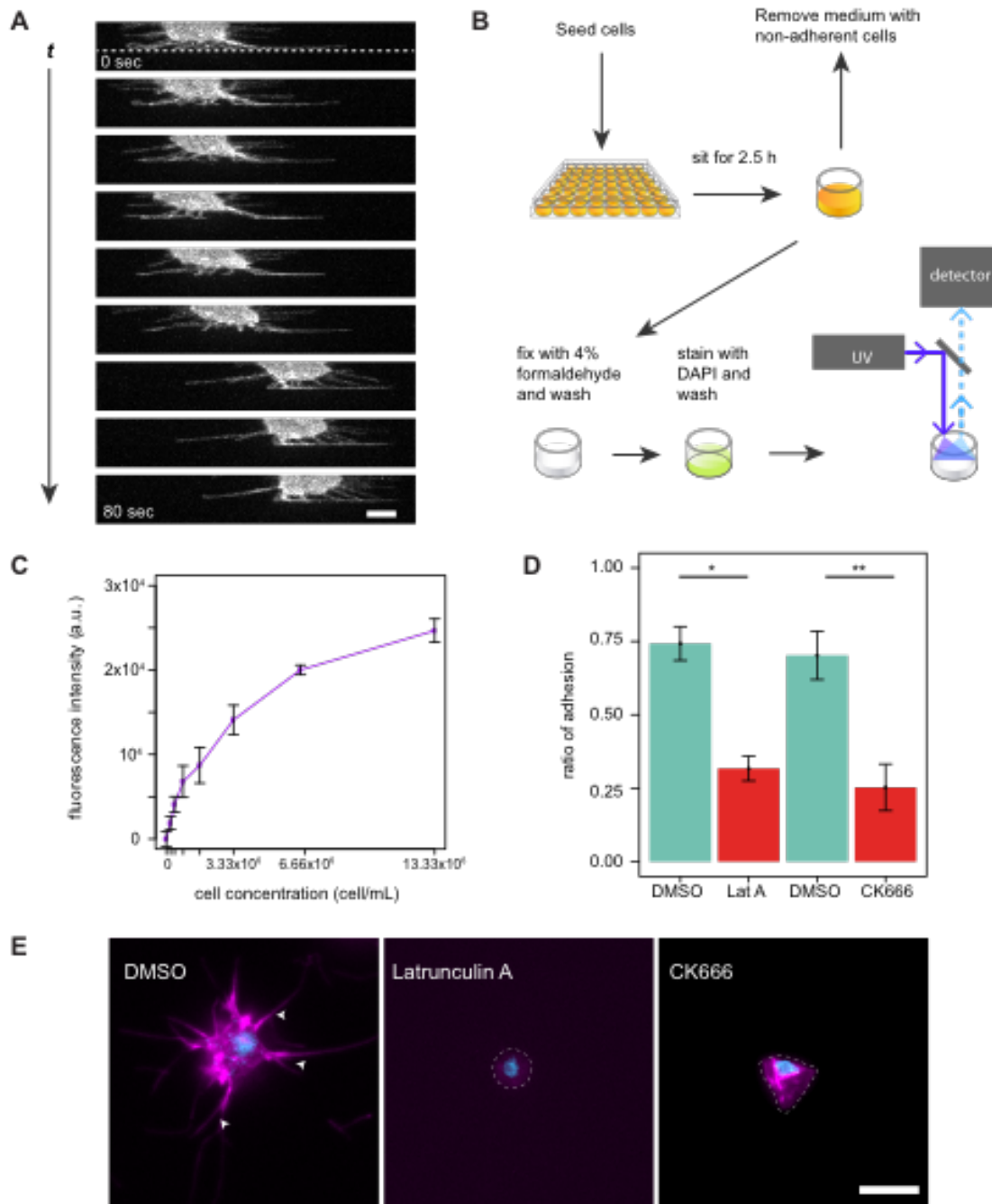
It is important to mention that filopodia are a major element of *Capsaspora* morphology. Indeed, in culture conditions, the cell body of *Capsaspora* is usually in suspension, and only the filopodia remain in direct contact with the surface ("Results I" section). We performed a time-lapse video with *Capsaspora* transfected with the membrane marker (NMM-mcherry) previously used and observed that filopodia not only sustain the cell over a surface but participate in its locomotion (Fig. 2A, S2). Such particularity places the filopodia as the primary interactor with the surface, and thus are candidates to be playing a role in adhesion.

To test whether filopodia act as an adhesion structure in *Capsaspora*, we developed an adhesion assay that allows us to measure the capacity of cells to adhere to a surface (Fig. 2B). This assay is inspired by previously published methods (Pierschbacher & Ruoslahti 1984; Busk et al. 1992). Briefly, it consists of fixing adherent cells on a multi-well plate after discarding non-adherent ones by removing the medium and using DAPI staining as a proxy for the number of cells that remain adhered (Fig. 2B, Materials and Methods). In order to test the sensitivity of this method, we fixed and measured fluorescence of increasing concentrations of cells. As expected, fluorescence intensity increased as we increased the concentration of cells (Fig. 2C). We also observed that with concentrations below  $1.6 \times 10^6$  cell/ml, the fluorescence signal decreased rapidly and approached zero, whereas with concentrations higher than  $6.6 \times 10^6$  cell/ml, the signal approached saturation, as the surface of the well got completely covered by cells (Fig. 2C and S3). To assess with confidence a decrease or an increase in cell adhesion we chose  $3.3 \times 10^6$  cell/ml as a starting cell concentration for further assays.

The actin cytoskeleton is intimately connected to IACs in animals. Filopodia are formed by actin filaments, and many of the actin regulators involved in filopodia formation, including Arp2/3, are found in *Capsaspora* (Sebé-Pedrós, Burkhardt, et al. 2013). In order to assess the similarity between cell-matrix adhesion in animals and cell-substrate adhesion in *Capsaspora*, we tested in *Capsaspora* the role of actin in adhesion. For that, we used commercially available inhibitors to disrupt the actin cytoskeleton. We observed that total depolymerisation of actin using Latrunculin A (Lat A) disrupted cell shape. Indeed, cells became round, small and lacked filopodia as well as showed no signal of actin filaments (Fig. 2E). On the other hand, inhibition of Arp2/3-dependent actin polymerization using CK666, did not disrupt actin filaments in the cell body, but it disrupted filopodia, suggesting that filopodia in *Capsaspora* are Arp2/3 dependent.

We then performed an adhesion assay in presence of LatA or CK666 to test whether filopodia are necessary for adhesion in *Capsaspora*. We performed three independent replicates and normalised the results to adhesion on untreated cells (adhesion set as 1). We observed that, in presence of 0.05 mM of Lat A, or 0.1 mM CK666, adhesion was reduced respect to its DMSO control (Fig. 2D). These results suggest that actin, including actin on filopodia is indeed necessary for cell adhesion in *Capsaspora*. Altogether, we successfully established a novel method to assess cell-substrate adhesion in *Capsaspora*, and using this method, we demonstrated that filopodia mediate the cell-substrate adhesion in *Capsaspora*.





### Figure 2. Adhesion of *Capsaspora*

(A) Time-lapse of a *Capsaspora* cell transfected with a membrane label (NMM-mcherry). Dotted line indicates surface. An image was taken every 10 seconds. (B) Scheme of the adhesion assay. (C) Fluorescence intensity measures of an adhesion assay on increasing cell concentration on untreated plates. (D) Ratio of adhesion of cells treated with several drugs relative to the adhesion of untreated cells. Bars represent mean  $\pm$  s.e.m. (n=3). \* p-value = 0.05, \*\* p-value < 0.05 (paired t-test). (E) Phalloidin (magenta) and DAPI (cyan) staining of floating cells from an adhesion assay under treatment with DMSO, 0.05mM LatA or 0.1 mM CK666. Arrowheads indicate filopodia. Scale bar= 5 $\mu$ m

## **Fibronectin is an adhesion substrate for *Capsaspora***

In animals, integrin dependent cell-matrix adhesion is mediated by specific interactions with ECM components, the most commonly studied being fibronectin, laminin and collagen (Geiger & Yamada 2011). To continue assessing similarity between animal cell-matrix adhesion and cell-substrate adhesion in *Capsaspora*, we tested whether *Capsaspora* has any adhesive preference for any of these three potential substrates. We performed an adhesion assay in plates previously treated with increasing concentrations of each of the three substrates (fibronectin, laminin or collagen). We also included bovine serum albumin (BSA) as another coating protein, because it is usually used in adhesion assays with animal cells as an agent to block unspecific adhesion sites (Yamada & Kennedy 1984; Bourdon & Ruoslahti 1989; Busk et al. 1992; Weinreb et al. 2004).

For easier visualisation of the effect of the coatings on adhesion ability, we normalised the fluorescence measures to that obtained with cells seeded on plastic without coating (adhesion set as 1, Fig. 3A). We observed that coating with increasing concentration of BSA progressively reduced adhesion of *Capsaspora* cells compared to untreated, indicating BSA blocked adhesion to plastic. Interestingly, we observed that coating with increasing concentration of fibronectin also increased cell adhesion. On the other hand, increasing laminin and collagen in the coating caused a decrease in adhesion compared to untreated surface. Repetition of the experiments yielded similar results (Fig. S4). This result indicates that fibronectin is recognised by *Capsaspora* as a substrate for adhesion, whereas laminin, collagen or BSA are not.

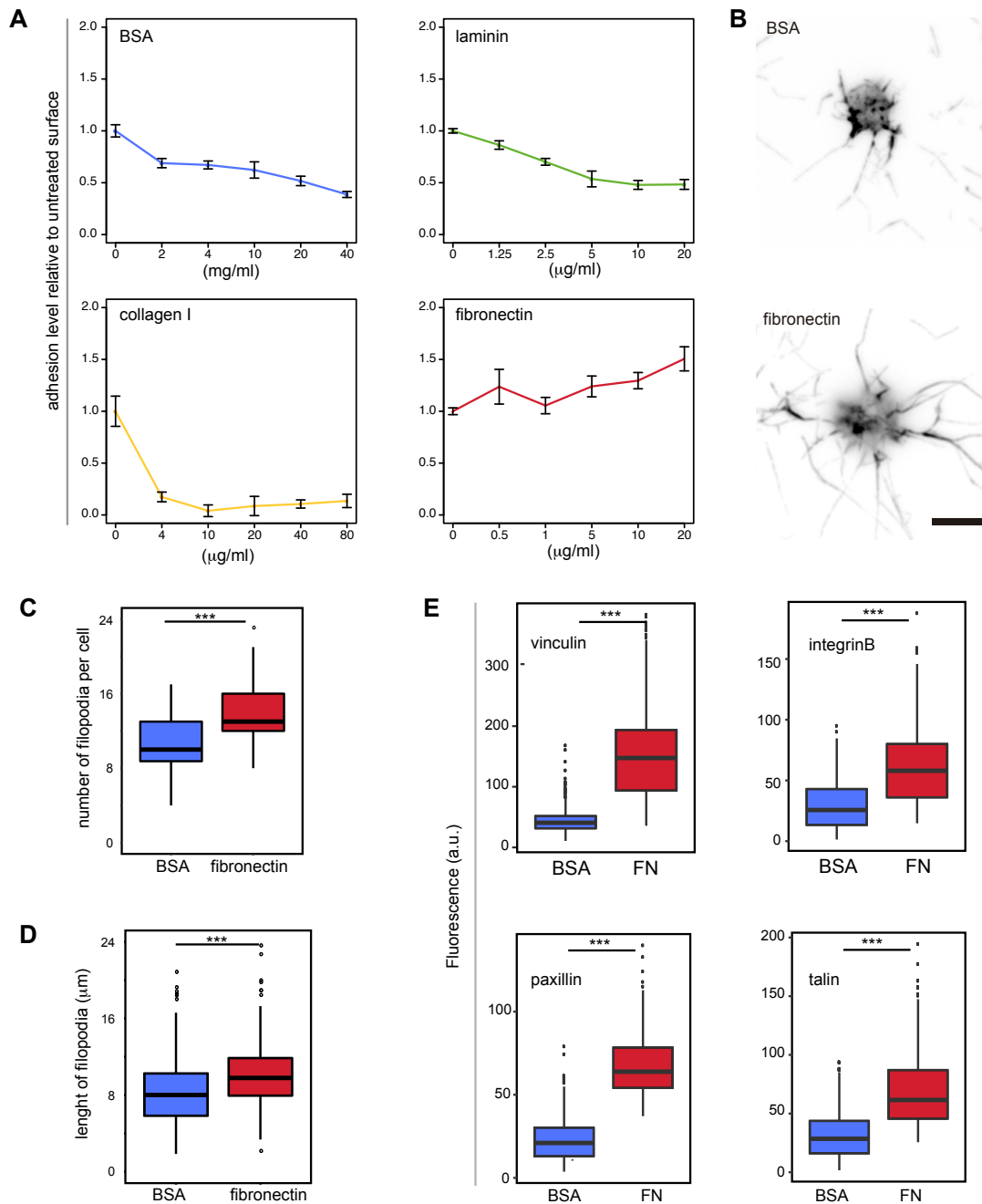
## **Fibronectin favors filopodia attachment and increases localisation of adhesome proteins in filopodia**

Given that filopodia are the adhesive structure in *Capsaspora*, we decided to check whether the morphology of the filopodia changes in presence of fibronectin (FN). For this, we measured the number and length of filopodia on fixed adherent cells in presence of either FN or BSA (Fig.3B). We observed that, in presence of FN, the number of the filopodia attached to the substrate was a higher than in BSA (Fig. 3C). Similarly, in presence of FN, the filopodia were longer than in BSA (Fig. 3D). This suggests that fibronectin promotes the anchoring of the filopodia to the surface.

Next, we wondered whether fibronectin would affect as well the localisation of any of the components of the integrin adhesome in *Capsaspora*. To check that, we immunostained cells plated on BSA or fibronectin with each of the antibodies separately (vinculin, integrin  $\beta$ , talin, paxillin) (Fig. 3E). Localisation was dotted in both cases (not shown). We then measured the overall fluorescence intensity along at least a 100 filopodia for each antibody, and we normalised it to the length of each filopodium. Surprisingly, we observed that the overall fluorescence intensity of the

integrin adhesome components was significantly higher on cells that were plated on fibronectin compared to cells plated on BSA. This increase in intensity indicates that more of these components of the integrin adhesome are recruited to filopodia in presence of fibronectin, which suggests that they are related to the anchoring role of filopodia to fibronectin.

Taken together, all of our results suggest that integrin adhesome components in *Capsaspora owczarzaki* play a pivotal role in cell-substrate adhesion, which share similarities with integrin-mediated cell-matrix adhesion in animals.

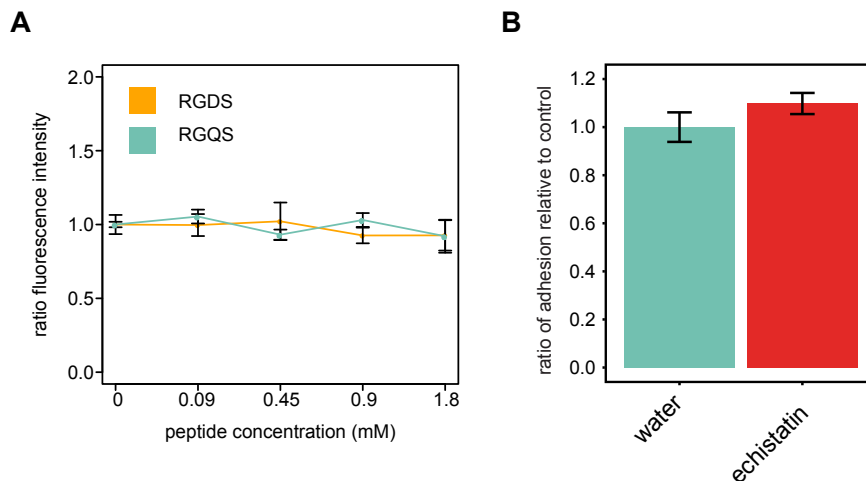


**Figure 3. Adhesion of *Capsaspora* on substrate coated with different proteins.**

(A) Adhesion assay on plates treated with different concentrations of BSA, laminin, collagen I and fibronectin. The data represent the mean  $\pm$  s.e.m (n=4). For each experiment, background is removed and data is normalised to the “untreated” condition for easier visualisation. (B) Phalloidin staining of cells seeded on fibronectin or BSA treated coverslip. Scale bar= 5µm. (C) Filopodia length in cells seeded on different substrate. (n=500). Mann-Whitney test p-value < 2.2e-16 (D) Number of filopodia per cell seeded on different substrate. (n=164). Mann-Whitney test p-value < 2.2e-16

## Adhesion to fibronectin is not mediated by an RGD motif as in animals

The tripeptide RGD is a main ligand of integrins in animals, which is found in proteins such as fibronectin (Humphries et al. 2006). Adding small soluble peptides containing this motif impairs adhesion to fibronectin-treated plates (Pierschbacher & Ruoslahti 1984). This motif is also found in some disintegrins, e.g. Echistatin, which are proteins found in snake venom that also impair integrin adhesion (Gan et al. 1988; Kumar et al. 1997). We thus tested whether RGD motif is involved in the recognition of fibronectin by *Capsaspora*, using these previously described inhibitors. We performed an adhesion assay on adherent cells in presence or absence of these molecules in plates previously treated with fibronectin (Fig. 4). The peptide RGDS, as well as a control peptide (RGQS) did not impair adhesion at the concentrations used (Fig. 4A). The disintegrin Echistatin, at the concentration used, neither altered the adhesion ability to fibronectin (Fig. 4B). This suggests that adhesion to fibronectin is not mediated by RGD recognition in *Capsaspora*.



**Figure 4. Adhesion capacity in presence of inhibitors of integrins**

(A) Adhesion to plate treated with fibronectin (20 $\mu$ g/ml) in presence of different concentrations of RGDS or RGQS (control) peptide. Data is mean  $\pm$  s.e.m. (n=4) (B) Adhesion to plate treated with fibronectin (20 $\mu$ g/ml) in presence of 0.25mM Echistatin . Data represents mean  $\pm$  s.e.m (n=4).

## **Discussion**

Overall, the results here presented suggest that the integrin adhesome of *Capsaspora* has a role in adhesive function as in animals.

In animals, integrin adhesome proteins were first identified by their localisation to focal adhesions, an area in close contact with the surface of cells in culture and associated with actin filaments. Focal adhesions are usually found localised ventrally, although in migrating cells, adhesions start assembling (nascent adhesions) at the leading edge, on filopodia and lamellipodia, which are formed by actin filaments. As a first step to assess the function of the adhesome proteins in *Capsaspora* we looked at their cell localisation. Antibody staining revealed specific dots in filopodia colocalising with actin, as well as all over the cell body. At the same time, we showed that filopodia represent the structure that provides adhesion in *Capsaspora*.

Thus, given that filopodia are the adhesion structure in *Capsaspora* and that integrin and integrin-related proteins are localised in filopodia together with actin, we can infer a similar role for the integrin adhesome in *Capsaspora* as in animals. If genome editing is ever developed in *Capsaspora*, knock-out experiments on the integrin adhesome will be key to confirm or not our hypothesis.

Homologues of animal ECM proteins have not been found in *Capsaspora*, nevertheless, the fibronectin domains type II and type III, and laminin domains G1 and G2 have been identified in several proteins (Suga et al. 2013, Hehenberguer et al. 2017) and are also upregulated in the aggregative stage (Sebé-Pedrós, Irimia, et al. 2013). To test whether *Capsaspora* can bind to animal ECM, we used laminin, fibronectin and collagen I. Our experiments showed that laminin and collagen did not affect adhesion, however fibronectin did. Thus, there is a possibility that the fibronectin type-III domain in one of *Capsaspora* proteins is the original ligand for the same adhesion receptor that recognises fibronectin in *Capsaspora*. To test this hypothesis, a plausible experiment will be to add these protein domains produced as recombinant proteins into the medium and test whether they disrupt adhesion to the fibronectin-treated substrate.

Importantly, *Capsaspora*, was originally found inside a snail (Stibbs et al. 1979; Hertel et al. 2002), thus it could be the case that the original capacity to adhere to certain motif that happens to be conserved in animal fibronectin may be used as an opportunity to adhere to a potential fibronectin-like protein in this animal.

The finding of proteins containing fibronectin domains in the *Capsaspora* genome raised the possibility of an ECM produced by *Capsaspora* that maybe interact with the integrin machinery already identified (Suga et al. 2013). Actually, a matrix-like structure was later identified by SEM imaging in aggregates (Sebé-Pedrós, Irimia, et al. 2013), although its composition has not been discovered. It has been found as

well that cells forming aggregates retain their filopodia (Florenza, personal communication), suggesting that filopodia could be as well involved in maintaining the aggregate structure. Our results suggest that the *Capsaspora* integrin adhesome, has a role in adhesion, specifically to fibronectin, and involving filopodia. Therefore, the possibility of a *Capsaspora* protein being secreted and recognised by a receptor via filopodia, appears as a plausible explanation for aggregate formation. If interaction of integrins with a secreted protein indeed occurs, and this interaction allows the clustering of cells, we could tell that *Capsaspora* has a cell-matrix adhesion system. With this, we could infer that the integrins in the ancestor of animals could have had also a cell-matrix adhesion role. Checking whether integrins are involved in aggregate formation would add more light into this.

Recognition of fibronectin, which is an ECM protein in animals, by *Capsaspora*, does not necessarily mean that the motif recognised by *Capsaspora* would be in a protein playing a “matrix” role. The original protein to which *Capsaspora* could bind, might as well be found in a cell-surface protein. If this was the case, we could refer to this protein and the integrin adhesome in *Capsaspora* as a cell-adhesion mechanism, implying that during emergence of animals, an ancestral cell-adhesion system that involves a “fibronectin-like region” could have been co-opted into a matrix-adhesion system. In order to shed light into this, it will be key to elucidate which is the specific region recognised by *Capsaspora* in animal fibronectin and find out whether *Capsaspora* produces proteins containing that region.

Under the hypothesis of *Capsaspora* integrins being involved in adhesion, we checked whether inhibition of adhesion to fibronectin with an RGD containing peptide (RGDS or the disintegrin Echistatin) would happen in *Capsaspora*, as it is the case with animal integrins (Pierschbacher & Ruoslahti 1984). Neither of them affected the binding in the conditions tested.

A potential explanation is that the sequence of the RGD-containing peptide affect the affinity of integrins for it. Some studies have shown that the aminoacids surrounding the RGD motif affect its affinity (Pierschbacher & Ruoslahti 1984; Ozawa et al. 2016; Kapp et al. 2017). Therefore we can not discard the possibility that the affinity for the peptide used is very low and our assay is not sensitive enough to detect it. Then, in order to observe a change in adhesion to fibronectin, maybe the molecule containing the RGD motif has to be a specific one for *Capsaspora*.

Another reason could be that *Capsaspora* integrin has affinity for another motif. Actually, different integrins in animals have been shown to recognise different regions than RGD in the same molecule. For instance, LDV and others motifs in fibronectin (Pankov & Yamada 2002; Humphries et al. 2006), or KQAGDV in fibrinogen (Springer et al. 2008). This means that if fibronectin is being recognised indeed by integrins in *Capsaspora*, it is reasonable to think that this is done by another motif different that RGD. In order to understand how this ligand (fibronectin) is recognised, trying available inhibitors for animal integrins

represents a costly and maybe unsuccessful task. Instead, trying different peptides from fibronectin, in a similar way as experiments performed by Pierschbacher & Ruoslahti 1984 to narrow down the actual region recognised by *Capsaspora*, is the best approach.

The localisation of talin, vinculin and paxillin in *Capsaspora* filopodia, and their increase in localisation on filopodia in presence of fibronectin suggest they have a role in the adhesion to this molecule. However, their potential role as adaptor proteins participating in the link between integrin and the actin cytoskeleton in *Capsaspora* still needs to be assessed. To prove direct interactions among these proteins it is necessary to perform other assays, like Co-IP or protein binding assays.



## **Material and Methods**

### **Cell strain and growth conditions**

*Capsaspora owczarzaki* cell cultures (strain ATCC®30864) were grown axenically in 25 cm<sup>2</sup> culture flasks (Falcon® VWR, #734-0044) with 5 mL ATCC medium 1034 (modified PYNFH medium).

### **Custom polyclonal antibodies for *Capsaspora* proteins**

IntegrinB2-antigen was a polypeptide produced in *E.coli* corresponding to a region in the stalk of the extracellular domain of the integrin B2 (CAOG\_005058). The talin antigen was a polypeptide produced in *E.coli* corresponding to 500 aminoacids in the C-terminus of the talin protein (CAOG\_009384). The vinculin-antigen was a polypeptide corresponding to the C-terminus of one of the vinculin homologs (CAOG\_005123). The paxillin-antigen was the whole protein (CAOG\_006505). Antigen for integrin B was produced and purified by the Biomolecular Screening and Protein Technologies Unit of CRG (Center for Genomic Regulation), and the antibody was produced in guinea-pig by TebuBio. Antigens for antibodies against talin, vinculin and paxillin, and their corresponding polyclonal antibodies (in rat) were produced by Genecust.

All sequences are in supplementary data.

### **Western Blot**

*Capsaspora* cytoplasmic extract was performed as follows: adherent cells were grown as explained above. Around  $5.5 \times 10^8$  cells were pelleted, washed once with PBS 1X and kept at -80°C. Cells were resuspended in lysis buffer (Tris-HCl 50mM pH 8.8, NaCl 150mM, SDS 0.1%, EDTA 5mM, EGTA 1mM, NP-40 1%, MgCl<sub>2</sub> 1mM, CaCl<sub>2</sub> 1mM, DTT 1mM, PMSF 0.5mM, half Complete for 3ml buffer) and kept on ice for 10 min. Then they were sonicated (amplitude 10%, 3 pulses 15 sec, 45 sec between pulses). The extract was centrifuged at 4°C for 30 min at 20,000 x g and the supernatant was kept as the soluble fraction.

For SDS-PAGE, 1-5 µg of purified antigen or 15 µg protein extract were loaded in a precast 4-20% acrylamide gel (BioRad 456-1094). Gel with antigens was run 20 min at 40V and 40 min at 90 V. Gel with cell extract was run 30 min at 40V, 30min at 60V and 30min at 100V. Proteins and antigens were wet-transferred in TGS 1X (Sigma T7777-1L) with methanol 20% to a nitrocellulose membrane (Amersham 10600004) at 30 V 4°C overnight. After transfer, membranes were incubated with blocking solution (Tween 0.1% and 5% powder milk in PBS 1X) at RT for 1 hour. Membranes were incubated with primary antibodies for 2 hours at RT in the following conditions: anti-intB2 1:1,000 in PBS-Tween with 1% milk, anti-talin 1:1,000 in PBS-Tween with 1% milk, anti-vinculin 1:1,000 in PBS-Tween and anti-paxillin 1:100 in PBS-Tween. Then they were washed 4 times with PBS-Tween and incubated with secondary antibodies for 2 hours at RT as follows: anti-guineapig 1:10,000 and anti rat 1:5000 in in PBS-Tween. They were washed again 4 times with

PBS-Tween. Detection was performed by chemiluminescence (SuperSignal™ West Pico Chemiluminescent Substrate 34078 Thermofisher).

### **Membrane labelling**

*Capsaspora* was transfected with the NMM-mcherry construct as in Parra-Acero et al. 2018, and visualised by confocal microscopy in an Andor Revolution XD Spinning Disk microscope equipped with an Andor Ixon 897E Dual Mode EM-CCD camera.. For time-lapse, images were taken as a z-stack, every 10 sec. A maximum intensity projection of an orthogonal view is represented. Images were edited using Fiji Imaging Software version 2.0.0-rc-44/1.50e (Schindelin et al. 2012).

### **Adhesion assay: general protocol**

Adherent confluent cells were obtained by seeding  $1 \times 10^7$  cells in a 25 cm<sup>2</sup> culture flask containing 5 mL growth medium, incubated overnight at 23°C. Cells were then scraped and harvested by centrifugation at 5000 x g and resuspended in medium without fetal bovine serum (FBS) (we observed that plating in presence of FBS showed heterogeneous results). Serial dilutions were performed to test different concentrations of cells. In the rest of the assays, cell concentration was fixed to  $3.3 \times 10^6$  cells mL<sup>-1</sup>. 300 µL of each solution was plated per well in a 48 well/plate. Cells were left to sit for 2.5 hours at 23°C, then, the liquid, containing floating cells, was removed and adherent cells were fixed with 110 µL of 4% formaldehyde for 10 min at room temperature (RT). Fixed cells were washed once with 150 µL of 1X PBS and stained with DAPI 50 µg mL<sup>-1</sup> for 10 min at RT and darkness. After this, cells were washed twice with 150 µL of 1X PBS, then, 150 µL of 1XPBS were added and the fluorescence was immediately detected on a plate reader (TECAN infinite 200). Fluorescence of DAPI (350nm excitation /470nm emission) was read using the i-control software. Detection was performed from the top, set to 5x5 reads per well (circle filled) and gain was set to optimal automatically by the software for each independent experiment.

### **Coating untreated plates with BSA or ECM proteins**

Untreated plates (PLC32048, Labclinics) were coated with different concentration of fibronectin (Sigma F1141-2MG), laminin (Sigma L2020-1MG), collagen I (Sigma C3867-1VL) or BSA (Sigma A3294-10G). For all proteins, solutions of the desired concentration were prepared by diluting the stock solutions in PBS 1X. 150 µL of each solution was plated per well. Fibronectin-coated plates were incubated overnight at 4°C and washed once with PBS 1X before plating the cells. Laminin-coated plates were incubated 2 hours at 37°C and washed 3 times with PBS 1X before placing the cells. Collagen I-coated plates were incubated 1hour at RT and washed once with PBS1X before plating the cells. BSA-coated plates were incubated 1 hour at room temperature (18°C culture room) and washed once with PBS1X before plating the cells.

**Adhesion assay: Actin drugs**

Latrunculin A (Sigma-Aldrich L5163-100UG) was diluted in DMSO (Sigma-Aldrich 41639-500ML) at 20mM. CK666 (SML0006-5MG) was dissolved in DMSO at 10mM. Cells were harvested as in the general protocol and resuspended in medium without FBS at  $3.3 \times 10^6$  cells/mL, then drugs were added directly to the resuspended cells to reach 50  $\mu$ M of LatA and 100  $\mu$ M of CK666. Cells in presence of drugs or DMSO were incubated 10 min at RT before plating in the wells. Adhesion assay was then performed as in the general protocol.

**Adhesion assay: Integrin inhibitors**

Peptides RGDS (Sigma A9041) and RGQS (Sigma A5686) were resuspended directly in medium without FBS at final concentration of 2 mM. Serial dilutions were prepared in order to test different concentrations. Cells were harvested from a confluent culture and washed once with medium without FBS. Then cells were resuspended again in medium without FBS and added to the solutions with inhibitor. Final concentration of cells in each conditions was  $3.3 \times 10^6$  cells/mL. Before plating the cells, the plates were first coated with 20ug/ml of fibronectin following the procedure already explained. Adhesion assay was then performed as in the general protocol.

**Immunostaining of adhesome proteins**

Cells were scraped from a confluent culture, centrifuged at  $5,000 \times g$  7 min and washed once with medium without FBS. They were then resuspended in medium without FBS. 3 mL of cells at  $1.5 \times 10^5$  cells/mL were added on top of a cover slide placed inside a 6 well/plate and incubated at 23°C for 2.5 hours (coverslides were previously coated with 150ul of BSA 4% or 300ul of FN 20  $\mu$ g/mL, left overnight at 4°C and washed with 600  $\mu$ L PBS1X). After the incubation, the liquid was removed, cells were fixed with 200  $\mu$ L of 4%FA for 10 min at RT. Cells were then washed with PBS 1X and incubated 2h at RT with 100 $\mu$ L of primary antibody diluted in blocking solution (1% BSA, 0.1% Triton in PBS 1X) at the following concentrations: anti-Int (9ug/mL), anti-talin (5  $\mu$ g/mL), anti-vinculin (10  $\mu$ g/mL), anti-paxillin (10  $\mu$ g/mL). Cells were washed with blocking solution (once quick, twice for 10 min) and incubated 1 hour at RT with 100  $\mu$ L of secondary antibody diluted in blocking solution as follows: anti-gp Dylight (Thermo Scientific SA5-10094) 1:1000 from the stock solution, anti-rat 488 1:2,000 from the stock solution). Cells were washed as before with PBS 1X and incubated with phalloidin texas red 1:100 for 15 min in darkness. They were mounted on glass slides with ProlonGold. Negative controls lacking primary antibody were performed in the same conditions.

For co-immunostaining, cells were seeded the day before,  $2 \times 10^6$  cells were added on top of a coverslide in a 6 well/chamber (coated with FN as above) and left to sit overnight at 23°C. Then they were washed with PBS 0.5 X and fixed with formaldehyde 4% for 5 min. Cells were washed again and incubated 30 min with

blocking solution. Primary antibody incubation included anti-Int (9 $\mu$ g/mL) plus anti-talin (5  $\mu$ g/mL), or anti-Int (9  $\mu$ g/mL) plus anti-vinculin (20  $\mu$ g/mL), for 1 hour at RT. Cells were washed with blocking solution twice for 15 min and then incubated with secondary antibody, anti-guineapig Dylight 488 (1:1000 from stock solution) plus anti-rat texas red (1:1000 from stock solution). Cells were then washed 20 min with PBS and stained with phalloidin 350 (1:100 from stock) for 15 min at RT. Cells were washed again and mounted on glass slides with Dako.

### **Staining of drug-treated cells**

Volume containing floating cells from an adhesion assay in presence of Lat A (0.05mM), CK666 (0.1 mM) and DMSO (corresponding volume) was taken. Cells were fixed directly in this medium with 4% formaldehyde for 10 min. Cells were then pelleted by repeated centrifugations of 5 min at 3000 g were the liquid was carefully removed. Cells were washed with PBS and stained with Phalloidin 1:100 and DAPI 1:500 (from 5mg/ml stock) for 15min in darkness at RT. Cells were washed with PBS for 5 mins and centrifuged at 3000 x g 10 min. Prolongold was added to the sample, which was resuspended carefully and mounted on glass-slides.

### **Imaging of fixed cells**

Cells treated with drugs and immunostained cells for adhesome proteins were imaged with a 63x objective in a Zeiss Axio Observer Z.1 epifluorescence inverted microscope equipped with LED illumination and Axiocam 503 mono.

All images were edited using Fiji Imaging Software version 2.0.0-rc-44/1.50e (Schindelin et al. 2012).

### **Morphology and fluorescent intensity measurements**

Filopodia were selected manually from the surface plane on images of fixed cells, and the imaging software returned number and length. Only filopodia connected to the cell body were counted.

IAPs intensity was measured along manually selected filopodia on the imaging software. First, background was removed from each image individually. Then a line was drawn on top the filopodia (wide enough to cover all signal) in the phalloidin staining channel and this area was transferred to the corresponding IAP staining channel, from which total intensity was acquired. The intensities were the normalised to the length of the corresponding filopodia.

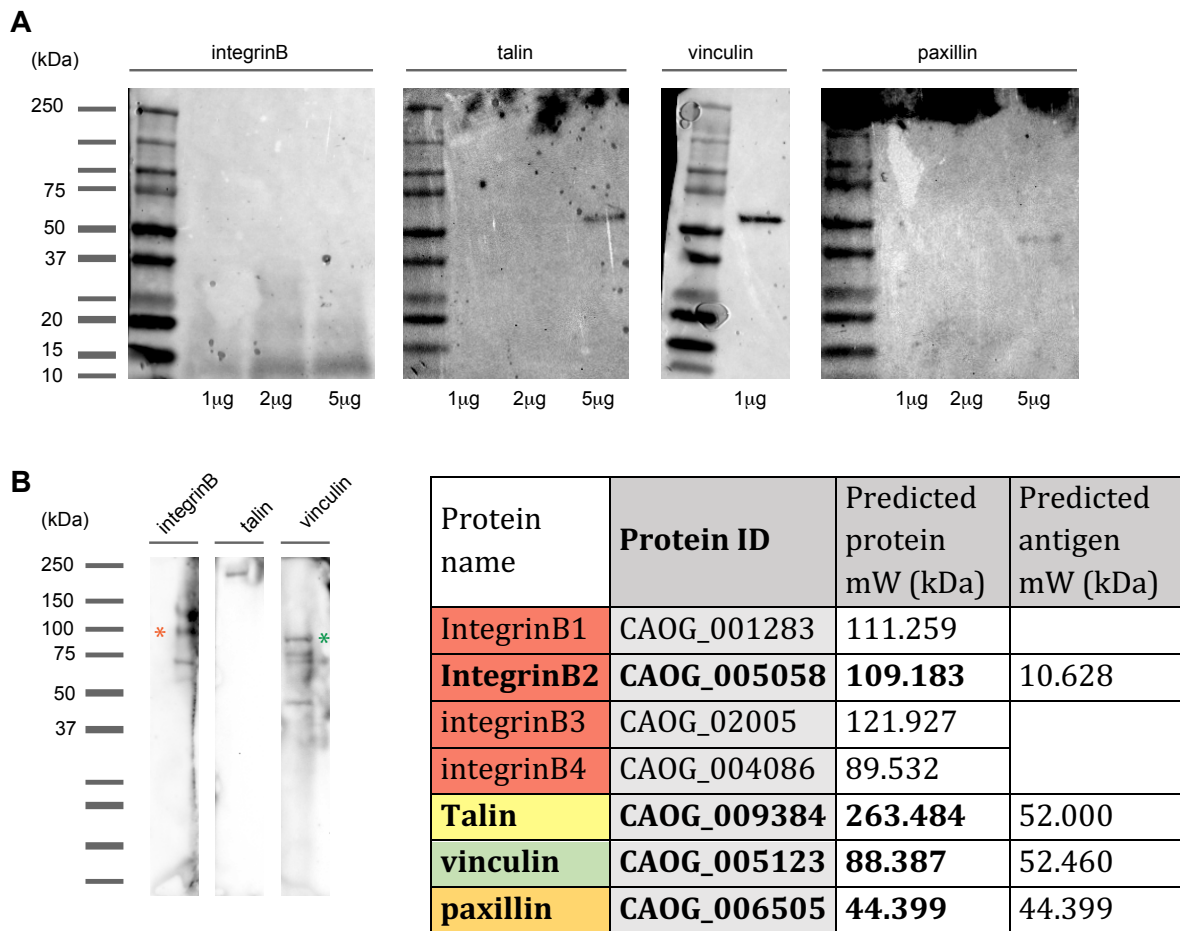
### **Statistical Analysis**

Results from drug effect on adhesion assay are shown as mean  $\pm$  standard error of the mean (s.e.m) as from 3 independent experiments. The significance of difference in the mean was tested using the parametric t-test for paired samples.

Results from measurements on morphology of filopodia and intensity of IAPs staining are represented as box-plots. The significance for each condition on the measurements was tested using the non-parametric Mann-Whitney U test

(Wilcoxon Rank-Sum test) for independent samples. All statistical analyses were performed using the R Stats Package version 3.3.1 (R Core Team, 2016).

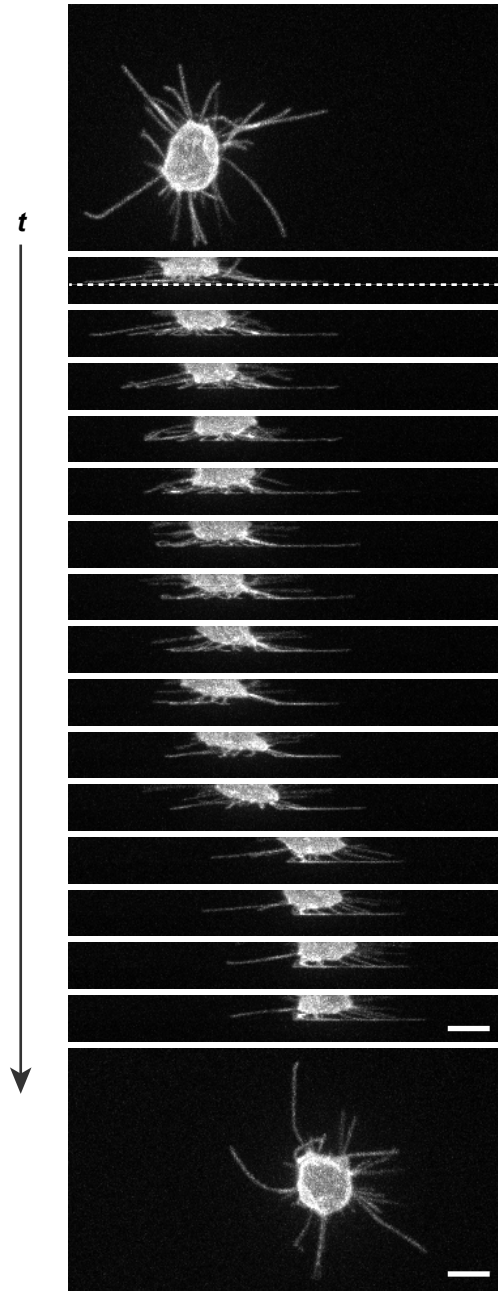
## Supplementary information



**Fig. S1. Antibody testing.** (A) Antibodies were tested by WB against their purified antigens. All of them recognised a band of the expected size of their antigen. (B) Antibodies were tested against total protein extract from adherent cells. Anti-*cowc\_IntB* detects a few bands that coincide with the expected sizes of the *intB2* homologs. Orange asterisk points the band expected to be integrin B2 (for which the antigen as designed). Anti-*cowc\_Tal* detects a single band, smaller than the expected size (263kDa). Anti-*cowc\_Vin* detects a band of the approximate expected size (green asterisk), plus additional bands of smaller size.

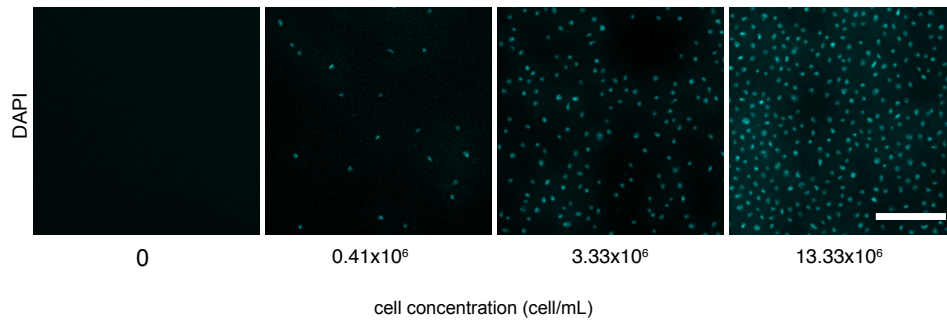
**Table 1. Antigen sequences**

antibody	Protein ID	protein Mw (kDa)	antigen Mw (kDa)	antigen sequence
anti-integrinB	CAOG_05058	109.183	10.628	QCVC DALHAGPACGCVKGVCPVSGGVRCNGG DCDPICGICTCPPGKTGPACDCDTVAHPCPTG NSTSGVVLPCSGQGTCLQSSATQCGICLCNRD PLTGTPLYTGSDCS
anti-talin	CAOG_9384	263.484	52.000	VTGAASSQEALATAAKSSVDTLSHLSDACKR GATSISSRDSNAQELLLNAVKDVAALADLIG STKTAAGRSVNDPAMEGLKENAKGMVNNIS QLVKVVKSVEDEASRGVRALESAIEAIGTELK VLESPEAPKRDASPEELVAATKMVTTSTAKI VSAANSNRQEEVVAANMARKALTDLMQYG KGAALKADTPDKQARVTVAVRDAANNAKA MLEAVYNALGHPTADAKNDVTNRSKKVA VADVVDAAKLLKGGDDYVDPEDPNVIAENELL AAAAAIEAAARKLADLKPRETPRAANEDLNF EEQILEAAKAIATATSALVKAAGAAQKELVST GKIDFTKGTAKYHENAMWSEGLVSAAKGVA AATGSLCDAANTAVQGEASQERLVSSAKQVA SSTAQLVVACRVKADANSKTQSRLNQAASM VKSATDELVKSEAAVFNQPDANITVDQRF VGSIAQVIAAQELILKQERELEKARKQLADIR KGQYQNK
anti-vinculin	CAOG_05123	88.387	52.460	LAERIKANPSDEVAQARFAELMDELPRELRL LEKALADDAIHAQMAVFANVAEPLSAIVQAA QSGNAADVNAAGVELQSQATLVKASRTVAS NAPDSEVSKEINTLSKQLEDLVPQIVVAARLV AANPDDQAARANLDLLMKSWSKVARLNE LSEQVAQPHAFLEVAERTIAAEVAKAKAAVT AQDKPSFDKAVKNIKATAARAGRLAAAEK NTDDADFRKKMAERRARIEASINGLEPTMN KAFTSRNAADIDAAVQPVTSSVSELKKEIAQS QGIEASGSGGAGAAGAGSSQAGSASSTAIAN EVRKSEEKQVAEIAAAAAGVSPQAVASHPISIA AGNLKLVASRWDAKNNALVQAADKISEKMR TMAAFMQPNNKDMIDMAKSMASEVAEIV KLAKAAAEQCSDRRLLKANLLQLCDKIPTISTQ LRIIASVKAANPSDDAETQLIAGSKNLMV TEIVKGTAAASLKFSSVASTANVALQWKRK ALGH
anti-paxillin	CAOG_06505	44.399	44.399	MDELDALLKDLQGGKPV DGAAAPAAAAA EPAAAVNPAIAGLQKDSRSSVDDLADLQS FKPQIKKIESTSAPAPQRASVDELLADLQSTG TLRNSVVRSTPSAADATRPKSGVADLDSVMA SLQDFKVEGASRPASMPAAGGAPAAAGGSG KLDSILNSLQSEM TSMGVD TARKGDCAACGK GIVGQVVTALGRTWHVEHFVCFQCRKPLGTT NFFEHESNPYCEKDFHELFSQRCAYCNGPV DRCIHALGKTWHPDHFCSQCGKNFEGGGF MERD GKAYCEEDYFNMFAPKCGGCDKAIMA DCISALGYQWHPNCFVCAECKGFNGGSFFE HEGKPFCE THYHAQSGSLCSCSQK PITGRCVT ALNKKYHPEHFVCSFCMKQLQKGTFKDENG KPYCHQCHVKLFG

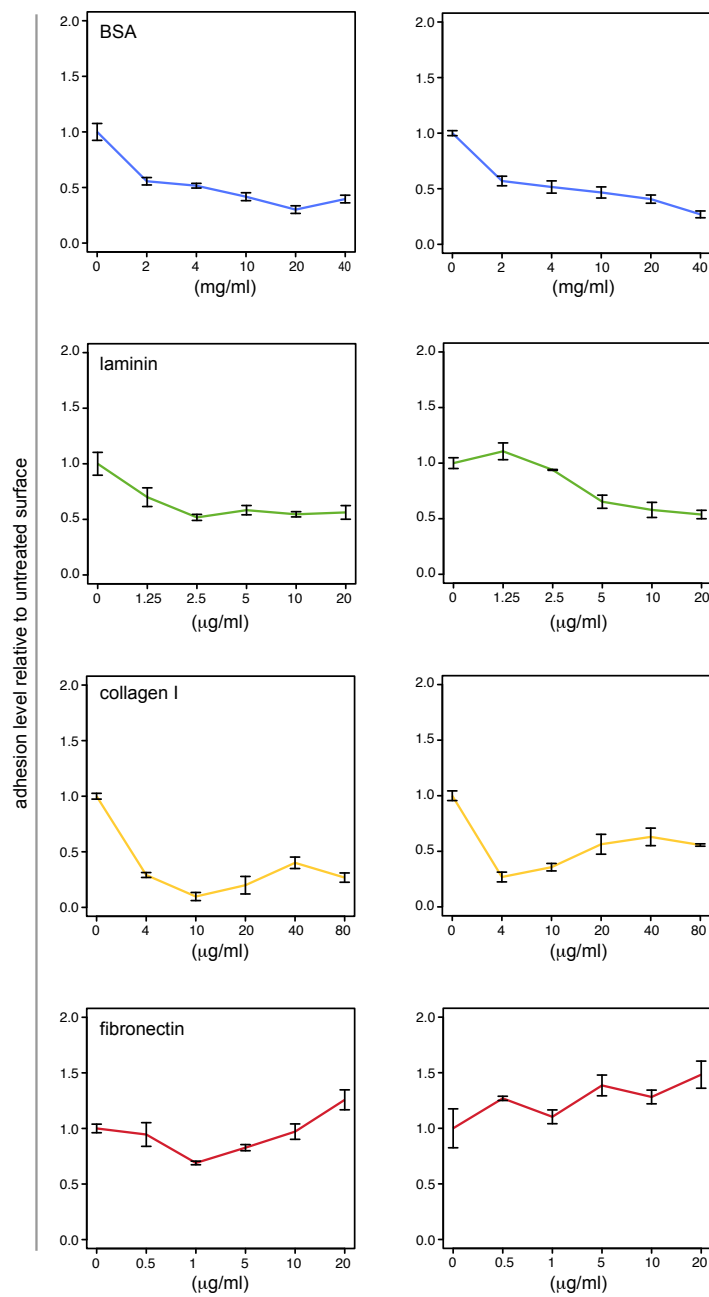


**Fig. S2. Time-lapse of a *Capsaspora* cell transfected with a membrane label (NMM-mcherry).** Longer fragment from time-lapse in Fig. 10A . Dotted line indicates surface. An image was taken every 10 seconds. Scale bar= 5 $\mu$ m





**Fig. S3.** Imaging of fixed and stained cells on the bottom of a plate during an adhesion assay in which different concentration of cells were plated. Scale bar= 5 $\mu$ m



**Fig. S4.** Two more replicate experiments on adhesion assay with different coatings. Supplementary to Fig. 3A.



## **DISCUSSION**



## **1. *Capsaspora owczarzaki* is a key organism to understand the origin of animals**

*Capsaspora owczarzaki* (hereafter *Capsaspora*) was discovered as a symbiont of the freshwater snail *Biomphalaria glabrata*, host of the human parasite *Schistosoma mansoni* (Stibbs 1979). It was found while studying explants from strains of this snail that were resistant to the infection by *S. mansoni*. Its putative role in causing resistance to the parasite motivated the study of its morphology and its behaviour both in culture conditions and together with the parasite (Owczarzak et al. 1980). By virtue of its morphology (an amoeba with long filopodia) it was tentatively placed as a member of the genus *Nuclearia* (Owczarzak et al. 1980). Further phylogenetic studies suggested it was not a nucleariid (Zettler et al. 2001) but seemed to be more closely related to another group (Hertel et al. 2002) named Mesomycetozoa or Ichthyosporea, (a group of organisms which diverged between animals and fungi [Mendoza 2002]). Its position was later defined as an independent lineage closely related to choanoflagellates and ichthyosporeans (Ruiz-Trillo 2004). Further studies placed it in a clade named Filasterea, as a sister group to choanoflagellates and animals (Schalchian-Tabrizi 2008, Ruiz-Trillo et al. 2008, Torruella et al. 2012).

The study of close unicellular relatives of animals and fungi offered a new perspective to understand multicellularity (King 2004, Steenkamp et al. 2006, Ruiz-Trillo 2007). This led to the UNICORN (unicellular opisthokont research initiative), which would involve the sequencing of several unicellular opisthokonts (the group formed by animals, fungi and their unicellular relatives) in order to clarify their phylogenetic position and to know their gene content (Ruiz-Trillo et al. 2007), which included the sequencing of *Capsaspora*. The position of *Capsaspora* as a close unicellular relative of animals, sister to choanoflagellates and animals makes it an ideal organism to be analysed in order to the origin of animals.

The analysis of *Capsaspora* genome surprisingly revealed a wide repertoire of genes related to multicellularity functions in animals (Suga et al. 2013): proteins involved in the integrin adhesion complex (Sebé-Pedrós et al. 2010), and in signalling, like the elements of the Hippo pathway and tyrosine kinases (Suga et al. 2012, Sebé-Pedrós et al. 2012), and a big repertoire of transcription factors (Sebé-Pedrós 2011, de Mendoza et al. 2013). Comparative genomics have helped to decipher the repertoire of genes that were present in the ancestor of animals, but the putative function of these genes remained a mystery.

Understanding the role that multicellularity-related genes have in their unicellular context would give us clues into the function they could have had in the ancestor of animals, and how they were co-opted into a multicellular lifestyle. For this to happen, the biological features of the closest unicellular relatives must be studied, as well as molecular tools to perform functional studies must be developed. This

need pushed our efforts towards developing molecular tools to work with *Capsaspora*, which is the main work done during this thesis.

## **2. Turning *Capsaspora* into a model organism to study the origin of animals**

### **2.1 Defining the life cycle of *Capsaspora***

The easiness by which *Capsaspora* can be cultured (it was placed after its discovery in the American Type Culture Collection, ATCC, as an axenic culture) made it an ideal candidate for experimental studies. Both adherent and cystic stages had previously been observed as the only stages of *Capsaspora* in its original description (Hertel 2002), but they had not been described with detail. The first thing we did was to characterise the life cycle of *Capsaspora* in culture conditions. For this we applied a combination of microscopy techniques, flow cytometry and transcriptomics, to gain insight into the life cycle at a molecular level (R1). This study resulted in the detailed characterisation of the filopodial (sometimes referred to “adherent” throughout this thesis) and cystic stages. Additionally, we characterised a new life-stage: multicellular aggregates (R1). The detailed analysis of *Capsaspora* life cycle in culture conditions led to the systematisation of the conditions necessary to obtain cells in either of the three stages synchronically (adherent, aggregative and cystic), which is the basis for many other studies. This was an essential step towards developing *Capsaspora* into an experimentally tractable organism.

### **2.2 Functional genomics to dissect *Capsaspora* life cycle regulation**

The capacity to obtain cells in the desired stage in a systematic way, allowed us to obtain large quantities of RNA, DNA or proteins from each stage in a reproducible way. This allowed further high-throughput analyses in *Capsaspora*, like transcriptomics (RNA-seq, shown in R1), chromatin states (ChIP-seq and ATAC-seq, in R2), and proteome dynamics (proteomics and phosphoproteomics, Sebé-Pedrós 2016b). At the same time, the definition of the highest proliferative phase in the beginning of the adherent stage helped the development of a reproducible transfection protocol (R3).

The transcriptomic analysis of *Capsaspora* life cycle revealed that each stage has a distinct expression profile (R1). The upregulation in the aggregative stage of certain genes, whose homologs in animals are essential for multicellular behaviour, like the integrin adhesion machinery and domains of extracellular matrix proteins, suggested that they could be involved in the formation of the aggregates. If techniques to analyse the function of specific genes were available, we would be able to find out whether integrins participate in adhering to an ECM, are involved in mediating cell-cell contact, or are otherwise playing another role.

The epigenome of *Capsaspora* was also analysed, revealing that it shares certain transcription regulatory mechanisms with animals and that different life stages are associated with the dynamic modifications of its epigenome (R2). *Capsaspora* shares some histone post-translational modifications with animals and has a repertoire of lincRNA (long intergenic non-coding RNA), which are regulatory elements in the animal genomes. It has only proximal cis-regulatory elements and only type-II promoters (defined by specific histone marks [Lenhard et al. 2012]). This shows that other promoter types (type-I and type-III) and distal enhancers, which *Capsaspora* lacks, are possibly animal innovations. The study of *Capsaspora* epigenome thus helped to decipher the regulatory mechanisms that probably participated in animal evolution. In this study, we also inferred the regulatory network of Brachyury (a transcription factor involved in animal development, specifically during gastrulation [Showell et al. 2004]). Many of its target genes were found to be orthologous to the Brachyury network of mouse, and those were enriched in actin cytoskeleton regulation and motility functions. This finding suggested there was a Brachyury regulatory network already present in the ancestor of animals involved in cell migration (R2). However, in absence of functional experiments, the exact role of Brachyury in *Capsaspora* and its targets cannot be yet addressed. To sum up, functional genomics with *Capsaspora* have increased our knowledge on the putative role of certain genes which are important for animal multicellularity, but still, functional experiments that would decipher the actual function in *Capsaspora* are missing. This highlights the need for developing genetic tools in *Capsaspora*.

### **2.3 Developing genetic tools in *Capsaspora***

The first step to perform genetic experiments is to have a method to efficiently deliver DNA inside the cell. So, given that *Capsaspora* was a promising organism to study the function of multicellularity-related genes, due to its phylogenetic position and its genome content, we worked in developing a method to transfect *Capsaspora*. These efforts resulted in the transfection method described in R3. The transfection efficiency was analysed by the expression of fluorescent proteins and resulted to be good enough to observe labelled structures in the cell (nucleus, actin cytoskeleton and filopodia). This achievement makes possible the future study of genes related to multicellularity, for instance by localising *in vivo* the protein with a tag and studying its localisation; or by performing over-expression experiments and studying its effects in the cell. This function can also be studied by generating knock-outs, for which another method has to be developed. In this case, the fact that transfection of *Capsaspora* is reproducible, turns the developing of knock-out methods in *Capsaspora* a conceivable task. Thus, this transfection method represents a major achievement towards turning *Capsaspora* into an experimentally tractable organism to study multicellularity-related genes.



The possibility to express fluorescent proteins in *Capsaspora* cells allowed us to visualise this organism with a new perspective (using fluorescent microscopy rather than transmitted light microscopy). When originally described by Stibbs et al. 1979, cells were found to be round and filopodiated, and this form was stated to be the motile form, which “moves by extending and retracting long filopodia and changes direction frequently”. Labelling of the membrane of *Capsaspora* cells with a fluorescent marker (R3) allowed us to visualise these filopodia indeed retracting, with unprecedented detail. Additionally, by obtaining a “lateral” view of the cell after performing a z-stack, we observed that the cell is separated from the substrate and attached to it by those filopodia (R3). This result confirmed what was originally stated in the original description (Fig 3. In Stibbs 1979) and seemed to go unnoticed in further observations. The possibility to label the membrane live and image the cell with high resolution allowed as well to observe a *Capsaspora* cell moving over the substrate with the same “lateral” perspective (R4). This as well confirmed that the cell uses filopodia for locomotion, as Stibbs et al. 1979 stated previously, but without further details. Furthermore, we could observe that the actin cytoskeleton forms a “basket” structure (R3), something previously unknown. Thus, the development of the transfection method led to a better understanding of the biological features of *Capsaspora*, which is essential for further functional experiments. This, indeed, inspired the development of an experimental set up to study specific characteristics of *Capsaspora*.

#### **2.4 Setting up a framework to analyse the function of the integrin adhesome**

One of the most promising discoveries about the gene repertoire of *Capsaspora* was the presence in its genome of homologs of the integrin adhesion machinery, which is the main system that provides linkage to the extracellular matrix in animals (Sebé-Pedrós et al. 2010). Because the function of *Capsaspora* integrin adhesome can shed light into the ancestral unicellular role of this machinery and how it could be co-opted for multicellularity, we started its functional characterisation. In absence of other tools, at first, we developed specific antibodies for localising several members of this machinery. The localisation of those proteins in filopodia plus the visualisation of *Capsaspora* filopodia as the major interactor with the substrate (R4) led to the hypothesis that filopodia have an adhesive function. These observations consequently promoted the development of an adhesion assay and led to the analysis of filopodia morphology. The results of our experiments suggested that these proteins are involved in adhesion to fibronectin via filopodia. This is the first step towards a more complete understanding of the role of these proteins in *Capsaspora*. With the transfection tool available, we tried over-expression but the results were inconclusive. A knock-out method to study the function of the integrin adhesome proteins from another perspective is not yet available for *Capsaspora*. When this technology is developed, and these proteins can be truncated directly in the genome, we will be able to study its effect on adhesion within the framework we already set up: the adhesion assay and the measurements on filopodia morphology

plus localisation of integrin adhesome proteins. Thus, this framework completes the requirements to convert *Capsaspora* an experimentally tractable organism in which the function of integrin adhesion machinery genes can be studied.

## **2.5 *Capsaspora* is a model among other unicellular relatives of animals**

While advances were being made towards converting *Capsaspora* into a tractable organism to study the origin of animals, in parallel, the same was happening with other unicellular relatives of animals (choanoflagellates and ichthyosporeans).

Among choanoflagellates, the life cycle of *Salpingoeca rosetta* (a colony forming choanoflagellate) has been analysed transcriptomically (Fairclough et al. 2013). Application of forward genetics has shown that formation of colonies is dependent on the gene *rosetteless*, a c-type lectin (Levin et al. 2014) which could be involved in adhering the cells to an ECM in the center of the colonies (also called *rosettes*). Recently, a transfection method has also been developed for *S. rosetta* (Booth et al. 2018). This method allowed visualisation of the choanoflagellate cellular structures by fluorescently-tagged proteins, showing that the biological features of this organism can be studied *in vivo*. Additionally, it allowed to shed some light into the function of *S. rosetta* septins, whose homologs in animals are involved in cytokinesis and cell polarity, and are hypothesised to regulate *rosette* development in *S. rosetta*. They localised these septins with a fluorescent tag, and observed that a mutated version of the same protein mis-localised, revealing a putative functional domain.

Among ichthyosporeans, the life cycle of *Creolimax fragrantissima*, which grows a coenocyte (multinucleated structure) that releases single amoebas, has been characterised transcriptomically (de Mendoza 2015). It can also be transfected (Suga and Ruiz-Trillo, 2013) and this tool was originally used to describe the synchronic nuclear division in the coenocyte. Recently, it has been used to study the regulation of the tyrosine kinase Src in *C.fragrantissima* (Suga and Miller, 2018) Src is a tyrosine kinase involved in signalling which is regulated by the tyrosine kinase Csk in animals, but this Csk is absent in *C.fragrantissima*. In this case, they showed that overexpression of a tyrosine phosphatase (PTP-3) rescued the phenotype caused by overexpression of Src. This suggested that PTP-3 could have been co-opted to regulate the phosphorylation of Src (Suga and Miller, 2018) in *C.fragrantissima*.

These represent the first steps to approaching the function of proteins related with multicellularity in a unicellular context in two different unicellular relatives of animals.

Overall, the characterisation of the life cycle and the development of transfection tools in *S. rosetta* and *C.fragrantissima* convert them into two experimentally tractable organisms. Along with *Capsaspora*, these three organisms are complementary for the study of the function of multicellularity-related genes or

mechanisms in a unicellular context. Thus, *S. rosetta* is a good model for colony formation by adhesion mechanisms and incomplete cytokinesis, *C. fragrantissima* is useful to understand cellularisation processes as well as signalling; and *Capsaspora* is useful to study integrin mediated adhesion and cell migration.

### **3. Outlook**

This work resulted into major steps towards turning *Capsaspora* into a model organism. Some challenges remain open, but also new possibilities appear.

#### **3.1 *Capsaspora* as a model organism: open challenges**

It is clear that Knock-out methods by homologous recombination or CRISPR/Cas9 will be a breakthrough, since they will allow to positively assign an adhesive function to integrin adhesome of *Capsaspora* or reject it, so efforts in this direction are needed. During this work, I also participated in developing the genome editing tool CRISPR/Cas9 for *Capsaspora*. This tool derived from a bacterial defence system that has been successfully used to edit the genome of numerous organisms, including the animal models *Caenorhabditis elegans* (Friedland 2013), *Danio rerio* (Hwang 2013), and *Drosophila melanogaster* (Gratz 2013), as well as some unicellular organisms, for which transfection tools were already available, such as *Plasmodium falciparum* (Ghorbal 2014), *Trypanosoma cruzi* (Peng 2015) or *Cryptosporidium parvum* (Vinayak 2015). The general approach is to use two components: Cas9, an endonuclease; and a single-guide RNA that would bear the target for Cas9. In order to try this method with *Capsaspora* I started the preparation of the plasmids encoding Cas9 and the RNA guide. Other labs are trying to develop this methodology in *Capsaspora*, based on the transfection method here presented.

In parallel, achieving stable transfection is also needed, as it would ease the screening for positively transfected cells and ensure the expression of the transfected gene after each cell division. Using a selectable marker would ensure stable transfection. During this work I have also been involved antibiotic testing, in order to select the appropriate one for selection, however, we do not have a selection method so far.

#### **3.2: Future perspectives**

The work developed during this project can be divided into two main points. On one side, we could propose hypotheses regarding the function in *Capsaspora* of genes putatively involved in multicellularity mechanisms, such as the role of Brachyury into regulating a network of developmental-related genes, or the role of the integrin adhesome proteins in adhesion to a substrate. On the other side, a significant part of this work has been developing assays and molecular tools to provide the basis for functional experiments with *Capsaspora* that will allow testing these hypotheses.

To test the role in adhesion of *Capsaspora* integrin adhesome proteins, these are possible approaches:

- 1- To use the adhesion assay to test different peptides that would alter adhesion abilities to narrow down which is the ligand that *Capsaspora* recognises.
- 2- To perform loss-of-function experiments to test whether integrin adhesome members in *Capsaspora* are indeed playing a role in adhesion. For this we would test whether the cells bearing the mutation altered its ability to bind to fibronectin. Additionally, to explore whether the various integrin subunits in *Capsaspora* are redundant.
- 3- To perform loss-of-function experiments on the integrin adhesome proteins and the putative ECM, to test its role in aggregate formation.

Additionally, the methods here developed can be used to explore other aspects of *Capsaspora*:

- 1- To obtain mutant lines expressing constitutively a marker (for the nucleus or the membrane) to study with more detail aspects like cell division or filopodia movements.
- 2- To tag proteins directly in the genome. For instance, tagging transcription factors to perform chromatin-immunoprecipitation with an antibody for the tag. This would overcome the inconveniences of developing specific antibodies to dissect the downstream network of multiple transcription factors.
- 3- To perform loss-of-function experiments to dissect the role of additional proteins. For instance, to test whether Brachyury targets are indeed involved in migration. This can be performed in a *Capsaspora* line that expresses constitutively the membrane tag here developed, to analyse its locomotion abilities.

Overall, new research possibilities are opened thanks to this work.



## **CONCLUSIONS**



The main conclusions of the present work are the following:

1. *Capsaspora* life cycle in culture conditions includes a filopodia stage, a cystic stage and a newly described stage: aggregation of cells into clusters.
2. Molecular characterisation of the life cycle in *Capsaspora* has revealed that each stage has a distinct transcriptomic profile. Specifically, genes related with integrin adhesion are upregulated in the aggregative stage.
3. Transitions during the life cycle are governed by dynamic changes in the chromatin states and histone marks. *Capsaspora* and animals have cis-regulatory elements and type-II promoters, and share transcription factor regulatory networks. Distal enhancers and different promoter types, however, are likely an animal innovation.
4. The transfection protocol presented is a tool that opens the possibility to modify genetically *Capsaspora* for the first time.
5. The possibility to modify *Capsaspora* genetically allowed to observe biological features previously unknown: actin filaments form a basket-like structure in the cell body of un-filopodiated cells and filopodia can be used as a mean of locomotion over a surface, sustaining the cell body separated from it.
6. The adhesion assay presented represents a straight-forward way of assessing a biological property, adhesion, in *Capsaspora*.
7. Localisation of several *Capsaspora* proteins conserved in the animal integrin adhesome (integrin B, talin, vinculin, paxillin) and a first insight into adhesive properties of *Capsaspora* revealed a possible role of these proteins in adhesion to fibronectin, via filopodia.
8. The hypothesis of a pre-metazoan integrin adhesome being functional in a similar way as in animals is further supported by the potential role in adhesion that *Capsaspora* adhesome proteins have shown.
9. Full implication in adhesion of each of the studied proteins remains to be tested by loss-of-function genetic experiments, something easier to achieve thanks to the development of a transfection protocol and an experimental setting.



10. The detailed description of the life-cycle, the setting up of a tool to modify genetically this organism, and the setting up of a framework in which to test a biological property, in this case adhesion; convert *Capsaspora* into a proper subject of experimentation.
11. Altogether, this work represents the first step to turning *Capsaspora* into a unicellular model organism to study the function of proteins involved in animal multicellularity.

## REFERENCES



- Adams, J.C., Chiquet-Ehrismann, R. & Tucker, R.P., 2015. The evolution of tenascins and fibronectin. *Cell Adhesion & Migration*, 9(1–2), pp.22–33.
- Adl, S.M. et al., 2012. The Revised Classification of Eukaryotes. *Journal of Eukaryotic Microbiology*, 59(5), pp.429–514.
- Akeson, A.L. & Woods, C.W., 1993. A fluorometric assay for the quantitation of cell adherence to endothelial cells. *Journal of Immunological Methods*, 163(2), pp.181–185.
- Baldauf, S.L., 1999. A Search for the Origins of Animals and Fungi: Comparing and Combining Molecular Data. *The American Naturalist*, 154(S4), pp.S178–S188.
- Barczyk, M., Carracedo, S. & Gullberg, D., 2010. Integrins. *Cell and Tissue Research*, 339(1), pp.269–280.
- Bays, J.L. & DeMali, K.A., 2017. Vinculin in cell–cell and cell–matrix adhesions. *Cellular and Molecular Life Sciences*, 74(16), pp.2999–3009.
- Bökel, C. & Brown, N.H., 2002. Integrins in Development. *Developmental Cell*, 3(3), pp.311–321.
- Booth, D.S., Szmids-Middleton, H. & King, N., 2018. A robust method for transfection in choanoflagellates illuminates their cell biology and the ancestry of animal septins. *bioRxiv*.
- Bourdon, M.A. & Ruoslahti, E., 1989. Tenascin mediates cell attachment through an RGD-dependent receptor. *The Journal of Cell Biology*, 108(3), pp.1149–1155.
- Brown, L.E., Sprecher, S.L. & Keller, L.R., 1991. Introduction of exogenous DNA into *Chlamydomonas reinhardtii* by electroporation. *Molecular and Cellular Biology*, 11(4), pp.2328–2332.
- Brown, M.W. et al., 2013. Phylogenomics demonstrates that breviate flagellates are related to opisthokonts and apusomonads. *Proceedings of the Royal Society B: Biological Sciences*, 280(1769), pp.20131755–20131755.
- Buerli, T. et al., 2007. Efficient transfection of DNA or shRNA vectors into neurons using magnetofection. *Nature Protocols*, 2(12), pp.3090–3101.
- Bulgakova, N.A., Klapholz, B. & Brown, N.H., 2012. Cell adhesion in *Drosophila*: versatility of cadherin and integrin complexes during development. *Current Opinion in Cell Biology*, 24(5), pp.702–712.
- Burridge, K., 1988. Focal Adhesions: Transmembrane Junctions Between The Extracellular Matrix And The Cytoskeleton. *Annual Review of Cell and Developmental Biology*, 4(1), pp.487–525.
- Busk, M., Pytela, R. & Sheppard, D., 1992. Characterization of the integrin alpha v beta 6 as a fibronectin-binding protein. *The Journal of biological chemistry*, 267(9), pp.5790–6.
- Calderwood, D.A., Campbell, I.D. & Critchley, D.R., 2013. Talins and kindlins: partners in integrin-mediated adhesion. *Nature Reviews Molecular Cell Biology*, 14(8), pp.503–517.
- Campbell, I.D. & Humphries, M.J., 2011. Integrin Structure, Activation, and Interactions. *Cold Spring Harbor Perspectives in Biology*, 3(3), pp.a004994–a004994.

- Carr, M. et al., 2017. A six-gene phylogeny provides new insights into choanoflagellate evolution. *Molecular Phylogenetics and Evolution*, 107, pp.166–178.
- Carr, M. et al., 2008. Molecular phylogeny of choanoflagellates, the sister group to Metazoa. *Proceedings of the National Academy of Sciences*, 105(43), pp.16641–16646.
- Cavalier-Smith, T. et al., 1996. Sponge phylogeny, animal monophyly, and the origin of the nervous system: 18S rRNA evidence. *Canadian Journal of Zoology*, 74(11), pp.2031–2045.
- Deakin, N.O. & Turner, C.E., 2008. Paxillin comes of age. *Journal of Cell Science*, 121(15), pp.2435–2444.
- Ensenauer, R. et al., 2011. Efficient and gentle siRNA delivery by magnetofection. *Biotech Histochemistry*, 86(4), pp.226–231.
- Felgner, P.L. et al., 1987. Lipofection: a highly efficient, lipid-mediated DNA-transfection procedure. *Proceedings of the National Academy of Sciences*, 84(21), pp.7413–7417.
- Ferrer-Bonet, M. & Ruiz-Trillo, I., 2017. Capsaspora owczarzaki. *Current Biology*, 27(17), pp.R829–R830.
- Friedland, A. E., et al., 2013. Heritable genome editing in *C. elegans* via a CRISPR-Cas9 system. *Nature Methods*, 10(8), pp. 741-743
- Gan, Z.R. et al., 1988. Echistatin. A potent platelet aggregation inhibitor from the venom of the viper, *Echis carinatus*. *Journal of Biological Chemistry*, 263(36), pp.19827–19832.
- Gaudet, P. et al., 2007. Transformation of *Dictyostelium discoideum* with plasmid DNA. *Nature Protocols*, 2(6), pp.1317–1324.
- Geiger, B. & Yamada, K.M., 2011. Molecular Architecture and Function of Matrix Adhesions. *Cold Spring Harbor Perspectives in Biology*, 3(5), pp.a005033–a005033.
- Geiger, T. & Zaidel-Bar, R., 2012. Opening the floodgates: proteomics and the integrin adhesome. *Current Opinion in Cell Biology*, 24(5), pp.562–568.
- Ghorbal, M. et al., 2014. Genome editing in the human malaria parasite *Plasmodium falciparum* using the CRISPR-Cas9 system. *Nature Biotechnology*, 32(8), pp.819-821.
- Glockling, S.L., Marshall, W.L. & Gleason, F.H., 2013. Phylogenetic interpretations and ecological potentials of the Mesomycetozoa (Ichthyosporea). *Fungal Ecology*, 6(4), pp.237–247.
- Gomaa, F. et al., 2017. Toward establishing model organisms for marine protists: Successful transfection protocols for *Parabodo caudatus* (Kinetoplastida: Excavata). *Environmental Microbiology*, 19(9), pp.3487–3499.
- Graham, F.L. & van der Eb, A.J., 1973. A new technique for the assay of infectivity of human adenovirus 5 DNA. *Virology*, 52(2), pp.456–467.
- Gratz, S. J., et al., 2013. Genome Engineering of *Drosophila* with the CRISPR RNA-Guided Cas9 Nuclease. *Genetics*, 194(4), pp. 1029-1035.

- Grau-Bové, X. et al., 2017. Dynamics of genomic innovation in the unicellular ancestry of animals. *eLife*, (6:e26036).
- Grosjean, F. et al., 2006. Multiple glycerol shocks increase the calcium phosphate transfection of non-synchronized CHO cells. *Biotechnology Letters*, 28(22), pp.1827–1833.
- Gumbiner, B.M., 1996. Cell Adhesion: The Molecular Basis of Tissue Architecture and Morphogenesis. *Cell*, 84(3), pp.345–357.
- Guo, L. et al., 2017. Optimizing conditions for calcium phosphate mediated transient transfection. *Saudi Journal of Biological Sciences*, 24(3), pp.622–629.
- Halbleib, J.M. & Nelson, W.J., 2006. Cadherins in development: cell adhesion, sorting, and tissue morphogenesis. *Genes & Development*, 20(23), pp.3199–3214.
- Hassett, B.T., López, J.A. & Gradinger, R., 2015. Two New Species of Marine Saprotrophic Sphaeroformids in the Mesomycetozoa Isolated From the Sub-Arctic Bering Sea. *Protist*, 166(3), pp.310–322.
- Hehenberger, E. et al., 2017. Novel Predators Reshape Holozoan Phylogeny and Reveal the Presence of a Two-Component Signaling System in the Ancestor of Animals. *Current Biology*, 27(13), p.2043–2050.e6.
- Hertel, L.A., Bayne, C.J. & Loker, E.S., 2002. The symbiont *Capsaspora owczarzaki*, nov. gen. nov. sp., isolated from three strains of the pulmonate snail *Biomphalaria glabrata* is related to members of the Mesomycetozoa. *International Journal for Parasitology*, 32(9), pp.1183–1191.
- Horton, E.R. et al., 2015. Definition of a consensus integrin adhesome and its dynamics during adhesion complex assembly and disassembly. *Nature Cell Biology*, 17(12), pp.1577–1587.
- Horton, E.R. et al., 2016. The integrin adhesome network at a glance. *Journal of Cell Science*, 129(22), pp.4159–4163.
- Howard, P.K., Ahern, K.G. & Firtel, R.A., 1988. Establishment of a transient expression system for *Dictyostelium discoideum*. *Nucleic Acids Research*, 16(6), pp.2613–2623.
- Humphries, J.D., Byron, A. & Humphries, M.J., 2006. Integrin ligands at a glance. *Journal of cell science*, 119(Pt 19), pp.3901–3.
- Huttenlocher, A. & Horwitz, A.R., 2011. Integrins in Cell Migration. *Cold Spring Harbor Perspectives in Biology*, 3(9), pp.a005074–a005074.
- Hwang, W. Y., et al., 2013. Efficient genome editing in zebrafish using a CRISPR-Cas system. *Nature biotechnology*, 31(3), pp. 227–229.
- Hynes, R.O., 1987. Integrins: A family of cell surface receptors. *Cell*, 48(4), pp.549–554.
- Hynes, R.O., 2002. Integrins: Bidirectional, allosteric signaling machines. *Cell*, 110(6), pp.673–687.
- Hynes, R.O., 1992. Integrins: Versatility, modulation, and signaling in cell adhesion. *Cell*, 69(1), pp.11–25.
- Jiang, M. & Chen, G., 2006. High Ca<sup>2+</sup>-phosphate transfection efficiency in low-density neuronal cultures. *Nature Protocols*, 1(2), pp.695–700.

- Jordan, M., Schallhorn, A. & Wurm, F.M., 1996. Transfecting Mammalian Cells: Optimization of Critical Parameters Affecting Calcium-Phosphate Precipitate Formation. *Nucleic Acids Research*, 24(4), pp.596–601.
- Jordan, M. & Wurm, F., 2004. Transfection of adherent and suspended cells by calcium phosphate. *Methods (San Diego, Calif.)*, 33(2), pp.136–43.
- Kanchanawong, P. et al., 2010. Nanoscale architecture of integrin-based cell adhesions. *Nature*, 468(7323), pp.580–584.
- Kapler, G.M., Coburn, C.M. & Beverley, S.M., 1990. Stable transfection of the human parasite *Leishmania major* delineates a 30-kilobase region sufficient for extrachromosomal replication and expression. *Molecular and Cellular Biology*, 10(3), pp.1084–1094.
- Kapp, T.G. et al., 2017. A Comprehensive Evaluation of the Activity and Selectivity Profile of Ligands for RGD-binding Integrins. *Scientific Reports*, 7(1), p.39805.
- Kim, T.K. & Eberwine, J.H., 2010. Mammalian cell transfection: the present and the future. *Analytical and Bioanalytical Chemistry*, 397(8), pp.3173–3178.
- King, N. et al., 2008. The genome of the choanoflagellate *Monosiga brevicollis* and the origin of metazoans. *Nature*, 451(7180), pp.783–8.
- King, N., 2004. The Unicellular Ancestry of Animal Development. *Developmental Cell*, 7(3), pp.313–325.
- King, N. & Carroll, S.B., 2001. A receptor tyrosine kinase from choanoflagellates: Molecular insights into early animal evolution. *Proceedings of the National Academy of Sciences*, 98(26), pp.15032–15037.
- Kingston, R.E. et al., 2003. Section I. Transfection of DNA into Eukaryotic Cells. In F. M. Ausubel et al., eds. *Current Protocols in Molecular Biology*. John Wiley & Sons, p. 9.1.1-9.1.11.
- Klapholz, B. et al., 2015. Alternative Mechanisms for Talin to Mediate Integrin Function. *Current Biology*, 25(7), pp.847–857.
- Klapholz, B. & Brown, N.H., 2017. Talin - the master of integrin adhesions. *Journal of cell science*, 130(15), pp.2435–2446.
- Kotnik, T. et al., 2015. Electroporation-based applications in biotechnology. *Trends in Biotechnology*, 33(8), pp.480–488.
- Kumar, C.C. et al., 1997. Biochemical characterization of the binding of echistatin to integrin  $\alpha\beta 3$  receptor. *The Journal of pharmacology and experimental therapeutics*, 283(2), pp.843–853.
- Lang, B.F. et al., 2002. The Closest Unicellular Relatives of Animals. *Current Biology*, 12(20), pp.1773–1778.
- Legate, K.R. & Fassler, R., 2009. Mechanisms that regulate adaptor binding to - integrin cytoplasmic tails. *Journal of Cell Science*, 122(2), pp.187–198.
- Lenhard, B., Sandelin, A. & Carninci, P., 2012. Metazoan promoters: emerging characteristics and insights into transcriptional regulation. *Nature Reviews Genetics*, 13(4), pp. 233-245.
- Lerche, K. & Hallmann, A., 2013. Stable nuclear transformation of *Eudorina elegans*. *BMC biotechnology*, 13(11), pp.1–19.

- Lerche, K. & Hallmann, A., 2014. Stable nuclear transformation of *Pandorina morum*. *BMC biotechnology*, 14(65), pp.1–16.
- Levin, T.C. et al., 2014. The rosetteless gene controls development in the choanoflagellate *S. rosetta*. *eLife*, 3, p.e04070.
- Liu, D., Ren, T. & Gao, X., 2003. Cationic Transfection Lipids. *Current Medicinal Chemistry*, 10(14), pp.1307–1315.
- Maartens, A.P. & Brown, N.H., 2015. Anchors and Signals. In *Current Topics in Developmental Biology vol.112*. Elsevier, pp. 233–272.
- Mattila, P.K. & Lappalainen, P., 2008. Filopodia: molecular architecture and cellular functions. *Nature Reviews Molecular Cell Biology*, 9(6), pp.446–454.
- Maurer-Stroh, S., Eisenhaber, B. & Eisenhaber, F., 2002. N-terminal N-myristoylation of proteins: refinement of the sequence motif and its taxon-specific differences. *Journal of molecular biology*, 317, pp.523–540.
- Mellor, H., 2010. The role of formins in filopodia formation. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research*, 1803(2), pp.191–200.
- de Mendoza, A. et al., 2013. Transcription factor evolution in eukaryotes and the assembly of the regulatory toolkit in multicellular lineages. *Proceedings of the National Academy of Sciences of the United States of America*, 110(50)
- de Mendoza, A. et al., 2015. Complex transcriptional regulation and independent evolution of fungal-like traits in a relative of animals. *eLife*, 4(8), p.085201.
- Mendoza, L., Taylor, J.W. & Ajello, L., 2002. The Class Mesomycetozoa: A Heterogeneous Group of Microorganisms at the Animal-Fungal Boundary. *Annual Review of Microbiology*, 56(1), pp.315–344.
- Miyahara, M. et al., 2013. Highly Efficient Transformation of the Diatom *Phaeodactylum tricornutum* by Multi-Pulse Electroporation. *Bioscience, Biotechnology, and Biochemistry*, 77(4), pp.874–876.
- Nagae, M. et al., 2012. Crystal structure of  $\alpha 5\beta 1$  integrin ectodomain: Atomic details of the fibronectin receptor. *The Journal of Cell Biology*, 197(1), pp.131–140.
- Nagai, T. et al., 2002. A variant of yellow fluorescent protein with fast and efficient maturation for cell-biological applications. *Nature Biotechnology*, 20(1), pp.87–90.
- Nellen, W., Silan, C. & Firtel, R.A., 1984. DNA-mediated transformation in *Dictyostelium discoideum*: regulated expression of an actin gene fusion. *Molecular and cellular biology*, 4(12), pp.2890–2898.
- Neumann, E. et al., 1982. Gene transfer into mouse lymphoma cells by electroporation in high electric fields. *The EMBO journal*, 1(7), pp.841–845.
- Nichols, S.A. et al., 2012. Origin of metazoan cadherin diversity and the antiquity of the classical cadherin/ -catenin complex. *Proceedings of the National Academy of Sciences*, 109(32), pp.13046–13051.
- Niu, Y.-F. et al., 2012. Transformation of diatom *Phaeodactylum tricornutum* by electroporation and establishment of inducible selection marker. *BioTechniques*, 52(6), pp.1–3.
- Owczarzak, A., Stibbs, H.H. & Bayne, C.J., 1980. The destruction of *Schistosoma*



- masoni mother sporocysts in vitro by amoebae isolated from *Biomphalaria glabrata*: an ultrastructural study. *Journal of Invertebrate Pathology*, 35(1), pp.26–33.
- Ozawa, A. et al., 2016. Molecular Basis of the Ligand Binding Specificity of  $\alpha\beta 8$  Integrin. *Journal of Biological Chemistry*, 291(22), pp.11551–11565.
- Pankov, R. & Yamada, K.M., 2002. Fibronectin at a glance. *Journal of cell science*, 115(Pt 20), pp.3861–3.
- Paps, J. & Holland, P.W.H., 2018. Reconstruction of the ancestral metazoan genome reveals an increase in genomic novelty. *Nature Communications*, 9(1), p.1730.
- Parsons, J.T., Horwitz, A.R. & Schwartz, M.A., 2010. Cell adhesion: integrating cytoskeletal dynamics and cellular tension. *Nature Reviews Molecular Cell Biology*, 11(9), pp.633–643.
- Peng, D. et al., 2015. CRISPR-Cas9-Mediated Single-Gene and Gene Family Disruption in *Trypanosoma cruzi*. *mBio* 6(1), pp. e02097-14
- Pierschbacher, M.D. & Ruoslahti, E., 1984. Cell attachment activity of fibronectin can be duplicated by small synthetic fragments of the molecule. *Nature*, 309(5963), pp.30–33.
- Plank, C. et al., 2003. The magnetofection method: Using magnetic force to enhance gene delivery. *Biological Chemistry*, 384(5), pp.737–747.
- Potter, H., 1988. Electroporation in biology: Methods, applications, and instrumentation. *Analytical Biochemistry*, 174(2), pp.361–373.
- Richter, D.J. et al., 2018. Gene family innovation, conservation and loss on the animal stem lineage. *eLife*, 7, pp.1–43.
- Richter, D.J. & King, N., 2013. The Genomic and Cellular Foundations of Animal Origins. *Annual Review of Genetics*, 47(1), pp.509–537.
- Riedl, J. et al., 2008. Lifeact: a versatile marker to visualize F-actin. *Nature methods*, 5(7), pp.1–8.
- Rokas, A., 2008. The molecular origins of multicellular transitions. *Current Opinion in Genetics and Development*, 18(6), pp.472–478.
- Rozario, T. & DeSimone, D.W., 2010. The extracellular matrix in development and morphogenesis: A dynamic view. *Developmental Biology*, 341(1), pp.126–140.
- Ruiz-Trillo, I. et al., 2004. *Capsaspora owczarzaki* is an independent opisthokont lineage. *Current Biology*, 14(22), pp.R946–R947.
- Ruiz-Trillo, I. et al., 2007. The origins of multicellularity: a multi-taxon genome initiative. *Trends in genetics*, 23(3), pp. 113-118
- Scherer, F. et al., 2002. Magnetofection: enhancing and targeting gene delivery by magnetic force in vitro and in vivo. *Gene Therapy*, 9, pp.102–109.
- Schiedlmeier, B. et al., 1994. Nuclear transformation of *Volvox carteri*. *Proceedings of the National Academy of Sciences of the United States of America*, 91(11), pp.5080–5084.
- Schindelin, J. et al., 2012. Fiji: An open source platform for biological image analysis. *Nature Methods*, 9(7), pp.676–682.
- Schultheiss, K.P., Suga, H. & Miller, W.T., 2012. Lack of Csk-Mediated Negative

- Regulation in a Unicellular Src Kinase. *Biochemistry*, 51(41), pp.8267–8277.
- Sebé-Pedrós, A. et al., 2012. Premetazoan Origin of the Hippo Signaling Pathway. *Cell Reports*, 1(1), pp.13-20.
- Sebé-Pedrós, A. et al., 2013. Early evolution of the T-box transcription factor family. *Proceedings of the National Academy of Sciences*, 110(40), pp.16050–16055.
- Sebé-Pedrós, A. et al., 2011. Unexpected Repertoire of Metazoan Transcription Factors in the Unicellular Holozoan *Capsaspora owczarzaki*. *Molecular Biology and Evolution*, 28(3), pp.1241–1254.
- Sebé-Pedrós, A. et al., 2010. Ancient origin of the integrin-mediated adhesion and signaling machinery. *Proceedings of the National Academy of Sciences*, 107(22), pp.10142–10147.
- Sebé-Pedrós, A., Burkhardt, P., et al., 2013. Insights into the Origin of Metazoan Filopodia and Microvilli. *Molecular Biology and Evolution*, 30(9), pp.2013–2023.
- Sebé-Pedrós, A., Irimia, M., et al., 2013. Regulated aggregative multicellularity in a close unicellular relative of metazoa. *eLife*, 2(2:e01287).
- Sebé-Pedrós, A., Ballaré, C., et al., 2016. The Dynamic Regulatory Genome of *Capsaspora* and the Origin of Animal Multicellularity. *Cell*, 165(5), pp.1224–1237.
- Sebé-Pedrós, A., Peña, M.I., et al., 2016. High-Throughput Proteomics Reveals the Unicellular Roots of Animal Phosphosignaling and Cell Differentiation. *Developmental Cell*, 39(2), pp.186–197.
- Sebé-Pedrós, A., Degnan, B.M. & Ruiz-Trillo, I., 2017. The origin of Metazoa: a unicellular perspective. *Nature Reviews Genetics*, 18(8), pp.498–512.
- Shalchian-Tabrizi, K. et al., 2008. Multigene Phylogeny of Choanozoa and the Origin of Animals R. Aramayo, ed. *PLoS ONE*, 3(5), p.e2098.
- Showell, C., Binder, O. & Conlon, F. L., 2004. T-box genes in early embryogenesis, *Developmental Dynamics*, 229(1), pp.201-218
- Sigal, C.T. et al., 1994. Amino-terminal basic residues of Src mediate membrane binding through electrostatic interaction with acidic phospholipids. *Proceedings of the National Academy of Sciences of the United States of America*, 91(25), pp.12253–12257.
- Springer, T.A., Zhu, J. & Xiao, T., 2008. Structural basis for distinctive recognition of fibrinogen  $\gamma$ C peptide by the platelet integrin  $\alpha$  IIb  $\beta$  3. *The Journal of Cell Biology*, 182(4), pp.791–800.
- Steenkamp, E.T., Wright, J. & Baldauf, S. L., 2006. The protistan origins of animals and fungi. *Molecular Biology and Evolution*, 23(1), pp. 93-106.
- Stibbs, H.H. et al., 1979. Schistosome sporocyst-killing amoebae isolated from *Biomphalaria glabrata*. *Journal of Invertebrate Pathology*, 33(2), pp.159–170.
- Suga, H. et al., 2012. Genomic Survey of Premetazoans Shows Deep Conservation of Cytoplasmic Tyrosine Kinases and Multiple Radiations of Receptor Tyrosine Kinases. *Science Signaling*, 5(222), pp.ra35-ra35.
- Suga, H. et al., 2013. The *Capsaspora* genome reveals a complex unicellular

- prehistory of animals. *Nature Communications*, 4(1), p.2325.
- Suga, H. & Ruiz-Trillo, I., 2013. Development of ichthyosporeans sheds light on the origin of metazoan multicellularity. *Developmental Biology*, 377(1), pp.284–292.
- Takada, Y., Ye, X. & Simon, S., 2007. The integrins. *Genome Biology*, 8(5).
- Tong, S.M., 1997. Heterotrophic flagellates and other protists from Southampton Water, U.K. *Ophelia*, 47(2), pp.71–131.
- Torruella, G. et al., 2012. Phylogenetic relationships within the Opisthokonta based on phylogenomic analyses of conserved single-copy protein domains. *Molecular Biology and Evolution*, 29(2), pp. 531-544.
- Torruella, G. et al., 2015. Phylogenomics Reveals Convergent Evolution of Lifestyles in Close Relatives of Animals and Fungi. *Current Biology*, 25(18), pp.2404–2410.
- Vicente-Manzanares, M. & Horwitz, A.R., 2011. Adhesion dynamics at a glance. *Journal of Cell Science*, 124(23), pp.3923–3927.
- Vinayak, S. et al., 2015. Genetic modification of the diarrhoeal pathogen *Cryptosporidium parvum*. *Nature* 523, pp. 477-480.
- Webb, D.J., 2005. Paxillin phosphorylation sites mapped by mass spectrometry. *Journal of Cell Science*, 118(21), pp.4925–4929.
- Wehrle-Haller, B., 2012. Structure and function of focal adhesions. *Current Opinion in Cell Biology*, 24(1), pp.116–124.
- Weinreb, P.H. et al., 2004. Function-blocking Integrin  $\alpha v \beta 6$  Monoclonal Antibodies. *Journal of Biological Chemistry*, 279(17), pp.17875–17887.
- Whittaker, C.A. & Hynes, R.O., 2002. Distribution and Evolution of von Willebrand/Integrin A Domains: Widely Dispersed Domains with Roles in Cell Adhesion and Elsewhere. *Molecular Biology of the Cell*, 13(October), pp.3369–3387.
- Wilson, S.P. & Smith, L.A., 1997. Addition of glycerol during DNA exposure enhances calcium phosphate transfection. *Anal.Biochem.*, 246(1), pp.148–150.
- Xiong, J.-P., 2002. Crystal Structure of the Extracellular Segment of Integrin alpha Vbeta 3 in Complex with an Arg-Gly-Asp Ligand. *Science*, 296(5565), pp.151–155.
- Yamada, K.M. & Kennedy, D.W., 1984. Dualistic nature of adhesive protein function: Fibronectin and its biologically active peptide fragments can autoinhibit fibronectin function. *Journal of Cell Biology*, 99(1 I), pp.29–36.
- Yamada, M. & Sekiguchi, K., 2015. Molecular Basis of Laminin–Integrin Interactions. In *Current Topics in Membranes*. Elsevier Ltd, pp. 197–229.
- Zaidel-Bar, R. et al., 2007. Functional atlas of the integrin adhesome. *Nature Cell Biology*, 9(8), pp.858–867.
- Zamir, E. & Geiger, B., 2001. Components of cell-matrix adhesions. *Journal of cell science*, 114(Pt 20), pp.3577–9.
- Zettler, L. A. A. et al., 2001. The Nucleariid Amoebae: More Protists at the Animal-Fungal Boundary. *The Journal of Eukaryotic Microbiology*, 48, pp. 293–297.

Ziegler, W.H., Liddington, R.C. & Critchley, D.R., 2006. The structure and regulation of vinculin. *Trends in Cell Biology*, 16(9), pp.453–460.