# The origin of multicellularity in animals: a functional approach from a unicellular perspective

## Developing *Capsaspora owczarzaki* as an emerging model

Núria Ros i Rocher

TESI DOCTORAL UPF / ANY 2018

DIRECTOR DE LA TESI

Dr. Iñaki Ruiz-Trillo

Institute of Evolutionary Biology (CSIC-UPF)

DEPARTAMENT DE CIÈNCIES EXPERIMENTALS I DE LA SALUT


*A la Irena, el meu yang*

*Barcelona, 21 de setembre del 2018*

# *Acknowledgements*

Siguem realistes. Sabeu perfectament que podria escriure una altra tesi sencera només amb els agraïments, perquè en tots els sentits aquesta tesi sento que té parts de tots vosaltres i no hagué estat possible, ni tan increïble, sense totes les persones que han compartit amb mi, sens dubte, una de les meves millors etapes de creixement tan professional com personal.

*Interestingly*, aquest viatge va començar amb un *abstract* de no més d'unes 5 o 6 línies que parlava dels opisthokonts, de l'origen dels animals i d'organismes unicel·lulars. Els opisthokonts? – vaig pensar... i alguna cosa em va dir que provés sort. Així, cap a l'abril-maig de ja fa uns quants anys, vaig entrar per segon cop a l'IBE i vaig conèixer a l'Iñaki. T'he de dir, Iñaki, que em vas convèncer de seguida. Així vaig viure els meus tres primers mesos – molt intensos ja – a l'MCG. I d'alguna manera, no en vaig tenir pas prou.

Iñaki, ja saps tot el que penso, i resumir tot el que has fet per mi aquests anys amb un "moltes gràcies" se'm queda curt, moooooolt curt. Gràcies per conèixer-me de debò i per donar-me totes les eines possibles per cremar l'ebullició d'energia, les ganes d'aprendre i l'entusiasme inevitable que he viscut aquests anys a l'MCG.

També per tooota la paciència i el suport incondicional que sempre m'has brindat, sobretot quan venia amb alguna idea o proposta boja, o preocupada perquè no funcionava res. Sobretot gràcies per totes les vegades que he vingut amb la llibreta temible de westerns: tot i ensenyar-te westerns negres o amb bandes rares sempre has confiat amb mi. Moltes gràcies per ensenyar-me i formar-me en molts aspectes, i per animar-me sempre a seguir endavant. I sobretot, per compartir amb mi el teu entusiasme per la recerca, la docència, els viatges i estades, congressos, cursos i outreach, que m'han inspirat i que se'ns dubte han calat molt dins meu i m'han fet millor persona. En aquest punt un se n'adona que no escull la recerca, sinó que la recerca t'escull i t'atrapa, i forma part de tu.

*Next*, sento cert paral·lelisme entre la transició cap a la multicel·lularitat animal i la transició que jo vaig viure a l'MCG. El *repertoire* ancestral ha estat molt important durant els meus inicis. Així doncs, gràcies a la "*old generation":* Guifré, Àlex i sobretot a l'Arnau, pel suport i pels teus famosos *sota-caballo-rei.* I evidentment, moltíssimes gràcies a la Núria (Ponsy), pel *mentoring* quan estava de pràctiques i per la teva energia positiva sempre.

Aquest *repertoire* ancestral va donar lloc a una gran expansió i duplicació de factors reguladors essencials per a la meva tesi. Sens dubte, el *core* factor ha estat la Meri (Merisisísiisisiiisisisma). Meri, ets l'ànima de l'MCG. Gràcies per donar-me seguretat, confiança i sobretot pel teu *sargentisme* sempre necessari per (intentar) mantenir l'ordre del lab. Per la teva paciència infinita i per no enfadar-te quan em polia dos caixes de flasks de cop (dels grans eeh!).

Saltem als downstream regulators, els meus megacompanys doctorands. Ha estat un veritable honor poder viure i compartir aquesta experiència amb vosaltres i el que pugui escriure aquí mai reflectirà verídicament tot el que sento i hem viscut. La vostra energia i suport han estat un dels meus motors principals per tirar endavant cada dia. Moltes gràcies al Xavi, per ser un exemple en tot, i al David, pel teu bon rotllo i positivisme. Helena (Parri), gracias por enloquecer conmigo, por emocionarte conmigo y por desesperarte conmigo. Hemos reído, hemos llorado y al final, lo hemos conseguido! Ya sabes que esta tesis no habría sido posible sin ti, en muchos sentidos. Alicia, qué puedo decirte que no sepas. Gracias por tu cuqismo infinito, por tu mimosismo y por tus megaabrazos cuando más los he necesitado. Edu, gràcies per les teves unixrespostes i sobretot pel teu *viejovenisme* que m'ha divertit tan tantíssim. Sisplau, no deixis mai de contagiar el teu riure! Alberto, por escucharme, ayudarme, reírte y compartir tus dudas, inquietudes y emociones. Gracias por todas las conversaciones filosóficas y por todos tus desplazos de silla hasta mi otra punta del laboratorio. Konstantina, gracias por tu energia positiva increíble, por tus ganas de aprender y volar y por compartir tu entusiasmo, no lo cambies nunca! And to Ola. Thanks for sharing your passion for life, your emotions, your questions and for your never-ending understanding. You know it means a lot to me.

I, *importantly*, sense els upstream regulators un doctorat no és possible. Mil milions de gràcies als postdocs de l'MCG. Gràcies Mati i Cristina, pels primers anys. Andrej, thanks

for questioning and putting everything in doubt. Omy, you are awesome! Thanks for your advises and for telling me repeatedly that *"it's gonna be fine".* Dan, thanks for the snails, for the coffees and for being able to steal a smile from my face everyday. Michelle, thanks a lot for your patience, for sharing your time with me especially these last months, and for encouraging me to keep on fighting! I finalmente, ya sabes que mi viaje durante estos 4 años no hubieran sido lo mismo sin ti. Gracias por animarme, por compartir tu interés y pasión por la biología, la ciencia y los fenómenos extraños, y sobretodo, por echarme la bronca por cada segundo que tardaba en levantarme si me caía. Gracias Sebas.

I moltes gràcies també als molts altres membres del lab: l'energia de Maria F., l'esperit animador de Maria R., i tot el suport de Gema i Cecília. I infinites gràcies a Xavi F., i a l'Alfonso, per la vostra lluita i energia durant el temps que em compartit. I a tots els altres estudiants que han fet estades al lab.

Moltíssimes gràcies als meus altres companys de l'IBE: Sílvia, Míriam, Anna, Pepi, Lídia, Héctor, Alfonso, Pellia, Adrián, Gabi, María, Vivien, Helena, Ana, Oliver, Gisela, Carolina, Santi i tota la llista infinita d'altres companys dels altres labs. Moltes gràcies a tots per arrencar-me un somriure cada dia, per animar-me tant i pel bon rotllo. Per compartir barbacoes, excursions a la muntanya i la passió per la ciència i la natura! Moltes gràcies al Marc, per la teva bona energia i per compartir durant dos anys les classes més divertides de la UPF. I moltíssimes gràcies al capità, Jose. Gràcies per animar-me tan aquests últims mesos i per ser l'organitzador i capità al que seguirem incondicionalment, al mar i a la muntanya.

Tornant als inicis, mil milions de gràcies a tu Salvi. Per obrir-me les portes a l'IBE amb el *Calotriton arnoldi* ja fa més de deu anys, per donar-me l'oportunitat de veure't en la teva *salsa*, tant al camp com a les aules, i de vegades, al lab,(sempre ben equipat) i per ser un exemple per mi en molts sentits. I sobretot pel teu gran suport i per animar-me cada dia amb un *Capsaspooooooooooooooooora* com Déu mana!

Moltíssimes gràcies a l'Elena. Se'm fa difícil resumir-ho tot amb un moltes gràcies només. Moltes gràcies pel *mentoring* des dels meus primers inicis, per seguir-me i ajudar-me en tots els meus dubtes i per recordar-me lo important que és escoltar-se a un mateix.

I moltíssimes gràcies al Txús, per tenir sempre un somriure guardat per mi, escoltar-me i animar-me tant!

De Barcelona m'he mogut una mica, sent un dels meus viatges més transcendentals el que vaig fer a Harvard. First, I would like to really thank you Mansi, for opening the doors of your lab, for sharing your enthusiasm for science and for teaching me *"in situ"* all those crazy protocols for the cutest worms and many other things behind. Thanks to everyone in the lab: Andree, Andrew, Lorenzo, Marce, Vincent and the undergrad students. And my most sincere thank you to Emily. Emily, thanks a lot for sharing all the true "american experiences" with me, for becoming one of my best friends and for making this friendship last at the other site of the ocean. I miss you a lot!

Many thanks to my roommates: Gabe, Arianna, Sarah and *Chewie*. For making my stage an adventure, and to Naomi, for making it possible. I per descomptat, moltíssimes gràcies al Gerard. Gràcies per compartir la teva curiositat i per animar-me tant. I sobretot per presentar-me i integrar-me en el teu grup d'amics increïbles: Cris, Marta, Uri i Ricardo, amb qui vaig compartir grans moments, com grans delícies americanes, pelis al cine de *Somerville* (amb cervesa!) i xerrades científiques al·lucinants.

Moltíssimes gràcies també al Roger i família, per compartir amb mi la meva última etapa als Estats Units i fer-me sentir com a casa. Gràcies Roger per escoltar-me i compartir les teves aventures, i inspirar-me per seguir treballant en ciència amb passió, i sobretot compartir-ho amb els qui més estimo.

El meu altre gran viatge va ser a l'altra punta de món, al Japó. Thanks a lot to Hiroshi and to the people in the lab, especially to Arisa and Takaaki, for inviting us to an unforgettable life-changing experience.

*Finally*, moltíssimes gràcies a la resta de l'equip de l'IBE i UPF, la maquinària que fa funcionar-ho tot. Moltes gràcies a la gent de secretaria, d'administració i de comunicació!

I, tot i que durant aquests 4 anys realment he tingut uns horaris bastant/molt bojos, afortunadament sempre he trobat algun moment per estar amb vosaltres.

Moltíssimes gràcies als meus amics de la uni, a les desUVicades i al KB. Gràcies per tot aquest recorregut de tants anys i per seguir endavant!

I què seria jo sense vosaltres nenes? Infinites gràcies a la Claus i a la Nurs, gràcies per tot l'amor, amistat i comprensió en els moments més difícils, per la muntanya els barrancs i per fer-nos inseparables. Pel vostre suport incondicional sempre. Us estimo moltíssim noies!

Moltíssimes gràcies a la *colla Agnès* – ja va sent hora que ens bategem amb algun nom no? Moltes gràcies a la Xènia, al Parra (Albert) i al Joan. Moltes gràcies per animar-me, donar-me suport i per haver crescut junts tots aquests anys. Tot i que sempre us fiqueu amb mi, us perdono. I mil milions de gràcies a l'Agnès. Gràcies per animar-me i per haver-te convertit en una de les *pomes* més importants. I el que ens queda!

Moltes gràcies a la Fina i al Ricard, per tota l'ajuda que he sentit i la vostra gran amistat. Gràcies pel vostre interès sempre en el meu creixement des dels meus inicis i per emocionar-vos en cadascuna de les etapes que he anat superant. Tot això quedarà en mi per sempre.

I moltes gràcies a la Gemma i a l'Esteve, per creure en mi, pel vostre entusiasme i per haver participat en aquesta gran aventura.

*I a tota la meva família.*

A tots els meus avis, moltes gràcies per ensenyar-me els valors que són realment importants. Moltes gràcies iaia, per ensenyar-me el valor de la força de la família. Moltes gràcies a tu, avi, per ser un exemple de lluita, per ensenyar-me el valor de la importància del temps i a espavilar-me. I moltes gràcies a tu, àvia, per ensenyar-me el valor de la dignitat i la fermesa i a mirar sempre endavant. Ets l'àngel de tots i sé que sempre estaràs present. Us trobo molt a faltar!

Moltes gràcies per tot el suport de les meves tietes i tiets. Gràcies a tots per motivar-me des de que era petita, per fer-me sentir tan estimada i per compartir totes i cadascuna de les meves etapes de vida, i tot el que ens queda per viure.

I moltes gràcies a les meves cosines i cosins, especialment al Manel i al Joan. Moltes gràcies per compartir amb mi les inquietuds i dubtes de la vida, per escoltar-me i per estimar-me tant. Això no ho perdrem mai.

*******

Als meus pares i a la Irena.
Papa, gràcies per no rendir-te, per seguir endavant i per lluitar incondicionalment i incansablement per nosaltres. Gràcies per fer-me riure tant, per entendre'm tant i per totes les teves reflexions de vida. Mama, com que saps perfectament el que penso no fa falta dir gaire més oi? Moltes gràcies pel teu incentiu i amor, i per haver-me motivat a que sempre vulgui anar més enllà. Gràcies als dos per inculcar-me la passió per la natura, el valor de l'esforç, de l'automotivació, del no parar, i de trobar en mi mateixa aquesta espurna que manté viva tota l'energia que fa que m'aixequi cada dia contenta i amb ganes de viure! I sobretot, gràcies per explotar des de ben petita la meva curiositat per les coses.

Moltes gràcies a la meva germana petita, a tu, Irena. Moltes gràcies per riure amb totes les meves tonteries, per tenir paciència i entendre'm tant, quan vols. I moltes gràcies per haver-me fet de germana gran quan més ho he necessitat. Ets la meva única certesa, i ja saps que t'estimo moltíssim.

*******

I per últim, infinites gràcies a tu, Alberto. Moltes gràcies per la teva paciència i comprensió, sobretot aquests últims mesos, i per escoltar-me, entendrem, i haver-me acompanyat en tot aquest recorregut. Per haver-me animat a seguir, i per recordar-me que mai m'he de rendir. I per compartir-ho tot. Ara i sempre.

*Sense tots vosaltres no seria la persona que sóc i res d'això hagués estat possible.*

## Abstract

The origin of animal multicellularity is a major evolutionary question. Recent genome data from the closest extant unicellular relatives of animals revealed that they actually possess and express a complex repertoire of genes related to animal development and multicellularity. Thus, assessing the functions of those genes in those unicellular relatives is key to gaining insight into to how they were co-opted at the onset of animals. However, such analyses have been hampered by the lack of genetic tools. Overall, this thesis advances our understanding of how co-option worked by using two different approaches. First, I provide a reliable transfection method for the filasterean *Capsaspora owczarzaki,* the close unicellular relative of animals with the richest repertoire of genes related to transcriptional regulation. This accomplishment converts *Capsaspora* into a unique experimentally tractable organism to investigate the origin and evolution of animal multicellularity. Then, I provide evidence of a remarkable degree of conservation between several transcription factor (TFs) networks in *Capsaspora*, suggesting that complex regulatory networks of TFs existed in the unicellular ancestor of animals.

## Resum

*L'origen de la multicel·lularitat dels animals és un gran enigma evolutiu. Dades genòmiques recents dels parents unicel·lulars més propers als animals han revelat que codifiquen i expressen un repertori complex de gens relacionats amb el desenvolupament i amb la multicel·lularitat animal. Per aquest motiu, investigar les funcions d'aquests gens en aquests parents unicel·lulars és clau per entendre com es van co-optar a l'origen dels animals. No obstant, aquestes anàlisis han estat obstaculitzades per la falta d'eines genètiques. En conjunt, aquesta tesi avança el nostre coneixement de com la co-opció va funcionar de dues maneres. En primer lloc, aporto un protocol fiable de transfecció per al filasteri Capsaspora owczarzaki, el parent unicel·lular més proper als animals amb el repertori genètic més ric en gens relacionats amb la regulació transcripcional. Això converteix Capsaspora en un organisme tractable experimentalment únic per investigar l'origen i l'evolució dels animals. Després, aporto evidència d'un grau considerable de conservació entre diverses xarxes de factors de transcripció a Capsaspora, suggerint que xarxes de factors de transcripció complexes van existir en l'ancestre unicel·lular dels animals.*

# *Preface*

Nature has amazed us since the beginning of civilizations. In an eternal pursuit of knowledge to satiate our intrinsic curiosity we have been tirelessly seeking to understand the origin and evolution of life: Where do we come from? How did life emerge? Which are the mechanisms that drove (and are driving) evolution?...

In antiquity, we tried to solve these questions cranking out myths and the pillars of religion. Later on, especially during the scientific revolution in the seventeenth century, we developed the fundamentals of modern sciences, such as astronomy, chemistry and biology. At that time, the paradigm of geocentrism was replaced by heliocentrism and novel principles of the cosmos were slowly introduced, shaking our vision of the world and the universe. While these new concepts were pointing out the stars, Antonie P. van Leeuwenhoek self-designed a single-lense microscope and shed light into a completely new universe of until then invisible creatures, originally referred to as *animalcules* (tiny animals). His observations and discoveries pioneered an interdisciplinary novel field of research that connects uncounted disciplines, such as biochemistry, cell biology, ecology and evolution, and bequeathed a legacy that paved the path to the fundamentals of microbiology.

Leeuwenhoek performed the very first observations of protists, a. k. a. unicellular eukaryotes. Protists are widely distributed in almost every habitat on Earth: from free-living forms in aquatic (freshwater and marine), terrestrial (soil crusts and forest litter) (Venter et al., 2018) or extreme environments (Oliverio et al., 2018), to other forms directly related to other species in commensal, parasitic or symbiotic relationships (Wilcox and Hollocher, 2018). Protists are major contributors to the active biosphere and to biological and geochemical processes (de Vargas et al., 2015; Debroas et al., 2015; Weisse et al., 2016). For instance, macrophytes and planktonic algae are key communities of marine plankton and act as primary producers for carbon fixation and photosynthesis (Reitsema et al., 2018). The composition, structure and interactions of protist communities also serve as bioindicators of environmental impacts and to assess the biological quality of different ecosystems.

Examples include diatoms, foraminifera, ciliates and testate amoeba as indicators for eutrophication, heavy metal contamination or chemical pollution (Desrosiers et al., 2013). Moreover, their close relationship to animals as vectors of infectious diseases, such as the malaria parasite *Plasmodium falciparum* or *Trypanosoma sp.* in Trypanosomiasis*,* put them in the spotlight for biomedical research.

Despite protists being mostly unicellular, they present a vast range of morphologies, lifestyles and behaviours, bringing cellular complexity to the extremes. There are species that are able to reach up to few centimeters, such as the giant alga *Ventricaria ventricosa,* or being as small as less than a micrometer, such as the minute green alga *Ostreococcus tauri*. Some protists, such as dinoflagellates, possess elaborate weapons shot by harpoon-like secretory organelles (Gavelis et al., 2017). Others, such as Diatoms, use silica to build their colorful shells (frustules) and Xenophyophores are even able to form intracellular barite crystals (Gooday and Nott, 1982). Other species present multicellular-like stages at some point during their life cycle, such as the aggregative behaviour of the social amoeba *Dictyostelium discoideum* (Konijn and Raper, 1961) or the colonial *rosettes* formed by the choanoflagellate *Salpingoeca rosetta* (Dayel et al., 2011; Levin et al., 2014).

Historically, the first tentative classification of protists in the Eukaryotic Tree of Life (EToL) grouped them together in a hodgepodge separated from Animals and Plants: the Kingdom Protista (Haeckel, 1866). This view soon changed with the advent of new DNA sequencing and molecular technologies, including Next-Generation Sequencing (NGS) and phylogenetics, that helped overcome the bias of morphology-based taxonomies and refined their evolutionary relationships (Medlin et al., 1988; Woese, 1996; Woese and Fox, 1977). Later on, high-throughput sequencing technologies promoted environmental DNA metabarcoding, which is nowadays revealing novel lineages of protists at different taxonomic levels, highlighting their ubiquitous phylogenetic distribution and still hidden diversity in the EToL (Del Campo et al., 2014; del Campo and Massana, 2011; Dìez et al., 2001; López-García et al., 2001; Pawlowski et al., 2016, 2012; Venter et al., 2018). Actually, protists are scattered across the EToL, being closely related to animals, plants, fungi and other multicellular groups of eukaryotes (Adl et al., 2012; Corliss, 1984; Keeling et al., 2005). These evolutionary relationships provide an ideal phylogenetic framework to study unicellular-to-multicellular transitions.

Multicellularity is, in fact, one of the major evolutionary transitions in the history of life. It entailed temporal and spatial regulatory programs of cell division and differentiation that gave rise to a wide variety of cell types and complex body plans (Szathmáry, 2015; Szathmáry and Smith, 1995). Among all the unicellular-to-multicellular transitions, the transition to animal multicellularity is undoubtedly the most unique one (Cavalier-Smith, 2017). However, the underlying genetic basis and mechanisms that drove this transition are still a long-standing question in biology. To fully understand how the transition occurred, it is key to investigate how distinct animal groups are related to each other and to their closest unicellular relatives, and how the genes involved in (and relevant to) their multicellularity evolved.

In fact, animals (Metazoa) and fungi are closely related to a heterogeneous assembly of protist lineages, altogether forming a highly diverse eukaryotic supergroup known as Opisthokonta (Brown et al., 2009; Cavalier-Smith, 1987; Del Campo et al., 2014; Ruiz-Trillo et al., 2004, 2006, 2008a; Shalchian-Tabrizi et al., 2008; Steenkamp et al., 2006; Torruella et al., 2012, 2015). Animals and their unicellular relatives Choanoflagellatea, Filasterea and Teretosporea form the Holzoa clade (Lang et al., 2002), whereas Fungi and their unicellular relatives Nucleariidae and Fontinculae form the Holomycota/Nucletmycea clade (Adl et al., 2012; Brown et al., 2009; Liu et al., 2009).

Choanoflagellatea is the sister-group to Metazoa consisting of flagellated protists that feed on prey using a collar structure made by microvilli (Cavalier-Smith and Chao, 2003; Medina et al., 2003; Ruiz-Trillo et al., 2008b, 2006, 2004; Steenkamp et al., 2006). Interestingly, some species are able to form multicellular-like structures of clonal colonies (Leadbeater, 2015; Levin et al., 2014). Filasterea is a group of filopodiated amoeba, composed to date by four species (Hehenberger et al., 2017; Ruiz-Trillo et al., 2008b, 2006, 2004; Shalchian-Tabrizi et al., 2008; Torruella et al., 2015). An example is *Capsaspora owczarzaki,* capable of forming aggregate structures of independent cells (Sebé-Pedrós et al., 2013). Finally, the group Teretosporea (composed by Ichthyosporea and Corallochytrea) includes fungus-like protists, some of them being parasites of animals and presenting a coenocytic development with palintomic cell division (Glockling et al., 2013; Marshall et al., 2008; Mendoza et al., 2002; Torruella et al., 2015).

Recent genome data from the choanoflagellates *Monosiga brevicollis* (King et al., 2008) and *Salpingoeca rosetta* (Fairclough et al., 2013), the filasterean *Capsaspora owczarzaki* (Suga et al., 2013) and the Ichthyosporean *Creolimax fragrantissima* (de Mendoza et al., 2015) showed they possess a complex repertoire of genes previously thought to be animal-specific and complex transcriptional regulatory control of their life stages. This included an unexpected conservation of genes related to typical multicellular functions, such as cell cycle control, cell growth, programmed cell death, specialized cell types, adhesion and transcriptional regulation (Fairclough et al., 2013; King et al., 2008; Sebé-Pedrós et al., 2016, 2011, 2010; Suga et al., 2013). Altogether, this indicates that the single-celled ancestor of animals was genetically complex and that co-option, tinkering and increased regulation, rather than gene innovation itself, were probably important driving forces for the origin of animals (Sebé-Pedrós et al., 2017, 2016). In this context, studying the function of genes related to important multicellular functions, such as transcriptional regulation and cell adhesion, in unicellular holozoans is key to understanding how they were co-opted during the transition to animal multicellularity.

In the present thesis, I address the transition to animal multicellularity from a functional perspective by: 1) developing a transfection protocol to turn the filasterean amoeba *Capsaspora owczarzaki* into an experimentally tractable system *(section 3.1. Transfection of Capsaspora owczarzaki, a close unicellular relative of animals),* 2) studying the evolution and diversification of *Runx1*, *Runx2* and *NF-κB* transcription factors families through a taxon-rich paneukaryotic survey and 3) assessing the role of *Runx1*, *Runx2* and *NF-κB* animal homologs in *Capsaspora owczarzaki* through localization and Chromatin Immunoprecipitation coupled with high-throughput sequencing (ChIP-seq) experiments *(section 3.2. Evolution of Runx and NF-κB developmental Transcription Factor families and the origin of animal multicellularity).* Altogether, the present thesis places *Capsaspora* as a good model to understand the function of key genes in a unicellular context and offers new insights into the evolution of animal multicellularity.

# Table of contents

# List of Figures and Tables

*Section 1*

# Introduction

*A unicellular perspective on*
*the transition to multicellularity in animals*

Life on Earth began more than three thousand million years ago (Mya), and since then it has conquered inconceivable ecological niches. From the stunning tropical environments that are bursting hotspots of biodiversity to the cold and solitary tundra in the poles or the fathomless depths in the abyss of the oceans. This astonishing environmental diversity has favoured life embracing refined and complex forms, such as the extremophile *Pyrococcus sp.* colonizing submarine hydrothermal vents or the tireless hummingbirds inhabiting cloud forests. But still, life has ultimately evolved from (and exists as) a single unit: a cell.

After the origin of the eukaryotic cell the biosphere underwent a complete upheaval, especially at the end of the Proterozoic eon when the first multicellular life forms appeared. From then on, multicellularity evolved independently in at least 25 lineages from all three domains of life and led to the diversification of a huge diversity of multicellular entities (Grosberg and Strathmann, 2007). Incredibly, simple groups of cells are able to form mat structures in Cyanobacteria (Rossetti et al., 2010) and Myxobacteria (Velicer and Vos, 2009) or even reproductive aggregates in slime molds (Du et al., 2015) and ciliates (Olive and Blanton, 1980). Complexity increases in differentiated multicellular organisms, like the volvocine green algae and the large seaweeds of red and brown algae, and is taken to extremes in land plants, fungi and animals (Bonner, 1998; Becker and Marin, 2009; Knoll, 2011; Umen, 2014; Cavalier-Smith, 2017). Multicellularity has been undoubtedly one of the major transitions in evolution contributing to the extraordinary diversity of life forms that exist on Earth.

The present introduction will be a journey to gain insight into the transition to animal multicellularity from a unicellular perspective. I will start by introducing the evolutionary transition to multicellularity, defining what multicellularity is and introducing the types and mechanisms of evolving multicellularity. Next, I will focus on the uniqueness of animal multicellularity and present the closest unicellular relatives of animals. Here, I will pinpoint their relevance as emerging model systems, presenting the knowledge gained from studying their genomes and life cycle transitions through comparative analyses. Moreover, I will stress the need for developing genetic tools among them for molecular and functional-based analyses. Finally, I will present *Capsaspora owczarzaki,* on which my work has been based, as an emerging model system to address the evolutionary transition to animal multicellularity.

## *Section 1.1*

# Evolutionary transitions to multicellularity

During billions of years, the planet biosphere has been reshaped as a result of a series of “major evolutionary transitions”, where pre-existing simpler entities evolved into more complex ones (Szathmáry and Smith, 1995; Szathmáry, 2015). At the beginning, genes were collected in protocells, a process that constitutes the origin of life as we know it. Next, unlinked replicators gave rise to chromosomes; prokaryotes formed a symbiotic relationship to constitute the eukaryotic cell, and unicellular eukaryotes preceded the origin of multicellular organisms. Not all living systems have been subjected to all of these transitions because there is no reason to expect continuous increases in complexity. Nevertheless, some of these transitions have occurred multiple times in evolution (Anderson, 1984; Duffy et al., 2000; Szathmáry, 2015). Multicellularity has long been viewed as this type of major transition, and it is a clear example of a transition that has occurred repeatedly, crossing over one of the major steps in the evolution of life’s complexity (Szathmáry and Smith, 1995; Szathmáry, 2015). Multicellularity evolved independently and repeatedly in at least 25 lineages from all the three domains of life, representing a wide spectrum of simple to complex organizations, especially in all major eukaryotic groups **(Figure 1)** (King, 2004; Grosberg and Strathmann, 2007; Knoll, 2011; Adl et al., 2012).

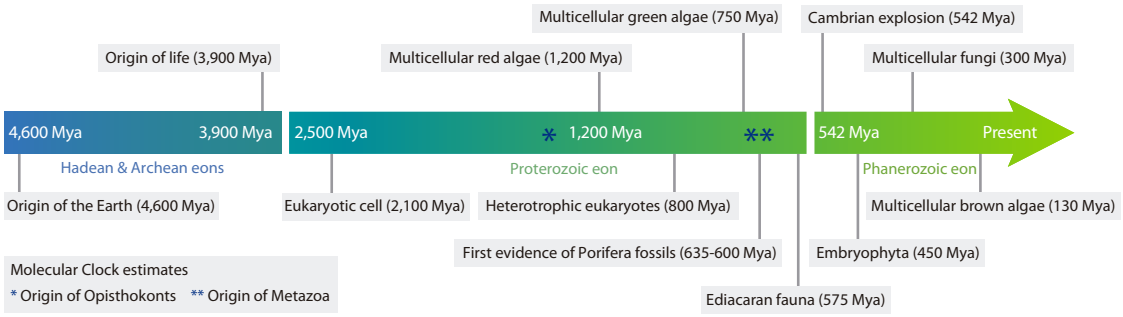



**Figure 1. A time-line of the origins of the major multicellular eukaryotic lineages.** Estimations are based on fossil record and molecular clock estimates. Time units are Millions of years ago (Mya). Adapted from Sebé-Pedrós et al., 2017 and Grau-Bové, 2017.

Simple multicellular structures are found in the early fossil records of both bacterial and eukaryotic lineages, being the first evidence of multicellular assemblies from around 3 to 3.5 billion years ago (Gya), corresponding to mat structures of cyanobacteria-like prokaryotes (Schopf, 1993).

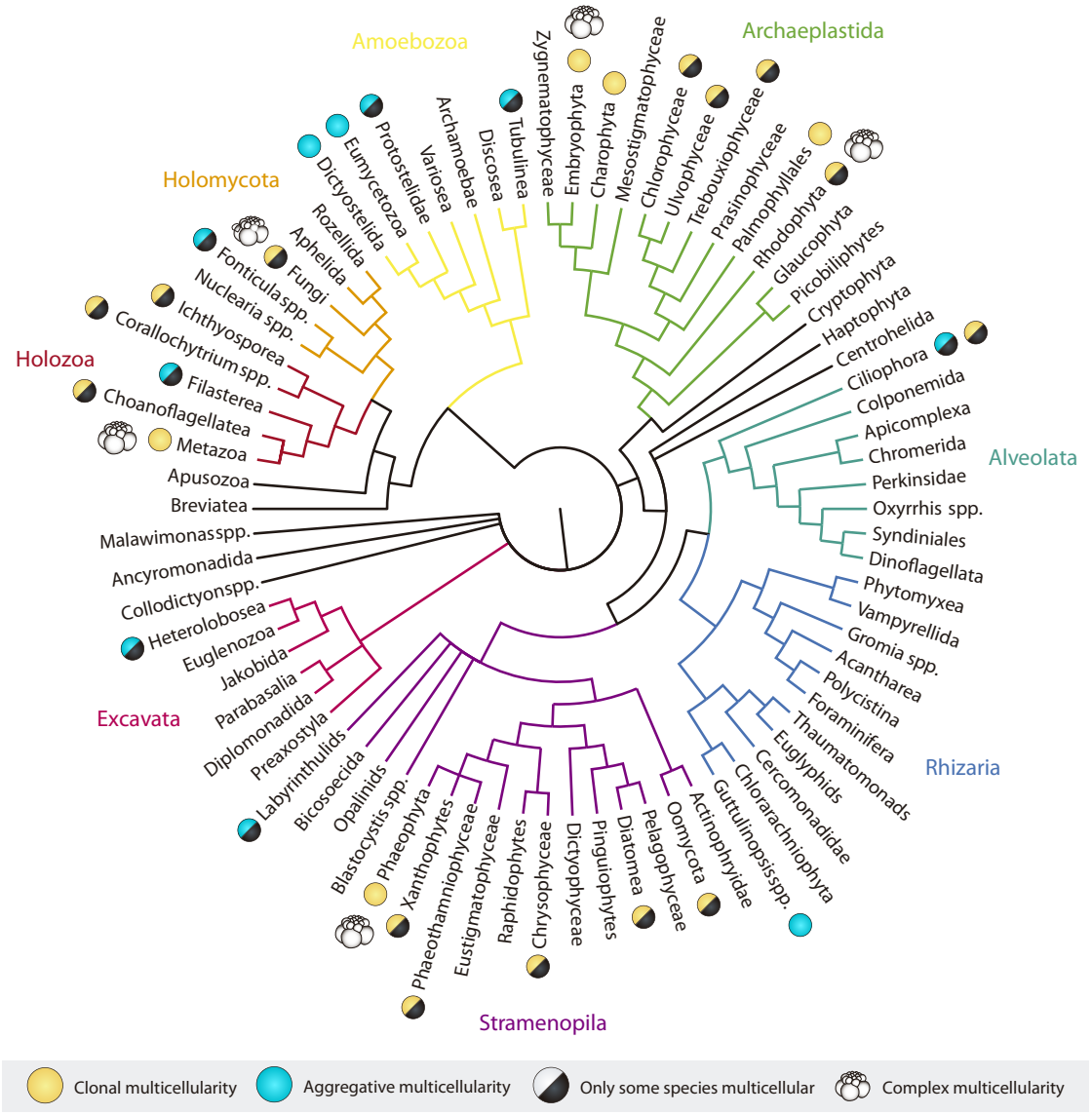



**Figure 2. The multiple origins of multicellularity.** Overview of the phylogenetic distribution of multicellularity among eukaryotes representing a consensus composite based on several recent phylogenomic studies. The multiple origins of multicellular forms and their modalities (clonal, aggregative or complex) are highlighted. Some lineages, such as animals (Metazoa) and land plants (Embryophyta), are entirely multicellular, whereas other lineages have only a few multicellular species, with the majority being unicellular. Adapted from Sebé-Pedrós et al., 2017.

Despite evidence for a certain level of cell differentiation existing around 2 Gya (Tomitani et al., 2006), it is after the origin of eukaryotes (1-1.9 Gya according to molecular clocks; Eme et al., 2014; or at least 2.1 Gya according to the fossil record; Knoll, 2014) that most of the known multicellular lineages and cell differentiation appeared within this group (Grosberg and Strathmann, 2007; Knoll and Hewitt, 2011). Multicellularity shows, thus, a discontinuous phylogenetic distribution in the Tree of Life **(Figure 2)**.

Multicellularity can be defined as a feature of a living system formed by the assembly of multiple cells that build up a cohesive, physically integrated and coordinated entity. In a multicellular context cells present at least some degree of cell-to-cell adhesion and cell-to-cell communication and their individual identities change, because they belong to a larger inclusive entity that shifts the nature of their individuality (Buss, 1987; Szathmáry and Smith, 1995). Multicellularity shows a wide range of organizations, from simple groups of cells to complex differentiated multicellular organisms (Bonner, 1998).

### 1.1.1. Types of multicellularity and their occurrence in the tree of life

Multicellularity can be "simple" or "complex", according to differences in cell differentiation, intracellular communication and direct contact with the external environment (Bonner, 2004; Butterfield, 2000; Knoll, 2011; Knoll and Hewitt, 2011; Lane and Martin, 2010; Schlichting, 2003).

Simple multicellularity consists of an assembly of identical cells, including filaments, clusters, balls, sheets or mats, that arise via mitotic cell division from a single progenitor. In general, these structures are formed by undifferentiated cell assemblages with some exceptions (*i.e.,* complete germ-soma differentiation with division of labour in *Volvox*; Umen, 2014). They can also include a limited degree of cell-cell communication and cell adhesion but necessarily entail direct contact with the external environment, at least during phases of the life cycle characterized by nutrient acquisition and active metabolism (Knoll and Hewitt, 2011). Simple multicellular organisms can be found both in prokaryotes, such as mat structures in cyanobacteria, and in eukaryotes: from the colonies of choanoflagellates and the diverse green algae structures like Volvocales (spherical colonies) or Charales (filamentous) to the aggregative structures of the

amoebozoan slime molds, the nucleariid *Fonticula* or the ciliate *Sorogena* (Bonner, 2000a; Knoll and Hewitt, 2011).

In contrast, complex multicellularity is a rarer event in evolution and is mostly restricted to eukaryotes where it appeared 6-7 times: in Metazoa (animals), in Ascomycota and Basidiomycota (fungi), in Embryophyta (land plants), in Phaeophyta (brown algae), and in Bangiophyceae and Florideophyceae (red algae) (Knoll, 2011). Complex multicellularity not only entails cell-cell adhesion but also a higher level of intercellular integration and communication, and commonly, tissue differentiation through networks of regulatory genes that tightly control spatial and temporal differentiation programs. Programmed cell death occurs in a number of these groups and individuals generally display a three-dimensional organization, in which only some cells are in direct contact with the external environment. This organization is critically important for organismic function, and includes both molecular conduits for cell-cell communication and tissues to facilitate bulk transport for oxygen, nutrients and signalling molecules required by internal as well as external cells (Beaumont, 2009; Knoll and Hewitt, 2011; Schlichting, 2003).

Hence, complex multicellular organisms have evolved structures that circumvent the limitations of diffusion, considered key to the evolutionary success of complex multicellular life (Knoll and Hewitt, 2011).

### 1.1.2. Mechanisms to evolve multicellularity

Multicellularity can appear via two different mechanisms that vary in the level of relatedness of the cells and the level of complexity they achieve: clonality or aggregation **(Figure 3).**

In most lineages, clonal multicellularity arises through successive rounds of clonal cell division from a single founder cell (spore or zygote) with incomplete cytokinesis (*i.e.,* division of the cytoplasm of the parental cell into two daughter cells). This results in a genetically identical population of cells that are stable **(Figure 3A)**. This kind of multicellularity acquisition is present in complex multicellular lineages.

For example, besides animals, plants and fungi, the division products of colonial organisms such as in the green alga *Volvox sp.,* are surrounded by a cohesive jelly that maintains them held together. In the case of *Chlorococcales,* the products of cell division are confined within the mother cell and, at some point, daughter flagellated cells are freed.

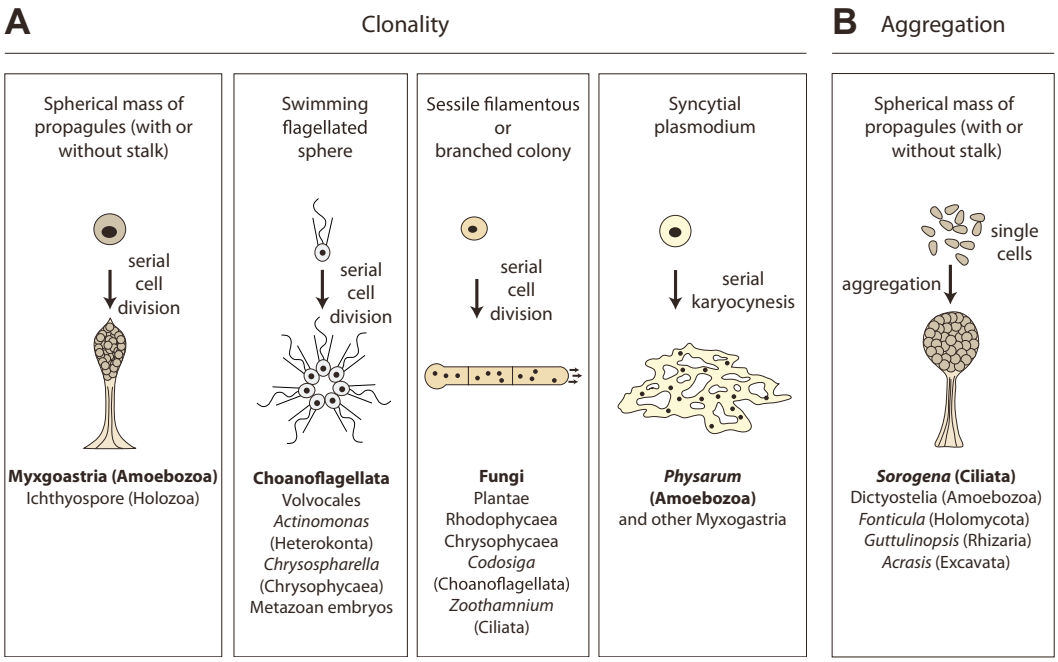



**Figure 3. Clonal and aggregative multicellularity.** (A) Clonal multicellularity gives rise to diverse multicellular forms. (B) Aggregative multicellularity gives rise to spherical masses of spores or cysts, sometimes atop a stalk. The organism indicated is sown in bold, and other organisms with similar forms of multicellularity are listed below. Adapted from Brunet and King, 2017.

However, other multicellular forms develop when two independent and genetically distinct cells bind or aggregate to each other **(Figure 3B)** (Grosberg and Strathmann, 2007). This aggregative multicellularity results in a heterogeneous population of cells evolutionarily more unstable than a clonal multicellular entity, as it is subjected to a slightly different fitness challenge (Aanen et al., 2008; Newman, 2012). Indeed, aggregative multicellularity is transient and implies lower cell relatedness and lower complexity levels, but still can present some level of organisation and differentiation. Some of these multicellular forms are triggered under stress or starvation, such as in the amoebae *Acrasis* or *Copromyxa,* in which starving cells crawl on top of each other and differentiate into spores or cysts (Du et al., 2015).

Others are formed during their reproductive stage building spore-bearing fruiting bodies, such as the ones present in slime molds. In fact, the social amoeba *Dictyostelium discoideum* (Schaap, 2011) is a great example of one of the most sophisticated structures of aggregative multicellularity, displaying a freely moving “slug” stage and up to five different cell types (Du et al., 2015). In the case of the nucleariid *Fonticula alba,* cells secrete an extracellular-matrix-like material to support the spore mass (Brown et al., 2009; Du et al., 2015).

All the multicellular aquatic organisms began their multicellularity by the products of cell division failing to separate, while most terrestrial microorganisms involve some form of motile aggregation of cells or nuclei in a multinucleate syncytium (Bonner, 1998). There are some apparent exceptions, such as the actinomycetes and a few cyanobacteria (*e.g.*, some species of *Stigonema*, which live in air in moist environments).

## *Section 1.2*

## Gene content for evolving multicellularity

The underlying molecular mechanisms that drove the transitions to multicellularity must have necessarily relied on at least some specific pathways ensuring physical, functional and reproductive stability and trade-off between these fitness components, especially related to survival and reproduction (Michod et al., 2005).

First, the physical unit of the multicellular entity requires tightly regulated mechanisms of cell-cell recognition and adhesion (Abedin and King, 2010). This is generally accomplished by extracellular adhesive molecules, such as pectins, hemicelluloses or glycoprotein-based bonds, or specific molecules in the surface of the cells (*e.g.,* transmembrane proteins), such as cadherins or other specialized cell adhesion molecules *(CAMs),* that maintain tight junctions between cells. Second, individual cells need to act in coordination and communicate between themselves through cell-cell signalling pathways. Finally, cell growth, proliferation and differentiation should be tightly regulated at least temporarily (simple multicellularity) and spatially (complex multicellularity) under precisely defined transcriptional programs (Degnan et al., 2009; Sebé-Pedrós et al., 2011; Suga et al., 2013).

Transitions to multicellularity, especially the evolution of complex multicellularity, additionally entailed an increase in the number of genes involved in these mechanisms, especially cell differentiation, cell-cell communication and transcriptional regulation. Probably, increases in complexity (number of cell types, body size, life-cycle stages, genes or protein domains) are linked to gene innovation, as well as tinkering and expansion of genetic material (Rokas, 2008a). Indeed, most of these cellular mechanisms were already present in the unicellular ancestor of multicellular lineages. This suggests that co-option or recycling of existing mechanisms, in addition to gene innovation of multicellular-specific genes, was probably an important driving force for multicellularity.

For instance, it has been shown that genes related to cell adhesion (*e.g., integrins* and *cadherins*), cell signalling and cell-cell communication (*e.g., tyrosine kinases*) predate the origin of Metazoa (Abedin and King, 2010; King et al., 2008, 2003; Sebé-Pedrós et al., 2010; Suga et al., 2013).

Similarly, genes related to extracellular matrix components, present in volvocine algae, evolved from a pre-existing genetic toolkit in their unicellular ancestors (Prochnik et al., 2010). And other genes related to cell differentiation and developmental pathways, present in the unicellular *Dictyostelium discoideum* (Schaap, 2011) and in *Volvox carteri* (Matt and Umen, 2016; Nedelcu, 2009; Prochnik et al., 2010), have also evolved from their unicellular ancestors.

## *Section 1.3*

## Selective advantages of evolving multicellularity

Being multicellular can provide opportunistic benefits and several selective advantages in a hostile environment. First of all, multicellularity is an effective way to increase size. Compared to a single-celled counterpart, a multicellular organism is able to sustain larger organismic sizes in a more efficient way as compared to hypertrophic cell growth (Bonner, 2000b; King, 2004). This overcomes most of the physico-chemical limitations, such as surface-volume ratio or higher diffusion rates in the cytoplasm and cell membranes, that might impose an upper limit on the size of a single-celled organism, despite some exceptions, such as the giant algae *Ventricaria ventricosa*, one of the largest unicellular species (Bisson and Beilby, 2002; Ryser et al., 1999). An increased size can represent as well a defensive mechanism to escape from heterotrophic predators (Stanley, 1973). Moreover, large sizes are more effective in nutrient storage (Bonner, 2000b; Kirk, 2003).

Second, multicellularity confers more environmental resilience, both over predators and over external conditions. In fact, it can produce an internal, more stable environment and therefore be more effectively shielded and resilient to the vagaries of the external environment (Bonner, 1998; Gerhart and Kirscher, 1997).

Third, functional specialization and division of labour is another advantage of multicellular organisms. Multicellular organisms can simultaneously partition complementary tasks among different cells. This adds to the selective advantages of the organisms under some circumstances and allows functional specialization and increased cell cooperation (Ispolatov et al., 2012; Michod, 2007; Szathmáry and Smith, 1995). Related to this, multicellularity allows for metabolic cooperation between key metabolic processes that cannot concurrently occur within a single cell. For example, photosynthesis interferes with nitrogen fixation because nitrogenase does not effectively catalyse fixation in the presence of oxygen (Kaiser, 2001).

Forth, multicellularity is also a more competitive way to carry the genes over generations and maintain a more genetic uniformity through programmed regulation of cell division (Butterfield, 2009; Dawkins, 1976; Grosberg and Strathmann, 2007; Wolpert and Szathmáry, 2002), especially when it is linked with the evolution of specialized reproductive cell types (germ lines and somatic cells). This also improves the trade-off associated with large body reproduction and results in a higher hereditary potential (Buss, 1987; Michod, 2006; Michod and Herron, 2006; Szathmáry, 2015; Szathmáry and Smith, 1997).

Finally, multicellularity opens new ecological opportunities and improve the efficiency of food consumption, such as by the secretion of digestive enzymes (Grosberg and Strathmann, 2007). There are, as well, other non-adaptive scenarios related to thermodynamics (Otsuka, 2008), genetic drift (Lynch and Conery, 2003) or other environmental effects, that have been proposed to explain an increased complexity during the evolution of multicellularity.

## *Section 1.4*

# The origin of animal multicellularity

Among the major groups of complex multicellular organisms, animals stand out for their vast morphological richness. Animals are a monophyletic group of heterotrophic organisms classified in 35 different phyla with around 1.3 million of species described, despite estimations raise the potential number of species up to 10 million species (Pawlowski et al., 2012). They present more genetic and phenotypic complexity than other eukaryotic lineages, being the group displaying the widest variety of cell types with refined coordination and physiological systems structured in tissues and organs (Bell and Mooers, 1997; Rokas, 2008a). Moreover, animals undergo coordinated embryogenesis (Valentine et al., 1991) and present highly regulated developmental programs.

### 1.4.1. Dating animal origins: fossil record and molecular clock-based estimates

Despite the enigmatic causes that might have driven the transition to animal multicellularity, there is no exact consensus on when animals appeared (Antcliffe et al., 2014; Budd and Jensen, 2017; Knoll and Hewitt, 2011).

Fossil record inferences have tentatively dated the origin of early-branching metazoans, including sponges, placozoans and cnidarians, back to the Ediacaran period (at the end of the Neoproterozoic era) around 632-600 Mya (Cohen et al., 2009; Love et al., 2009; Narbonne, 2005; Narbonne and Gehling, 2003; Yin et al., 2015). In fact, most of the rock fossils of the Ediacaran preserve evidence of some of the earliest complex macroscopic organisms, many of which have been interpreted as animals. To date, around 200 Ediacaran macrofossil taxa have been described (Fedonkin et al., 2007) and classified according to their morphological similarities in “morphogroups” representing grades of organism (Erwin et al., 2011; Grazhdankin, 2014).

A recent study by Dunn et al., analysed the affinities of three widely recognised morphogroups (the rangeomorphs, dickinsoniomorphs and erniettomorphs) with animals from a developmental point of view (Dunn et al., 2018). Interestingly, rangeomorphs and dickinsoniomorphs presented congruence with aspects of metazoan development, and allowed the identification of Dickinsonia **(Figure 4A),** Andiva, Yorgia and the rangeomorphs as early metazoans (Dunn et al., 2018). Interestingly, this supports recent suggestions of considerable developmental complexity in early-branching metazoans (Ferrier, 2015). Some other examples of early-branching animal fossil records include the well-established Porifera *Protohertzina anabarica* from around 535 Mya (Hamdi et al., 1989) and some earlier but ambiguous sponges, such as *Eocyathispongia qiania* from around 600 Mya (Yin et al., 2015) **(Figure 4B)** and pre-Marinoan limestones specimens from around 635 Mya (Antcliffe et al., 2014; Maloof et al., 2010). Ctenophora are relatively absent from the fossil record, with only a few confident fossils dated 540-580 Mya, such as the Eoandromeda genus **(Figure 4C)** (Ou et al., 2015; Tang et al., 2011). Nevertheless, the first unequivocal eumetazoan[1] (*i.e.,* animals with well-defined tissues and organs) fossil evidence dates back around 541-565 Mya (Cunningham et al., 2017; Erwin et al., 2011) and the extraordinary radiation of metazoan forms tentatively occurred around 542 Mya, during the Cambrian explosion.

Relaxed molecular clock-based studies from phylogenomic and fossil data of crown Metazoa do not match the estimations of fossil record-based interpretations (Cunningham et al., 2017), instead yielding older estimates for animal origins to around 700-800 Mya (Erwin et al., 2011), or 755-838 Mya (Sperling et al., 2010) or 650-833 Mya **(Figure 4D)** (dos Reis et al., 2015). This implies that animals would have begun diversifying over 100 million years before the first definitive metazoan fossil evidence in the Cambrian. Despite the discrepancy between molecular clock estimates and rock record interpretations, biomarker evidence suggests that animals may have been present by 635 Mya, and reasonably convincing fossil evidence would date from 565 Mya onwards **(Figure 4D)**.

[1] The term “eumetazoans” was originally coined to (and is used here to) encompass all animals except sponges (Hyman, 1940). Cnidarians and ctenophores, together with bilaterians, are also often referred to as “eumetazoans”, excluding placozoans. Nevertheless, the presence of epithelia in *Trichoplax adhaerens* has led to the definition of placozoans plus all eumetazoans as the “epitheliozoa” (Sperling et al., 2009).

Nevertheless, molecular clock-based analyses estimate the radiations of eumetazoans around 626-746 Mya; bilaterians around 596-688 Mya; Deuterostomia 587-662 Mya; and Protostomia 578-653 Mya (dos Reis et al., 2015).

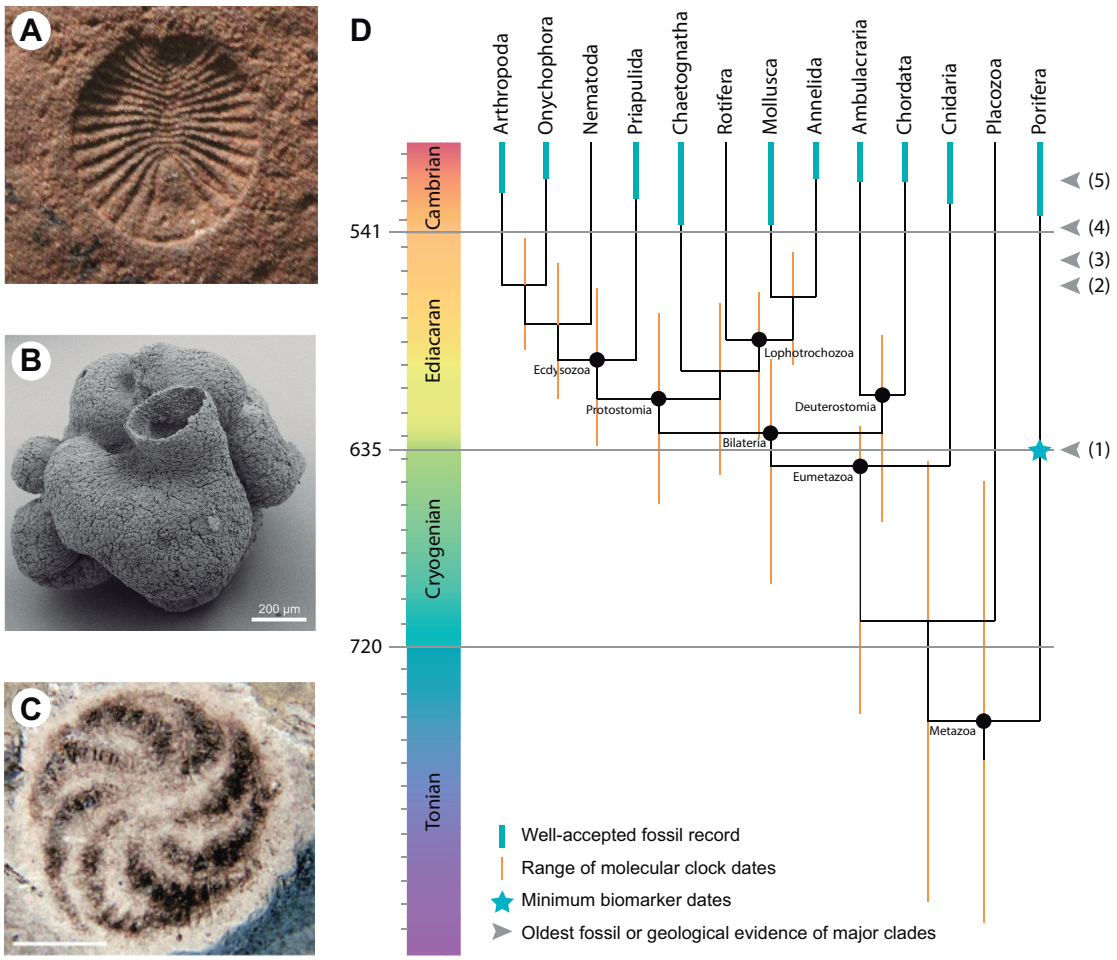



**Figure 4. Examples of animal fossil specimens and the mismatch between the fossil and molecular clock records of early animal evolution.** (A) Dickinsonia, a representative of soft-bodied Ediacaran macrobiota. Scale bar: 5 mm (Cunningham et al., 2017). (B) Scanning electron micrograph of *Eocyathispongia qiania* phosphatized fossil (600 Mya) displays multiple independent characters of modern adult sponges. Scale bar: 200 µm (Yin et al., 2015). (C) Carbonaceous compressions of *Eoandromeda octobrachiata* (JK10909). Scale bar: 5 mm (Tang et al., 2011). (D) The phylogeny follows dos Reis et al., 2015. Note that ctenophores, the phylogenetic position of which is contentious, were not included. Blue bars represent well-accepted reports of fossils that can be assigned to extant animal phyla, which are limited to the Cambrian; ranges mainly follow Erwin et al., 2011. Orange bars represent the range of molecular clock estimates of the origins of major clades obtained in dos Reis et al., 2015. Note that the origin of eumetazoans is inferred to predate the Ediacaran and the origins of bilaterians, protostomes, deuterostomes, ecdysozoans and lophotrocozoans are inferred to predate the Cambrian. The righthand column shows the first evidence for major clades in the geological record: (1) Metazoa 1/4 635 Mya, possible biomarker evidence; alternatively, 565 Mya eumetazoan trace fossils; (2) Eumetazoa 1/4 565 Mya, trace fossils; (3) Bilateria 1/4 555 Mya, trace fossils, (3) Protostomia 1/4 540 Mya, helcionellids, protoconodonts, (5) Deuterostomia 1/4 530 Mya, echinoderm plates. Adapted from Cunningham et al., 2017.

### 1.4.2. Geochemical context

In the past decades, the lack of pre-Cambrian eumetazoan fossils and the sudden and late diversification of animals was reasoned to be related to changes in the Earth's geochemical environment. The end of the Neoproterozoic was especially subjected to key environmental and biological transitions from a geochemical point of view. In particular, water masses were enriched in oxygen and more predictably gradually oxic than in previous eras (Knoll, 2011).

This transition gradually shifted marine environments from a sulfidic to a more ferruginous composition, favouring a less toxic milieu which promoted eukaryotic diversification to new environments (Knoll, 2011). In parallel, this transition was influenced by the second significant increase in oxygen levels in the atmosphere, which has also been linked to animal origins (Lyons et al., 2014a). This view, known as the "oxygen control hypothesis" or OCH (Knoll, 1992), argues that the raise of oxygen levels and the maintenance of a constant oxygen concentration in the atmosphere created the necessary permissive environment of oxygen availability, reaching the relatively high oxygen level requirements to support oxidative metabolism and animal life (Nursall, 1959). In fact, the synthesis of collagen-based extracellular matrices necessary to sustain multicellular tissues in animals require high oxygen concentrations in the environment (Towe, 1970). The OCH was quantified by Berkner and Marshall, which argued that 1% of present atmospheric levels of oxygen passed the "critical threshold" that allowed aerobic metabolism and, subsequently, animals to evolve at the end of the Proterozoic Eon (Berkner and Marshall, 1965). Shortly afterwards, geological evidence suggested that 1% oxygen levels were reached as far as back to 1.2 (Cloud, 1965) to 2 Gya (Cloud, 1976), and thus were not sufficient for animal evolution. Instead, oxygen levels perhaps closer to 3% were required for the emergence of animals (Cloud, 1968). While other inferences suggested 6-10% oxygen levels as the minimum oxygen requirements for animals (Erwin et al., 2011; Mills et al., 2014; Runnegar, 1991), other estimates suggested oxygen levels could go as low as 1-3% (Runnegar, 1991). Nevertheless, more recent analyses have shown that collagen synthesis can also occur at low oxygen concentrations, although with lower efficiency (Mills and Canfield, 2014). Moreover, certain animals are able to grow under low oxygen levels, such as the demosponge *Halichondria panicea* at 0.5%-4% oxygen concentration (Mills et al., 2014) or some bilaterians at 0.3% oxygen concentration (Mills and Canfield, 2014).

Thus, the shift from an initially reducing atmosphere to a modern oxidizing environment with constant oxygen levels were not directly sufficient for animal origins, it might have suppressed the proliferation of large multicellular life (Knoll and Carroll, 1999; Nursall, 1959). Other external geochemical constraints possibly also played a role in animal evolution, such as global surface temperature changes (Schwartzman, 2002) or salinity (Knauth, 2005). However, they still cannot fully explain the huge diversity of animal species (Knoll and Carroll, 1999; Sperling et al., 2013).

### 1.4.3. Ecological context

In combination with the geochemical context, ecological factors were also possibly key for the emergence of animals, and especially might have had an impact on reshaping animal feeding modes and their morphological features (Knoll and Carroll, 1999; Sperling et al., 2013). Animals evolved in an environment teeming with bacteria and have lived since then in close association with bacteria throughout their evolutionary history. As mentioned above, one critical factor that positively influenced animal origins was the series of oxygenation events that raised and prolonged oxygen levels both in the atmosphere and in the ocean. These events were partly driven by the photosynthetic activity of marine cyanobacteria (Alegado and King, 2014; Kasting and Siefert, 2002; Lyons et al., 2014b; Schirrmeister et al., 2015). Moreover, animal-bacteria interactions may have inadvertently shaped their biology as well. These include predation on bacteria, harbouring bacterial commensals, and infection with bacterial pathogens.

The earliest fossil evidence recording the existence of bacteria (and possibly archaea) dates back around 3.45 billion years, in the form of layered macroscopic sedimentary structures known as “stromatolites” (Allwood et al., 2007; Grotzinger and Knoll, 1999; Kaiser, 2001; Mojzsis et al., 1996; Rosing, 1999; Walter et al., 1980). Interestingly, the emergence and diversification of animals coincided with a sudden and rapid decline in the abundance of stromatolites in the Cambrian, leading some authors to hypothesize that bacteria involved in stromatolite formation could have been prey targets of early animals (Alegado and King, 2014; Awramik, 1971; Garrett, 1970; Walter and Heys, 1985).

Nearly all animals have stable interactions with bacteria. For example, larvae produced by some early-branching animals, including sponges and cnidarians, respond to uncharacterized compounds released by environmental bacteria (Webster et al., 2004; Woollacott and Hadfield, 1996). In addition, commensal bacteria can regulate development and gut morphogenesis in some animals (Bates et al., 2006; Mazmanian et al., 2005; McFall-Ngai and Ruby, 2000; Montgomery and McFall-Ngai, 1994). Interestingly, bacterial cues are able to regulate a developmental switch in choanoflagellates, the closest relatives of animals. Bacterial-induced colony formation allows the choanoflagellate species *Salpingoeca rosetta* to exhibit enhanced prey capture abilities.

Moreover, bacteria also play a key role as trigger for meiotic sex in *S. rosetta* (Levin et al., 2014; Woznica et al., 2017, 2016; Woznica and King, 2018). The bacterivorous nature of choanoflagellates also suggests that the progenitors of animals were likely bacterivores. Hence, bacterivory may have also impacted genome evolution at the onset of animals by lateral gene transfer (LGT) mechanisms (Alegado and King, 2014; Ford Doolittle, 1998). Other adaptive advantages of being in a close relationship with bacteria during mass extinctions would be extending the nutritional capacity of animal hosts by commensal bacteria. This may have allowed animals to adapt to changing environments and expand into new ecological niches (Alegado and King, 2014).

### 1.4.4. Theories on the origin of Metazoa

The origin and evolution of animal multicellularity has been a long-debated and challenging question in biology. For over a century, Haeckel's "Gastraea theory", inspired by the early stages of animal development, was the first widely accepted model for animal origins (Haeckel, 1874). The Gastraea theory suggests that the transition from unicellularity to multicellularity occurred in two consecutive stages **(Figure 5A).** First, unicellular flagellates formed a hollow ball-shaped (volvox-like) colony of identical cells, named *Blastea* for its resemblance to the Blastula, an early stage in animal embryogenesis. At the second stage, the homogeneous *Blastea* invaginated at the posterior pole, forming a solid double walled cup-like structure, the precursor of the primary gut.

After this important event, the *Blastea* underwent primary cell differentiation (spatial cell differentiation), the first step in the evolution of Metazoa (Haeckel, 1874). Thus, the animal ancestor was a blastula-like colony of uniform cells that gradually evolved cell differentiation.

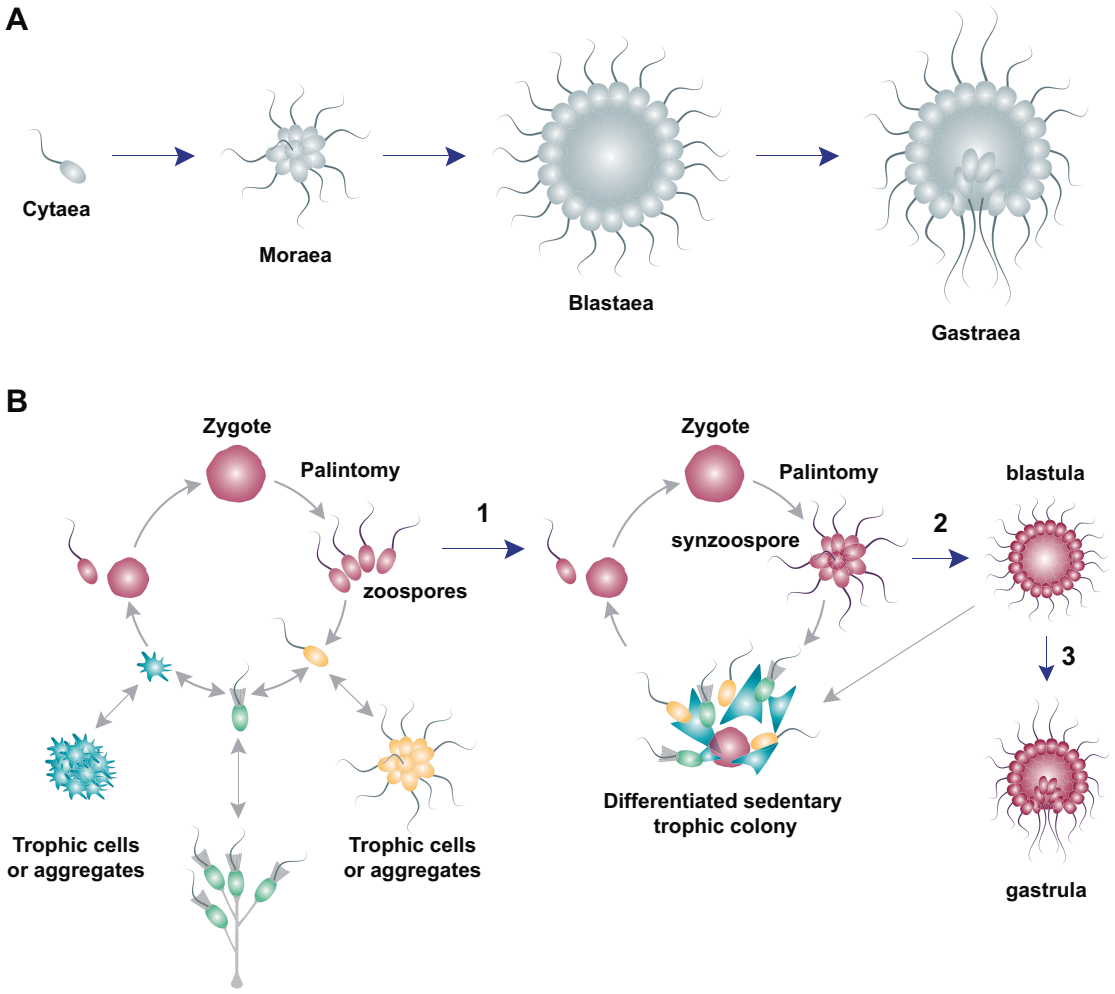



**Figure 5. The Gastraea and Synzoospore scenarios.** (A) The Gastraea theory assumes gradual modification of a colony of uniform cells. Primary cell differentiation occurs with the formation of functional primary gut, the evolutionary precursor of the endoderm. (B) The Synzoospore theory envisions the metazoan ancestor as a protist with a complex life cycle that includes monotomously dividing trophic cells (or cellular aggregates), hypertrophic growth of gametes, and their subsequent palintomic cleavage producing non-feeding dispersal zoospores. The transition to multicellulairty occurs with (1) integration of trophic cells into a differentiated colonial body (transition towards spatial cell differentiation and the synzoospore), (2) integration of zoospores into the uniform synzoospore, the primary lecitotrophic dispersal larva of the animals capable of carrying primary building material (a multicellular synzoospore, the blastula) and (3) the emergence of feeding larvae in the evolution of Eumetazoa. This last step allows the development of the primary gut as an adaptation to catch the multiceled prey, and entails a transition toward neoteny and loss of the sedentary filtrator stage. Blue arrows mark hypothetical evolutionary transitions, gray arrows designate the life cycle. Adapted from Mikhailov et al., 2009.

In the later years, Haeckel's Gastraea theory was gradually reshaped. Lankester proposed that the ancestral colonial protozoan was a solid, flagellated protozoan-like animal (*e.g., Pandorina*) (Lankester, 1877, 1873) and Metschnikoff suggested that the ancestral colonial organism, similar to a choanoflagellate, resembled the existing Proterospongia (Metchnikoff, 1886). Modern authors have further adapted this idea into the Choanoblastea theory, supported by the striking morphological resemblance between colonial choanoflagellates and the choanocyte chamber of sponges (Fairclough et al., 2010; Grosberg and Strathmann, 2007; Willmer, 1990). This view postulates that the metazoan multicellular development initially occurred through a series of incomplete cell divisions from a flagellated blastea, or "choanoblastea", originated from a choanoflagellate colony. At the early phase of the transition to multicellularity, cells composing the choanoblastea were undifferentiated, and by successive generations they started to develop different cell types. Those differentiated cell types led to the blastula formation which represented the origin of a colony that started to gradually become more and more complex, until the formation of the Urmetazoan, the single common ancestor of all metazoans (Nielsen, 2008).

Recent knowledge from the genomes and complex lifestyles of choanoflagellates and other unicellular relatives of animals shows that there is evidence of regulated temporal cell differentiation prior the emergence of animals (de Mendoza et al., 2015; Fairclough et al., 2013; Sebé-Pedrós et al., 2016a, 2013a). This, together with the genetic complexity observed in those lineages, indicates that the nature of the unicellular ancestor of animals was relatively plastic with a complex life cycle, multiple cell types and a complex genetic repertoire of genes related to multicellular functions **(Figure 6)** (Sebé-Pedrós et al., 2017; Torruella et al., 2015). These observations completely renewed the view of animal multicellularity, originally conceived as a "synzoospore" in the so called "Synzoospore theory" **(Figure 5B)** (Mikhailov et al., 2009). The Synzoospore theory integrates pre-existing transient cell types into the body of an early metazoan, which already possessed a complex life cycle with a differentiated sedentary filter-feeding trophic stage, and a non-feeding blastula-like larva, the synzoospore (Mikhailov et al., 2009). Therefore, the animal ancestor was a composition of premetazoan cell types that appeared after a temporal-to-spatial switch in cell differentiation programs (Budd and Jensen, 2017; Mikhailov et al., 2009; Sebé-Pedrós et al., 2017).

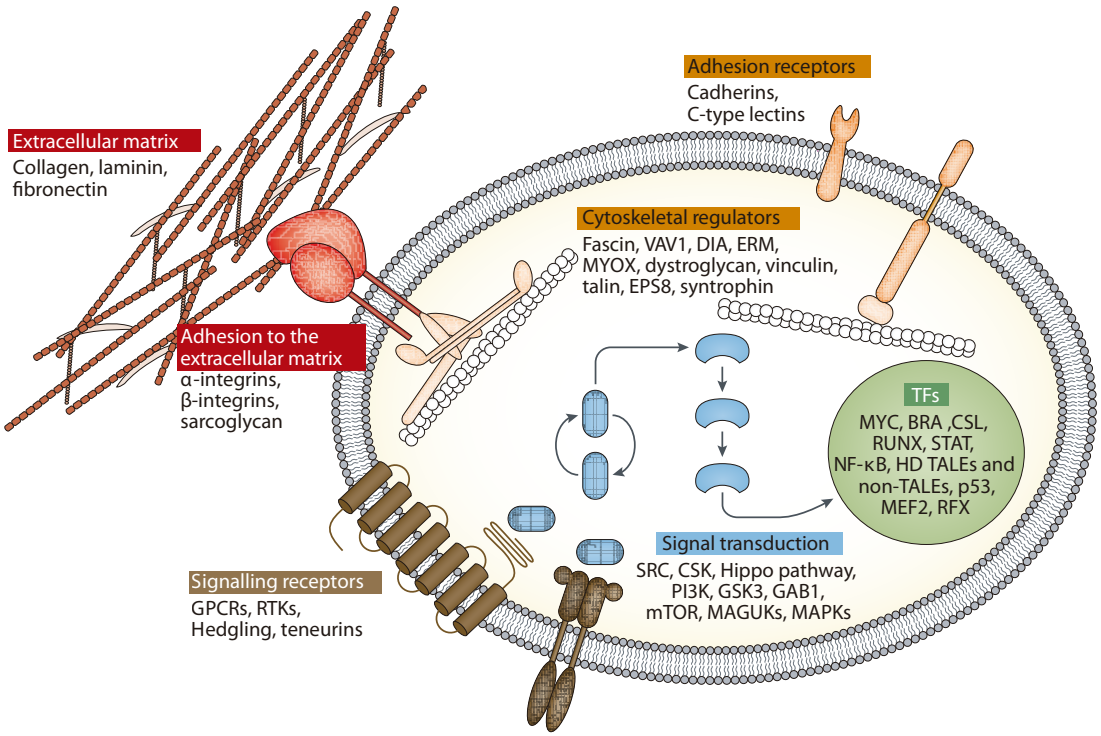



**Fig. 6. The premetazoan genetic toolkit: an inferred gene repertoire of the unicellular ancestor of animals.** Many genes that are important for the metazoan multicellularity-related functions - such as adhesion, signalling and transcriptional regulation - evolved in a unicellular context and were present in the unicellular ancestor of animals. The inference is based on the presence of homologes of these metazoan genes in the genomes of unicellular relatives of animals. *BRA: Brachyury*; *DIA: Diaphanous*; *EPS8: Epidermal Growth Factor Receptor Tyrosine Kinase substrate 8; GAB1: GRB2-associated binding protein; GPCRs: G-protein coupled receptors; GSK3: Glycogen Synthase Kinase 3; HD: Homeodomain; MAGUK: membrane-associated Guanylate Kinases; MAPKs: Mitogen-activated protein Kinases; MEF2: Myocyte-specific enhancer factor 2; mTOR: mechanistic target of Rapamycin; MYOX: Myosin X*; *NF-κB: Nuclear Factor-KappaB; PI3K: Phosphatydilinositol 3-Kinase; RTKs: Receptor Tyrosine Kinases; STAT: Signal Transducer and activator of transcription; TALEs: Three aminoacid loop extensions.* Adapted from Sebé-Pedrós et al., 2017.

### 1.4.5. Origin of metazoan genetic developmental toolkits

Animal multicellularity entailed the diversification of architecturally complex body plans built from a diverse collection of differentiated cell types generated by cell-type specific gene expression programs (Bonner, 1998; Carroll, 2000; Raff, 1996). To attain this tantalizing complexity, animals require the existence of coordinated developmental genetic programs, key to control cell differentiation, cell-cell communication and cell adhesion processes (Larroux et al., 2008).

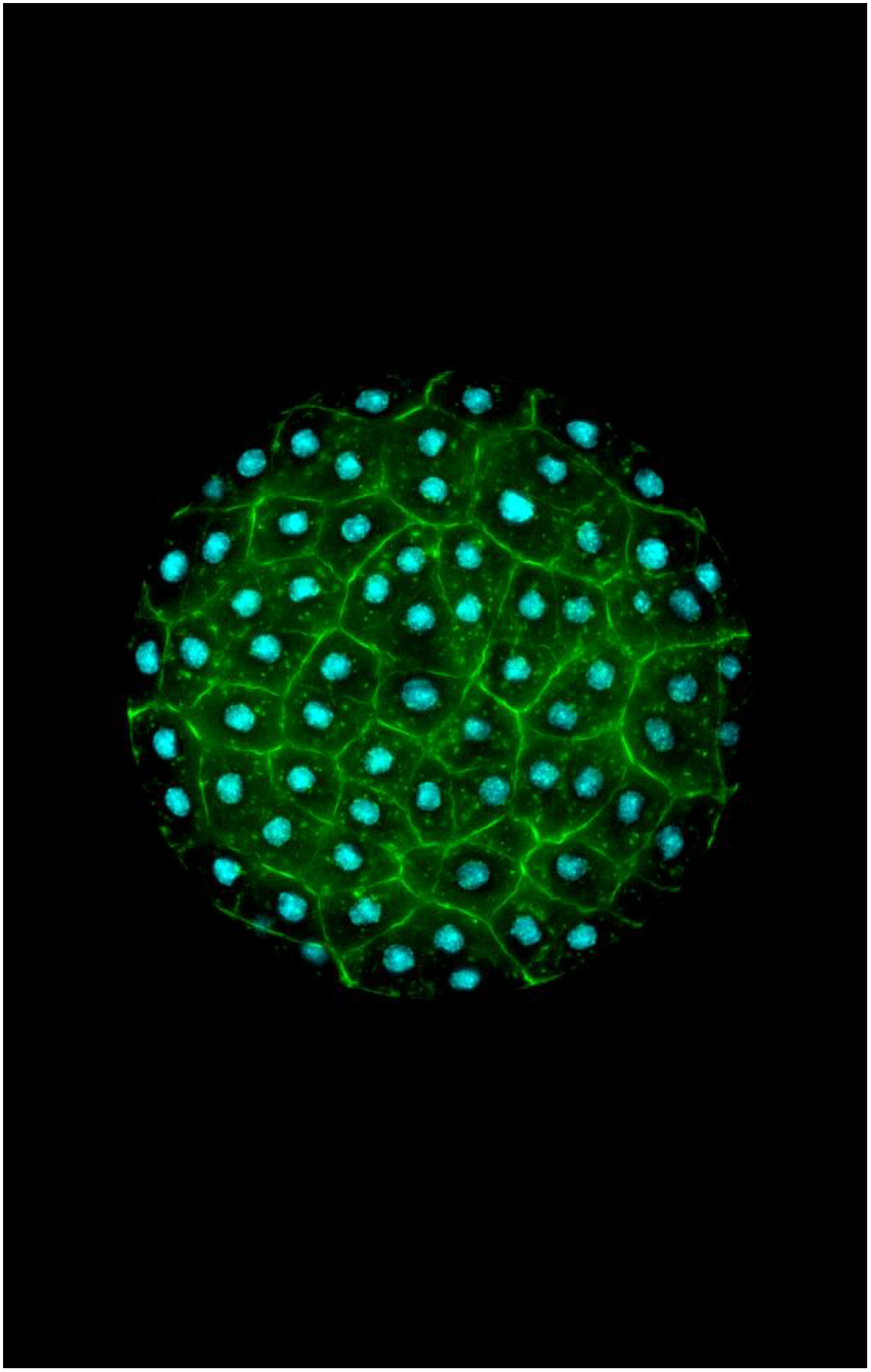
The genetic toolkit for animal development comprises a set of few hundred genes from a few dozen gene families which underwent extensive gene duplications and evolved new roles for establishing the patterning of animal body plans (Carroll, 2001, 2000; Knoll and Carroll, 1999). Examples include components of the *Hox* transcription factors, *Wnt* and *receptor tyrosine kinases* cell signalling families and *cadherins* and *integrins* involved in cell adhesion (Rokas, 2008b).

Developmental transcription factors (TFs) appear to be one of the most crucial aspects of animal development and have been associated to the origin of complex morphological features, such as the ones observed in bilaterians. The large majority of TFs networks that lie at the basis of bilaterian developmental gene regulatory networks are also present in early-branching animals (including sponges, placozoans and cnidarians), and thus aroused much earlier in evolution (Degnan et al., 2009; King et al., 2008; Larroux et al., 2008, 2006; Putnam et al., 2007; Richards and Degnan, 2009; Srivastava et al., 2010, 2008). Nevertheless, most TFs families appear to have expanded during early eumetazoan evolution. In fact, comparisons among early-branching animals show an appreciable increase of TFs family complexity, which qualitatively and quantitatively changes across lineages, being more complex in cnidarians relative to poriferans or placozoans (Degnan et al., 2009; Levin et al., 2016; Putnam et al., 2007; Ryan et al., 2013; Srivastava et al., 2010). In fact, it has been hypothesized that higher genome complexity requires an over-proportional increase in regulatory elements: the number of TFs per genome correlates over-proportionally with the number of genes in genomes, resulting in a higher proportion of TFs in larger genomes (Babu et al., 2004).

The increase in the number of regulatory proteins in general and of TFs in particular has been connected to phenotypic innovations and the evolution of more complex organisms (de Mendoza et al., 2013; Schmitz et al., 2016). Taken together, the emergence and expansion of TF networks are related to increases in morphological complexity and number of cell types (Schmitz et al., 2016; Sebé-Pedrós et al., 2018).

Some of the key TFs families of the metazoan developmental toolkit include members of the *basic helix-loop-helix (bHLH), myocite enhancer factors 2 (Mef2), Fox, Sox, T-box, Ets, nuclear receptor (NR), Rel/Nuclear Factor-KappaB (NF-κB), basic-region leucine zipper (bZIP), Smad* families and a range of homeobox-containing classes, including *ANTP, Prd-like, Pax, POU, LIM-HD,* Six and three-amino acid-loop extension *(TALE)* (Degnan et al., 2009).

The study of the origins and assembly of key developmental TFs may be carried out through careful comparisons between animals and their extant closest unicellular relatives. In fact, several genes involved in the animal developmental toolkit have been identified in the unicellular relatives of animals over the last decade (de Mendoza et al., 2013; Grau-Bové et al., 2017; King et al., 2008; Putnam et al., 2007; Sebé-Pedrós et al., 2011; Srivastava et al., 2010; Suga et al., 2013). For example, the choanoflagellate *Monosiga brevicollis* (Fairclough et al., 2013; King et al., 2008) possesses *cadherins, integrins, receptor tyrosine kinases* and *laminins*, as well as the *Myc* TF network. Further comparative genomics analyses in filastereans and ichthyosporeans revealed they possess a complex repertoire of developmental TFs networks, like *Brachyury*, the *Myc-Max* network, *Runx* and *p53* (de Mendoza et al., 2015; Sebé-Pedrós et al., 2011). Thus, ancient regulatory genes evolved prior the divergence of animals and were possibly co-opted for the various developmental roles associated with animal morphology.



## *Section 1.5*

# The striking diversity among unicellular holozoans, the closest relatives of animals

Animals are closely related to a heterogeneous assembly of protist lineages, altogether forming the group Holozoa (Lang et al., 2002) within the Opisthokonta eukaryotic supergroup (Cavalier-Smith, 1987; Ruiz-Trillo et al., 2008, 2004; Shalchian-Tabrizi et al., 2008; Torruella et al., 2015, 2012). Unicellular Holozoa lineages are Choanoflagellatea, Filasterea, Ichthyosporea[2] and the still enigmatic Corallochytrea clade, composed to date of one single species *(see* **Figure 7** *on next page).* Recently, three new holozoan flagellated species have been described, reshaping the phylogenetic framework of this group.

In this section, I will introduce the inconspicuous diversity among the unicellular relatives of animals. I will examine their distribution and ecological heterogeneousness, focus on their morphological features, life cycles and multicellular-like behaviours and present their genetic diversity and phylogenetic relationships. In a later section *(1.6. Emerging model systems among unicellular holozoans),* I will pinpoint the importance of developing them as experimentally tractable systems to address the origin and evolution of animal multicellularity.

[2] Image on previous page credits: *Sphaeroforma arctica,* a member of the Ichthyosporea, stained with Phalloidin (green) and DAPI (blue). *"Sphaeroforma arctica"* by Omaya Dudin (CC by 2.0).

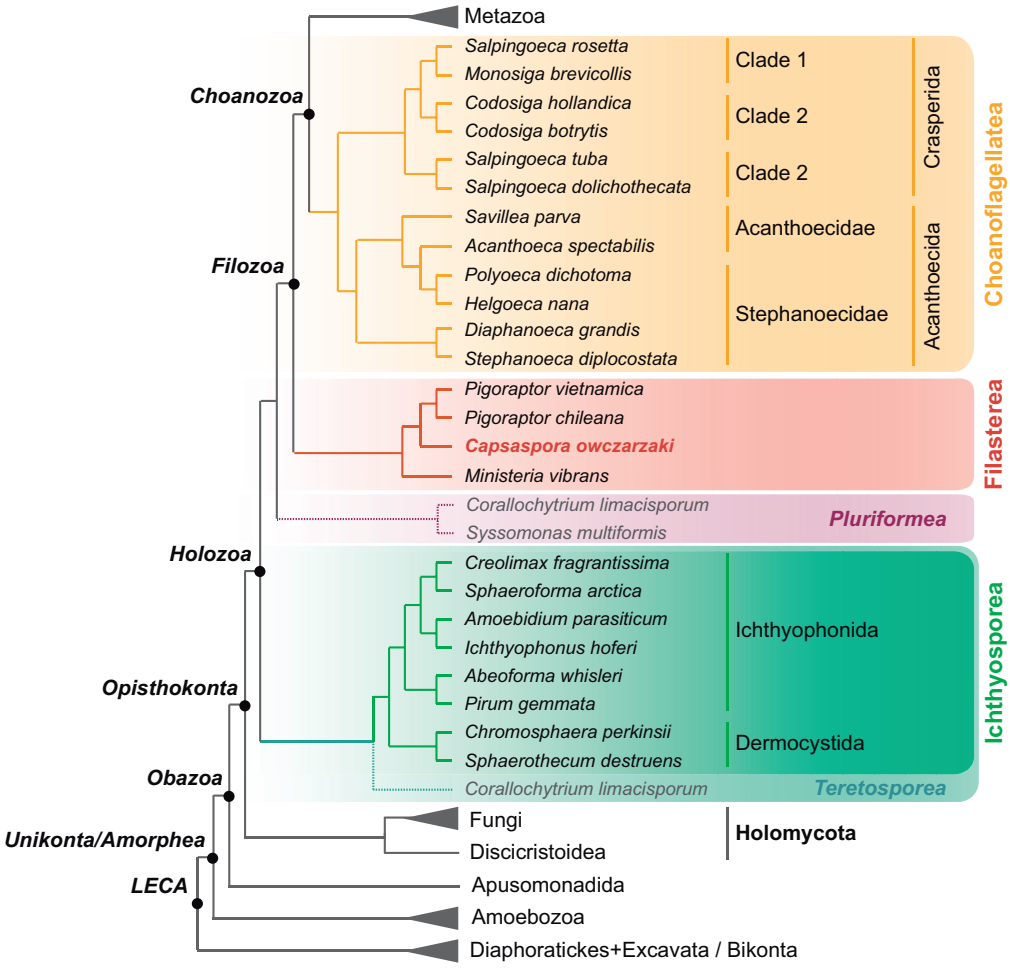



**Fig. 7. Phylogenetic classification of unicellular Holozoa within the eukaryotes.** Note the changing position of *Corallochytrium limacisporum* grouping with Ichthyosporeans forming the Teretosporea clade (Torruella et al., 2015; Grau-Bové et al., 2017) or with the recently described *Syssomonas multiformis* forming the Pluriformea clade (Hehenberger et al., 2017). *LECA: Last Eukaryotic Common Ancestor.*

### 1.5.1. Choanoflagellatea

Choanoflagellates, *a.k.a.,* Choanomonada (Adl et al., 2012, 2005), is a group of free-living heterotrophic flagellates that has been a longstanding mystery for microbiologists. Dating back to the nineteenth century, they were first morphologically associated with animals due to their striking resemblance to choanocytes, a specific cell type of sponges **(Figure 8)** (James-Clark, 1866). This similarity gave rise to the hypotheses of a close relationship between animals and choanoflagellates and animals evolving from a choanoflagellate-like ancestor (Cavalier-Smith, 2017; Cavalier-Smith and Chao, 2003; Lang et al., 2002; Ruiz-Trillo et al., 2008; Steenkamp et al., 2006; Wainright et al., 1993).

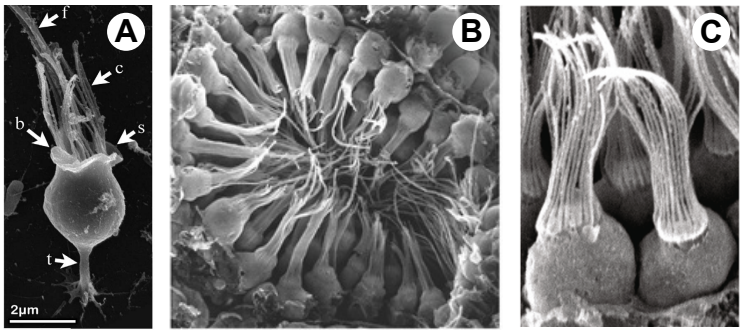



**Figure 8. Choanoflagellates and its resemblance to sponge choanocytes.** (A) SEM image of *Salpingoeca rosetta* thecate cell. Key: flagellum (f), collar (c), bacterium (b), skirt (s) and theca (t) (image credit Dayel and King, 2014). (B) *Oscarella carmela* choanocyte chamber. Like choanoflagellates, sponge collar cells (choanocytes) also have a single apical flagellum and collar of microvilli that they use to capture bacteria (image credit Scott Nichols). (C) Sponge choanocytes (image credit Peña et al., 2016).

In the past years, phylogenomic analyses confirmed their placement in the tree of life as the closest living unicellular relatives of animals (Lang et al., 2002; Ruiz-Trillo et al., 2008, 2004; Wainright et al., 1993; Zettler et al., 2001). Choanoflagellates and animals have been proposed to belong to the monophyletic Apoikozoa clade (Budd and Jensen, 2017), referred to, in informal terms, as *choanimals* (Fairclough et al., 2013). More recently, as an alternative to these terms, Brunet and King redefined and proposed the term Choanozoa as the clade containing the most recent common ancestor of animals and choanoflagellates (the Urchoanozoan), along with all of its descendants **(Figure 7)** (Brunet and King, 2017).

Choanoflagellates include around 250 species of spherical/ovoid protists, widely distributed in a range of aquatic environments: from marine water columns or abyssal plains, to freshwater or anoxic/hypoxic brackish waters (del Campo and Massana, 2011; del Campo and Ruiz-Trillo, 2013; Nitsche et al., 2007; Wylezich et al., 2012). They are also considered to be major contributors to aquatic microbial food webs (Arndt et al., 2000; King, 2005).

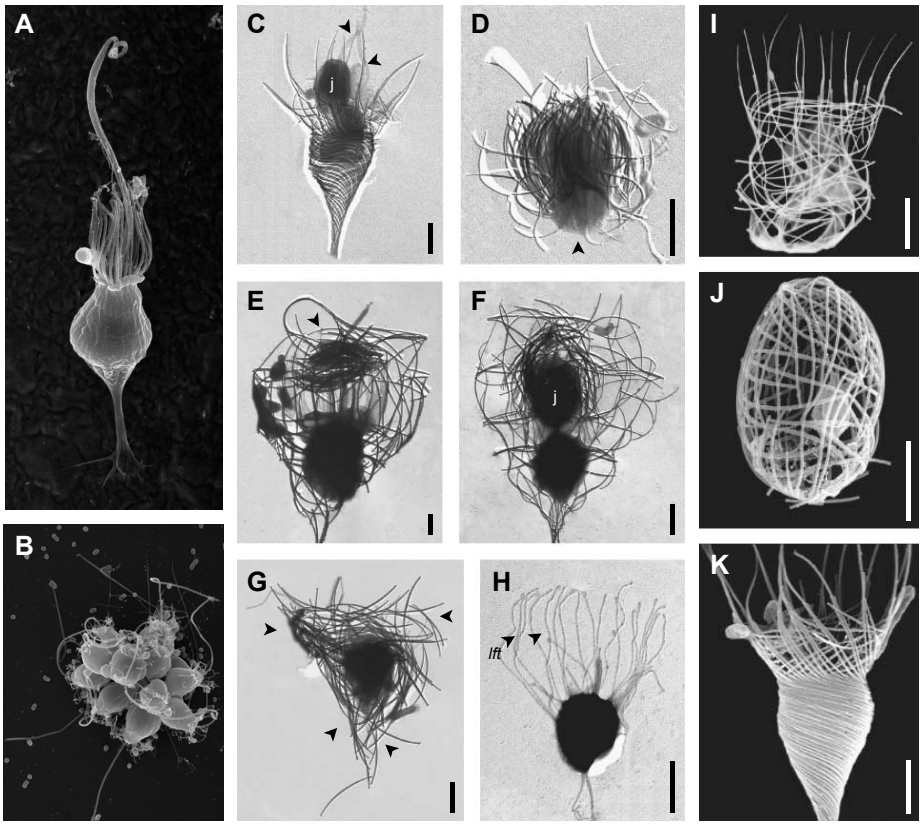



**Figure 9. Diversity of morphologies in choanoflagellates.** (A-B) *Salpingoeca rosetta* thecate cell and rosette colony, respectively (image credits Mark Dayel and Nicole King). (C-K) Examples of loricate choanoflagellates. (C-D) *Acanthoeca spectabilis,* a distinctive member of the nudiform clade. (C) Recently divided cell showing juvenile (j) and sister cell, both with flagellum (arrows). (D) Juvenile cell with covering of vertical bundels of costal strips (arrow). (E-H) *Stephanoeca diplocostata,* an example of tectiform lorica replication. (E) Cell with substantial accumulation of costal strips at top of the collar (arrow). (F) Recently divided cell showing inverted juvenile (j) emerging from parent lorica with covering of costal strips. (G) Recently released juvenile with bundles of strips in vertical and transverse planes (arrows). (H) Juvenile with extended lorica forming tentacles, *lft*. Siliceous costae has been removed with hydrofluoric acid. (I-K) Scanning electron microscopy (SEM) images of nudiform loricae showing helical costae and anterior spines of *Helgoeca nana* (I); *Savillea micropora* (J) and *Acanthoeca spectabilis* (K). Images I-K from Leadbeater et al., 2009.

Morphologically, they present a “collar-bearing” structure, similar to a funnel-shaped collar of interconnected microvilli (a specialized actin-based filopodia structure) at the base of a flagellum **(Figure 9A)** (Leadbeater, 2015). This microvilli-flagellum complex is used to propel forward the cell through the medium as well as to facilitate phagocytosis by a precise coordination of whipping movements that create currents leading to the collar (Dayel and King, 2014). Some of their representatives are able to form clonal colonies **(Figure 9A-B)** and one group is remarkable for its siliceous basket-like covering **(Figure 9C-K).** Moreover, they are capable of both asexual and sexual reproduction (Levin and King, 2013).

#### 1.5.1.1. Classification of Choanoflagellatea

Choanoflagellates are classified in two main monophyletic clades: Acanthoecida (including Stephanoecidae and Acanthoecidae) and Craspedida. This recent classification overcomes the limitations of morphology-based taxonomy at the species and genera level and still agrees with traditional classifications (Carr et al., 2017; Nitsche et al., 2011).

Acanthoecida, *a.k.a.,* loricate choanoflagellates, is a group of mostly marine and brackish water choanoflagellates (with few freshwater records) that present a distinctive siliceous extracellular structure forming a rigid basket-like cage that surrounds the cell, known as *lorica* **(Figure 9C-K)**. Most of the representatives of this group belong to the Stephanoecidae (around 150 species) and present “tectiform” loricae with rings that facilitate their pelagic lifestyle **(Figure 9E-K)** (Carr et al., 2008). Daughter cells receive a full set of strips from the parent cell that are immediately assembled into a new lorica, therefore they are never naked. Few other members (between 5-6 species) belong to the Acanthoecidae, and present a “nudiform” lorica that comprises longitudinal and helical horizontal costae **(Figure 9C-D, I-K)** (Leadbeater, 2015). Both groups present as many variations as there are species, differing in their lorica assembly, construction, patterns and sizes (Leadbeater, 2015).

Craspedida is a group of choanoflagellates with exclusively organic coverings divided in three different clades (Clades 1, 2 and 3). This classification was described using a

phylogenetic approach and only the representatives of Clade 3 share the same morphology (Carr et al., 2017). In general, their basic life cycle comprises a sedentary interphase cell, with mainly feeding functions, followed by cell division that produces transitory motile cells with dispersive functions. The flagellar apparatus is ideally suited for this dual functional role. Some species produce a specialized organic structure that binds cells to the substrate, the *theca* (especially in Salpingoecidae morphology) **(Figure 9A)**. Others (non-thecated), produce coverings of glycocalix or sheath (Codosigidae morphology). Interestingly, other species are able to produce multicellular-like structures of clonal cell colonies, such as the *rosette* colonies produced by *Salpingoeca rosetta* **(Figure 9B and 10)** (Carr et al., 2017, 2008; Fairclough et al., 2013; Nitsche et al., 2011). Other clades of choanoflagellates have been defined from environmental surveys pointing to their wide distribution and still hidden diversity to uncover (del Campo and Ruiz-Trillo, 2013).

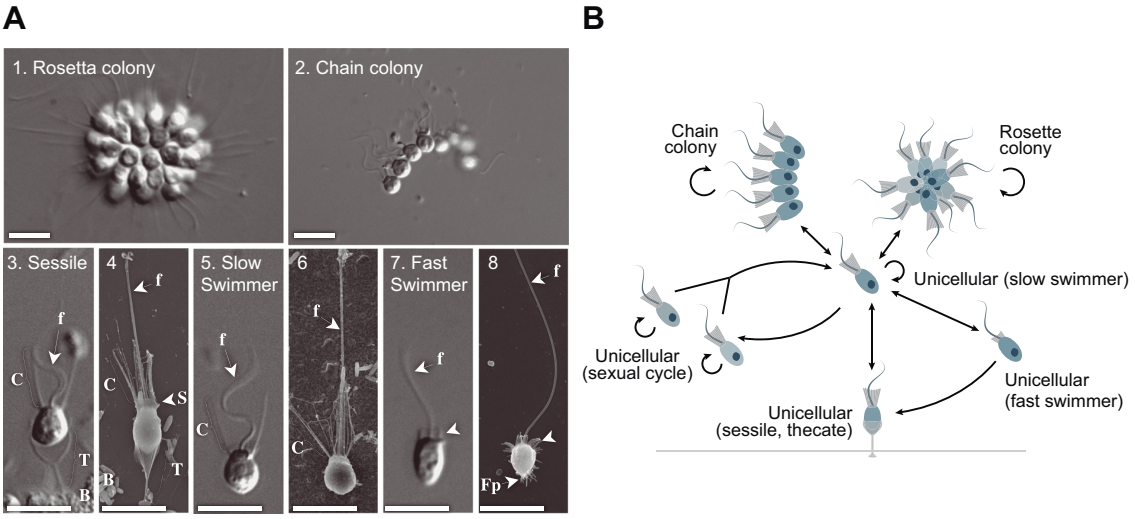



**Figure 10. *Salpingoeca rosetta* morphologies and life cycle.** (A) Five distinct cell morphologies observed in *S. rosetta*. Images in 1-3, 5 and 7 correspond to DIC microsocopy; 4, 6 and 8, correspond to SEM images. (3-4) Sessile forms; (5-6) Slow swimmers; (7-8) Fast swimmers. Key: flagellum (f); collar (C); theca (t); skirt (s); filopodia (Fp); bacteria (B). Scale bars: 5 µm. Images from Dayel et al., 2011. (B) Life cycle transitions of *S. rosetta.*

Currently, there are two available choanoflagellate genomes, both of them craspedids: the *Monosiga brevicollis* (King et al., 2008) and the colony-forming *S. rosetta* (Fairclough et al., 2013). *S. rosetta* has been thoroughly characterized from the point of view of transcriptomics and cell biology *(see section 1.6.1. Salpingoeca rosetta, a benchmark among choanoflagellates).* Moreover, recent analysis included 19 new transcriptomes from a broad representation of different choanoflagellate species (Richter et al., 2018).

### 1.5.2. Filasterea

Filasterea is an holozoan lineage composed of only four known species to date **(Figure 11)**: the endosymbiont *Capsaspora owczarzaki* (Owczarzak et al., 1980; Stibbs et al., 1979), the free-living heterotroph *Ministeria vibrans* (Shalchian-Tabrizi et al., 2008) and the more recently described predatory flagellates *Pigoraptor vietnamica* and *Pigoraptor chileana* (Hehenberger et al., 2017). Together with animals and choanoflagellates, they form the Filozoa clade **(Figure 7)** (Shalchian-Tabrizi et al., 2008).

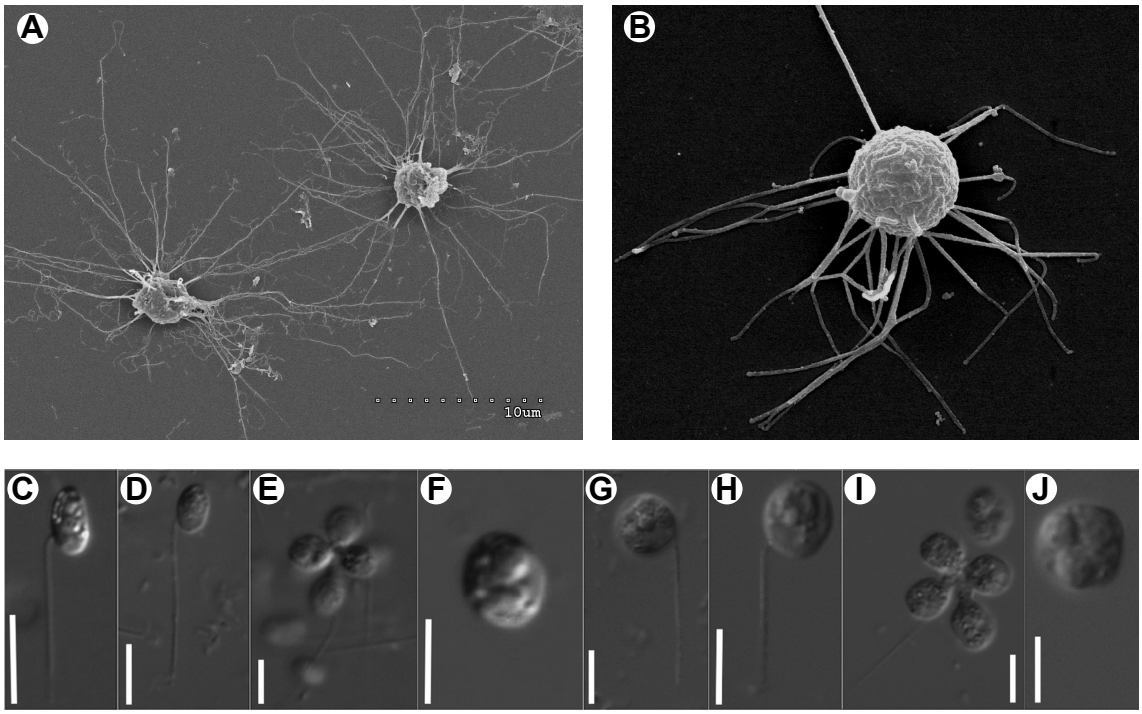



**Figure 11. Diversity of morphologies in filastereans.** (A) SEM image of *Capsaspora owczarzaki ("Capsaspora owczarzaki"* by Multicellgenome Lab, CC by 2.0). (B) SEM image of *Ministeria vibrans* (*"Ministeria vibrans"* by Multicellgenome Lab, CC by 2.0). (C-F) Light micrographs of *Pigoraptor vietnamica* life cycle stages: two different flagellated cells (C and D), cell cluster (E), and cyst (F). (G-J) Light micrographs of *Pigoraptor chileana* life cycle stages: two different flagellated cells (G and H), cell cluster (I), and cyst (J). Scale bars represent 10 μm; (C-J) images from Hehenberger et al., 2017.

*Capsaspora owczarzaki* is the member of filastereans that has been more deeply studied, as the complete genome, transcriptome, proteome, phosphoproteome and epigenome are available **(Figure 11A)** (Sebé-Pedrós et al., 2016b, 2016a, 2013b, 2011, 2010; Sebé-Pedrós and Ruiz-Trillo, 2010; Suga et al., 2013).

An exhaustive description of the molecular and morphological characterization of *C. owczarzaki* will be presented in *section 1.7. Capsaspora owczarzaki.*

*Ministeria vibrans* is a small heterotrophic, filopodiated, spherical amoeba of around 4 µm in diameter found as a marine free-living species (Cavalier-Smith and Chao, 2003; Tong, 1997). It presents several fine, stiff, slender radiating arms of equal length, and a thicker stalk flagellum frequently extended by a cytoplasmic thread of variable length that helps cells attach to surfaces **(Figure 11B)** (Torruella et al., 2015). It also extends a pseudopodium to feed on bacteria, especially motile rods. *Ministeria vibrans* cells sometimes vibrate when adhered to surfaces, a fact that gives the species its name, although the function of this behaviour is still unknown. Under culture conditions, it grows slowly in a minimal medium containing a mixture of live bacteria, a fact that still complicates further experimental manipulation *in vitro*. For this reason, the life cycle of *M. vibrans* still remains unknown. Nevertheless, there is transcriptomic data available (Torruella et al., 2015), and its genome is currently being obtained in our lab. Apparently, *Ministeria vibrans* has a sibling species belonging to the same genus, *Ministeria marisola* (Patterson et al., 1993). Although this species has never been molecularly characterized to date and cultures are still not available, it presents morphological similarities to *M. vibrans*.

Recently, two additional filasterean species were isolated from freshwater environments: *Pigoraptor vietnamica* and *Pigoraptor chileana* (Hehenberger et al., 2017). Both *Pigoraptors* are predatory flagellates that feed on large eukaryotic prey (similar size to themselves) and are able as well to engulf bacteria and small detritus. They also seem to be morphologically plastic and present complex life cycles with different stages, including immobile cysts and multicellular clusters of several cells (Hehenberger et al., 2017). *Pigoraptor vietnamica* flagellated cells are elongated with an oval-like shape and they measure around 5-12 µm in diameter **(Figure 11C-F).** *Pigoraptor chileana* flagellated cells are roundish and slightly bigger, measuring around 6-14 µm in diameter. Both motile swimming cell types present a smooth flagellum that emerges from the middle-lateral point of the cell **(Figure 11G-J)**. They are also naked and sometimes they produce a short, thin pseudopodium. Besides its morphological description, transcriptomic sequencing of both species revealed by homology searches that they possess genes related to extracellular matrix (ECM) adhesion, as well as cell-cell adhesion components.

Moreover, they both contain nearly a complete integrin adhesome system with the associated signalling and cell adhesion components, such as the integrin-linked kinase *(ILK), PINCH and Parvin* (*a.k.a., IPP complex*), as well as several *T-box* domain-containing transcripts related to *Brachyury* TF family (Hehenberger et al., 2017).

With the new *Pigoraptors* data, Small Subunit (SSU) rRNA gene phylogenies relates another group of abundant environmental sequences with filastereans, the previously defined Marine Opisthokonts 1 (MAOP1) clade (del Campo et al., 2015; del Campo and Ruiz-Trillo, 2013; Hehenberger et al., 2017). Further investigations need to be performed to gain more insights into this group of unicellular holozoans.

### 1.5.3. Ichthyosporea

Ichthyosporea, *a.k.a.,* Mesomycetozoea (Mendoza et al., 2002), is a group of osmotrophic and saprotrophic fungus-like protists frequently multinucleated and sometimes with a single posterior flagellum, especially in dispersal forms **(Figure 7)**. They were formerly referred to as the DRIP clade, an acronym for the original four members of the group: *Dermocystidium*, the "*rosette* agent", *Ichthyophonus* and *Psorospermium* (Cavalier-Smith, 1998; Ragan et al., 1996).

Most members of Ichthyosporea have been found in commensal, mutualistic or parasitic relationships with aquatic (both freshwater and marine) and terrestrial animals, and therefore have been directly isolated from different animal tissues, especially guts of molluscs and arthropods (Glockling et al., 2013). Few species have been identified as free-living as well (Hassett et al., 2015) and the presence and abundance of other putative lineages of ichthyosporeans in unsampled environmental surveys suggests that there are still other undescribed free-living members to be analysed (del Campo et al., 2015; del Campo and Massana, 2011; del Campo and Ruiz-Trillo, 2013; Savin et al., 2004; Stougaard et al., 2002; Takishita et al., 2007, 2006, 2005). Over the past decade, the number of taxa in Ichthyosporea has increased considerably, reaching to date about 40 characterized species, although around half of them are phylotypes (Feldman et al., 2005; Lohr et al., 2010; Lord et al., 2012; Marshall et al., 2008; Marshall and Berbee, 2011).

Moreover, many taxa in the Eccrinales and Amoebidiales originally identified as fungi have been phylogenetically classified as ichthyosporeans as well, increasing even more its diversity and richness (Benny and O'Donnell, 2000; Cafaro, 2005; Mendoza et al., 2002).

### 1.5.3.1. Classification of Ichthyosporea

Ichthyosporea are divided in two groups, Ichthyophonida and Dermocystida (Adl et al., 2012; Cavalier-Smith, 1998; Glockling et al., 2013; Mendoza et al., 2002, 2001), consistent to distinct phenotypic traits related to their morphology and life cycle (Glockling et al., 2013; Mendoza et al., 2002) and supported as well by phylogenetic analyses, according to which both groups are monophyletic (Marshall and Berbee, 2011).

The Ichthyophonida is the most species-rich clade according to environmental surveys (del Campo and Ruiz-Trillo, 2013). Most of its members are associated to either vertebrate or invertebrate animals, such as *Amoebidium parasiticum*, *Ichthyophonus hoferi*, *Creolimax fragrantissima*, *Pirum gemmata*, *Abeoforma whisleri*, *Sphaeroforma tapetis*, *Sphaeroforma arctica* **(Figure 12)** (Glockling et al., 2013; Marshall et al., 2008; Marshall and Berbee, 2011; Mendoza et al., 2002).

There are also free-living species such as the marine saprotrophic *Sphaeroforma sirkka* and *Sphaeroforma napiecek* (Hassett et al., 2015). In general, most of the Ichthyophonida present a broadly conserved developmental mode consisting of large, multinucleated spherical or ovoid-shaped coenocytes, often with a large central vacuole or multiple vacuoles (*a.k.a.* sporangia or sporocyst) and a thick cell wall. At some point, this stage releases multiple motile amoebas (sometimes referred to as spores, zoospores, endospores or schizonts) by cellularization of the internal nuclei *(see* **Figure 13** *on next page).*

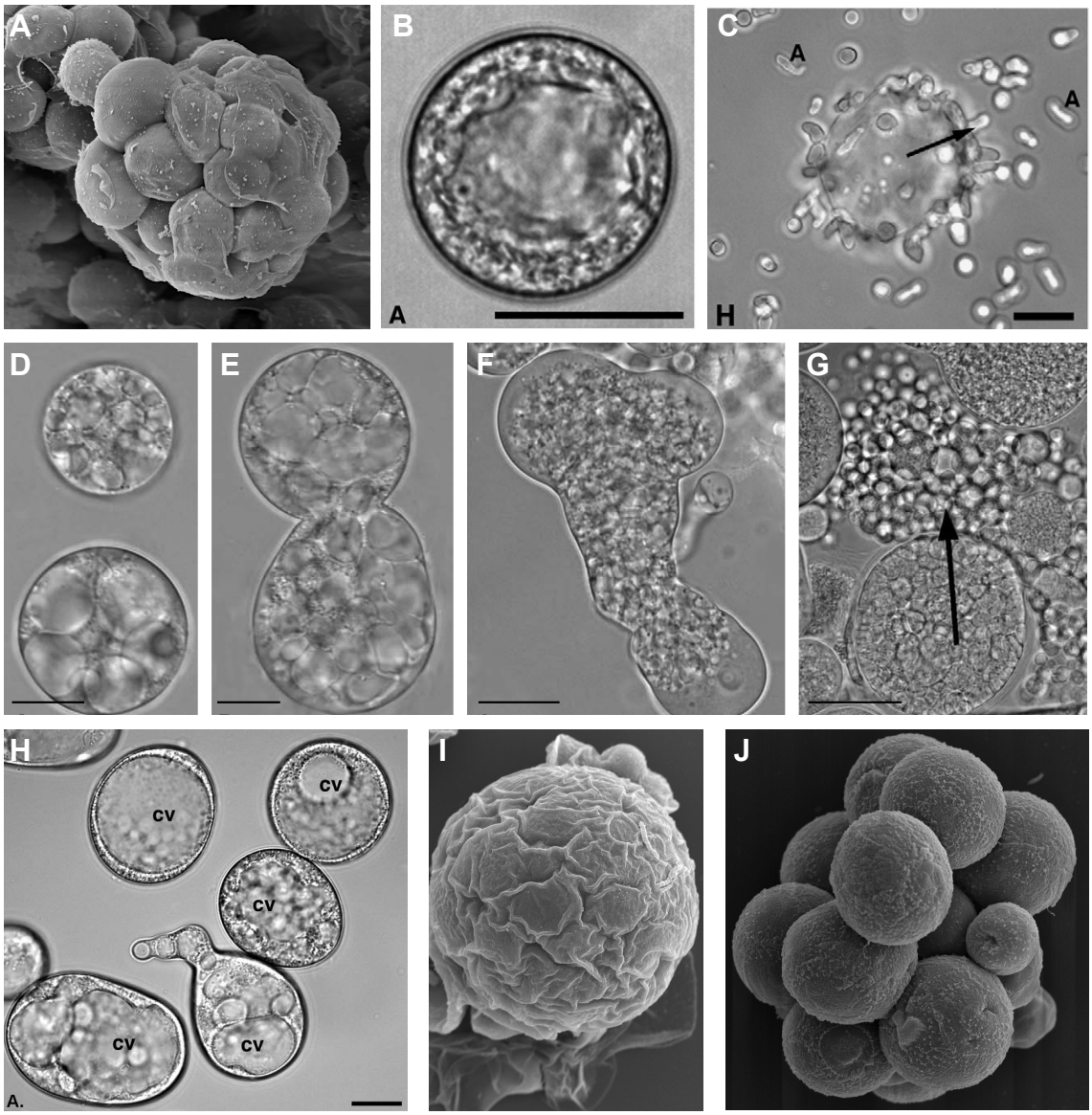



**Figure 12. Diversity of the Ichthyophonida.** (A-C) Diversity of cell morphologies in *Creolimax fragrantissima.* (A) SEM image of *Creolimax fragrantissima* (*"Creolimax fragrantissima"* by Multicellgenome Lab, CC. (B) Spherical cell with a single large vacuole. (C) Amoeba escaping through pores or tears in parental cell wall (arrow) and continuing to crawl along slide. Image credits in (B-C) from Marshall et al., 2008. (D-G) Examples of diversity of cell shapes of *Abeoforma whisleri.* (D) A commonly observed spherical form with prominent vacuoles that occupy a large proportion of the cell volume. (E) Example of an irregularly shaped cell that may have been in transition between a plasmodial and spherical stage. (F) Large plasmodial cell. (G) Reproduction via release of large numbers of endospores (arrow) through an opening in the parent cell wall. (H) Vegetative cells of *Pirum gemmata.* Cells are round in cross section and pear shaped in longitudinal view with large, often convoluted, central vacuoles (CV). Scale bar: 20 µm. Image credits in (D-H) from Marshall and Berbee, 2011.(I-J) SEM images of *Sphaeroforma arctica* single cell (I) and colony (J) (*"Sphaeroforma arctica"* by Arnau Sebé-Pedrós, CC by 2.0).

Amoebas are frequently spherical or limax-shaped, lacking a flagellum, and will typically disperse to establish a new colony and start the cycle again (Mendoza et al., 2002). Interestingly, some species exhibit fungal-like traits, such as chitin walls present in elongated amoebae of *A. parasiticum* (Mendoza et al., 2002; Torruella et al., 2015) or the hyphal structures of *I. hoferi* (Mendoza et al., 2002). Others exhibit a wide range of complex phenotypes, such as *A. whisleri*. While some cells present motile pseudopodia, hyphal or plasmodial structures, the amoeboid cell types can divide without reaching the coenocytic stage (Marshall and Berbee, 2011).

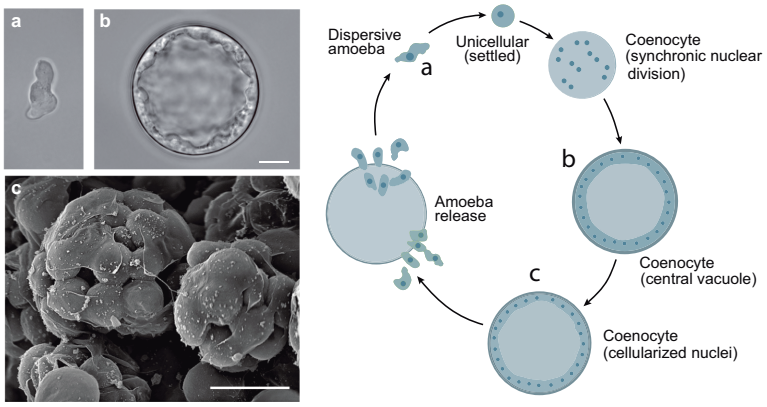



**Figure 13. Life cycle of *Creolimax fragrantissima*.** Single-celled motile amoebas settle and start a coenocytic outgrowth with synchronized nuclear division. The nuclei are gradually displaced towards the cell periphery as a central vacuole grows. Then, individual nuclei are cellularized and released as dispersive amoebas. Adapted from Sebé-Pedrós et al., 2017.

The Dermocystida (*a.k.a.* Rhinosporideacae) is composed of strictly parasitic species typically associated with skin, gills, mucus membranes and visceral organs of vertebrate hosts. A classic example is *Sphaerothecum destruens,* previously described as the "*rosette* agent", a well-described fish pathogen known to cause anaemia and lymphocytosis in salmon (Glockling et al., 2013; Mendoza et al., 2002). Other examples include *Amphibiocystidium* and *Amphibiothecum*, which appear to be closely related to Dermocystidium species (Feldman et al., 2005; González-Hernández et al., 2010; Pereira et al., 2005), *Rhinosporidium* (Herr et al., 1999) and the recently described *Chromosphaera perkinsii* (Grau-Bové et al., 2017).

Similar to Ichthyophonida, most of its members present roughly conserved developmental modes, including a spherical sporangium lacking the central vacuole. This stage also releases dispersive motile zoospores, which frequently present a single posterior flagellum. In general, the Dermocystida are less well characterized from a molecular point of view, especially due to their strictly parasitic nature that intricates establishing them in monoaxenic cultures (Glockling et al., 2013).

To date, there are six available ichthyosporean genomes: *S. arctica* (Grau-Bové et al., 2017), *C. fragrantissima* (de Mendoza et al., 2015), *I. hoferi* (Torruella et al., 2015), *P. gemmata* and *A. whisleri* and the dermocystid *C. perkinsii* (Grau-Bové et al., 2017). Currently, transcriptomic data is available for several species, including the dermocystids *S. destruens* (Torruella et al., 2015) and *C. perkinsii* (Grau-Bové et al., 2017) and the ichthyophonids *C. fragrantissima* (de Mendoza et al., 2015), *A. whisleri* and *P. gemmata* (Torruella et al., 2015), *Amoebidium parasiticum* (Torruella et al., 2012) and *S. arctica* (Ondracka et al., 2018; Torruella et al., 2012).

### 1.5.4. Corallochytrea

The Corallochytrea clade includes a single described species, the enigmatic *Corallochytrium limacisporum* **(Figure 14)** (Raghu-kumar, 1987). *C. limacisporum* is a small spherical-shaped free-living osmotroph, originally isolated from marine coral reefs in the Arabian Sea, and more recently in Hawaii (Raghu-kumar, 1987; Torruella et al., 2015). Despite initial descriptions pointed out that it lost its flagellum secondarily (Cavalier-Smith, 1998), recent comparative transcriptomic analysis revealed that at least it expresses most of the required flagellar genetic toolkit (Torruella et al., 2015).

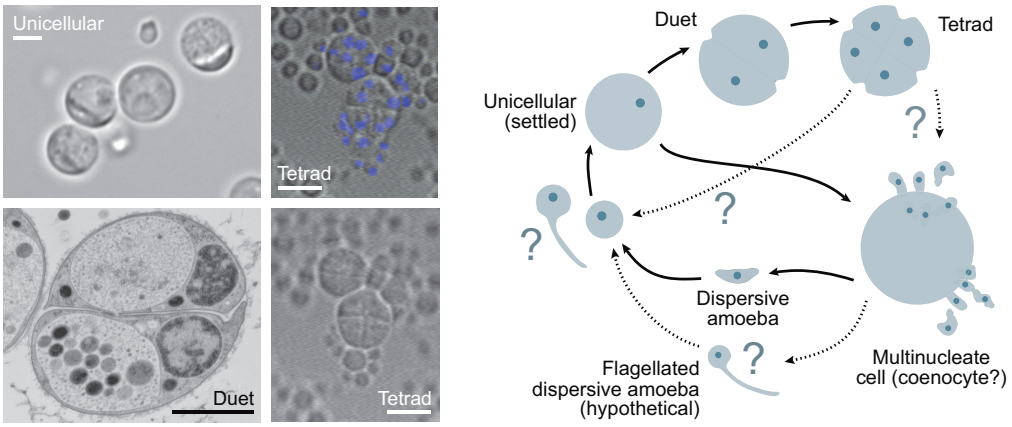



**Figure 14. *Corallochytrium limacisporum* and its life cycle transitions.** Clonal outgrowths from settled amoebas are similar to *C. fragrantissima's*, but the existence of a multinucleate, vacuolated coenocyte is unclear. Sometimes, individual cells undergo (confocal microscopy) serial binary palintomic division to form cell duets (TEM picture), tetrads (pictured with confocal microscopy and DAPI nuclear staining; upper right), etc. A flagellated stage (possibly dispersive) has been hypothesized. Scale bars are 1µm. Adapted from Grau-Bové et al., 2017.

*C. limacisporum* presents a complex developmental mode that in some aspects resembles the one present in Ichthyosporea **(Figure 14).** Usually, its life cycle starts with a uninucleated cell that undergoes a number of binary cell divisions without cytokinesis, until it releases amoeboid limax-like cells that disperse and form new colonies (Raghu-kumar, 1987). However, it is still not clear whether during its life cycle *C. limacisporum* presents a coenocytic stage. Interestingly, cell division sometimes occurs by palintomic cleavage, a developmental pattern involving serial cell divisions without accompanying size increase, originating Y-shaped junctions. This gives rise to several tiny cells, possibly reproductive propagules (Butterfield, 2011; Chen et al., 2014; Xiao et al., 2012). With all these complex developmental modes, it presents a variety of morphologies: single-cells, diads and tetrads.

Its phylogenetic position has been and still is controversial **(Figure 7).** Originally, it was classified as a thraustochytrid based on morphological similarities to other members of this group (Raghu-kumar, 1987). Later on, it was reclassified as a fungus due to its lysine catabolism (Sumathi et al., 2006), as sister-group of choanoflagellates based on SSU rRNA gene phylogenies (Cavalier-Smith and Allsopp, 1996; Mendoza et al., 2002; Zettler et al., 2001) placed with ichthyosporeans (Ruiz-Trillo et al., 2004; Steenkamp et al., 2006) and branching between filastereans and choanoflagellates (Paps et al., 2013).

Finally, in a more recent and taxon-rich phylogenomic analysis, *C. limacisporum* was classified as sister group to Ichthyosporea within Holozoa, forming a monophyletic group named Teretosporea, referring to the prevalence of “rounded spores” in both *C. limacisporum* and ichthyosporeans (Grau-Bové et al., 2017; Torruella et al., 2015). Teretosporea members interestingly present similarities in their lifestyles consisting of osmotrophic feeding modes and coenocytic development followed by the release of daughter propagules which in some cases present amoeboid and/or flagellated forms. However, the addition of the newly described species *Syssomonas multiformis* (Hehenberger et al., 2017) placed *C. limacisporum* in an independent Holozoa clade with *S. multiformis,* named Pluriformea, referring to the variety of forms found within the members of the group. Therefore, its position needs to be further confirmed.

### 1.5.5. *Syssomonas multiformis*

The newly described *Syssomonas multiformis* is a freshwater-dwelling predatory flagellate that feeds on large eukaryotic preys, especially on heterotrophic chrysomonads and bodonids (Hehenberger et al., 2017). It can engulf as well bacteria, small detritus and cytoplasmic content of eukaryotic cells. Morphologically, it presents a complex developmental mode that includes several cell types **(Figure 15).** In fact, *S. multiformis* exists as amoeboflagellate and amoeboid cells. Moreover, it presents motile swimming stage cells, predominantly roundish and naked, which measure around 7-14 µm in diameter. A smooth flagellum emerges from the middle-lateral point of these cells, ending by short *acroneme* (*i.e.,* the slender section of the flagellum) and directed backwards. Similarly to the filasterean *Pigoraptors* it can form as well clusters of multiple cells and spherical cysts.

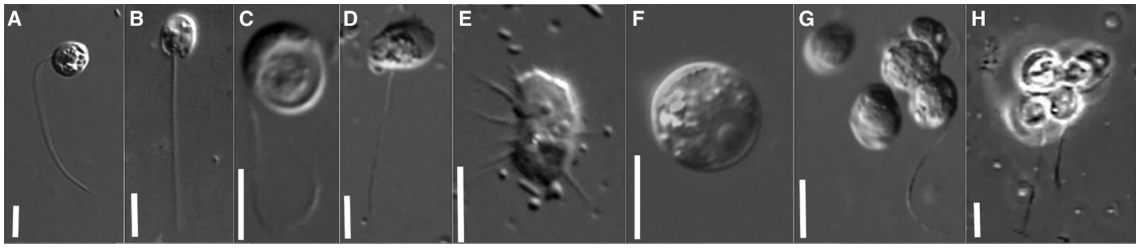



**Figure 15. Morphology of *Syssomonas multiformis* showing multiple life-cycle stages.** Light micrographs of *Syssomonas multiformis* stages: flagellated cell with laterally emerging flagellum (A), flagellated cell with posteriorly orientated flagellum (B), flagellated cell with large food vacuole (C), amoeboflagellate (D), amoeboid stage (E), cyst (F), and two different cell clusters (G and H). Scale bars represent 10 µm. Images from Hehenberger et al., 2017.

Recent transcriptomic analyses revealed *S. multiformis* presents some components of the integrin-based adhesion system. It actually presents a reduced set of components compared to filastereans but includes others not found in earlier-branching ichthyosporeans, such as in *C. fragrantissima.* Moreover, it also contains Fibronectin-3 domain-encoding transcripts and one T-box domain-encoding transcript.

*S. multiformis* phylogenetic position remains unclear **(Figure 7).** A recent study relates it with *C. limacisporum* forming the Pluriformea clade, a putative new holozoan lineage that branches between Filasterea and Ichthyosporea (Hehenberger et al., 2017).

In fact, both *C. limacisporum* and *S. multiformis* share some similarities, such as the capacity to form amoebae at some point during their life cycle, presumably exist as flagellated cells (*C. limacisporum* possesses the flagellar genetic toolkit), and the putative ability to form a syncytium before dividing into progeny. However, they differ both in habitat and feeding modes and the study included a smaller dataset and a limited taxon sampling with reduced representation of dermocystids and outgroup taxa (Hehenberger et al., 2017). Therefore, *S. multiformis* life cycle and phylogenetic position need to be further investigated.

## *Section 1.6*

## Unicellular holozoans as emerging model systems

Traditional model systems in biology have been widely used in research to address key biological questions, especially in the fields of developmental biology and biomedical research. Classically, they have thoroughly included animal or animal-related systems, such as mouse, the fruit fly genus *Drosophila,* the nematode *Caenorhabditis elegans*, zebrafish and mammalian and insect cell lines and few other non-animal systems such as yeast, the flowering plant *Arabidopsis thaliana* or the gram-negative bacteria *Escherichia coli.*

The recent advent of molecular and sequencing tools has opened new avenues of research, uncovering an unlimited number of novel biological questions. However, traditional model systems still include only a handful of species representing a limited taxonomic diversity**,** with a clear animal-bias (especially bilaterian-bias). This progress leap confronts the need of new systems, entailing an obligate paradigmatic change: instead of trying to answer the questions that traditional model systems can outline, we rather contemplate new relevant biological questions themselves and assess which model is more convenient to investigate.

Novel questions require suitable and phylogenetically better-placed non-traditional models, or *emerging model systems* (Goldstein and King, 2016). Fortunately, most of the tools developed for traditional model systems can, in principle, be “easily” adapted to non-model systems. Examples include the acoel *Hofstenia miamia,* the earliest lineage of animals with bilateral symmetry, as a model to study the evolution of the mechanisms of regeneration (Raz et al., 2017; Srivastava et al., 2014); the sea anemone *Aiptasia pallida*, as a model to study coral bleaching (Weis et al., 2008) or the green alga *Chlamydomonas reinhardtii,* as a model to study cell cycle control (Brown et al., 1991; Grossman et al., 2003; Harris, 2001; Rasala and Mayfield, 2011).

This is also applicable to address the longstanding question of the origin of animals. In fact, our understanding on the transition to animal multicellularity and the nature of the unicellular ancestor of animals in the past years has been empowered thanks to the study and characterisation of the genomes and transcriptomes of their extant closest unicellular relatives. Their complex genetic repertoires possessing genes related to multicellular functions in animals, the fact that they present complex developmental modes with multicellular-like structures and their phylogenetic position as close unicellular relatives to animals makes unicellular holozoans ideal systems to study the evolutionary origins of multicellular development.

### 1.6.1. *Salpingoeca rosetta,* a benchmark among choanoflagellates

After more than one decade, the choanoflagellate *Salpingoeca rosetta* has become a reference model to address the evolution to multicellular animals (Hoffmeyer and Burkhardt, 2016). *S. rosetta* is able to produce multicellular-like structures of spherical clonal colonies by incomplete cell division, known as *rosettes,* triggered and enhanced by a bacterial sulfonolipid produced by the prey *Algoriphagus machipongonensis* **(Figure 10)** (Alegado et al., 2012; Fairclough et al., 2010; Woznica et al., 2016).

A comparative transcriptomic analysis of multiple cell stages revealed that these multicellular colonies and the unicellular swimming cells from which they develop are enriched in genes exclusively shared between choanoflagellates and Metazoa. Colonies presented a *Salpingoeca*-specific profile (Fairclough et al., 2013). The authors hypothesized that early colony development is based on genomic features that originated at the shared ancestor of choanoflagellates and Metazoa, while specific colonial cell type would be a choanoflagellate innovation. This result is supported by the identification of upregulated *Septin* genes in the *rosette* colonies (GTPases that regulate cytokinesis in fungi and Metazoa), which the authors linked to a mode of incomplete cytokinesis also found in Metazoa (Fairclough et al., 2013, 2010).

Moreover, recent studies revealed *S. rosetta* is able to undergo meiosis and genetic recombination, suggesting that mechanisms used for gamete recognition and fusion in the ancestor of animals may have been co-opted for somatic cell adhesion during the evolution of animal multicellularity (Levin and King, 2013).

This opens new avenues of research for both reconstructing the evolution of sex and establishing classical genetics in choanoflagellates. Finally, the experimental tractability of *S. rosetta* is pushing techniques that can be further adapted for other choanoflagellate species (Booth et al., 2018).

### 1.6.2. *Creolimax fragrantissima* and *Sphaeroforma arctica,* two emerging models within Ichthyosporea

Ichthyosporeans are of significant interest for their close association with animals and for their complex developmental life cycle. Interestingly, the coenocytic growth of most of the ichthyosporeans has been suggested to be evolutionary related to animal multicellularity, because of its resemblance to the coenocytes and/or syncytia exhibited by some animal embryos (Bonner, 2000a; Grosberg and Strathmann, 2007; Suga and Ruiz-Trillo, 2013).

The ichthyophonid *Creolimax fragrantissima* follows the prototypical developmental mode of ichthyophonids, with two clear cell types **(Figure 13)** (Marshall et al., 2008). First, small, spherical zoospores of around 6-8 µm in diameter develop a central vacuole and grow in size. After multiple rounds of coenocytic nuclear division, reaching maturation around 25-60 µm in diameter, cells cellularize and release several motile amoebae, that disperse, encyst and restart the cycle again (Marshall et al., 2008). The transcriptomic profile of both developmental cell types has been investigated in a comparative analysis with other holozoans, and demonstrated that *C. fragrantissima* has a program of transcriptionally regulated cell type differentiation (de Mendoza et al., 2015). Unexpectedly, this osmotroph upregulates animal-like gene toolkits in the amoeboid stage, including developmental TFs and adhesion genes involved in the integrin adhesome.

Multinucleated coenocytes instead, appear to have transcriptomic profiles analogous to the proliferative undifferentiated animal cell types, such as stem cells. Therefore, *C. fragrantissima* probably co-opted ancestral gene regulatory programs to develop a novel osmotrophic feeding mode, absent in non-ichthyosporean holozoans.

Overall, this study provided direct evidence of the plasticity of cell type evolution across holozoan lineages, supporting a scenario of recurrent recruitment of co-regulated expression programs to support the emergence of novel cell types and developmental programs (Newman, 2012). Moreover, nuclear division is synchronized within the coenocytic cell of *C. fragrantissima*, and nuclei are arranged beneath the cell surface of the colony (Suga and Ruiz-Trillo, 2013).

Interestingly, the same process also occurs in *Sphaeroforma arctica*, another ichthyophonid (Suga and Ruiz-Trillo, 2013). This process was recently described by Ondracka et al., showing that *S. arctica* presents a highly regular and experimentally tractable coenocytic cell cycle (Ondracka et al., 2018). Interestingly, cells present a synchronous coenocytic growth, in which cells grow from 1 to 64-128 nuclei before cellularization periodically driven by a clock-like mechanism that operates independently of the cell volume and growth rate. Thus, it is distinct from the regulation of the cell cycle in multinucleate filamentous fungi, that present asynchronous nuclear divisions within the coenocyte and individual nuclei control local cytoplasm growth (Anderson et al., 2013; Gladfelter et al., 2006; Nair et al., 2010). Nevertheless, it putatively evokes the synchronous cell cycles in early animal embryos. Moreover, a recent study reported the first evidence for microRNA (miRNA) genes and homologs of the animal miRNA biogenesis machinery (including Drosha and Pasha) in *S. arctica,* suggesting that the origin of animal miRNAs and the microprocessor complex predated the origin of animals (Bråte et al., in press).

Although these investigations undoubtedly represent advances, molecular and functional-based analyses to understand the function of key genes encoding proteins important for metazoan development and multicellularity, have been limited by the lack of genetic tools (*e.g.,* transfection and gene silencing). Hence, it is key to develop unicellular holozoans as experimentally tractable systems for functional approaches.

### 1.6.3. Genetic tools for gene function studies

Genetic tools used in molecular biology enable the study of the function of genes and gene products in cells by enhancing (overexpression), downregulating (knock-down) or inhibiting (knock-out or silencing) their gene expression (Bosher and Labouesse, 2000; Prelich, 2012; Ran et al., 2013). The first necessary step to achieve this is the introduction of foreign nucleic acids into a cell using chemical, physical or biological approaches in a process known as “transfection”. Transfection allows the delivery of DNA in the form of plasmid (circular) or linear molecules, or RNA in the form of messenger RNA (mRNA), small interfering RNA (siRNA) or microRNAs (miRNAs) into a cell (Kim and Eberwine, 2010; Recillas-Targa, 2006; Washbourne and McAllister, 2002). With the proper signalling tags, the proteins encoded by these molecules can be specifically delivered in subcellular locations and transiently or stably alter the function of the gene of interest.

In a transient transfection, transfected nucleic acids do not self-replicate nor integrate into the genome. Thus, they generally persist for a short period of time in the cells, being lost after few generations. In contrast, the hallmark of stably transfected cells is that foreign nucleic acids are integrated into the genome, replicated and transmitted to the next generation, generating stable cell lines and allowing long-term studies (Recillas-Targa, 2006). Stable transfection starts with a transient transfection, the first step used to introduce foreign DNA, and its infrequent but serendipitous integration to the genome. Generally, integration happens through homologous recombination processes between homologous regions flanking the gene of interest and the genome. Moreover, this DNA generally encodes a gene that allows the small proportion of transfected cells to grow under particular conditions and, thus, be further “selected” from the non-transfected or transiently-transfected populations using selectable markers (*e.g.,* conditioned media or antibiotics). Despite the vast range of currently available transfection methods, not all of them are equally suitable nor can be applied to all types of cells or experiments. This entails huge variation regarding achieved transfection efficiencies, viability of the cells, or persistence and level of gene expression. Determination of the more suitable methodology for a specific application will depend on several factors, such as cell type or organism to be transfected, compromise between cell viability and transfection efficiencies, time consumption and laboratory costs.

### 1.6.3.1. Chemical, physical and biological transfection methods

Chemical transfection methods use carrier molecules to overcome the cell-membrane barrier and favour nucleic acid uptake by the cell, generally by endocytosis. They consist of the interaction of negatively charged nucleic acids with positively charged carrier molecules, such as polymers, lipids or salts, enabling nucleic acids to come into contact with the negatively charged membrane components, incorporating them into the cell, and later being released into the cytoplasm. Currently, there is a wide range of chemical methods used to efficiently transfect eukaryotic cells, and some of them even combine with physical methods to increase nucleic acids uptake.

The first chemical transfection method, initially developed in the late 1950s, used hyperosmotic and polycationic proteins to promote DNA uptake into cells (for a review, see Felgner, 1990). Few years later, Diethylaminoethyl-dextran (DEAE-dextran) was developed to introduce poliovirus RNA and SV40 (Kumar et al., 2018; Schenborn and Goiffon, 2000; Vaheri and Pagano, 1965) and polyomavirus DNA (McCutchan and Pagano, 1968; Thorne et al., 1968) into cells. DEAE-dextran-mediated transfection forms nucleic acid-polymer complexes with a carbohydrate polymer (*i.e.,* a polycationic derivate of dextran) that putatively serve as a bridge between the negatively charged nucleic acids and the negatively charged surface of the cell. After complexes internalization, nucleic acids are transported into the nucleus (Holter et al., 1989; Lieber et al., 1987; Ryser, 1967; Yang and Yang, 1997). With slight modifications, this procedure continues to be widely used for transfection of cultured cells with viral genomes and recombinant plasmids (Kumar et al., 2018). Despite it being one of the simplest transfection methods and requires low cost and time, a major drawback is that achieved transfection efficiencies are generally low for a range of cell types, and cytotoxicity can compromise cell viability in the long term. Thus, it is not generally suitable to generate stable lines.

Another popular chemical-based transfection method is calcium-phosphate precipitation (Graham and van der Eb, 1973). This method has been widely used to transfect mammalian cells and the amoebozoan *Dictyostelium discoideum* (Gaudet et al., 2007; Nellen et al., 1984). In this method, DNA is mixed with calcium chloride in a buffered saline/phosphate solution. Proper incubation times and temperature promote the formation of DNA-calcium phosphate coprecipitates which adhere to the surface of cells

('Calcium phosphate – mediated transfection of eukaryotic cells', 2005). Contrary to DEAE-dextran, calcium phosphate precipitation is widely used to generate stably-transfected cell lines, allowing for long-term gene expression studies. Nevertheless, cytotoxicity should be also considered in the long-term as well. Moreover, coprecipitate (or crystal) formation is critical for a successful transfection. In fact, crystal size is highly sensitive to slight changes in pH, temperature, incubation times and buffer salt concentrations, which can compromise successful transfection (Chowdhury et al., 2003; Jordan et al., 1996; Jordan and Wurm, 2004). Transfection efficiency is also low compared to other chemical transfection methods.

This is not the case in lipid-based transfection, or lipofection, the most commonly used transfection method in mammalian cells (Felgner et al., 1987; Holmen et al., 1995). Similarly to previous methods, lipofection takes advantage of the positive nature of cationic transfection lipids, which consist of a positively charged head group (*e.g.,* an amine), a flexible linker group (*e.g.,* an ester or ether) and two or more hydrophobic tail groups. Cationic lipids are mixed with a neutral lipid, such as 1,2-Dioleoyl-sn-glycero-3-phosphoethanolamine (DOPE) (*a.k.a.,* helper lipid), in a solution that favours the formation of unilamellar liposome vesicles, carrying a net positive charge. Nucleic acids are then adsorbed to these vesicles, which will in turn be absorbed by the cellular membrane (Cardarelli et al., 2016). Lipofection is known to transfect a wide range of cell types (mainly adherent cell lines) with high efficiency, and with relatively low cost. Among its advantages, lipofection successfully delivers nucleic acids of all sizes, and is applicable for both transient and stable transfection. Nevertheless, it presents some caveats. For example, in most primary cell lines and in suspension cell lines, lipofection presents low transfection efficiencies. Moreover, it is generally more cytotoxic than other methods in the long term.

Magnetofection is a recently developed procedure that combines chemical and physical approaches, and it has been reported to efficiently transfect neurons (Buerli et al., 2007; Dobson, 2006; Ensenauer et al., 2011; Plank et al., 2003; Scherer et al., 2002). In fact, physical transfection methods use physical or mechanical forces to enable the direct transfer of nucleic acids into the cytoplasm or nucleus. In this case, Magnetofection uses lipid-based transfection or cationic-based nucleic acids coating magnetic nanoparticles which, under the proper magnetic field (physical force), are concentrated and transported into cells, favouring cell surface contact and promoting cellular uptake.

Among its advantages, Magnetofection is relatively fast and is recommended for primary cells and hard-to-transfect adherent cells.

Microinjection is a physical transfection method which uses a micromanipulator to directly inject nucleic acids or proteins in the cytoplasm or directly into the nucleus of single cells, such as oocytes or embryonic stem cells (Chow et al., 2016). A major advantage is the high efficiency of this method; nevertheless, it is very time consuming, expensive and requires certain operator skills. Cell viability can be compromised as well. Microinjection is commonly used to generate transgenic organisms (Ivics et al., 2014).

Biolistic particle delivery (*a.k.a.,* particle bombardment) is another physical transfection method mainly used for genetic vaccination and agricultural applications. This approach transfers microparticles (*e.g.,* gold or tungsten) coated with nucleic acids into cells by a particular driving force, such as gas pressure or voltage discharge between two electrodes. Among its advantages, this technique is fast, relatively simple and allows transfection of large numbers of genes. Nevertheless, cell mortality is considerably high. Particle bombardment has been widely used to efficiently transfect volvocine green algae, such as *Volvox carteri, Gonium pectoral, Eudorina elegans* and *Pandorina morum* (Lerche and Hallmann, 2014, 2013, 2009; Schiedlmeier et al., 1994).

The most frequently used physical transfection approach is electroporation. Electroporation applies high voltage pulses of electricity or electric shocks to introduce DNA into cells suspended in an electroporation buffer (Neumann et al., 1982; Potter et al., 1984; Potter and Heller, 2010). The electrical pulse creates a differential potential across cell membranes, as well as charged membrane components, and induces temporary pores in the cell membrane for DNA uptake. In general, transfection efficiencies using electroporation are considerably high. However, its main drawbacks are low efficiency in primary cells and high mortality rates. Thus, this technique requires finetuning and optimization of pulse strength, pulse duration and pulse repetition for each cell type used. Electroporation has been widely used to transfect eukaryotic cells, such as the unicellular green alga *Chlamydomonas reinhardtii* (Brown et al., 1991).

Related to this, a novel electroporation-based methodology, known as Nucleofection, has recently been developed to deliver DNA directly into the nucleus of cells (Caro et al., 2012; Janse et al., 2006; Vinayak et al., 2015).

Nucleofection is based on a Nucleofector™ device, that delivers unique electrical parameters, and Nucleofector™ Kits, which contain cell type-specific buffer solutions, cuvettes and pipettes. This technique is especially recommended for cells or cell lines that are difficult to transfect. Moreover, this robust transfection methodology maintains high cell viability. Its main caveats are that the specific electrical settings are determined in distinct pre-programmed settings, which cannot be directly optimized by the user and require service attention. Moreover, its cost is relatively high.

Other physical transfection methods include laser-mediated transfection (*a.k.a., optoporation* or *phototransfection*), in which transient pores are formed by irradiating cell membranes using a pulse laser (Barrett et al., 2006; Schneckenburger et al., 2002; Shirahata et al., 2001; Yao et al., 2008). The osmotic differential created between the transfection medium and the cytosol of the cells promotes nucleic acids uptake into the cells. Laser-mediated transfection allows to specifically localize the pores at any location of the cells and it is generally used in very small cells. Similar to phototransfection, other methods use ultrasound-mediated transfection *(sonoporation)* to efficiently deliver nucleic acids into cells (Kim et al., 1996; Tomizawa et al., 2013).

Finally, biological methods use genetically engineered viruses (*e.g.,* retroviruses, adenoviruses, adenoassociated) as gene delivery vehicles, taking advantage of their capacities to carry foreign nucleic acids with a high efficiency, especially in primary cells and cell lines (Pfeifer and Verma, 2001). These methods consist in generating recombinant virus containing the transgene by molecular cloning and amplification of viral particles in a packaging cell line. After virus isolation, purification and titration, one can proceed to infect the target cell type (usually mammalian cell lines or insect cell lines). Interestingly, virus-mediated transfection can allow the integration of the transgene in host genome, and therefore achieve stable transfection. Nevertheless, viral-mediated transfection is relatively time consuming, presents elevated laboratory costs and limitations of insert sizes. Moreover, cell types to transfect should have specific viral receptors to be successfully infected.

### 1.6.3.2. Gene function modulation and genome editing approaches

Transfection allows studying the function of genes and gene products, either by enhancing, diminishing or inhibiting specific gene expression in cells. Examples include overexpression, targeted gene therapy for disease treatment, siRNA knock-down procedures and CRISPR/Cas9-mediated gene silencing.

Gene regulation studies are a common approach to investigate the activity of gene regulatory elements. Using overexpression or siRNA knock-down of gene products, a particular signalling pathway can be manipulated, a fact that allows the investigation of gene function. Generally, measurable reporter genes, such as fluorescent proteins, are frequently used in overexpression experiments to enzymatically, colorimetrically or fluorescently assay the effect of the activity of the target gene.

Other experimental approaches to estimate the effect of altering the function of target genes in siRNA knock-down is through the examination of transcriptional variation, such as reverse transcriptase polymerase chain reaction (RT-PCR), RNAse protection assays or Northern blotting.

A recently developed approach to ultimately achieve permanent gene silencing is the CRISPR/Cas9 technology (Ran et al., 2013). In this method, the RNA-targeted Cas9 nuclease is used to mediate genome alteration from the microbial *clustered regulatory interspaced short palindromic repeats* (CRISPR) adaptive immune system. CRISPR/Cas9 efficiently performs genome engineering in eukaryotic cells by simply specifying a 20-nucleotide target sequence within its guide RNA via *non-homologous end-joining (NHEJ)* or *homology-directed repair (HDR)* for functional studies (Ran et al., 2013). This system overcomes off-target effects of RNAi-mediated protein depletion (as well as incomplete and reversible protein depletion) and has been widely used to disrupt gene function in various cell lines (Adli, 2018; Janssen et al., 2018; Ran et al., 2013)

### 1.6.4. Development of genetic tools among unicellular holozoans

Several years ago, the first functional link between genotype and phenotype was set out in choanoflagellates by establishing forward genetics in *Salpingoeca rosetta,* notable for its experimental tractability relative to other choanoflagellate species (de Mendoza and Ruiz-Trillo, 2014; Levin et al., 2014). In this study, the authors induced random mutagenesis by exposing *S. rosetta* cells to either electromagnetic spectrum (EMS) or X-rays. Clonal lines of potential mutants were further established by isolating individual cells through serial dilution and exposing them to the *rosette*-inducing *A. machipongonensis* conditioned medium. A highly penetrant mutant defective in *rosette* development, named *rosetteless,* allowed the identification of a *C-type lectin* gene, an exclusive choanoflagellate and metazoan gene family, as essential for the formation and establishment of *rosette* colonies (Levin et al., 2014).

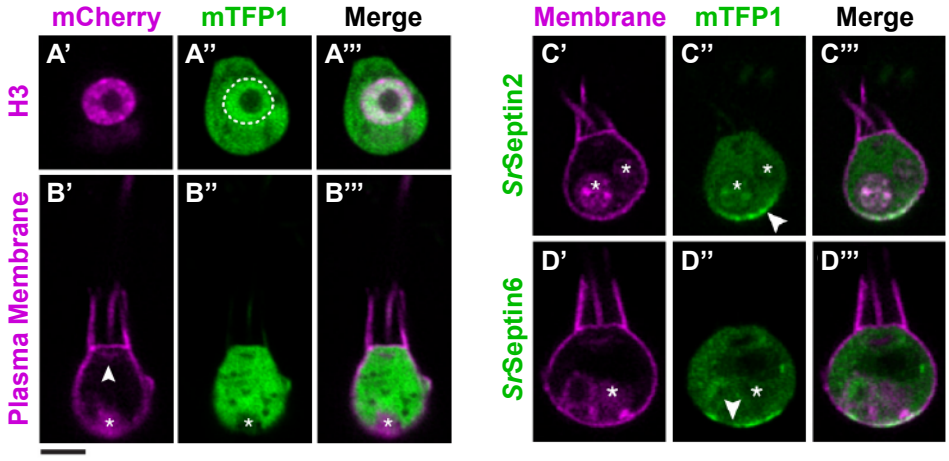



**Figure 16. Fluorescent subcellular markers expressed from reporter plasmids in live *S. rosetta* cells.** Twenty-four hours post-transfection imaging by superresolution microscopy with a Zeiss LSM 880 Airyscan. (A) A fusion of mCherry to the C-terminus of Histone H3 was confined to the nucleus. (B) A plasma membrane marker fusing a geranyl-geranylation sequence (PTSG_00306) to the C-terminus of mCherry outlined the entire cell shape, including the collar, flagellum, and cell body. The membrane marker also weakly highlighted the Golgi (arrowhead). The food vacuole (asterisk) was often visualized due to autofluorescence from ingested bacteria or through accumulation of the fluorescent markers in the food vacuole, perhaps through autophagy. (C) mTFP1-*Sr*Septin2 co-transfected with a plasma membrane marker revealed *Sr*Septin2 distributed throughout the cytosol and enriched at the basal pole in *S. rosetta* cells (C'', arrowhead). (D) mTFP1-*Sr*Septin6 mirrored the enrichment of mTFP1-*Sr*Septin2 at the basal pole (D'', arrowhead). Scale bars: 2 µm. Adapted from Booth et al., 2018.

More recently, new avenues for exploring gene function in choanoflagellates are upcoming thanks to the development of a robust transfection method for delivering and expressing transgenes in *S. rosetta* (Booth et al., 2018). Cells are efficiently transfected using an electroporation-based method for DNA delivery by Nucleofection (Caro et al., 2012; Janse et al., 2006; Vinayak et al., 2015). Using a self-engineered panel of fluorescently-tagged endogenous proteins, *S. rosetta* organelles can be now visualized in live cells. This has enabled the first *in vivo* characterization of Septins, a class of cytoskeletal proteins which localized to the basal poles of cells, resembling Septin localization in animal epithelia. Altogether, this represents an expanded repertoire of new functional-based approaches for investigating gene function in *S. rosetta.*

Transient transfection has also been developed in the ichthyosporean *Creolimax fragrantissima* by electroporation (Suga and Ruiz-Trillo, 2013). A reporter vector with the *Histone 2B (H2B)* gene of *S. arctica,* another ichthyosporean, fused to a fluorescent protein under the control of the endogenous $\beta$-tubulin promoter of *C. fragrantissima* was used to successfully trace the nuclear divisions in a growing cell *in vivo* **(Figure 17A)**.

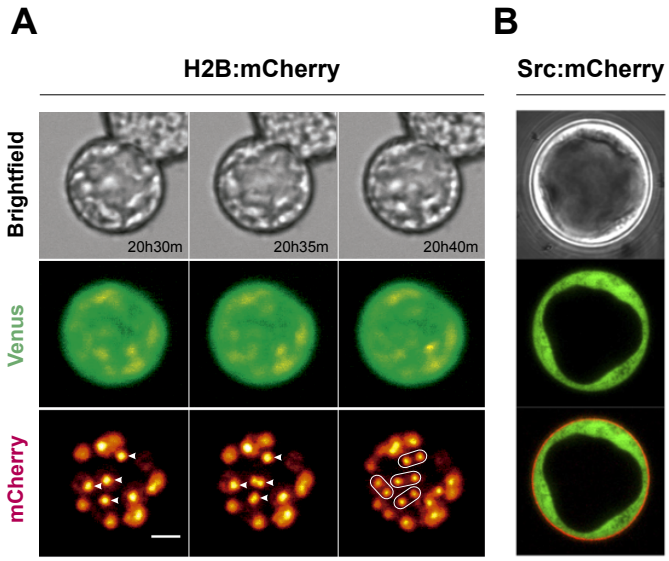



**Fig. 17. *C. fragrantissima* transfection reveals synchronized nuclear division and *Src* localization in live cells.** (A) Time-lapse movie showing synchronized nuclear divisions (arrows and ovals). Venus fluorescent protein is expressed in the cytoplasm and a fusion of *H2B* gene with mCherry in the nuclei. Adapted from Suga and Ruiz-Trillo, 2013. (B) *C. fragrantissima* cell co-transfected with Venus fluorescent protein and a fusion of mCherry with the endogenous *C. fragrantissima Src kinase (CfrSrc).* Note *CfrSrc* localisation in the cell membrane. Adapted from Suga and Miller, 2018.

Results indicated that *C. fragrantissima* colonies develop from a fully-grown multinucleated syncytium, in which nuclear divisions are strictly synchronized. Moreover, two strategies for gene silencing using RNA interference by small interference RNAs (siRNA) and Morpholinos probes to successfully knock down the expression of target genes in a dosage-dependent manner (Suga and Ruiz-Trillo, 2013).

The function of *c-Src* kinase of *C. fragrantissima* has also been recently analysed *in vivo* using overexpression experiments (Suga and Miller, 2018). *C. fragrantissima* possesses a single homolog of the animal *c-Src* kinase *(CfrSrc)* and 7 tyrosine-specific phosphatases (PTPs), but lacks any homolog of the animal *Csk* kinase, known to negatively regulate animal *Src* at the C-terminal site. This suggested that another mode of negative regulation may exist for *CfrSrc.* To further examine this, expression of both proteins was assessed throughout the lifecycle of *C. fragrantissima*. *CfrSrc* resulted to be active and expressed throughout its lifecycle in the cell membrane **(Figure 17B)**, whereas RNAseq data revealed that expression of *CfrPTP-3* is over 5 times higher than that of *CfrSrc* during the multinucleate stage, and nearly double in the amoeboid stage. Interestingly, the phosphatase *CfrPTP-3* suppressed *CfrSrc* activity *in vitro* and *in vivo*. Moreover, *C. fragrantissima* cells overexpressing *CfrSrc* showed growth defects and never reached the maturation stage to produce amoebae. Co-expression of *CfrPTP-3* rescued these phenotypes, suggesting that an existing tyrosine-specific phosphatase was co-opted for the role of *Src* regulation in the highly reduced kinome of *C. fragrantissima* (Suga and Miller, 2018).

Transfection has also been developed in our lab for two additional unicellular holozoan species: transient transfection in *A. whisleri*[3] and stable transfection in *C. limacisporum*[4]. Both species can be transiently transfected using an electroporation-based transfection protocol with a fluorescently tagged reporter cassette with endogenous genes. *C. limacisporum* transfectants can be stably maintained through selection with the antibiotic puromycin using a puromycin resistance cassette. Stable transfectant cell lines are further being characterized. Finally, efforts are underway to develop transfection in other ichthyosporean species, such as *S. arctica.*

[3] Protocol available in *protocols.io:* dx.doi.org/10.17504/protocols.io.hmvb466

[4] Protocol available in *protocols.io:* dx.doi.org/10.17504/protocols.io.r5ud86w

Altogether, this opens new avenues to other experimentally-tractable systems among unicellular holozoans and increase the mechanistic insights into the ancestry of animal cell biology. However, there has not been a representative experimentally tractable filasterean species to date.

Recent analyses in the filasterean *Capsaspora owczarzaki* provided important insights into the origins of animal multicellularity and the nature of their unicellular ancestor (Sebé-Pedrós et al., 2016b, 2016a; Suga et al., 2013). The *C. owczarzaki* genome encodes an unexpected set of TFs known to be involved in animal development, some of them secondarily lost in *S. rosetta* (Sebé-Pedrós et al., 2010; Suga et al., 2013). This puts *C. owczarzaki* in the spotlight as the closest relative of animals in which such genes can be studied.

## *Section 1.7*

## *Capsaspora owczarzaki*

In the late seventies, Stibbs *et al.* serendipitously observed an amoeba-like symbiont in three strains of the freshwater snail *Biomphalaria glabrata* (Owczarzak et al., 1980; Stibbs et al., 1979). Symbionts were repeatedly cultivated unintentionally from the snail hemolymph during several attempts to culture *B. glabrata* cells infected by the widespread human pathogen *Schistosoma mansoni.* Notably, the amoebae not only showed excellent growth *in vitro* but also proved to be able to rapidly ingest *Schistosoma* sporocysts. To further investigate the enigmatic symbiont, amoebae were successfully isolated using snail pericardial explants and mantle swabs, maintained axenically *in vitro* and deposited in the American Type Culture Collection (ATCC® 30864) (Owczarzak et al., 1980). Nearly twenty years later, symbionts were described as the novel genus and species *Capsaspora owczarzaki,* referring to their ability to destroy trematode sporocysts *in vitro* and after one of its discoverers (Hertel et al., 2002; Owczarzak et al., 1980).

Based on initial morphological observations, *Capsaspora owczarzaki,* hereafter *Capsaspora,* was initially classified as a *Nuclearia* sp. (Owczarzak et al., 1980). A further molecular analysis based on a 18S rDNA phylogeny reclassified *Capsaspora* within the Ichthyosporea (also known as Mesomycetozoea), a group comprising unicellular opisthokonts with parasitic lifestyles (Hertel et al., 2002). Deeper phylogenetic studies using a multigene phylogenetic approach and a wider taxon sampling confirmed *Capsaspora's* phylogenetic placement within Filasterea, an independent Opisthokonta lineage closely related to Metazoa and Choanoflagellatea **(Figure 7)** (Ruiz-Trillo et al., 2008, 2006, 2004; Shalchian-Tabrizi et al., 2008; Torruella et al., 2015).

### 1.7.1. *Capsaspora,* a filopodiated amoeba with a complex life cycle

*Capsaspora* is a small filopodiated amoeba of about 6-7 µm in diameter **(Figure 18A)** (cell body, excluding filopodia). Unlike its other Filasterea siblings, it does not present any flagellar structure. *Capsaspora* presents a single nucleus that measures one third to one half of the diameter of the cell with a central nucleolus (Stibbs et al., 1979) and several mitochondria with flattened cristae (Zettler et al., 2001). Cells generally present numerous phagosomes, several lipid vacuoles, some glycogen granules and a Golgi apparatus, but rarely rough and smooth endoplasmic reticulum **(Figure 18B)** (Stibbs et al., 1979).

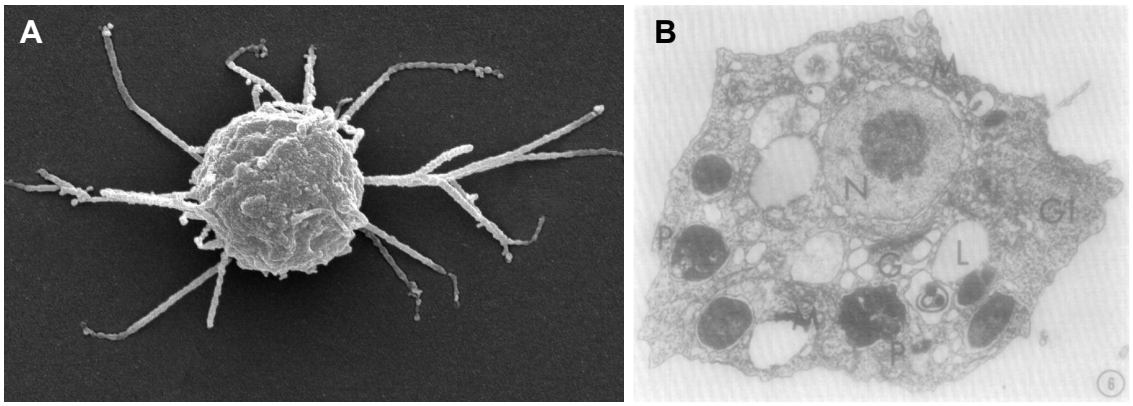



**Fig. 18. *Capsaspora owczarzaki.*** (A) SEM image of a filopodiated *C. owczarzaki* cell (*"Capsaspora owczarzaki"* by Multicellgenome Lab, CC). (B) Electron micrograph of *C. owczarzaki* in its trophozoite form. N: Nucleus with central nucleolus; G: Golgi complex; Gl: Glycogen; P: Phagosome; M: mitochondria; L: lipid vacuole. Image credit from Stibbs et al., 1979.

Under culture conditions, *Capsaspora* presents three morphologically different life stages **(Figure 18)** that are differentially regulated at the transcriptomic, proteomic and phosphoproteomic levels *(see section 1.7.2. The Capsaspora genome and its regulated life cycle transitions)* (Sebé-Pedrós et al., 2016b, 2016a, 2013b; Stibbs et al., 1979). In the *adherent or filopodial stage*, *Capsaspora's* cell body is flattened, elongated and extends several long and thin filopodia, which enable the amoeba to aimlessly crawl and inspect the environment, attached to the substrate **(Figure 19A)** (Stibbs et al., 1979). At this stage, cells proliferate exponentially, rapidly dividing (in less than a minute) into two daughter cells of approximately equal size every 6 to 8 hours.

Adherent stage cells can be induced from a fresh dilution of a confluent culture and grown for 2-4 days. In response to crowding or stress, usually after 6-7 days of growth onwards, adherent cultures detach and encyst by retracting the filopodia **(Figure 19B)**. At this *cystic or floating stage*, the cell body shrinks to 3-5 µm in diameter and presents a double wall, features that resemble dormant/resistance cell types **(Figure 19C)** (Stibbs et al., 1979). Interestingly, *Capsaspora* can actively form a multicellular-like structure by the aggregation of independent cells. In this *aggregative stage*, cells secrete a cohesive extracellular material to putatively maintain the aggregates **(Figure 19D)**, which can be sometimes observed in crowded cultures or induced by agitation of adherent cells **(Figure 19E-F)** (Sebé-Pedrós et al., 2013b). This observation represented the first example of an aggregative behaviour in a close unicellular relative of animals (Sebé-Pedrós et al., 2013b).

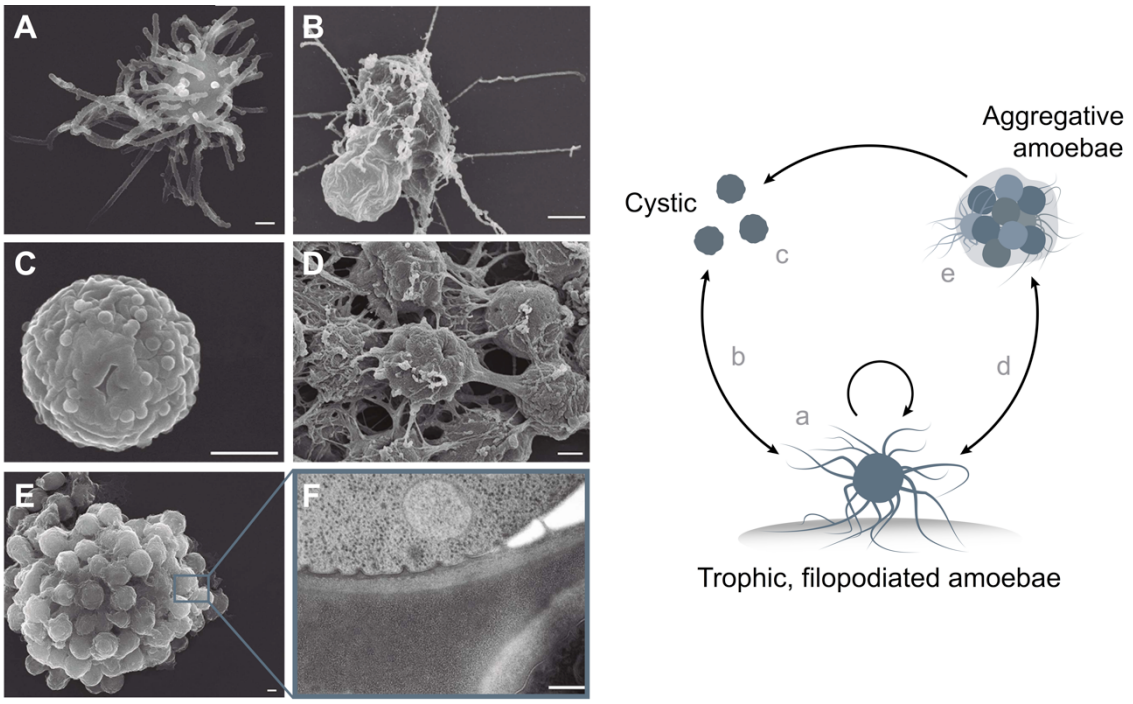



**Fig. 19. *Capsaspora* and its life cycle transitions.** (A) Filopodiated (adherent) stage cells, amoebas with long filopodia. (B) Transition from filopodial to cystic stage: cells retract filopodia. (C) Cystic stage cells are rounded cysts without filopodia and slightly smaller than filopodial cells. (D) Transition from filopodial to aggregative stage: cells attach to each other and an extracellular matrix appears. (E) Mature aggregate. (F) Transmission EM image showing adjacent cells in the aggregate separated by extracellular matrix. Arrows indicate the observed stage inter-conversions. Scale bars: 1 µm, except in panel F: 200 nm. Modified from Sebé-Pedrós et al., 2013b.

### 1.7.2. The *Capsaspora* genome

The striking temporal differentiation at the morphological level **(Figure 19)** and its phylogenetic placement as a close unicellular relative of animals put *Capsaspora* in the spotlight for further analyses. In fact, the sequencing of the genome of *Capsaspora* and analyses of the regulation of its life cycle completely tweaked our conception of the nature of the unicellular ancestor of animals (Sebé-Pedrós et al., 2017, 2016a, 2016b, 2013b; Suga et al., 2013).

The *Capsaspora* genome[5] contains an unexpected repertoire of genes previously thought to be animal-specific and, most importantly, directly related to multicellular functions in animals. These includes members of the complete *integrin adhesome system* (Sebé-Pedrós et al., 2010; Sebé-Pedrós and Ruiz-Trillo, 2010), essential for cell adhesion to the extracellular matrix in animals; members of important signalling cascades, such as the *Myc-Max* network (Young et al., 2011) and developmental TFs, such as members of the *T-box, Runx* and *NF-κB* TF families (de Mendoza et al., 2013; Sebé-Pedrós et al., 2013a, 2011), key to temporal and spatial cell differentiation during early embryonic development in animals.

### 1.7.3. *Capsaspora* and its regulated life cycle transitions

*Capsaspora* life stages presented different transcriptomic profiles: in the *adherent stage*, genes related to signalling cascades, such as tyrosine kinases and G-protein-coupled receptors; transcriptional regulation, such as the *Basic Leucine Zipper (bZIP)* superfamily; and cell metabolism, such as protein synthesis and DNA replication machinery; were significantly upregulated (Sebé-Pedrós et al., 2013b). In the *cystic stage*, cells significantly upregulated genes involved in vesicle transport and autophagy, suggesting a high cytosolic rearrangement and protein turnover, possibly in response to

[5] *Capsaspora's* genome is relatively small, of around 28 Mb. It encodes 8,657 protein-coding genes, which account around 58.7% of the genome. Its genome is compact, around 309.5 genes per Mb. Interestingly, it presents a 196.9 Kb Mitochondrial genome, which is 12 times bigger than the one of other eukaryotes (Suga et al., 2013).

starvation or stress (Kiel, 2010). Other protein domains involved in the ubiquitin pathway (*e.g., UQ_con, zf-RING2* and *Cullin* domains) and in synaptic cell-cell communication (*e.g., SNARE*, *synaptobrevin* and *syntaxin*), as well as specific TF families (*e.g., bHLH* TFs), were also significantly upregulated (Sebé-Pedrós et al., 2013b). Nevertheless, genes associated with myosin transport, DNA replication, protein translation and metabolic activities (especially mitochondrial energy production) are significantly downregulated. Interestingly, aggregates upregulated genes related to multicellular behaviours in animals, such as the complete set of the *integrin adhesome system* and associated signalling and cell-adhesion proteins, such as *laminins* and *tyrosine kinases* (Hamazaki et al., 1998; Lewandowska et al., 1991; Sebé-Pedrós et al., 2013b). Other proteins involved in multicellular functions such as the IPP (ILK-PINCH-Parvin) complex, G-protein α-13 (Gong et al., 2010) and fibronectin_3 domains (known to interact with integrins) as well as genes involved in mitosis and in the tubulin cytoskeleton were also upregulated at this stage. Altogether, this suggested that the molecular toolkit associated with animal multicellularity could function both in aggregative and clonal multicellular contexts.

Analysis of the proteome and phosphoproteome of *Capsaspora* in the different life stages also confirmed a temporally regulated profile that correlated with transcriptomic data (Sebé-Pedrós et al., 2016a). Especially in the aggregative stage, there was a significant enrichment of genes shared with Metazoa and choanoflagellates. Moreover, its dynamic pattern of protein phosphorylation also was argued to be an holozoan distinctive feature on the basis of the emergence of tyrosine kinase signalling genes in the ancestor of Holozoa (Suga et al., 2012).

The regulatory functions of *Capsaspora's* genome have also been characterized in a cell type-specific manner. *Capsaspora* life cycle transitions resulted to be associated to changes in chromatin states as well as to differential long intergenic non-coding RNAs (lincRNAs) expression and dynamic *cis*-regulatory sites, revealing a dynamic regulation of the chromatin states that correlated with gene expression. Interestingly, *Capsaspora* presents several novel specific histone modifications, some of them being the fastest evolving components of the histone code (Sebé-Pedrós et al., 2016a).

Moreover, *Capsaspora* lacks both animal promoter types (distal regulatory sites) and signatures of animal enhancers. Contrary to animals, the regulatory sites in *Capsaspora* are significantly smaller, proximal and more uniformly distributed (Sebé-Pedrós et al., 2016a).

### 1.7.4. Conserved transcription factor networks in *Capsaspora*

*Capsaspora* is the closest unicellular relative of animals with the known largest repertoire of metazoan-like TFs, such as *Brachyury, Myc* and *Runx* TF families, key for animal development (Sebé-Pedrós et al., 2011). The *Capsaspora* ortholog of *Brachyury (CoBra),* a developmental TF involved in animal gastrulation, was recently analysed in an integrative cell-type specific analysis (Sebé-Pedrós et al., 2016a). The *CoBra* downstream regulatory network included genes involved in establishment of cell polarity, phagocytosis, metabolism, transcription factors, and G-protein coupled receptor (GPCR) signalling genes. Moreover, 63 shared orthologs between inferred *CoBra* targets and those known for mouse *Brachyury* were enriched in actin cytoskeleton and amoeboidal cell-motility functions (Sebé-Pedrós et al., 2016a). Thus, similarly to its animal homolog, *CoBra* controls genes related to cell migration in the filopodial and aggregative stages, an essential cellular function later used in animal gastrulation. *CoBra* also demonstrated a functional conservation in another study using heterologous expression in *Xenopus laevis* (Sebé-Pedrós et al., 2013a). The molecular phenotype of *CoBra*, and that of homologs of early branching metazoans (*i.e.,* the cnidarian *Nematostella vectensis* and the calcareous sponge *Sycon ciliatum*), were analysed in a comparative approach. Interestingly, *CoBra* could partially rescue the phenotype of *Xenopus laevis* embryos, although it activated target genes known to be regulated by other T-box gene classes, not by *Brachyury* (Sebé-Pedrós et al., 2013a). Altogether, this suggested that animal *Brachyury* TF network is remarkably conserved in *Capsaspora.*

Other TF networks have also been analysed in *Capsaspora* (Sebé-Pedrós et al., 2016a). A blind motif-enrichment analysis of ATAC-defined sites revealed three significantly enriched nucleotide motifs, which possessed ~90% similarity to known binding motifs for animal *Runx, NFAT/NF-κB* and *Myc* TFs, orthologs that had been previously described in *Capsaspora* (Sebé-Pedrós et al., 2011).

Interestingly, *Capsaspora Myc* ortholog was strongly associated with regulatory sites with higher ATAC-seq signal in the filopodial stage. As previously described, the filopodial stage corresponds to the proliferative stage in *Capsaspora,* correlating with the well-studied proliferative role of *Myc* as a proto-oncogene in animals. Moreover, *Capsaspora Myc* downstream regulatory genes were related to ribosome biogenesis and translation, similar to what is described in animal *Myc* networks (Sebé-Pedrós et al., 2016a).

Altogether, the results of the TFs analysed suggest that *Capsaspora* presents relatively complex TF-TF regulatory interactions and that at least some of the TF downstream regulatory networks were already conserved in the unicellular ancestor of animals. Subsequent remodelling and expansion of TF networks were probably associated with increased complexity of the animal lineage. Moreover, some of these TF networks might have been co-opted during the transition, and acquired novel functions in animals (Richter and King, 2013; Sebé-Pedrós et al., 2017).

## *Summary of the Introduction*

1. Multicellularity evolved independently and repeatedly in at least 25 lineages from all three domains of life representing a wide spectrum of organizations, from simple colonies of cells (simple multicellularity) to complex differentiated multicellular organisms (complex multicellularity), especially in all major eukaryotic groups.
2. Genetic programs linked to cell adhesion, cell-cell communication and cell differentiation, are key for the establishment and maintenance of multicellularity.
3. Multicellularity confers several selective advantages: it is an effective way to increase body size, efficiently store nutrients, be more environmentally resilient (*e.g.,* escape from predation), it allows functional specialization by division of labour, it confers higher hereditary potential and opens new ecological opportunities.
4. Animal multicellularity is unique in the eukaryotic tree of life, presenting the widest range of complex architectural forms with cell differentiation and tissue organization regulated by cell-type specific regulatory programs. The biogeochemical context at the end of the Neoproterozoic era influenced animal diversification ~541-565 Mya.
5. Transcriptional regulation, and especially developmental transcription factor networks, are key for animal development and multicellularity, as they act as regulators of cell fate, cell patterning and cell differentiation programs.
6. Unicellular holozoans are the closest extant unicellular relatives of animals. Strikingly, they possess complex repertoires of genes previously thought to be animal-specific, and multicellular-like structures at some point during their life cycle. Comparative analyses of their genomes and regulated life cycle transitions can help us get insights into the nature of the unicellular ancestor.
7. Novel biological questions require suitable and phylogenetically better-placed emerging model systems to be fully addressed. The development of genetic tools among unicellular holozoans is key to performing molecular and functional-based analyses and better understand the origin of animal multicellularity.
8. *Capsaspora owczarzaki* is the closest unicellular relative of animals in which key genes related to transcriptional regulation and adhesion can be studied to date. It presents a differentially regulated life cycle at the transcriptomic, proteomic and phosphoproteomic levels.

*Section 2*

# Objectives

*A functional approach to address*
*the transition to multicellularity in animals*

## *Section 2*

# Objectives

The general framework of my thesis is to gain insights into the transition to animal multicellularity from a functional perspective, focusing on the filasterean *Capsaspora owczarzaki*. To this end, I have focused in the following objectives:

1. The development of a reliable and reproducible protocol to transiently transfect the filasterean *Capsaspora owczarzaki.*

2. The study of the evolution and diversification of *Runx* and *NF-κB* transcription factor families in eukaryotes and assessment of the role of *Runx1, Runx2* and *NF-κB* transcription factor animal homologs in *Capsaspora owczarzaki*.

*Section 3*

# Results

*Transfection of Capsaspora owczarzaki*
*and the evolution of transcription factor networks*

## *Impact and autorship report of the publications*

Among the two manuscripts here presented, Núria Ros i Rocher has been the sole first author in one of them, and co-first author in the other. The specific contributions of Núria Ros i Rocher to each publication are:

### Publication *(section 3.1)*

Parra-Acero H, **Ros-Rocher N,** Perez-Posada A, Kożyczkowska A, Sánchez-Pons N, Nakata A, Suga H, R. Najle S, Ruiz-Trillo I. *Transfection of Capsaspora owczarzaki, a close unicellular relative of animals* Development. 2018. 145(10): dev.162107. doi: 10.1242/dev.162107

- Impact Factor (2016): 5.843
- 5 Year Impact Factor: 6.2
- Authorship: co-first authorship with equal contributions with HPA.
- Contribution: Núria had a leading role in the design of the experiments, the realization of the experiments, the analysis and the interpreation of the data generated, and in the writting and editing of the manuscript and the figures.

### Unpublished result *(section 3.2) (not yet submitted)*

**Ros-Rocher N,** R. Najle S, M. Leger M, Yang A, T. Weirauch M, R. Hughes T, Ruiz-Trillo I. *Evolution of Runx and NF-κB developmental Transcription Factor families and the origin of animal multicellularity. Unpublished*.

- Impact Factor (2018): NA
- 5 Year Impact Factor: NA
- Authorship: first author
- Contribution: Núria had a leading role in the design of some of the experiments, the realization of the experiments, the analysis and the interpreation of the data generated, and in the writting and editing of the manuscript and the figures.

## *Section 3.1*

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

Parra-Acero H, **Ros-Rocher N,** Perez-Posada A, Kożyczkowska A, Sánchez-Pons N, Nakata A, Suga H, R. Najle S, Ruiz-Trillo I. Transfection of *Capsaspora owczarzaki,* a close unicellular relative of animals Development. 2018. 145(10): dev.162107. doi: 10.1242/dev.162107

## *Section 3.2*

# Evolution of *Runx* and *NF-κB* developmental Transcription Factor families and the origin of animal multicellularity

**Ros-Rocher N,** R. Najle S, M. Leger M, Yang A, T. Weirauch M, R. Hughes T, Ruiz-Trillo I. Evolution of *Runx* and *NF-κB* developmental Transcription Factor families and the origin of animal multicellularity. *Unpublished*.

# Evolution of *Runx* and *NF-κB* developmental Transcription Factor families and the origin of animal multicellularity

Núria Ros-Rocher[1], Sebastián R. Najle[1,2], Michelle M. Leger[1], Ally Yang[3], Matthew T. Weirauch[3,4], Timothy R. Hughes[3,5], Iñaki Ruiz-Trillo[1,6,7,*]

[1]*Institut de Biologia Evolutiva (CSIC-Universitat Pompeu Fabra), Passeig Marítim de la Barceloneta 37-49, 08003 Barcelona, Catalonia, Spain*

[2]*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*

[3]*The Donnelly Centre, University of Toronto, Toronto ON M5S 3E1, Canada*

[4]*Center for Autoimmune Genomics and Etiology (CAGE) and Divisions of Biomedical Informatics and Developmental Biology, Cincinnati Children's Hospital Medical Center, Cincinnati, OH 45229, USA; Department of Pediatrics, University of Cincinnati College of Medicine, Cincinnati, OH 45229, USA.*

[5]*Department of Molecular Genetics, University of Toronto, Toronto, ON M5S 1A8, Canada; Canadian Institute for Advanced Research, MaRS Centre, West Tower, 661 University Avenue, Suite 505, Toronto, ON M5G 1M1, Canada.*

[6]*Departament de Genètica, Microbiologia i Estadística, Facultat de Biologia, Institut de Recerca de la Biodiversitat (IRBio), Universitat de Barcelona (UB), Barcelona 08028, Catalonia, Spain*

[7]*ICREA, Passeig Lluís Companys 23, 08010, Barcelona, Catalonia, Spain*

**Correspondance and requests to I.R.-T. (email: inaki.ruiz@ibe.upf-csic.es)*

## *Abstract*

The origin of animal multicellularity is a major question in biology. Recently, new genome data from extant unicellular relatives of animals revealed that the single-celled animal ancestor possessed a complex repertoire of developmental transcription factors (TFs), key for animal multicellularity. This suggests a more ancient origin of complex regulatory networks prior to the emergence of animals. However, the role of TFs in a unicellular context and how they were possibly co-opted during the transition to animal multicellularity remains unsolved. Here, we investigated the evolution of *Runx1, Runx2* and *NF-κB* developmental transcription factors through a paneukaryotic bioinformatic analysis. We also assessed their functional role in *Capsaspora owczarzaki,* a close unicellular relative of animals. Interestingly, *Runx1, Runx2* and *NF-κB* proteins localise in the nucleus and/or in vesicle bodies in a cell-type specific manner across *C. owczarzaki* life stages. We additionally evaluated their binding preferences and downstream regulatory network through Chromatin Immunoprecipitation experiments.

***Keywords***

*Capsaspora owczarzaki, multicellularity, origin of Metazoa, transcription factors, development*

## INTRODUCTION

How animals emerged from their single-celled ancestor is an intriguing yet unsolved question in biology (Cavalier-Smith, 2017; King, 2004; Knoll, 2011; Richter and King, 2013; Rokas, 2008; Ruiz-Trillo et al., 2007; Sebé-Pedrós et al., 2017). Developmental programs linked to embryogenesis and cell type differentiation, ultimately orchestrated by developmental transcriptional factors (TFs), are key to animal multicellularity (Davidson and Erwin, 2006; Larroux et al., 2008). Strikingly, recent genome data of the closest extant unicellular relatives of animals revealed that they actually possess and express an unexpected repertoire of developmental TFs previously thought to be animal-specific (de Mendoza et al., 2015, 2013; Fairclough et al., 2013; King et al., 2008; Richter et al., 2018; Sebé-Pedrós et al., 2013a, 2011). This suggests that animal evolution was not solely dependent on gene innovation itself, but that co-option, tinkering and expansion of pre-existing transcription factor networks were probably key for the evolution of animal multicellularity (Arenas-Mena, 2017; King, 2004; Sebé-Pedrós et al., 2016a). Thus, to determine how key TF networks were co-opted it is crucial to study their regulatory capabilities among the extant closest unicellular relatives of animals. Some of the pre-metazoan developmental TF networks include representatives of the *Basic helix–loop–helix (bHLH), T-box, Rel/NF-κB, bZIP, Runx*, and a diversity of homeobox-containing classes (Degnan et al., 2009; Fairclough et al., 2013; Nichols et al., 2012; Sebé-Pedrós et al., 2013a, 2011, 2010). *Runx* and the *NF-κB* developmental TFs have myriad roles in development, cell fate determination, cell differentiation and stress responses in animals. Both TFs were identified in the filasterean amoeba *Capsaspora owczarzaki,* hereafter *Capsaspora,* one of the closest unicellular relatives to animals (Sebé-Pedrós et al., 2016a, 2016b, 2013a, 2011). *Runx* is a family of heterodimeric TFs with essential functions as master regulators of diverse developmental processes such as haematopoiesis and skeletogenesis in animals, and was initially thought to be an animal apomorphy (Coffman, 2003; Rennert et al., 2003; Robertson et al., 2009). The *Rel/NF-κB* signalling pathway in bilaterians is a well-described multicomponent pathway related to cell proliferation in immune responses and stress responses, such as pathogen stresses and environmental stresses (Hayden and Ghosh, 2004; Macian, 2005).

*Capsaspora* is an ideal organism to analyse TFs because it presents a complete repertoire of TFs, some of which were secondarily lost in other unicellular relatives of animals such as choanoflagellates (Sebé-Pedrós et al., 2011). Moreover, *Capsaspora* has three morphologically and transcriptionally different life stages under culture conditions, including a multicellular stage (Sebé-Pedrós et al., 2013b). *Capsaspora* grows exponentially in the “adherent stage”, in which motile amoebae extend long, thin filopodia and crawl attached to the substrate. In response to crowding or stress, *Capsaspora* transitions to

the “cystic stage”, in which cells float and encyst by retracting the filopodia. Finally, *Capsaspora* can actively form multicellular structures by the aggregation of independent cells into an “aggregative stage”, in which cells stick together with the aid of a cohesive extracellular matrix-like material. Interestingly, *Capsaspora’s Runx* and *NF-κB* homologs present different transcriptional levels and protein abundance across its different life stages, suggesting they act as complex regulatory networks with a putative role in transitioning between life stages (Sebé-Pedrós et al., 2016a, 2016b, 2013b; Stibbs et al., 1979). Moreover, they present different phosphorylation profiles across life stages (Sebé-Pedrós et al., 2016b). However, their downstream regulatory network has not been assessed *in vivo.*

Here we investigated the evolution and diversification of *Runx* and *NF-κB* TF families through a taxon-rich paneukaryotic survey and searched for the presence of their key domains in several eukaryotic genomes and transcriptomes, including the newly described filastereans *Pigoraptor chileana* and *Pigoraptor vietnamica* and *Syssomonas multiformis* (Hehenberger et al., 2017). We additionally assessed their role in *Capsaspora* through localisation and Chromatin Immunoprecipitation coupled with high-throughput sequencing (ChIP-seq) experiments, to gather additional evidence of their putative function as transcriptional regulators in a unicellular context. We found that *Capsaspora CoRunx2*, and likely *CoRunx1* and *CoNF-κB,* are capable of functioning as transcription factors. Their downstream regulatory network was primarily enriched with genes related to cell growth, response to stress and proliferation. Moreover, we found a remarkable degree of conservation between *Capsaspora* TF networks and their animal homologs, especially for *CoRunx2*. Thus, our results suggest that complex regulatory networks of transcription factors exist in *Capsaspora* and that they are temporally regulated across *Capsaspora* different life stages.

## RESULTS AND DISCUSSION

### *Runx, NF-κB* and related partners in unicellular holozoans

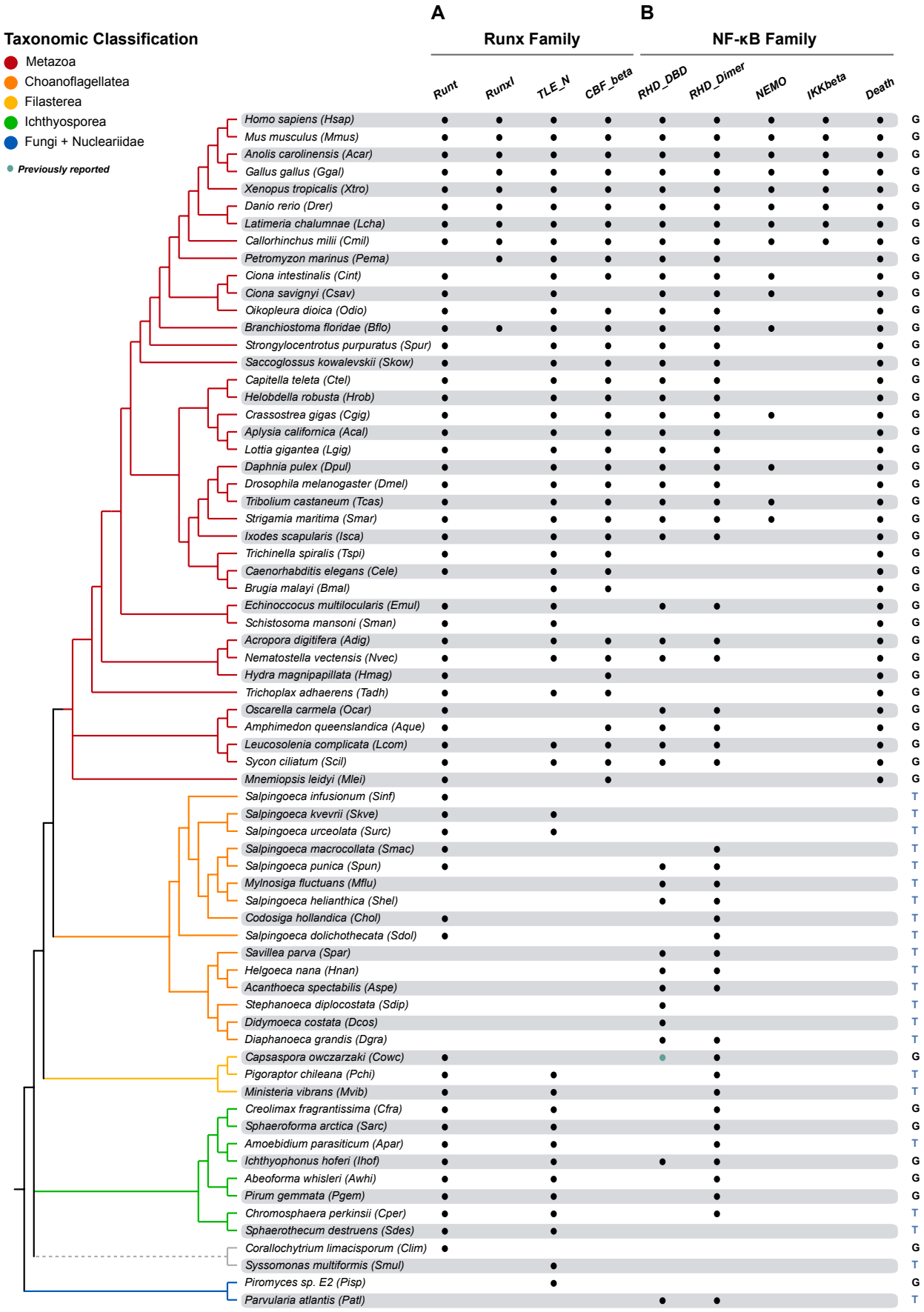
The *Runx* and the *NF-κB* TFs are two developmental TF families with evolutionarily conserved sequence-specific DNA-binding domains (DBDs): the Runt and the Rel-homology Domain (RHD), respectively. In contrast to Runt, which is exclusive to *Runx* TF family, RHD is found in several TF families, including *NF-κB* and the Nuclear factor of activated T-cells (*NFAT*) TF family. Thus, to detect the presence of *Runx* and *NF-κB* TFs we performed Hidden Markov Model (HMM) searches using HMMs for both DBDs over a taxon-rich paneukaryotic database of predicted proteomes, including newly sequenced protistan lineages **(Table S1)** (Hehenberger et al., 2017; Richter et al., 2018). We also searched for the C-terminal *Runx* Inhibitor domain related to *Runx* TF (RunxI), and the Rel-homology Dimerization domain and the C-terminal Death domain that are each found in some *NF-κB* proteins. After the primary search, we additionally re-

analysed significant hits by BLAST searches and performed a PfamScan search to further confirm their presence and to evaluate their domain architecture.

Among unicellular holozoans, we confirmed well-conserved Runt and RHD domains in some of the recently sequenced choanoflagellate transcriptomes (Richter et al., 2018), two Runt and a single RHD in the filasterean *Capsaspora* (Sebé-Pedrós et al., 2011), two Runt and two RHD in the filasterean *Ministeria vibrans* and in several ichthyosporeans, and a single Runt domain in *Corallochytrium limacisporum* **(Fig. 1, Fig. S1)** (de Mendoza et al., 2013). Interestingly, both domains were also identified in the newly described filasterean *Pigoraptor chileana* but neither Runt nor RHD were identified in *Pigoraptor vietnamica* nor in *Syssomonas multiformis* (Hehenberger et al., 2017). Moreover, we recovered a single RHD in the recently described nucleariid *Parvularia atlantis* (López-Escardó et al., 2018), which is closely related to fungi **(Fig. 1, Fig. S1**) (de Mendoza et al., 2013). Both the *Runx* Inhibitor domain and the *NF-κB*-related Death domain appeared to be metazoan innovations, RunxI appearing in chordates **(Fig. S2)** and the Death domain in non-bilaterian animals **(Fig. S3).** Our observations confirm that the Runt domain was gained at the base of Holozoa and *NF-κB* at the root of Opisthokonta, putatively being secondarily lost in Fungi. Both TFs families were significantly enriched at the origin of Metazoa, possibly through gene duplication events and diversification of protein domain architectures.

Furthermore, we found that *Runx* homologs among unicellular holozoans share with animals key DNA binding amino acids and Cysteine residues involved in redox binding affinity regulation **(Fig. S4).** A single *Runx* in the filastereans *Capsaspora* and *P. chileana,* in the ichthyosporeans *Creolimax fragrantissima, Sphaeroforma arctica, Abeoforma whisleri, Chromosphaera perkinsi* and *Sphaerothecum destruens* and in *C. limacisporum* shared the two highly conserved Cysteine residues with other animal *Runx*. Interestingly, all unicellular holozoan sequences lacked the *Runx* C-terminal inhibitor motif (WRPY) known to interact with Groucho/Transducin-like Enhancer-of-split (TLE) co-repressors in animals **(Fig. S2)** (Coffman, 2003; Levanon et al., 1998; Robertson et al., 2009). Nevertheless, sequences for some of the species included come from transcriptomic data, so we cannot completely exclude the possibility that this motif is present in some unicellular holozoans **(Supplementary File 1).**

We additionally extended the search to key domains from the evolutionarily conserved Groucho/TLE co-repressors and the Core binding Factor (CBFβ) **(Fig. 1A, Fig. S1A)**. Both protein families are related as direct interacting partners for *Runx*, and mediate its dual role as activator or repressor of key developmental pathways (Levanon et al., 1998; Robertson et al., 2009; Yarmus et al., 2006). Groucho/TLE co-repressors are known to directly interact with the C-terminal WRPY domain of *Runx* TFs, favouring repression,




*(Figure 1 legend on next page)*

**Figure 1. Table of domain presence of *Runx* and *NF-κB* TF families and related partners.** Columns represent all the PFAM domains analysed in this study. Black dots indicate the presence of an ortholog. The phylogenetic relationships are based on several recent phylogenomic studies (Brown et al., 2009; Hehenberger et al., 2017; Ruiz-Trillo et al., 2008; Torruella et al., 2015). Letters on the right depict the data source: genome (G) or transcriptome (T). Taxa are color-coded according to taxonomic assignment (indicated in the upper left legend).

through a highly conserved N-terminal Glutamine rich domain (TLE_N domain), which is unique to the Groucho/TLE family. The Core binding Factor (CBFβ) is an heterodimeric binding partner known to enhance *Runx* DNA binding affinity (Robertson et al., 2009; Sullivan et al., 2008; Wheeler et al., 2000). We identified the TLE_N domain in two choanoflagellate species, *Salpingoeca kvevrii* and *Salpingoeca urceolata,* in the filastereans *Pigoraptor chileana* and *Ministeria vibrans,* in all ichthyosporeans and in *Syssomonas multiformis.* We confirmed its absence in *Capsaspora* **(Fig. 1A, Fig. S1A)** (Sebé-Pedrós et al., 2011). We also identified a single TLE_N domain in the fungus *Piromyces* sp. E2, suggesting this domain originated at the root of Opisthokonta, despite having diversified at the onset of Metazoa. In contrast, CBFβ domain was found to be a clear metazoan innovation.

We found that *NF-κB* homologs in unicellular holzoans share with their animal homologs key domains for DNA-binding specificity, including a highly conserved specific recognition loop (RL) within the Rel-homology DBD, an immunoglobulin-like fold in a Rel-homology dimerization domain and a highly conserved basic Nuclear Localisation Sequence (NLS) **(Fig. S5).**

A Glycine-Serine rich region (GS), previously reported in *Capsaspora* (Sebé-Pedrós et al., 2011), was also present in the *NF-κB* homolog of *M. vibrans, P. chileana, C. perkinsi* and *P. atlantis,* but not in the *NF-κB* homolog of other ichthyosporean sequences nor in choanoflagellates **(Supplementary File 2).** Interestingly, we also identified in the *NF-κB* homolog of several filasterean, ichthyosporean and nucleariid taxa C-terminal Ankyrin repeats key for protein-protein interactions, that were previously thought to be exclusive to metazoan *NF-κB* **(Fig. S3).**

We finally extended the search to two additional binding partners of *NF-κB:* the *NF-κB* essential modulator (NEMO) and the Inhibitor of Kappa B Kinase (IKK) **(Fig. 1B, Fig. S1B).** They are both associated regulatory proteins from the *IKK* complex that activate *NF-κB* to initiate inflammatory responses, cell proliferation or cell differentiation in animals. Thus, we searched for NEMO and the NEMO-binding domain in IKK (IKKbetaNEMObind) **(Fig. 1B, Fig. S1B)**. Both NEMO and the IKK NEMO-binding domain were found to be specific to metazoans and were not detected in any of the unicellular holozoans evaluated.

### *Runx* and *NF-κB* evolution and diversification in Holozoa

To further confirm the evolutionary relationships of both *Runx* and *NF-κB* TF families, we performed a taxon-rich and updated maximum likelihood-based phylogenetic analyses for each TF family **(Fig. 2, Supplementary Files 1-6)**. The *Runx* phylogeny remained poorly resolved, with low nodal support throughout most parts of the tree **(Fig. 2B, Supplementary Files 5-6).**

For the *NF-κB* phylogeny, we included both *NF-κB* and *NFAT* TFs to elucidate the evolutionary relationships between these proteins and the RHD-containing proteins that we had identified in unicellular holozoans **(Fig. 2A, Supplementary Files 3-4).** As in previous phylogenies (Sebé-Pedrós et al. 2011), animal RHD-containing proteins formed a well-supported clade to the exclusion of the unicellular holozoan sequences. Within this clade, *NFAT* orthologs and *Rel/NF-κB* family proteins formed each a well-supported clade. In the absence of additional RHD-containing proteins that might provide an outgroup, we cannot exclude the possibility that the last holozoan common ancestor possessed both *NF-κB* and *NFAT* orthologs, and that the *NFAT* orthologs were subsequently lost in all extant unicellular holozoans. However, the most parsimonious explanation is a duplication of an ancestral Rel domain-like protein prior to the last metazoan common ancestor, followed by acquisitions of new domains in the paralogs leading to the emergence of modern *NF-κB* and *NFAT* proteins.

### *CoRunx1*, *CoRunx2* and *CoNF-κB* are dynamically localised in the nucleus in *Capsaspora*

The biological activity of some transcription factors is controlled through dynamic subcellular translocation, sometimes being shuttled between the cytoplasm and the nucleus upon activation through specific interactions and signalling by Nuclear Localisation Sequences (NLS). In animals, inactive *NF-κB* is sequestered in the cytoplasm by the physical interaction with its inhibitor, *IκB*. Upon activation, signal-mediated *IκB* degradation triggers the release and subsequent nuclear translocation of *NF-κB* (Huguet et al., 1997; Mikenberg et al., 2007; Oeckinghaus and Ghosh, 2009; Tay et al., 2010). Similarly, mammalian *Runx1* translocates between the nucleoplasm and vesicles. *Runx2* has been reported, in animals, mainly in nucleus and nucleoplasm and *Runx3* translocates between cytoplasm and nucleus upon activation through Pim-1 Kinase-mediated phosphorylation (Kim et al., 2008). In *Capsaspora,* expression of *Runx1 (CoRunx1), Runx2 (CoRunx2)* and *NF-κB (CoNF-κB)* homologs at the transcript and proteomic level was previously reported across its different life stages (Sebé-Pedrós et al., 2016b, 2013b). Transcriptomic expression levels and relative protein abundance of *CoNF-κB* indicate its predominantly expression at the cystic stage **(Fig. S6A-B).** *CoRunx1* protein levels resulted more abundant at the adherent and aggregative stages, whereas *CoRunx2* showed more protein abundance at the cystic stage **(Fig. S6A-B)**.

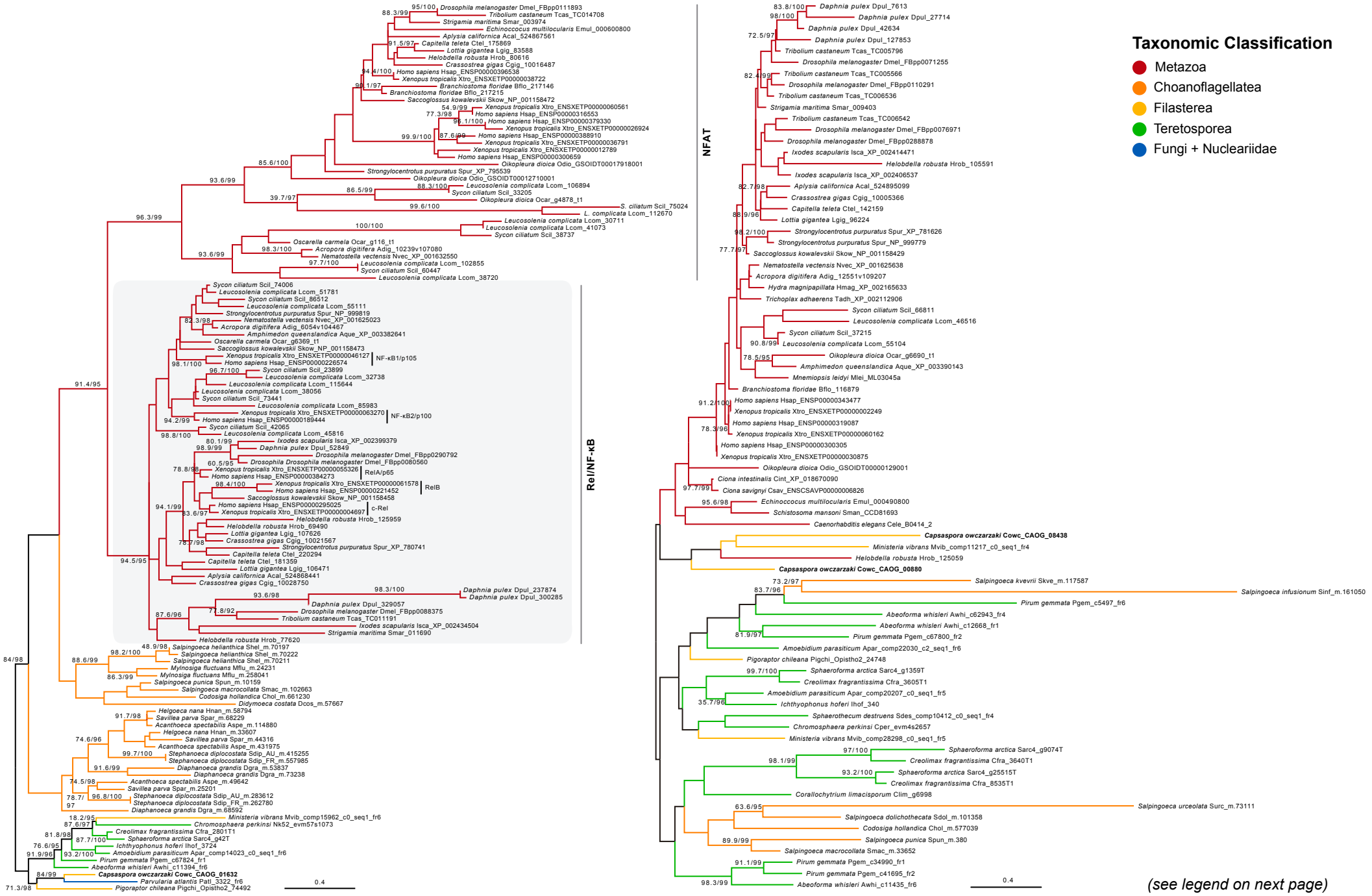

A
Rel-homology DNA-binding Domain
B
Runt DNA-binding Domain
NFAT
Rel/NF-kB
NF-κB1/p105
NF-κB2/p100
RelA/p65
RelB
c-Rel
Taxonomic Classification
Metazoa
Choanoflagellatea
Filasterea
Teretosporea
Fungi + Nucleariidae
0.4
0.4
(see legend on next page)

**Figure 2. IQ-Tree of Rel-homology and Runt DNA-binding Domains.** (A) IQ-Tree of Rel-homology DNA-binding Domain. (B) IQ-Tree of Runt DNA-binding Domain. Taxa are color-coded according to taxonomic assignment (indicated in the upper right legend).

In order to assess whether *CoRunx1, CoRunx2* and *CoNF-κB* homologs likely function as transcription factors, we designed custom antibodies to evaluate their endogenous expression and localisation (see Materials and Methods), and to perform other functional analyses, such as Chromatin Immunoprecipitation experiments coupled with high-throughput sequencing (ChIPseq). We first confirmed their differential protein abundance using *Capsaspora* protein extracts from different life stages (adherent, aggregative and cystic) by western blot analysis using custom antibodies. We observed similar protein abundance as previously reported from proteomic data, with *CoRunx2* being more abundant in the cystic stage **(Fig. S6C)** (Sebé-Pedrós et al., 2016b).

Next, we examined differential localisation of endogenous *CoRunx1, CoRunx2* and *CoNF-κB* by immunofluorescence assays from adherent and cystic *Capsaspora* life stages **(Fig. 3).** We did not localise their expression in the aggregative stage because of technical limitations and low expression levels **(Fig. S6C).** Interestingly, *CoRunx1* localised in small vesicles (seen as small bright dots around the nucleus) in adherent cells (4-day old culture) and in cystic stage cells (14-days old) **(Fig. 3A)** and in the nucleus and small vesicles in early cystic stage cells (6-days old) **(Fig. S7A).**

Not all cells showed the same signal intensity or presence, possibly due to asynchrony of the culture. In contrast, *CoRunx2* localised clearly in the nucleus both in adherent and cystic stages **(Fig. 3B and Fig. S7B).** Consistent with previously reported data and analysis of *CoNF-κB* protein abundance in *Capsaspora* extracts, *CoNF-κB* was only expressed in the cystic stage **(Fig. 3C, Fig. S7C)**, localising in small vesicles around the nucleus in the early cystic stage (6-day old), and more intensely in small vesicles and in the cytoplasm in late cystic stage **(Fig. 3C).** Altogether, our data indicates that *CoRunx2*, and possibly *CoRunx1* and *CoNF-κB*, could be capable of functioning as transcription factors in a dynamical manner across *Capsaspora* life stages.

## Effect of *CoRunx1*, *CoRunx2* and *CoNF-κB* overexpression in *Capsaspora*

To further confirm the localisation and the effect of overexpressing *CoRunx1, CoRunx2* and *CoNF-κB in vivo* we performed overexpression experiments in live cells using the recently developed transfection protocol for *Capsaspora* (Parra-Acero et al., 2018). To this end, we fused each gene in-frame to mCherry fluorescent protein using the *Capsaspora* expression vector backbone pONSY (Parra-Acero et al., 2018). We transfected *Capsaspora* cells with either

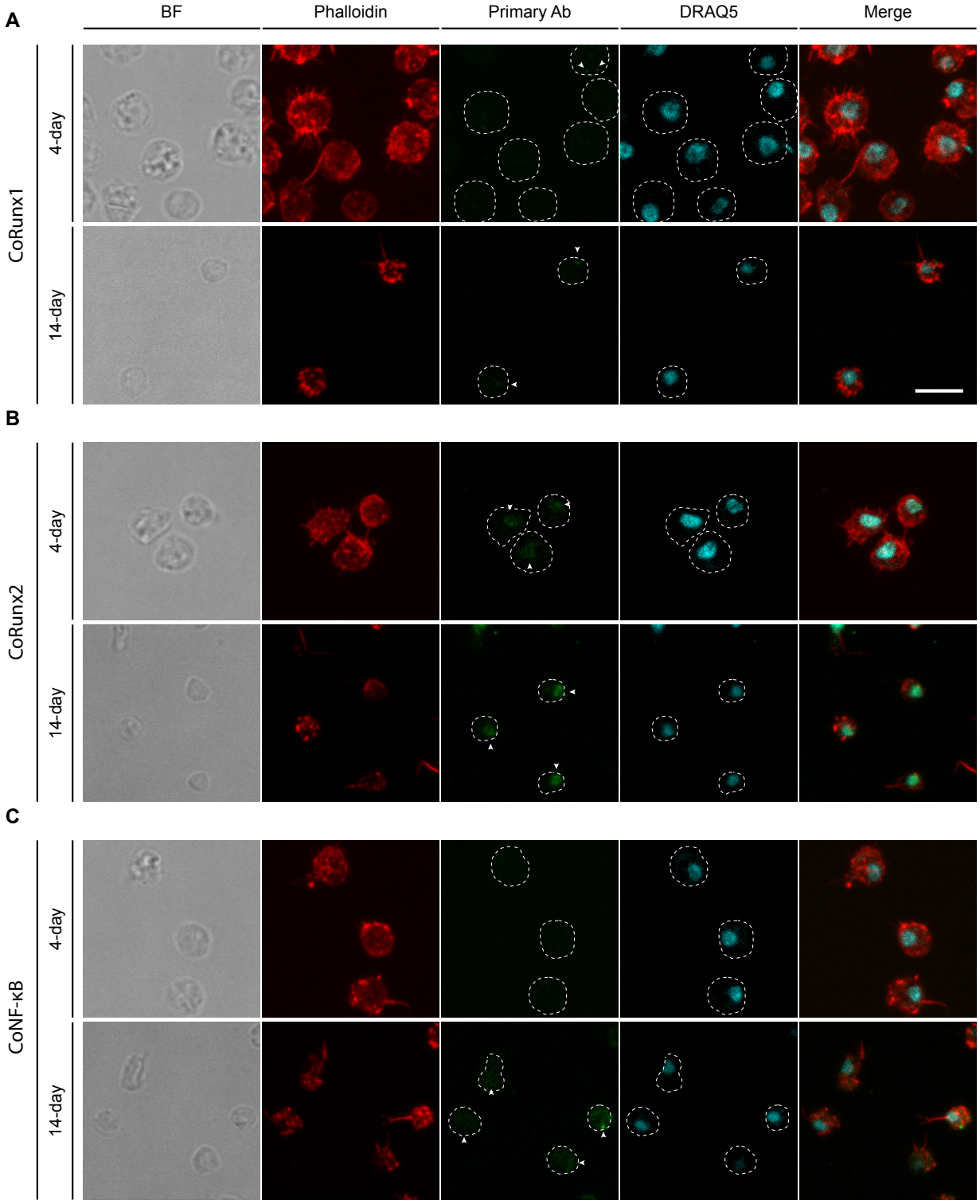



**Figure 3. Localisation of *CoRunx1*, *CoRunx2* and *CoNF-κB* n *Capsaspora.*** Immunofluorescence of 4-day adherent and 14-day cystic stage *Capsaspora* cells using custom *CoRunx1* (A), *CoRunx2* (B) and *CoNF-κB* (C) antibodies. Dashed line indicates cell body and arrows indicate antibody signal. Scale bar represents 5 μm.

pONSY-CoRunx1:mCherry, pONSY-CoRunx2:mCherry or pONSY-CoNF-κB:mCherry and used pONSY-mCherry transfected cells as a control **(Fig. S8).** Moreover, to benchmark each fluorescent marker with the nucleus, we co-transfected each construct with a cassette containing a fusion of the endogenous *Capsaspora* Histone 2B gene with Venus fluorescent protein (pONSY-CoH2B:Venus, from Parra et al., 2018) and compared its localisation in positive cells **(Fig. S9).** In total, we analysed 2.286 cells in three biological replicates and additionally examined differences in cell size (diameter) and "cell stage" (adherent or floating) **(Table S2).**

Live fluorescence imaging revealed CoRunx1:mCherry fusion localised to the cytoplasm and in small vesicles and, interestingly, it additionally localised to the nucleus in floating cells. CoRunx2:mCherry fusion consistently localised to the nucleus, both in adherent and in floating cells **(Fig. S8A-B).** This is consistent with the localisation of *Runx1* and *Runx2* animal homologs, where they are generally nuclear localised but they can be as well cytoplasmic under certain conditions (Deepak et al., 2011; Mandoli et al., 2014). In contrast, CoNF-κB:mCherry fusion was mainly located in small vesicles, in the cytoplasm and, in some cases, in the nucleus; both in adherent and floating cells. Similarly, animal *NF-κB* homologs are also located in the cytoplasm and translocated to the nucleus upon IKK-mediated activation **(Fig. S8A-B)** (Mikenberg et al., 2007; Oeckinghaus and Ghosh, 2009).

To further assess whether overexpression of either endogenous transcription factor was affecting *Capsaspora* morphologically, we compared differences in cell diameter (in μm) between positive and negative fluorescent cells transfected with each construct. We did not detect any significant differences between positive (6.092±0.990 μm) and negative cells (6.076±0.867 μm) in CoRunx1:mCherry ($P=0.8746$, Student's T-Test) and between positive (5.858±0.737 μm) and negative cells (5.932±0.809 μm) in CoRunx2:mCherry ($P=0.3478$, Student's T-Test) transfections, as well as in mCherry control ($P=0.5331$, Student's T-Test). Nevertheless, mean diameter of cells overexpressing CoNF-κB:mCherry fusion was 5.973±1.259 μm, which resulted to be significantly larger than a mean diameter of 5.616±0.831 μm in negative cells ($P=0.00567$, Student's T-Test) **(Fig. S8C).** Additionally, we examined differences in cell diameter between cells overexpressing CoRunx1:mCherry, CoRunx2:mCherry or CoNF-κB:mCherry fusions compared to mCherry control. mCherry positive cells measured a mean 5.503±1.001 μm in diameter. In general, cells overexpressing either fusions resulted in significant increased cell size compared to mCherry (CoRunx1:mCherry, $P=1.522e-05$; CoRunx2:mCherry, $P=0.002726$; and CoNF-κB:mCherry, $P=0.002476$; Student's T-Test). Overall, CoRunx1:mCherry, CoRunx2:mCherry and CoNF-κB:mCherry overexpression resulted in increased cell diameter of positive cells, compared to mCherry, suggesting that the effect of

overexpressing either TF in *Capsaspora* has an impact on cell size.

Finally, to assess whether overexpression of each construct could affect transitioning between stages, we compared differences in the percentage of floating cells between cells overexpressing CoRunx1:mCherry, CoRunx2:mCherry or CoNF-κB:mCherry fusion proteins with mCherry control **(Fig. S8D).**

Cells overexpressing CoRunx2:mCherry and CoNF-κB:mCherry floated significantly more than mCherry transfected cells (CoRunx2:mCherry, P=0.04608, CoNF-κB:mCherry, P=0.03615; Student's T-Test), whereas no significant differences were observed when comparing CoRunx1:mCherry overexpressing cells with mCherry (P=0.1353, Student's T-Test). This suggests a putative dynamic role of *CoRunx1, CoRunx2* and *CoNF-κB* across *Capsaspora* different life stages.

## Ancient developmental transcription factor networks in *Capsaspora*

To gain insights into the dynamics of *CoRunx1, CoRunx2* and *CoNF-κB* TF networks *in vivo,* we performed ChIP-seq experiments using custom antibodies against *CoRunx1, CoRunx2* and *CoNF-κB*. We assessed their binding preferences and their downstream regulatory genes across the different *Capsaspora* life stages (adherent, cystic and aggregative).

First, we examined the distribution of *CoRunx1, CoRunx2* and *CoNF-κB* binding sites across the *Capsaspora* genome **(Fig. 4A-F).** *CoRunx1* and *CoRunx2* distribution patterns were practically identical, and preferentially strongly enriched in promoter proximal regions and 5'UTR, closer to transcriptional start sites (TSS) in the aggregative stage **(Fig. 4A-D).** A number of high-confidence peaks were located in CDS regions at 2-day (37% and 47%) and 4-day (43% and 44%) adherent stages and 6-day (41% and 45%) and 14-day (49% and 41%) cystic stages, respectively. *CoNF-κB* showed a similar distribution to *CoRunx1* and *CoRunx2* at the adherent (2-day and 4-day) and cystic (14-day) stages and slightly different at the 6-day stage **(Fig. 4E-F).** However, *CoNF-κB* was enriched in regions closer to TSS (promoter proximal and 5'UTR) in the cystic stage (14-day) and, in contrast to *CoRunx1* and *CoRunx2*, presented a lower distribution of peaks around regions closer to TSS in the aggregative stage. The mock ChIP control (beads only control) did not recover any DNA in the 2-day adherent stage. In general, the control presented a similar distribution to *CoNF-κB* **(Fig. S10A-B)**.

We additionally used previously reported *Capsaspora* high-coverage nucleosome-free ATAC reads data (Sebé-Pedrós et al., 2016a) to cross-validate the distribution of high confident ChIP peaks in active regulatory sites in the *Capsaspora* genome. We calculated the distribution of the maximum ChIP peak signal around ATAC-defined regulatory sites across different *Capsaspora* life cycle stages (**Fig. S10C).** *CoRunx1* was enriched in regulatory sites at the 2-day, 14-

day and aggregative stages (**Fig. S10C)**, while *CoRunx2* was particularly enriched at the 2-day and aggregative stages (**Fig. S10C)**. However, *CoNF-κB* resulted in few peaks falling near ATAC-defined regulatory regions, being enriched only in the 14-day stage (**Fig. S10C)**. We recovered few peaks in the mock ChIP control **(Fig. S10C).** This suggests that at least *CoRunx1* and *CoRunx2* show dynamic changes in regulatory sites signal across life stages in *Capsaspora.*

To explore the nature of *CoRunx1, CoRunx2* and *CoNF-κB* bound regions *in vivo*, we performed *de novo* motif discovery to identify overrepresented sequence motifs across *Capsaspora* life stages. We recovered significantly overrepresented motifs for *CoRunx1* at the 6-day cystic (TAGGTATTAG) and aggregative stages (CACTCACTCA and CTCTCTCTCT) (2-day similar to aggregative but not significant). Interestingly, these motifs do not resemble the binding motifs of *Runx1* animal homologs (Bowers et al., 2010).

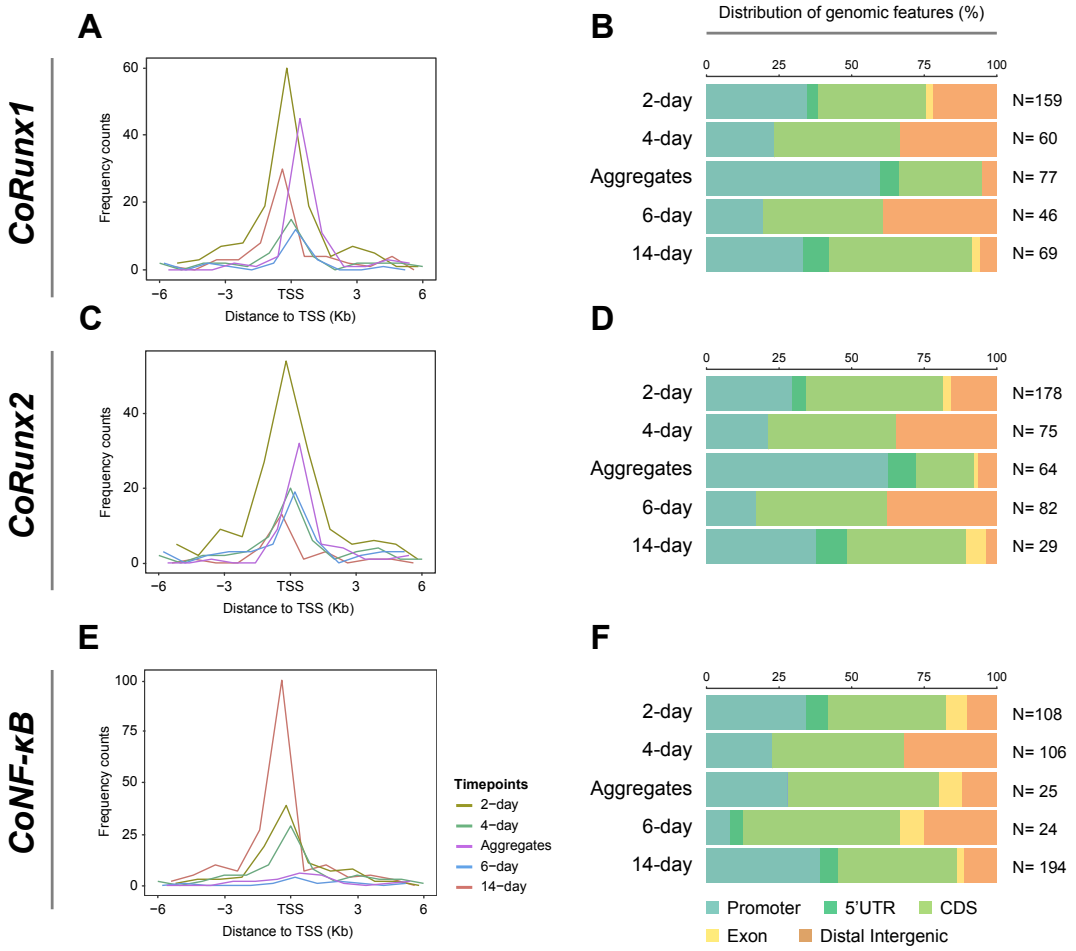
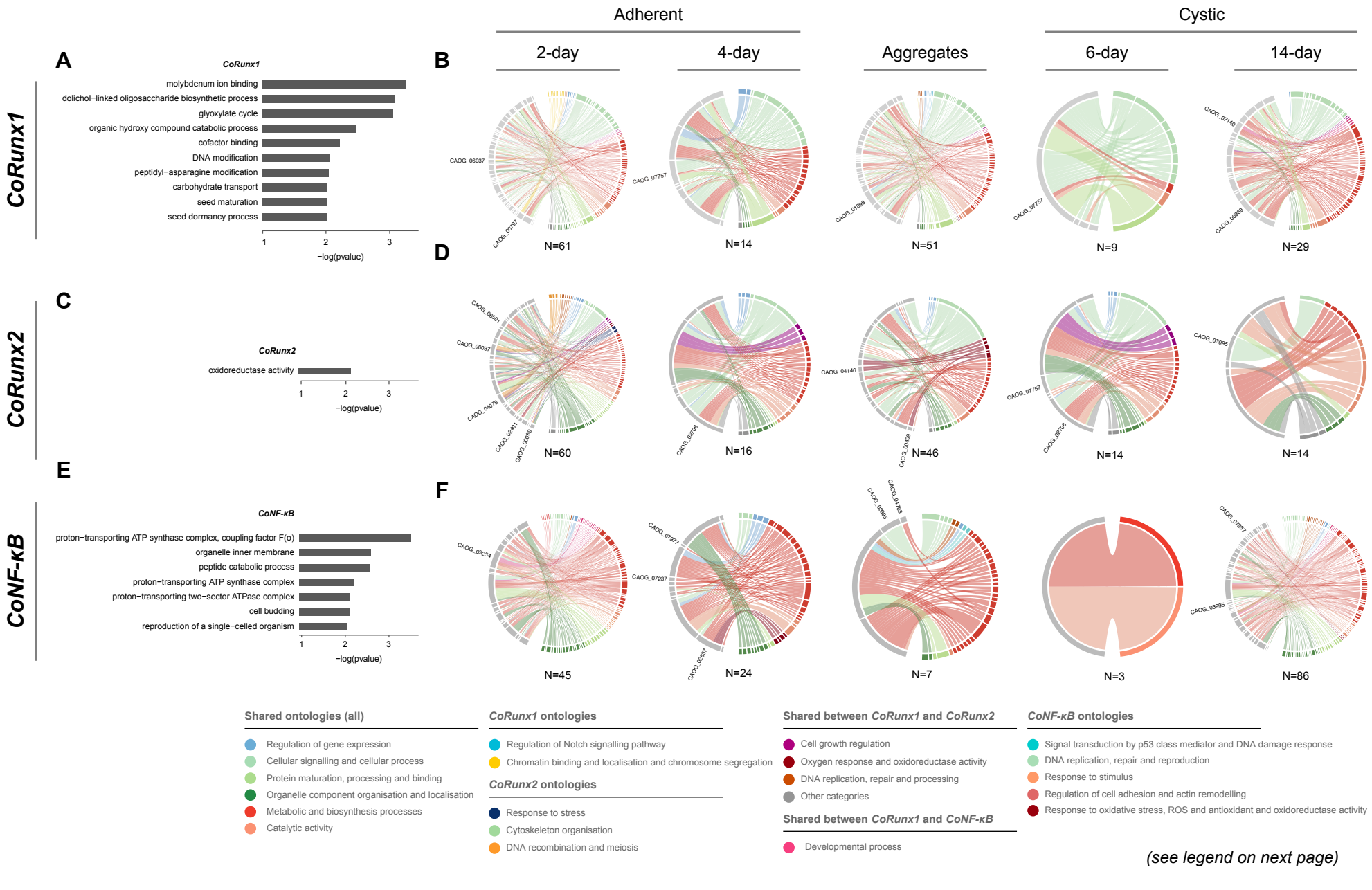



**Figure 4. Distribution of *CoRunx1, CoRunx2* and *CoNF-κB* high confidence ChIP peaks.** (A) Distribution of *CoRunx1* high confidence ChIP peaks near TSS across *Capsaspora* life stages. (B) Genomic distribution of *CoRunx1* high confidence ChIP peaks across *Capsaspora* life stages. (C-D) Same as (A-B) for *CoRunx2* ChIP peaks. (E-F) Same as (A-B) for *CoNF-κB* ChIP peaks. N in B, D, F indicate number of high-confidence peaks.

Strikingly, the most significantly overrepresented motif for *CoRunx2* corresponded to a perfect match to the animal *Runx2* consensus binding sequence AAACCGCA in the aggregative stage (Jolma et al., 2013; Wu et al., 2014). Nevertheless, we did not recover any significantly enriched motif in *CoNF-κB* samples.

We additionally determined *Capsaspora CoRunx1, CoRunx2* and *CoNF-κB* binding preferences using the universal Protein Binding Microarrays (PBMs), a common method for *in vitro* analysis of TFs sequence specificity (Berger et al., 2006; Berger and Bulyk, 2009, 2006, Weirauch et al., 2014a, 2013). Of these, only *CoRunx2* PBM yielded data satisfying our stringent success criteria, including statistically significant *E-scores* (at least one k-mer sequence with E-score≥0.45) on two independent arrays (denoted "ME" and "HK") and agreement in both k-mer data and motifs between both arrays (Berger et al., 2008; Berger and Bulyk, 2009, 2006; Weirauch et al., 2013). *CoRunx2* enriched motif corresponded to AAACCGCA, providing independent confirmation of the ChIP-seq data quality **(Fig. S10D).** This motif has been previously surveyed in *Capsaspora*, being associated to signalling receptor and transducer activities, DNA replication and repair and G protein-coupled receptors signalling gene ontologies (Sebé-Pedrós et al., 2016a). Thus, this suggests that *Capsaspora* Runx2 is a relatively complex TF network with regulatory functions related to similar processes in animals.

The *Capsaspora CoRunx1* downstream target genes were enriched in gene ontology (GO) terms associated with metabolic (catabolic) and biosynthetic processes, and regulation of cell growth and development **(Fig. 5A-B)**. Particularly, GO terms related to proliferation and development, such as chromatin binding and chromosome segregation (putative homolog to Centromere protein S (CAOG_6037) and to wings-apart like protein (*WALP*) involved in sister chromatid cohesion during mitosis in animals (CAOG_00797)) and cell signalling and regulation of gene expression (tyrosine protein kinase *Src* (CAOG_07757)) were enriched at the adherent stage (2-day and 4-day) and in the cystic stage (6-day) **(Fig. 5B)**. Genes associated to metabolic and biosynthetic processes, especially involved in lipid and carbohydrate metabolism, were enriched in the cystic stage (14-day). For example, *3-hydroxyacyl-CoA dehydrogenase* (CAOG_00369) and *isocitrate lyase* (CAOG_07140) were enriched at 14-day sample. Genes related to response to stress (*e.g.,* oxidative stress) and signalling pathways (*e.g.,* the *Notch* signalling pathway), both associated to CAOG_01898, and metabolic processes were enriched at the aggregative stage **(Fig. 5B).** Similarly to *Capsaspora CoRunx1*, animal *Runx1* is associated with cell growth, proliferation and cell cycle, DNA replication, recombination and repair, responses to genotoxic stress and biosynthesis (Bowers et al., 2010; Cai et al., 2015; Coffman, 2003; Kim et al., 2014; Mandoli et al., 2014; Michaud et al., 2008). Thus, our data suggests that there might have




*(see legend on next page)*

**Figure 5. GO enrichment of *Capsaspora CoRunx1, CoRunx2* and *CoNF-κB* high confidence ChIP peaks mapping regulatory regions.** (A) Enriched GO terms among genes associated with *CoRunx1* high confidence ChIP peaks mapping regulatory sites (P-value cutoff is 0.01). (H-I) Same as (G) for *CoRunx2* and *CoNF-κB* high confidence ChIP peaks, respectively. (H) Enriched GO terms among genes associated with *CoRunx1* peaks mapping regulatory sites across *Capsaspora* life stages (P-value cutoff is 0.05). GO terms in (H, J and L) are color-coded according to grouping GO term categories, indicated in the lower-right legend.

been a conserved *Runx1* downstream regulatory network with primary functions related to cell growth, proliferation and stress responses already present in premetazoan lineages.

The *Capsaspora CoRunx2* downstream target genes were enriched in GO terms associated with oxygen metabolism, primarily related to oxidoreductase, oxygen transport and dioxygenase activities **(Fig. 5C).** Interestingly, GO terms related to meiosis, DNA replication, repair and recombination, such as a putative Centromere protein S (CAOG_06037), a PMS2-like mismatch repair endonuclease (CAOG_04075) and a putative Replication protein A (RPA)-interacting protein (CAOG_00089) were enriched in the adherent (2-day) stage. At this stage, other genes related to cell growth regulation and metabolic (catabolic) processes, and response to oxygen and stress, such as *cytosolic malate dehydrogenase A* (CAOG_02401) and a putative *methyltransferase* (CAOG_06501) were also particularly enriched **(Fig. 5D)**. Cell growth and gene expression regulation, such as a histidine kinase (CAOG_02706), and nitrogen metabolism were particularly enriched at the 4-day adherent stage **(Fig. 5D)**.

*Tyrosine protein kinase Src* (CAOG_07757) and a histidine kinase (CAOG_02706), related to cell signalling, cell growth and metabolic and biosynthesis processes, were also enriched in the 6-day stage cystic stage **(Fig. 5D)**. In the 14-day cystic stage, GO terms were enriched in genes related to oxygen metabolism, such as an *oxidoreductase* (CAOG_03995). Similarly to *CoRunx1,* genes related to oxygen metabolism, such as oxidoreductase activities, gas transport, oxygen transport and binding, and cytoskeleton and microtubule organisation were enriched in the aggregative stage. Examples of these genes related to oxygen metabolism include a homolog to *gamma-butyrobetaine dioxygenase* (CAOG_00499), a member of the family of oxidoreductases, specifically those acting on paired donors, with $O_2$ as oxidant and incorporation or reduction of oxygen and epoxidase subunit A (CAOG_04146) **(Fig. 5D)**. Interestingly, the animal homolog *Runx2* has been associated with cell growth, cell fate determination and differentiation (Young et al., 2007) but has also been associated with oxidative stress responses (Byon et al., 2008). Thus, *Capsaspora CoRunx2* downstream target genes were enriched in GO terms associated with similar functions to

those of its animal homolog, primarily related to cell growth and oxygen metabolism.

The *Capsaspora CoNF-κB* downstream target network entailed genes involved in cell division (cytokinesis, reproduction), developmental processes such as sporulation and cell budding, cell responses to stimulus, such as nutrients or oxidative stress response (including to reactive oxygen species and antioxidant activity), transmembrane transport (ion binding, ion channel activity), and metabolic (catabolic) and biosynthetic processes **(Fig. 5E-F)**. Regulation of cell adhesion, developmental growth, response to external stimuli (alcohol, retinoic acid, mechanical stimulus and nutrients), and other related metabolic processes were enriched in the 2-day adherent stage. Interestingly, a Tenascin-like protein (CAOG_05254) was particularly related to general response to stimuli. Tenascins are involved in negative regulation of cell adhesion, having anti-adhesive properties in animals (Chiquet-Ehrismann and Chiquet, 2003). Transcription and translation, hydrogen peroxide catabolic processes, oxidoreductase activity (acting on peroxide as acceptor) and oxidative stress responses (including reactive oxygen species (ROS) metabolic processes), such as *Catalase HPII* (CAOG_02637), growth-inhibiting protein 1 (CAOG_07797) and TKL/IRAK protein kinase (CAG_07237), were enriched in the 4-day adherent stage. Interestingly, this *Catalase* was also enriched at the 14-day cystic stage. *Catalases* are essential to catalyse the decomposition of hydrogen peroxide to water and oxygen, and it has key roles in protecting the cell from oxidative damage by ROS (Vanderauwera et al., 2011). Moreover, genes were related to metabolic (catabolic) processes, such as hydrolase activity, mitosis (reproduction, cytokinesis) and transmembrane transport and cellular organisation was enriched at the 6-day and at the 14-day cystic stage. In aggregates, enriched genes were related to amine and lipid metabolism, such as *2-acylglycerol O-acyltransferase 2-A* (CAOG_04763), signal transduction (through *p53* animal homolog-mediated regulation), response to DNA damage and oxygen metabolism, such as an oxidoreductase (CAOG_03995) (also at 14-day cystic stage) **(Fig. 5F)**. This is consistent with animal *NF-κB* TFs' well-described roles in oxidative and stress responses and regulation of the p53 pathway (Hayden et al., 2006; Hayden and Ghosh, 2004; Mercurio and Manning, 1999; Morgan and Liu, 2011; Oeckinghaus and Ghosh, 2009; Webster and Perkins, 1999). Thus, similarly to its animal homolog, *Capsaspora CoNF-κB* downstream target genes were enriched in GO terms associated with cell growth and response to stimulus, such as oxidative stress response.

Globally, our results indicate that *Capsaspora CoRunx2*, and likely *CoRunx1* and *CoNF-κB*, are capable of functioning as transcription factors. Their downstream regulatory network of genes primarily related to cell growth, response to stress and proliferation, might have evolved before the origin of animals. Despite the tendency of TF networks to rapidly undergo functional changes during

evolution (Li and Johnson, 2010; Sorrells and Johnson, 2015), we found a remarkable degree of conservation between *Capsaspora* TF networks and their animal homologs (Sebé-Pedrós et al., 2016a), especially for *CoRunx2*, both in TFBS and in its downstream target genes. However, these results have been analysed in one biological replicate meaning that further confirmation is needed. Thus, analyses of a second biological replicate are underway to validate the overall findings. Nevertheless, these preliminary results suggest that complex regulatory networks of transcription factors exist in *Capsaspora* and are temporally regulated across its different life stages.

These core conserved TF networks, including the ones here depicted, were most likely integrated, expanded and remodelled during animal evolution, allowing for a tightly regulated spatiotemporal control of gene expression at the onset of animals (Peter and Davidson, 2011). Moreover, functional-based approaches, now available for other unicellular holozoans, together with other analyses of developmental TF networks in other unicellular holozoans and in early branching animals will be key to gaining insights into the early evolution of transcriptional regulatory networks at the onset of animals.

## MATERIALS AND METHODS

### Taxon sampling and domain analysis

Raw Hidden Markov Models (HMMs) of Rel Homology DNA-binding domain (RHD_DNA_bind v.21, PF00554), Rel Homology Dimerization domain (RHD_dimer v.4, PF16179), Inhibitor of Kappa B Kinase Beta NEMO binding domain (IKKbetaNEMObind v.7, PF12179), NF-Kappa B Essential Modulator (NEMO v.7, PF11577), Runt DNA-binding domain (Runt v.18, PF00853), *Runx* Inhibitor domain (RunxI v.10, PF08504), Groucho/TLE N-Terminal Q-rich domain (TLE_N v.14, PF03920), Core binding factor (CBF) beta subunit (CBF_beta v.16, PF02312) and Death domain (Death v.21, PF00531) were retrieved from Pfam v.31.0 (Finn et al., 2014; Punta et al., 2012), and used as queries in hmmscan (hmmer 3.1b2-2; Eddy, 1998; Söding, 2005) searches against a paneukaryotic predicted proteome database. Taxon sampling included 39 animals, 22 choanoflagellates, 4 filastereans, 9 teretosporeans and *Syssomonas multiformis*, 23 fungi, 2 discicristoideans, 1 apusozoan, 3 CRuMs, 3 ancyromonads, 4 amoebozoans, 7 land plants, 5 chlorophytes, 3 rhodophytes, 1 glaucophyte, 11 heterokonts, 6 alveolates, 2 rhizarians, 1 haptophyte,1 cryptophyte and 6 excavates **(Table S1).**

Using custom Perl scripts, the resulting output files were parsed and re-analysed using PfamScan v.1.5 (Mistry et al., 2007) and protein BLAST v.2.2.31 (Altschul, 1993; Altschul et al., 1990; Gish and States, 1993), and all sequences containing a Rel Homology

or Runt DNA-binding domains were retrieved. Sequences were verified using BLAST searches. The domain architecture of all retrieved sequences was inferred with PfamScan using the gathering threshold as cutoff value.

### Phylogenetic analysis

Sequences were aligned using MAFFT v7.299b e-ins-i (Katoh et al., 2005, 2002; Katoh and Standley, 2013) with the gap extension parameter set to 0, trimmed using BMGE v.1.0 (Criscuolo and Gribaldo, 2010) using the BLOSUM45 matrix, and alignments and trimming were verified by eye. Partial sequences with very few positions represented in the final trimmed alignment were excluded. Preliminary phylogenies were constructed using FastTree v. 2.1.9 (Price et al., 2010, 2009), and long-branching sequences were excluded. Final alignments were constructed using mafft e-ins-i and trimmed using BMGE as above, and final phylogenies were constructed using IQ-TREE multicore version 1.6.1 (Minh et al., 2013; Nguyen et al., 2015) using the ModelFinder option (Kalyaanamoorthy et al., 2017). ModelFinder recovered LG+G4 as the best fit for the Runt phylogeny, and LG+R5 as the best fit for the Rel homology domain-containing protein phylogeny according to the Bayesian Information Criterion.

### Cell strain, growth conditions and stage differentiation

*Capsaspora owczarzaki* cell cultures (strain ATCC® 30864) were grown axenically in 25 $cm^2$ culture flasks (Falcon® VWR, #734-0044) with 5 mL ATCC medium 1034 (modified PYNFH medium), hereafter growth medium, in a 23ºC incubator.

Different cell stages were obtained by starting *Capsaspora* cultures with $5*10^5$ cells in 75 $cm^2$ flasks with 15 mL growth medium (from an approximately $5*10^6$-$2*10^7$ cells $mL^{-1}$ at 80-90% confluence). Early and late filopodial stage cells were collected after 2 and 4 days growth, respectively. Aggregative stage cells were induced by agitating a 2-day filopodial stage culture at 50 rpm during 24 hours. Finally, early and late cystic stage cells were collected from 6-day-old and 14-day-old cultures, respectively.

### Construction of expression cassettes for *Capsaspora*

*Capsaspora* RNA was extracted using Trizol reagent (Invitrogen/ThermoFisher Scientific, #15596026). cDNA was obtained by RT-PCR using SuperScript® III Reverse Transcriptase (Invitrogen, #18080044) following manufacturer's instructions.

*Capsaspora* fusion protein constructs were built using pONSY-mCherry plasmid (Parra-Acero et al., 2018). pONSY-CoRunx1:mCherry was created by cloning *Capsaspora Runx1* (CAOG_00880) into pONSY-mCherry. *CoRunx1* was PCR amplified from cDNA using primers 1 and 2 **(Table S3)** and cloned into pONSY-mCherry multi cloning site using XmaI and EcoRV restriction enzymes.

pONSY-CoRunx2:mCherry was created by cloning *Capsaspora Runx2* (CAOG_08438) as before. *CoRunx2* was PCR amplified from cDNA using primers 7 and 8 **(Table S3)**. pONSY-CoNF-κB:mCherry was created by cloning *Capsaspora NF-κB* (CAOG_01632) as before. *CoNF-κB* was PCR amplified from cDNA using primers 13 and 15 **(Table S3)**.
All plasmids DNA were obtained using the plasmid GenElute TM Plasmid Midiprep Kit (Sigma, #NA0200-UKT), lyophilized and resuspended at an approximate concentration of 1 µg/µL in distilled water.

### Overexpression of candidate TFs in *Capsaspora*

*Capsaspora* cells were transfected as in Parra et al., 2018 with either pONSY-mCherry, pONSY-CoRunx1:mCherry, pONSY-CoRunx2:mCherry or pONSY-CoNF-κB:mCherry constructs. Each construct was transfected per duplicate in three independent experiments. Localisation of fluorescent proteins was evaluated 24h post-transfection. Controls were performed by co-transfecting pONSY-CoH2B:Venus with each construct to identify the nucleus and by staining transfected cells with DAPI.

### Imaging of transfected cells

Transfected cells were plated in µ-Slide 4-well glass bottom dish (Ibidi, #80427) with 750 µL of growth medium and grown overnight at 23ºC. For DAPI-staining control, cells were washed once with 200 µL 1X PBS, fixed with 4% Formaldehyde in 1X PBS during 15 min at RT, washed again as before and stained with 1:1000 DAPI (0.5 µg/mL) during 20 min at RT. After staining, cells were washed as before and covered using DAKO mounting media (SIGMA). Live and DAPI-stained cells were imaged in 0.7 µm width z-stack using a 63X oil objective Zeiss Axio Observer Z.1 epifluorescence inverted microscope equipped with LED illumination and Axiocam 503 mono. Images were analysed using Fiji Imaging Software version 2.0.0-rc-44/1.50e (Schindelin et al., 2012). ObjectJ Fiji Imaging software Plugin (ObjectJ-1.04c; https://sils.fnwi.uva.nl/bcb/objectj/index.html) was used for fluorescent cells quantification and diameter measurement

### Protein expression and purification

Recombinant proteins were produced by the Biomolecular Screening & Protein Technologies Unit from the Centre for Genomic Regulation (CRG). C-terminal protein fragments of CoRunx1 (711 aa, 75.8 kDa), CoRunx2 (518 aa, 54.2 kDa) and CoNF-κB (721 aa, 75.6 kDa) were cloned into pETM14 with a His-tag, expressed in *E. coli* BL21(DE3) strain and purified by Nickel affinity chromatography under native conditions with an Elution Buffer containing 50 mM Tris-HCl pH=7.4, 300 mM NaCl and 10% Glycerol. CoRunx1 and CoNF-κB pellets were additionally re-solubilized with GuHCl, purified by Nickel affinity chromatography in denaturing conditions and refolded in a second Nickel affinity chromatography under native conditions to avoid protein degradation products. Purified recombinant proteins Co-Runx1 (7mL, 1.4 mg/mL), Co-Runx2 (22.5 mL, 1.3 mg/mL) and CoNF-κB (7 mL, 1.47

mg/mL) were used for antibody production and as controls for Western Blot analysis.

**Recombinant protein Sequences:**

>CoRunx1_recombinant_protein
IKMTVDGPRDPRRRRPGSEMIASASEDDLS
DASGFASDLGSGTMSAEAFGSTASTGSPLS
GSSQQLAPGQTPPAQPNSAGTGGSFKSHR
TRSLSTTLRRCQSAPYMAPEFRRRTGSGTL
LPHDKPSSPRASSNNNDSLDPRPRSAFPSD
LADSEMQRFRSVSMGSSADSMAEDVFSPL
ESPRSPSVAALYHLQHSGVSSPSPDHSLLS
SMESLLAPLTTPAALSFNAADVVLSSGAMN
LVPDFTLVSLSVYEAPEGSACVATMRVGSA
GQVMTDVGVMFGNAIPTLDSVNLGIDGTVQ
CMFKVPSRSECGASLQRGNRLEVVGFYKL
ADRVCCSSNSLGFAFKDVEGAITERQLNAV
LHKLHLKNQQTQLAHSEPLQKQVLGGGAP
RNLLRQLRPLEDNLFARSQAVLQQCGMQH
TLNPIDLPIGPDGQISVSNVQSSGLTALHVA
AEFGWNKFVTQLVTAGASINQRDNFGNTAL
DWAFMSDQHVTCQLLQSAGGLFNMFAQPK
VTDDLQASFQQFTLSSPVLGQLPSPTTRIAS
SRSPRSPHLGSPLNFGGLSQNTSPPGQVLS
EIPEAQSSAQQSYQTQLQFLQQLQHHQLVE
QEQQQKQQQLLVQLMQQQQPQLQLQQQA
KQQAAHHNLVQHLLSMDMSASASGSSSRA
VSTSSLLSNASAVTSAPSMLPLDFTTTSDAH
LYSEPDLLALDLPSPNPSTLSMGELHF

>CoRunx2_recombinant_protein
PQSQPHHPQQPGMHQFQPPMPPAFNVGG
GHPQQQQVNSQHLSHPPHHQQPGQHGTP
LPTPMSATLHFNFPASARETAMLEESANVS
RCSMIVETQLTGLSQRAQNALNACTHSMLS
IGSCNEALIPLNEALWNLENAKRAFDAPTAG
FGAAYEELRQLYKNARESLVTTMDHLQNAI
HAQEKLHHQTTGGVPPSPGTLNRGGSSGL
GAGFHRLGLSGASTGSVQPSFGMTTAAAA
ASAQGFPVAVSGAAVGRPSIAGSDTGSFVV
PNRIGTASAMPGSTEPLVPTFAVELVEKSVR
ERLAPIGNASASSLFVTHFAFCRPSDKCMM
PADHLIAARDGVMLVGLFATNASAISEQLAH
TSLDTLATRSQSPSSSGGVISAAVHLTRILD
ATQISAGADSANMHHLQFLQLRTELGEIRIA
GPREQYAGMFDLLKVMRTFAGESAPPHAA
GSKPGLVLPGERPGLTIEPPSPHTAEAVATL
MNYAGTPRSVKRDAGMLEESASSESAASS
GSNKRIRMDA

>CoNF-κB_recombinant_protein
ESQYVEYTYLPAEAAVRNAELAARKRRRDD
SMRDFMDRFDGSDGGNGSGSGRGNNGG
HDGSDANNNGRGGGGGSSSSKGGDEPFN
FNSLIPMHQHKLHQLALSTVRAVQGFAASG
DARYLLALHRQLLAAPNENGDSPLHTAVAQ
GNLRSTMALLPLLAAEDLQSVNDMGETVLH
SAVIEKRAAIARLLLVAGADLGQSNARNFNR
NSLHYLARHGDRATAMAVFGVFGSAQAPP
ANTNTPAQAPAGETKPKPADLRLLARIQAQ
AIKALLACELETGATPAHLAIRGGHWHVFEA
CAKLAASAPIPKAAGSLLSMVAEKSSGHSLL
HSCVLANNEQAVRLLINLGASGNARDFGKN
TPLHLAARQGHIGIAALLVEAGATLSLNAVS
QTPLDVLTSEGSGLSRDQLRALVAVLRGEL
KYADMRGRPTLRMPTHAELHSTAAALTSAS
PGAVSLADFYAGKKASRSPAPLGASSSLLS
STGASAAGASAPTIAAVHAASATPVERTSM
NNDDDYVLLEKDAPYPVEQQPHGKRNKHS
HHRFTRSSHGSQDKDELKKDKDDPKKEKE
PKELSKFTLKEAFVDGTNFWELTRKFAGKK
KMASASTGEMEPLSPERPLSPTNAGSGAA
SPFNQAKEQVSPGAVPPTGLEKLVNKLMDA
SEATLSSQPAEAVTPEQKLAEKLEKLGLAPA
STTSAPPPHPKVAALNAQSVEDARKTSTHA
LYSVD

### Antibody production

Polyclonal antibodies were raised in rabbits by the Custom Antibody Service Unit (CAbS, CIBER-BBN, IQAC-CSIC) from the ICTS "NANBIOSIS". Rabbits were immunized monthly (6 times) using 100 µg/rabbit per immunization of CoRunx1, CoRunx2 or CoNF-κB recombinant proteins (conjugated with KLH) and affinity-purified from the final antiserums using CoRunx1, CoRunx2 or CoNF-κB recombinant proteins. The final concentrations of affinity-purified antibodies were 170 µg/mL, 350 µg/mL and 50 µg/mL. Antibodies were validated by Immunoprecipitation coupled to MS/MS and Western Blot.

### Antibody validation

For Western Blot analysis, we used *Capsaspora* total protein extracts and recombinant *Capsaspora Runx1, Runx2* and *NF-κB* proteins (see Protein expression and purification section). *Capsaspora* protein extracts were obtained from $5*10^8$-$3*10^{10}$ cells of the different cell stages (see *Cell strain and growth conditions*). Cells were scraped and harvested by centrifugation at 5.000g during 10 min at 4ºC. Cell pellets were washed twice with 20 mL 1X PBS (Sigma, #P5368-10 PAK), flash frozen using liquid Nitrogen and resuspended in the appropriate extraction buffer.

For total protein extraction, cells were resuspended in Total Extraction Buffer (50 mM Tris-HCl pH=8.8, 150 mM NaCl, 0.1% SDS, 5 mM EDTA, 1 mM EGTA, 1% Triton X-100, 1 mM MgCl2, 1 mM CaCl2, 1 mM DTT, 0.5 mM PMSF, cOmplete™ Mini EDTA-free protease inhibitor cocktail (#11 836 170 001, Roche) and PhosStop phosphatase inhibitor (#4906837001, Roche)) incubated 30 min incubation on ice and sonicated using Digital Sonifier sonicator (Model S-250D, BRANSON) 30''ON/30''OFF 20% amplitude 6-12 cycles. After sonication, samples were centrifuged at 20.000g during 30 min at 4ºC and the supernatant was collected and stored at -80ºC until further use.

Approximately 20 and 40 ng of *Capsaspora Runx1*, *Runx2* and *NF-κB* recombinant proteins were separated by SDS-PAGE and transferred to nitrocellulose membranes. Proteins were probed using anti-CoRunx1 (µg/mL), anti-CoRunx2 (µg/mL) and anti-CoNF-κB (50 µg/mL), detected with 1:2000 HRP-conjugated goat anti-rabbit IgG antibody (#P0448, Dako) and visualized with Supersignal WestPico Chemiluminescent substrate (#34078, ThermoFisherScientific) in a ChemiDoc™ Touch Gel Imaging System Transiluminator (#1708370, BIO-RAD).

### Immunofluorescence of *Capsaspora*

Cells collected from each timepoint were harvested at 4.500xg at 4ºC during 5 min, washed with 10 mL of 1X PBS (Sigma-Aldrich, #P5368-10PAK) and fixed with 10 mL 4% Formaldehyde (Sigma-Aldrich, #F8775-25mL) in 1X PBS during 15 min at RT in a rotor. After fixation, cells were washed with 10 mL 1X PBS and blocked for 1h at RT in a rotor with 2 mL Blocking solution (1% Bovine Serum Albumin (Sigma-Aldrich, #A3294-10G) and 0.1% Triton-X100 (Sigma-Aldrich, #X100-500ML) in 1X PBS).

Each sample was then divided into 200 µL aliquots and incubated o/n at 4ºC in a rotor with 1.5 µg/mL primary antibody (custom anti-CoRunx1, anti-CoRunx2 and anti-CoNF-κB). After incubation, samples were washed twice with 500 µL Blocking solution and incubated 1h at RT in a rotor with 1:2000 Alexa Fluor® 488 goat anti-rabbit IgG (Life Technologies 2 mg/mL, ref: A11008; lot: 1678787) secondary Antibody. Samples were washed once with 500 µL Blocking solution and once with 500 µL 1X PBS. Cells were then incubated with 1:100 Texas Red®-564 Phalloidin (Life Technologies, #T7471, lot:1463206) during 15 min at RT in a rotor, washed once with 1X PBS and resuspended in 20 µL of 1:100 5 mM DRAQ5 (Thermo Scientific, #62251) in 1X PBS. The preparation was overlaid with ProLon™ Gold antifade reagent (Invitrogen, #P36930, lot:1942345) over a pre-treated coverslide with Poly-L-Lysine (Sigma-Aldrich, #P4832) and sealed with nail polish.

Immunostained samples were imaged using a Leica TCS SP5 II inverted confocal microscope with a 63X immersion oil objective. Acquisition settings were adjusted using samples without primary antibody.

## Immunoprecipitation and Mass Spectrometry

Direct immunoprecipitation (IP) experiments were performed using 50 µL of Protein A beads (for rabbit IgG, Pure Proteome Millipore) per IP. Beads were first incubated with 2 uL preserum in Protein Extraction Buffer during 45 min at RT. Beads were washed twice with 500 µL PBS-Tween 0.1%. 2 mg of protein extracts from 4-day and 6-day adherent *Capsaspora* cultures were incubated with preserum:beads for 2.5 hours at 4ºC on a rotor. In parallel, 50 µL beads were incubated with 5 µL of final bleeding in Protein Extraction Buffer for 1 hour at 4ºC on a rotor. Beads were then washed twice as before. Precleared protein extracts from before were incubated with final-bleeding serum:beads overnight at 4ºC on a rotor, washed twice with 500 µL PBS-Tween 0.1%, washed three times with 200 mM Ammonium Bicarbonate (ABC) and resuspended in 60 µL Resuspension Buffer (6M Urea, 200 mM ABC). Immunocomplexes were reduced with 10 mM DTT during 1 hour at 37ºC in agitation, and alkylated with 20 mM Iodoacetamide (IAA) during 30 min at RT in the dark and under agitation. After alkylation, samples were diluted with 280 µL 200 mM ABC and digested on beads with 0.2 µg/µL Trypsin overnight at 37ºC in agitation. Beads were pulled-down and samples were acidified using 20 µL 100% Formic Acid. Samples were desalted using C18 stage tips (UltraMicroSpin Column, #SUM SS18V, The Nest Group, Inc., MA).

45% of each sample was analysed by LC-MS/MS using a 1-hour gradient in the Orbitrap Velos Pro. As a quality control, BSA controls were digested in parallel and ran between each sample to avoid carryover and assess the instrument performance. Samples were searched against *Capsaspora* database (September, 2015), using the search algorithm Mascot v2.5.1 (http://www.matrixscience.com/). Peptides

were filtered based on FDR, retaining only peptides showing an FDR lower than 5%.

### Chromatin Immunoprecipitation coupled with high-throughput sequencing (ChIPseq)

ChIPseq experiments were performed using custom antibodies against *CoRunx1, CoRunx2 and CoNF-κB*. Cells were scraped from $5*10^8$-$3*10^{10}$ cells of the different cell stages (see Cell strain and growth conditions) and harvested by centrifugation at 5.000g during 10 min at 4ºC. Cell pellets were washed twice with 20 mL 1X PBS and crosslinked in 1% Formaldehyde (Sigma-Aldrich, #F8775-4X25ML) in 1X PBS for 10 min at room temperature (RT). Crosslinking was quenched with 125 mM Glycine (Sigma-Aldrich, #50046-250G) for 5 min at RT. Cells were then harvested by centrifugation, washed twice as before and flash freezed with liquid Nitrogen. Cell pellets were resuspended in Lysis Buffer I (10 mM HEPES-KOH pH=7.9, 1.5 mM MgCl2, 10 mM KCl, 0.2% NP-40 and 0.5 mM PMSF, cOmplete™ Mini EDTA-free protease inhibitor cocktail and PhosStop phosphatase inhibitor), incubated on ice for 20 min and centrifuged at 8.500g during 10 min at 4ºC. Pelleted nuclei were resuspended in Lysis Buffer II (1%SDS, 10 mM EDTA, 50 mM Tris-HCl pH=8.1 and 0.5 mM PMSF, cOmplete™ Mini EDTA-free protease inhibitor cocktail and PhosStop phosphatase inhibitor), incubated 10 min on ice and sonicated 30''ON/30''OFF during 16 cycles using Diogenode Bioruptor Pico at 5ºC in order to generate ~200 bp fragments. Optimal sonication conditions for each stage were previously set up by testing a range of sonication cycles (from 12 to 22). Sonicated chromatin was transferred to eppendorf tubes, incubated on ice from 30 min to 1h and centrifuged at 20.000g for 20 min at 4ºC. Supernatants containing ready-to-use chromatin were kept at 4ºC until further use.

A 100 µL aliquot of each sample was reverse crosslinked for chromatin quality control analysis. 10% V of 5 M NaCl was added per sample and samples were incubated at 65ºC overnight. Samples were treated with 3 µL of RNAse A (SIGMA, # R6148-25ML) at 37ºC during 30 min and 8 µL of ≥10 mg/mL Proteinase-K (SIGMA, #P4850-1ML) at 50ºC during 1 hour. For DNA precipitation, 2 µL of 15 mg/mL GlycoBlue (Invitrogen, #AM9516) were added and 1/10 V 3M NaOAc, and 3V of cold 100% ETOH. Samples were incubated at -80ºC for minimum 30 min. Samples were centrifuged at 20.000g for 30 min at 4ºC. Pellets were washed twice with 1 mL cold 70% ETOH and air-dried before resuspension with 30 µL of distilled water. DNA samples were quantified using NANODROP and imaged on a 2% Agarose gel for fragment size validation.

For chromatin immunoprecipitation, we used an amount of chromatin equivalent to 60 µg of DNA per ChIP. 50 µL of Protein A beads (for rabbit IgG, Pure Proteome Millipore) per ChIP were transferred to a fresh tube (protein LoBind), washed twice with 500 µL of 1X PBS-Tween 0.1% and washed once with IP Buffer (16.7 mM Tris-HCl pH=8, 167 mM NaCl, 1.2 mM EDTA, 0.01% SDS, 1.1%

Triton X-100) plus Protease/Phosphatase Inhibitors. After washing, beads were blocked with 500 µL of blocking solution (IP Buffer, 5% BSA) during 1 hour at 4ºC on a rotor. Blocked beads were then resuspended in 1 mL IP Buffer and incubated with 10 µg of primary antibody for at least 4 hours at 4ºC on a rotor. 60 µg of DNA were added per ChIP and incubated overnight at 4ºC on a rotor. A no antibody negative control (mock ChIP) was performed per each timepoint. Immunocomplexes were washed once with 1 mL low salt Washing Buffer 1 (0.1% SDS, 1% Triton X-100, 2 mM EDTA, 20 mM Tris-HCl pH=8, 150 mM NaCl), once with 1 mL Washing Buffer 2 (0.1% SDS, 1% Triton X-100, 2 mM EDTA, 20 mM Tris-HCl pH=8, 500 mM NaCl), once with 1 mL low salt Washing Buffer 3 (0.25 M LiCl, 1% NP-40, 1% NaDOC, 1 mM EDTA, 10 mM Tris-HCl pH=8), three times with 1 mL TE Buffer (10 mM Tris-HCl pH=8, 1 mM EDTA) during 3 min at 4ºC on a rotor each. DNA complexes were eluted with 400 µL Elution Buffer (50 mM Tris-HCl pH=8, 10 mM EDTA) during 30 min at 65ºC, de-crosslinked overnight at 65ºC with 10% V 5M NaCl and treated with 6 µL of RNAse A (SIGMA, # R6148-25ML) at 37ºC during 30 min and 16 µL of ≥10 mg/mL Proteinase-K (SIGMA, #P4850-1ML) at 50ºC during 1 hour. Samples were purified using 1V Phenol:Chloroform:IAA (25:24:1) (#P3803-100ML, Sigma), centrifuged at 20.000g during 10 min at 4ºC and precipitated as before. The immunoprecipitated DNA was resuspended in 20 µL distilled water.

Samples were subjected to 50 bp paired-end deep sequencing in 1 sequencing Lane using the Illumina Hiseq2000 sequencer (high output mode) in the Ultrasequencing Unit at the Centre for Genomic Regulation (CRG). Between 9 and 16 M paired-end 50 bp reads were obtained per sample.

### ChIP-seq data analysis

ChIP-seq and input 50 bp paired-end Illumina sequencing reads were mapped into the *Capsaspora* reference genome (v.3) using Bowtie2 v.2.2.6 (Langmead et al., 2009) with -N 1 parameter (1 or 0 mismatches) and paired-end alignment mode. Duplicates reads were removed using samtools v.1.8 (Li et al., 2009) markdup with -r parameter. Peak calling was performed using MACS2 (Zhang et al., 2008) in paired-end mode (BAMPE) with a genome mappability of -g 24500000 as (Sebé-Pedrós et al., 2016a) and a q-value threshold of 0.01. High confidence peaks were selected according to relative fold-enrichment values thresholds set on the upper quartile distribution in each sample. Read density files were normalized by sequencing depth using deepTools v.3.1.0 (Ramírez et al., 2016) bamCoverage function with -e 200 (read extension, determined as the average of *d* parameters from MACS2 output) --scaleFactor 0.1 --normalizeUsing RPKM --binSize 1 parameters.

Bedtools toolset v.2.27.0 (Quinlan and Hall, 2010) was used for genome arithmetic analyses. Bedtools *intersect* tool was used to calculate overlaps between peaks and the different genomic features, as well as to

assign each *cis*-regulatory site to a particular gene. Bedtools *closest* tool was used to search for overlapping features, reporting the nearest genomic distance to TSS and ATAC-seq signals.

Gene ontology (GO) functional enrichment analyses were calculated using Ontologizer v.2.1. (Bauer et al., 2008) with the Parent-Child-Union method and a p-value cut-off of 0.05 from a previously generated gene ontology of 8,637 *Capsaspora* genes (Sebé-Pedrós et al., 2013b). Enrichment of GO terms mapped with mock (control) ChIP peaks were used to discard random enrichment genes in sample analysis.
*De novo* motif enrichment analysis of significant peaks in regulatory sites was performed using HOMER software v.4.10 (Heinz et al., 2010), with default parameters. *homerTools extract* tool was used to extract high confident peak regions of sequence from genomic FASTA files.

### Protein Binding Microarrays

PBM experiments were assayed over *Capsaspora CoRunx1, CoRunx2* and *CoNF-κB*. *CoRunx1* Pfam-defined DNA-binding domain (DBD) sequence plus a flanking region of approximately 50 amino acids was PCR amplified from cDNA using primers 3 and 4. *CoRunx1* DBD plus a flanking region of approximately 300 amino acids was PCR amplified from cDNA using primers 5 and 6 **(Table S3).** *CoRunx2* DBD plus a flanking region of approximately 50 amino acids was PCR amplified from cDNA using primers 10 and 11. *CoRunx2* Pfam-defined DBD plus a flanking region of 300 amino acids was PCR amplified from cDNA using primers 12 and 13 **(Table S3)**. *CoNF-κB* DBD plus a flanking region of approximately 50 amino acids was PCR amplified from cDNA using primers 15 and 16. *CoNF-κB* DBD plus a flanking region of 300 amino acids was PCR amplified from cDNA using primers 17 and 18 **(Table S3)**. The resulting cassettes were digested using AscI and SbfI restriction enzymes and cloned at the N-terminal of glutathione S-transferase (GST) in the pTH6838 expression vector linearized at AscI and SbfI sites. All plasmids DNA were obtained using the plasmid GenElute™ Plasmid Miniprep Kit (Sigma, #PLN350-1KT) and resuspended at an approximate concentration of 500 ng/µL in distilled water.

PBM laboratory methods were identical to those described previously (Lam et al., 2011; Weirauch et al., 2013). Each plasmid was analysed in duplicate on two different arrays with differing probe sequences. PBM data processing was performed as in (Berger et al., 2006; Berger and Bulyk, 2006; Weirauch et al., 2014b, 2013). Motif logos were created aligning the top ten scoring mean E-scores by WebLogo (Crooks et al., 2004; Schneider and Stephens, 1990).

## ACKNOWLEDGMENTS

We thank the Advanced Light Microscopy Unit of the CRG for support on confocal images acquisition. We thank Elena Casacuberta, Arnau Sebé-Pedrós and Omaya Dudin for fruitful discussions on experiments and data analysis.

## COMPETING INTERESTS

The authors declare no competing or financial interests.

## AUTHOR CONTRIBUTIONS

Conceptualisation: N.R.-R., S.R.N., I.R.-T.; Methodology: N.R.-R., S.R.N., M.M.L., Y.A; Formal analysis: N.R.-R., S.R.N., M.M.L.; Writing - original draft: N.R.-R.; Writing - review & editing: N.R.-R., M.M.L., I.R.-T.; Supervision: I.R.-T.; Funding acquisition: I.R.-T.

## FUNDING

This work was supported by a European Research Council Consolidator Grant (ERC-2012-Co-616960), and grants (BFU2014-57779-P and BFU2017-90114-P) from Ministerio de Economía y Competitividad (MINECO), Agencia Estatal de Investigación (AEI), and Fondo Europeo de Desarrollo Regional (FEDER) 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) and "Formación del Profesorado Universitario (FPU13/01840)" from Ministerio de Educación, Cultura y Deporte (MECD) PhD fellowship to N.R.-R.

## DATA AVAILABILITY

Supplementary Files (1-6) and Tables (S1-S2) have been deposited to *Figshare:* https://figshare.com/s/afc2e5c3d6d5661adf8c

## REFERENCES

**Altschul, S.F., 1993.** A protein alignment scoring system sensitive at all evolutionary distances. J. Mol. Evol. 36, 290–300.

**Altschul, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J., 1990.** Basic local alignment search tool. J. Mol. Biol. 215, 403–10.

**Arenas-Mena, C., 2017.** The origins of developmental gene regulation. Evol. Dev. 19, 96–107.

**Bauer, S., Grossmann, S., Vingron, M., Robinson, P.N., 2008.** Ontologizer 2.0 - A multifunctional tool for GO term enrichment analysis and data exploration. Bioinformatics 24, 1650–1651.

**Berger, M.F., Badis, G., Gehrke, A.R., Talukder, S., Philippakis, A.A., Peña-Castillo, L., Alleyne, T.M., Mnaimneh, S., Botvinnik, O.B., Chan, E.T., Khalid, F., Zhang, W., Newburger, D., Jaeger, S.A., Morris, Q.D., Bulyk, M.L., Hughes, T.R., 2008.** Variation in Homeodomain DNA Binding Revealed by High-Resolution Analysis of Sequence Preferences. Cell 133, 1266–1276.

**Berger, M.F., Bulyk, M.L., 2006.** Protein Binding Microarrays (PBMs) for the Rapid, High-Throughput Characterization of the Sequence Specificities of DNA Binding Proteins. Methods Mol. Biol. 338, 245–260.

**Berger, M.F., Bulyk, M.L., 2009.** Universal protein-binding microarrays for the comprehensive characterization of the

DNA-binding specificities of transcription factors. Nat. Protoc. 4, 393–411.

**Berger, M.F., Philippakis, A.A., Qureshi, A.M., He, F.S., Iii, P.W.E., Bulyk, M.L., 2006.** Compact, universal DNA microarrays to comprehensively determine transcription-factor binding site specificities 24, 1429–1435.

**Bowers, S.R., Calero-Nieto, F.J., Valeaux, S., Fernandez-Fuentes, N., Cockerill, P.N., 2010.** *Runx1* binds as a dimeric complex to overlapping *Runx1* sites within a palindromic element in the human GM-CSF enhancer. Nucleic Acids Res. 38, 6124–6134.

**Brown, M.W., Spiegel, F.W., Silberman, J.D., 2009.** Phylogeny of the 'forgotten' cellular slime mold, Fonticula alba, reveals a key evolutionary branch within Opisthokonta. Mol. Biol. Evol. 26, 2699–2709.

**Byon, H.C., Javed, A., Dai, Q., Kappes, J.C., Clemens, T.L., Darley-Usmar, V.M., McDonald, J.M., Chen, Y., 2008.** Oxidative stress induces vascular calcification through modulation of the osteogenic transcription factor *Runx2* by *AKT* signaling. J. Biol. Chem. 283, 15319–15327.

**Cai, X., Gao, L., Teng, L., Ge, J., Oo, Z.M., Kumar, A.R., Gilliland, D.G., Mason, P.J., Tan, K., Speck, N.A., 2015.** Runx1 Deficiency Decreases Ribosome Biogenesis and Confers Stress Resistance to Hematopoietic Stem and Progenitor Cells. Cell Stem Cell 17, 165–177.

**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.

**Chiquet-Ehrismann, R., Chiquet, M., 2003.** *Tenascins*: Regulation and putative functions during pathological stress. J. Pathol. 200, 488–499.

**Coffman, J.A., 2003.** *Runx* transcription factors and the developmental balance between cell proliferation and differentiation. Cell Biol. Int. 27, 315–324.

**Criscuolo, A., Gribaldo, S., 2010.** BMGE (Block Mapping and Gathering with Entropy): A new software for selection of phylogenetic informative regions from multiple sequence alignments. BMC Evol. Biol. 10.

**Crooks, G., Hon, G., Chandonia, J., Brenner, S., 2004.** WebLogo: a sequence logo generator. Genome Res. 14, 1188–1190.

**Davidson, E.H., Erwin, D.H., 2006.** Gene regulatory networks and the evolution of animal body plans. Science. 311, 796–797.

**de Mendoza, A., Sebé-Pedrós, A., Sestak, M.S., Matejcic, M., Torruella, G., Domazet-Loso, T., Ruiz-Trillo, I., 2013.** Transcription factor evolution in eukaryotes and the assembly of the regulatory toolkit in multicellular lineages. Proc. Natl. Acad. Sci. U. S. A. 4858–4866.

**de Mendoza, A., Suga, H., Permanyer, J.,**

**Irimia, M., Ruiz-Trillo, I., 2015.** Complex transcriptional regulation and independent evolution of fungal-like traits in a relative of animals. Elife 4, 1–26.

**Deepak, V., Zhang, Z., Meng, L., Zeng, X., Liu, W., 2011.** Reduced activity and cytoplasmic localization of *Runx2* is observed in C3h10t1/2 cells overexpressing Tbx3. Cell Biochem. Funct. 29, 348–350.

**Degnan, B.M., Vervoort, M., Larroux, C., Richards, G.S., 2009.** Early evolution of metazoan transcription factors. Curr. Opin. Genet. Dev. 19, 591–599.

**Eddy, S.R., 1998.** Profile hidden Markov models. Bioinformatics 14, 755–763.

**Fairclough, S.R., Chen, Z., Kramer, E., Zeng, Q., Young, S., Robertson, H.M., Begovic, E., Richter, D.J., Russ, C., Westbrook, M.J., Manning, G., Lang, B.F., Haas, B., Nusbaum, C., King, N., 2013.** Premetazoan genome evolution and the regulation of cell differentiation in the choanoflagellate Salpingoeca rosetta.

**Finn, R.D., Bateman, A., Clements, J., Coggill, P., Eberhardt, R.Y., Eddy, S.R., Heger, A., Hetherington, K., Holm, L., Mistry, J., Sonnhammer, E.L.L., Tate, J., Punta, M., 2014.** Pfam: The protein families database. Nucleic Acids Res. 42, 222–230.

**Gish, W., States, D.J., 1993.** Identification of protein coding regions by database similarity search. Nat. Genet. 3, 266–272.

**Hayden, M.S., Ghosh, S., 2004.** Signaling to *NF-κB.* Genes Dev. 18, 2195–2224.

**Hayden, M.S., West, A.P., Ghosh, S., 2006.** *NF-κB* and the immune response. Oncogene 25, 6758–6780.

**Hehenberger, E., Tikhonenkov, D. V., Kolisko, M., del Campo, J., Esaulov, A.S., Mylnikov, A.P., Keeling, P.J., 2017.** Novel Predators Reshape Holozoan Phylogeny and Reveal the Presence of a Two-Component Signaling System in the Ancestor of Animals. Curr. Biol. 27, 2043–2050.e6.

**Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y.C., Laslo, P., Cheng, J.X., Murre, C., Singh, H., 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.

**Huguet, C., Crepieux, P., Laudet, V., 1997.** *Rel/NF-κB* transcription factors and *IκB* inhibitors: Evolution from a unique common ancestor. Oncogene 15, 2965–2974.

**Jolma, A., Yan, J., Whitington, T., Toivonen, J., Nitta, K.R., Rastas, P., Morgunova, E., Enge, M., Taipale, M., Wei, G., Palin, K., Vaquerizas, J.M., Vincentelli, R., Luscombe, N.M., Hughes, T.R., Lemaire, P., Ukkonen, E., Kivioja, T., 2013.** DNA-Binding Specificities of Human Transcription Factors. Cell 152, 327–339.

**Kalyaanamoorthy, S., Minh, B.Q., Wong, T.K.F., Von Haeseler, A., Jermiin, L.S., 2017.** ModelFinder: Fast model selection for accurate phylogenetic

estimates. Nat. Methods 14, 587–589.

**Katoh, K., Kuma, K.I., Toh, H., Miyata, T., 2005.** MAFFT version 5: Improvement in accuracy of multiple sequence alignment. Nucleic Acids Res. 33, 511–518.

**Katoh, K., Misawa, K., Kuma, K., Miyata, T., 2002.** MAFFT : a novel method for rapid multiple sequence alignment based on fast Fourier transform 30, 3059–3066.

**Katoh, K., Standley, D.M., 2013.** MAFFT Multiple Sequence Alignment Software Version 7 : Improvements in Performance and Usability Article Fast Track 30, 772–780.

**Kim, H.R., Oh, B.C., Choi, J.K., Bae, S.C., 2008.** *Pim-1 kinase* phosphorylates and stabilizes *Runx3* and alters its subcellular localization. J. Cell. Biochem. 105, 1048–1058.

**Kim, W., Barron, D. a, San Martin, R., Chan, K.S., Tran, L.L., Yang, F., Ressler, S.J., Rowley, D.R., 2014.** *Runx1* is essential for mesenchymal stem cell proliferation and myofibroblast differentiation. Proc. Natl. Acad. Sci. U. S. A. 111, 16389–94.

**King, N., 2004.** The unicellular ancestry of animal development. Dev. Cell 7, 313–25.

**King, N., Westbrook, M.J., Young, S.L., Kuo, A., Abedin, M., Chapman, J., Fairclough, S., Hellsten, U., Isogai, Y., Letunic, I., Marr, M., Pincus, D., Putnam, N., Rokas, A., Wright, K.J., Zuzow, R., Dirks, W., Good, M., Goodstein, D., Lemons, D., Li, W., Lyons, J.B., Morris, A., Nichols, S., Richter, D.J., Salamov, A., Bork, P., Lim, W.A., Manning, G., Miller, W.T., McGinnis, W., Shapiro, H., Tjian, R., Grigoriev, I. V., Rokhsar, D., 2008.** The genome of the choanoflagellate *Monosiga brevicollis* and the origin of metazoans. Nature 451, 783–788.

**Knoll, A.H., 2011.** The Multiple Origins of Complex Multicellularity.

**Lam, K.N., Bakel, H. Van, Cote, A.G., Ven, A. Van Der, Hughes, T.R., 2011.** Sequence specificity is obtained from the majority of modular *C2H2* zinc-finger arrays. Nucleic Acids Res. 39, 4680–4690.

**Langmead, B., Trapnell, C., Pop, M., Salzberg, S.L., 2009.** Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. 10.

**Larroux, C., Luke, G.N., Koopman, P., Rokhsar, D.S., Shimeld, S.M., Degnan, B.M., 2008.** Genesis and expansion of metazoan transcription factor gene classes. Mol. Biol. Evol. 25, 980–996.

**Levanon, D., Goldstein, R.E., Bernstein, Y., Tang, H., Goldenberg, D., Stifani, S., Paroush, Z., Groner, Y., 1998.** Transcriptional repression by *AML1* and *LEF-1* is mediated by the *TLE/Groucho* corepressors. Proc. Natl. Acad. Sci. U. S. A. 95, 11590–11595.

**Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., 2009.** The Sequence Alignment/Map format and SAMtools. Bioinformatics 25, 2078–

2079.

**Li, H., Johnson, A.D., 2010.** Evolution of transcription networks-lessons from yeasts. Curr. Biol. 20, R746–R753.

**López-Escardó, D., López-García, P., Moreira, D., Ruiz-Trillo, I., Torruella, G., 2018.** Parvularia atlantis gen. et sp. nov., a Nucleariid Filose Amoeba (Holomycota, Opisthokonta). J. Eukaryot. Microbiol. 65, 170–179.

**Macian, F., 2005.** NFAT proteins: Key regulators of T-cell development and function. Nat. Rev. Immunol. 5, 472–484.

**Mandoli, A., Singh, A.A., Jansen, P.W.T.C., Wierenga, A.T.J., Riahi, H., Franci, G., Prange, K., Saeed, S., Vellenga, E., Vermeulen, M., Stunnenberg, H.G., Martens, J.H.A., 2014.** *CBFβ-MYH11/Runx1* together with a compendium of hematopoietic regulators, chromatin modifiers and basal transcription factors occupies self-renewal genes in inv(16) acute myeloid leukemia. Leukemia 28, 770–778.

**Mercurio, F., Manning, A.M., 1999.** *NF-κB* as a primary regulator of the stress response. Oncogene 18, 6163–6171.

**Michaud, J., Simpson, K.M., Escher, R., Buchet-Poyau, K., Beissbarth, T., Carmichael, C., Ritchie, M.E., Schütz, F., Cannon, P., Liu, M., Shen, X., Ito, Y., Raskind, W.H., Horwitz, M.S., Osato, M., Turner, D.R., Speed, T.P., Kavallaris, M., Smyth, G.K., Scott, H.S., 2008.** Integrative analysis of *Runx1* downstream pathways and target genes. BMC Genomics 9, 1–17.

**Mikenberg, I., Widera, D., Kaus, A., Kaltschmidt, B., Kaltschmidt, C., 2007.** Transcription factor *NF-κB* is transported to the nucleus via cytoplasmic dynein/dynactin motor complex in hippocampal neurons. PLoS One 2.

**Minh, B.Q., Nguyen, M.A.T., Von Haeseler, A., 2013.** Ultrafast approximation for phylogenetic bootstrap. Mol. Biol. Evol. 30, 1188–1195.

**Mistry, J., Bateman, A., Finn, R.D., 2007.** Predicting active site residue annotations in the Pfam database. BMC Bioinformatics 8, 1–14.

**Morgan, M.J., Liu, Z., 2011.** Crosstalk of reactive oxygen species and *NF-κB* signaling. Cell Res. 21, 103–115.

**Nguyen, L.T., Schmidt, H.A., Von Haeseler, A., Minh, B.Q., 2015.** IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol. Biol. Evol. 32, 268–274.

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

**Oeckinghaus, A., Ghosh, S., 2009.** The *NF-κB* Family of Transcription Factors and Its Regulation. Cold Spring Harb. Perspect. Biol. 1, a000034–a000034.

**Parra-Acero, H., Ros-Rocher, N., Perez-Posada, A., Kożyczkowska, A., Sánchez-Pons, N., Nakata, A., Suga, H., Najle, S.R., Ruiz-Trillo, I., 2018.**

Transfection of *Capsaspora owczarzaki*, a close unicellular relative of animals. Development 145, dev162107.

**Peter, I.S., Davidson, E.H., 2011.** Evolution of gene regulatory networks controlling body plan development. Cell 144, 970–985.

**Price, M.N., Dehal, P.S., Arkin, A.P., 2009.** Fasttree: Computing large minimum evolution trees with profiles instead of a distance matrix. Mol. Biol. Evol. 26, 1641–1650.

**Price, M.N., Dehal, P.S., Arkin, A.P., 2010.** FastTree 2 - Approximately maximum-likelihood trees for large alignments. PLoS One 5.

**Punta, M., Coggill, P.C., Eberhardt, R.Y., Mistry, J., Tate, J., Boursnell, C., Pang, N., Forslund, K., Ceric, G., Clements, J., Heger, A., Holm, L., Sonnhammer, E.L.L., Eddy, S.R., Bateman, A., Finn, R.D., 2012.** Pfam: The protein families database. Nucleic Acids Res. 40, 290–301.

**Quinlan, A.R., Hall, I.M., 2010.** BEDTools: A flexible suite of utilities for comparing genomic features. Bioinformatics 26, 841–842.

**Ramírez, F., Ryan, D.P., Grüning, B., Bhardwaj, V., Kilpert, F., Richter, A.S., Heyne, S., Dündar, F., Manke, T., 2016.** deepTools2: a next generation web server for deep-sequencing data analysis. Nucleic Acids Res. 44, W160–W165.

**Rennert, J., Coffman, J.A., Mushegian, A.R., Robertson, A.J., 2003.** The evolution of *Runx* genes I. A comparative study of sequences from phylogenetically diverse model organisms 11, 1–11.

**Richter, D.J., Fozouni, P., Eisen, M.B., King, N., 2018.** Gene family innovation, conservation and loss on the animal stem lineage. Elife 7, 1–43.

**Richter, D.J., King, N., 2013.** The Genomic and Cellular Foundations of Animal Origins. Annu. Rev. Genet. 47, 509–537.

**Robertson, A.J., Larroux, C., Degnan, B.M., Coffman, J. a, 2009.** The evolution of *Runx* genes II. The C-terminal *Groucho* recruitment motif is present in both eumetazoans and homoscleromorphs but absent in a haplosclerid demosponge. BMC Res. Notes 2, 59.

**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., Gray, M.W., 2007.** The origins of multicellularity: a multi-taxon genome initiative. Trends Genet. 23, 113–118.

**Ruiz-Trillo, I., Roger, A.J., Burger, G., Gray, M.W., Lang, B.F., 2008.** A phylogenomic investigation into the origin of Metazoa. Mol. Biol. Evol. 25, 664–672.

**Schindelin, J., Arganda-carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., Schmid, B., Tinevez, J., White, D.J., Hartenstein, V., Eliceiri, K.,**

**Tomancak, P., Cardona, A., 2012.** Fiji: an open-source platform for biological-image analysis 9.

**Schneider, T.D., Stephens, R.M., 1990.** Sequence logos: a new way to display consensus sequences. Nucleic Acids Res. 18, 6097–6100.

**Sebé-Pedrós, A., Ariza-Cosano, A., Weirauch, M.T., Leininger, S., Yang, A., Torruella, G., 2013a.** Early evolution of the *T-box* transcription factor family. Proc. Natl. Acad. Sci. U. S. A. 110, 16050–16055.

**Sebé-Pedrós, A., Ballaré, C., Parra-acero, H., Chiva, C., Tena, J.J., Sabido, E., Gómez-Skarmeta, J.L., Di Croce, L., Ruiz-Trillo, I., 2016a.** The Dynamic Regulatory Genome of Capsaspora and the Origin of Animal Multicellularity 1224–1237.

**Sebé-Pedrós, A., de Mendoza, A., Lang, B.F., Degnan, B.M., Ruiz-Trillo, I., 2011.** Unexpected repertoire of metazoan transcription factors in the unicellular holozoan *Capsaspora owczarzaki.* Mol. Biol. Evol. 28, 1241–54.

**Sebé-Pedrós, A., Degnan, B.M., Ruiz-trillo, I., 2017.** The origin of Metazoa: a unicellular perspective. Nat. Publ. Gr. 18, 498–512.

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

S**ebé-Pedrós, A., Peña, M.I., Capella-Gutiérrez, S., Antó, M., Gabaldón, T., Ruiz-Trillo, I., Sabidó, E., 2016b.** High-Throughput Proteomics Reveals the Unicellular Roots of Animal Phosphosignaling and Cell Differentiation. Dev. Cell 39, 186–197.

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

**Söding, J., 2005.** Protein homology detection by HMM-HMM comparison. Bioinformatics 21, 951–960.

**Sorrells, T.R., Johnson, A.D., 2015.** Making sense of transcription networks. Cell 161, 714–723.

**Stibbs, H.H., Owczarzak, A., Bayne, C.J., DeWan, P., 1979.** Schistosome sporocyst-killing amoebae isolated from *Biomphalaria glabrata.* J. Invertebr. Pathol. 33, 159–170.

**Sullivan, J.C., Sher, D., Eisenstein, M., Shigesada, K., Reitzel, A.M., Marlow, H., Levanon, D., Groner, Y., Finnerty, J.R., Gat, U., 2008.** The evolutionary origin of the *Runx/CBFβ* transcription factors - Studies of the most basal metazoans. BMC Evol. Biol. 8, 1–20.

**Tay, S., Hughey, J.J., Lee, T.K., Lipniacki, T., Quake, S.R., Covert, M.W., 2010.** Single-cell *NF-κB* dynamics reveal digital activation and analogue information processing. Nature 466, 267–271.

**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., Paley, R., Roger, A.J., Sitjà-Bobadilla, A., Donachie, S., Ruiz-Trillo, I., 2015.** Phylogenomics Reveals Convergent Evolution of Lifestyles in Close Relatives of Animals and Fungi. Curr. Biol. 25, 2404–2410.

**Vanderauwera, S., Suzuki, N., Miller, G., van de Cotte, B., Morsa, S., Ravanat, J.-L., Hegie, A., Triantaphylides, C., Shulaev, V., Van Montagu, M.C.E., Van Breusegem, F., Mittler, R., 2011.** Extranuclear protection of chromosomal DNA from oxidative stress. Proc. Natl. Acad. Sci. U. S. A. 108, 1711–1716.

**Webster, G.A., Perkins, N.D., 1999.** Transcriptional Cross Talk between *NF-κB* and *p53*. Mol. Cell. Biol. 19, 3485–3495.

**Weirauch, M.T., Cote, A., Norel, R., Annala, M., Zhao, Y., Riley, T.R., Saez-rodriguez, J., Cokelaer, T., Vedenko, A., Talukder, S., Consortium, D., Bussemaker, H.J., Morris, Q.D., Bulyk, M.L., Stolovitzky, G., Hughes, T.R., 2013.** Evaluation of methods for modeling transcription factor sequence specificity. Nat. Biotechnol. 31, 126–134.

**Weirauch, M.T., Yang, A., Albu, M., Cote, A.G., Montenegro-montero, A., Drewe, P., Najafabadi, H.S., Lambert, S.A., Mann, I., Cook, K., Zheng, H., Goity, A., Bakel, H. Van, Lozano, J., Galli, M., Lewsey, M.G., Huang, E., Mukherjee, T., 2014a.** Determination and Inference of Eukaryotic Transcription Factor Sequence Specificity. Cell 158, 1431–1443.

**Weirauch, M.T., Yang, A., Albu, M., Cote, A.G., Montenegro-Montero, A., Drewe, P., Najafabadi, H.S., Lambert, S.A., Mann, I., Cook, K., Zheng, H., Goity, A., Bakel, H. Van, Lozano, J., Galli, M., Lewsey, M.G., Huang, E., Mukherjee, T., 2014b.** Determination and Inference of Eukaryotic Transcription Factor Sequence Specificity. Cell 158, 1431–1443.

**Wheeler, J.C., Shigesada, K., Peter Gergen, J., Ito, Y., 2000.** Mechanisms of transcriptional regulation by Runt domain proteins. Semin. Cell Dev. Biol. 11, 369–375.

**Wu, H., Whitfield, T.W., Gordon, J.A.R., Dobson, J.R., Tai, P.W.L., van Wijnen, A.J., Stein, J.L., Stein, G.S., Lian, J.B., 2014.** Genomic occupancy of *Runx2* with global expression profiling identifies a novel dimension to control of osteoblastogenesis. Genome Biol. 15, 1–17.

**Yarmus, M., Woolf, E., Bernstein, Y., Fainaru, O., Negreanu, V., Levanon, D., Groner, Y., 2006.** Groucho/Transducin-like Enhancer-of-split (TLE)-dependent and -independent transcriptional regulation by *Runx3*. Proc. Natl. Acad. Sci. U. S. A. 103, 7384–7389.

**Young, D.W., Hassan, M.Q., Yang, X.-Q., Galindo, M., Javed, A., Zaidi, S.K., Furcinitti, P., Lapointe, D., Montecino, M., Lian, J.B., Stein, J.L., van Wijnen, A.J., Stein, G.S., 2007.** Mitotic retention of gene expression patterns by the cell

fate-determining transcription factor *Runx2*. Proc. Natl. Acad. Sci. U. S. A. 104, 3189–3194.

**Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nussbaum, C., Myers, R.M., Brown, M., Li, W., Shirley, X.S., 2008.** Model-based analysis of ChIP-Seq (MACS). Genome Biol. 9.

## SUPPLEMENTARY FIGURES

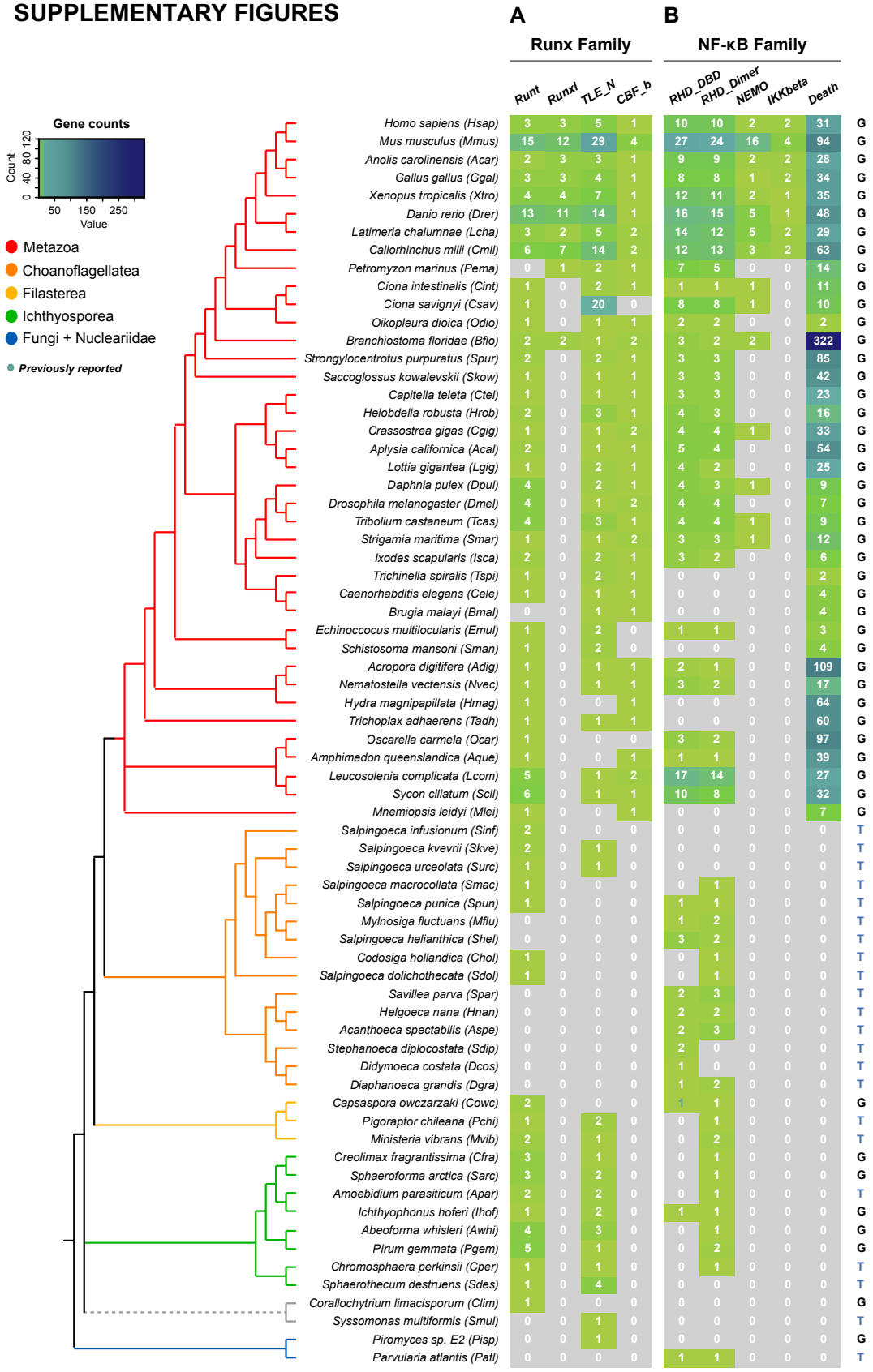
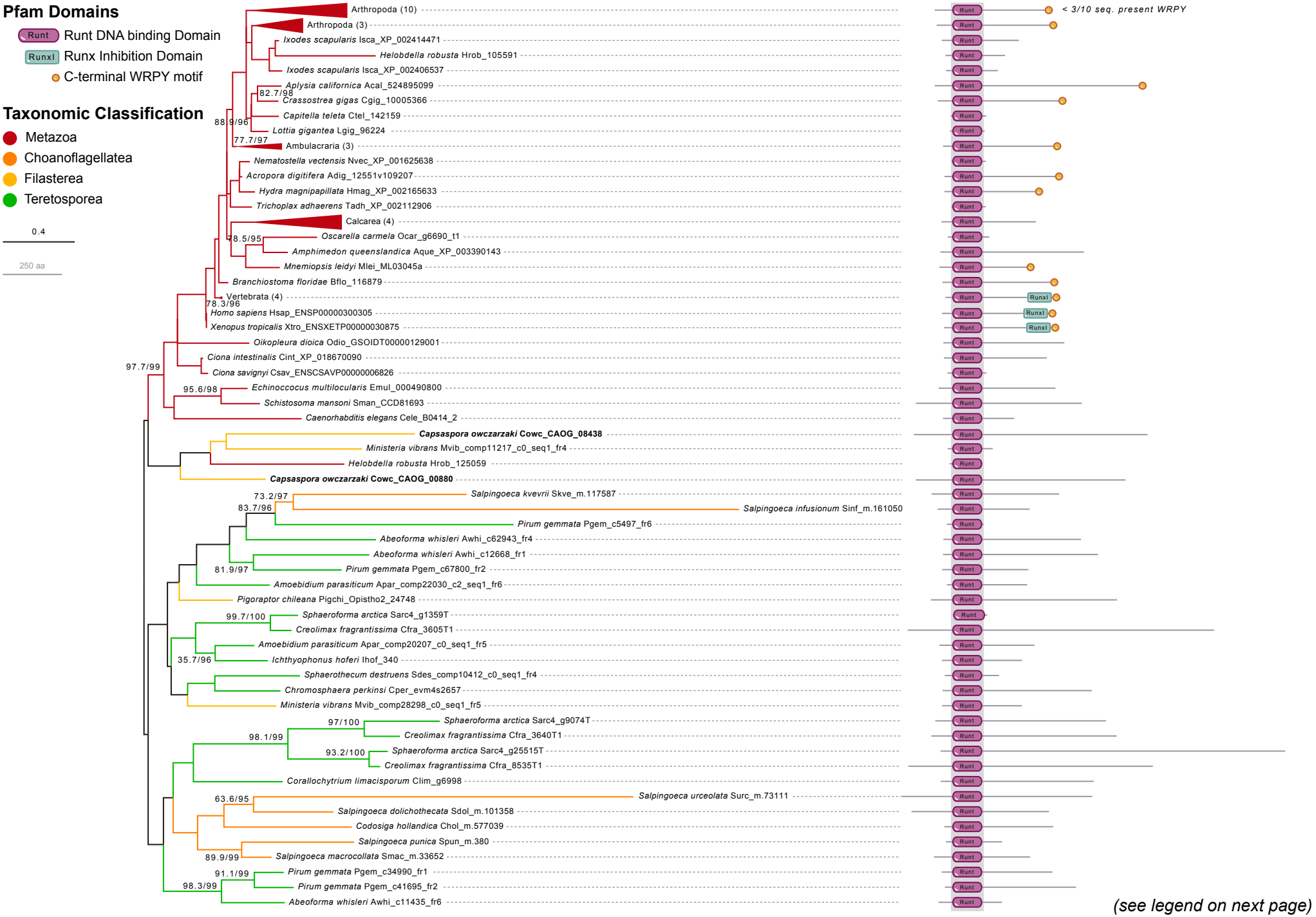



**Figure S1. *Runx* and *NF-κB* TF families and related partners domain presence in eukaryotes.** Number of domains identified by HMMsearch. Letters on the right depict the data source: genome (G) or transcriptome (T).Taxa are color-coded according to their taxonomic assignment (indicated in the upper-left legend).




(see legend on next page)

**Figure S2. IQ-Tree of Runt DNA-binding Domain and domain architecture.** Taxa are color-coded according to taxonomic assignment (indicated in the upper right legend).

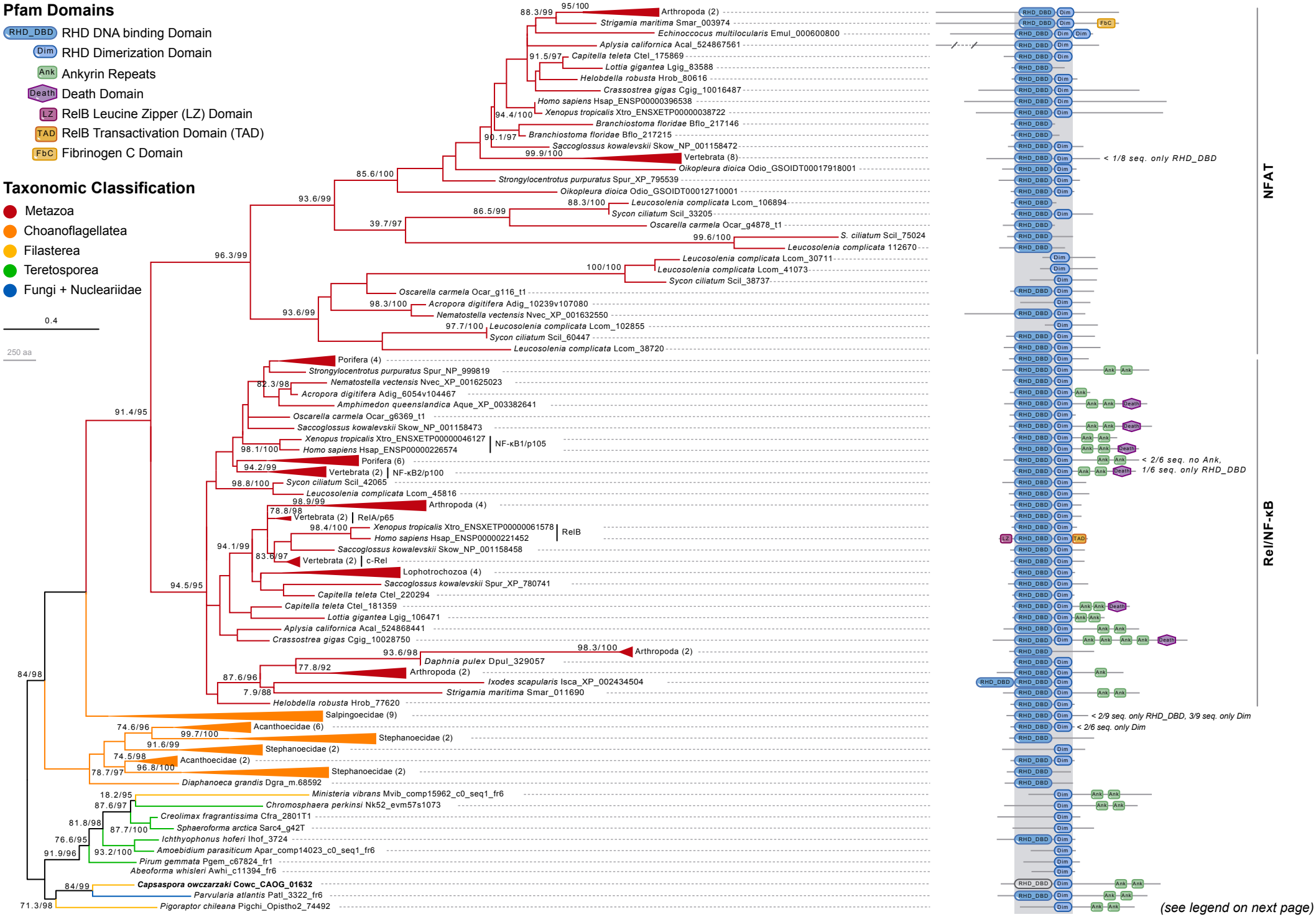

Pfam Domains
RHD_DBD RHD DNA binding Domain
Dim RHD Dimerization Domain
Ank Ankyrin Repeats
Death Death Domain
LZ RelB Leucine Zipper (LZ) Domain
TAD RelB Transactivation Domain (TAD)
FbC Fibrinogen C Domain
Taxonomic Classification
Metazoa
Choanoflagellatea
Filasterea
Teretosporea
Fungi + Nucleariidae
0.4
250 aa
Arthropoda (2)
Strigamia maritima Smar_003974
Echinococcus multilocularis Emul_000600800
Aplysia californica Acal_524867561
Capitella teleta Ctel_175869
Lottia gigantea Lgig_83588
Helobdella robusta Hrob_80616
Crassostrea gigas Cgig_10016487
Homo sapiens Hsap_ENSP00000396538
Xenopus tropicalis Xtro_ENSXETP00000038722
Branchiostoma floridae Bflo_217146
Branchiostoma floridae Bflo_217215
Saccoglossus kowalevskii Skow_NP_001158472
Vertebrata (8)
Oikopleura dioica Odio_GSOIDT00017918001
Strongylocentrotus purpuratus Spur_XP_795539
Oikopleura dioica Odio_GSOIDT00012710001
Leucosolenia complicata Lcom_106894
Sycon ciliatum Scil_33205
Oscarella carmela Ocar_g4878_t1
S. ciliatum Scil_75024
Leucosolenia complicata 112670
Leucosolenia complicata Lcom_30711
Leucosolenia complicata Lcom_41073
Sycon ciliatum Scil_38737
Oscarella carmela Ocar_g116_t1
Acropora digitifera Adig_10239v107080
Nematostella vectensis Nvec_XP_001632550
Leucosolenia complicata Lcom_102855
Sycon ciliatum Scil_60447
Leucosolenia complicata Lcom_38720
Porifera (4)
Strongylocentrotus purpuratus Spur_NP_999819
Nematostella vectensis Nvec_XP_001625023
Acropora digitifera Adig_6054v104467
Amphimedon queenslandica Aque_XP_003382641
Oscarella carmela Ocar_g6369_t1
Saccoglossus kowalevskii Skow_NP_001158473
Xenopus tropicalis Xtro_ENSXETP00000046127
Homo sapiens Hsap_ENSP00000226574
NF-κB1/p105
Porifera (6)
Vertebrata (2)
NF-κB2/p100
Sycon ciliatum Scil_42065
Leucosolenia complicata Lcom_45816
Arthropoda (4)
Vertebrata (2)
RelA/p65
Xenopus tropicalis Xtro_ENSXETP00000061578
Homo sapiens Hsap_ENSP00000221452
RelB
Saccoglossus kowalevskii Skow_NP_001158458
Vertebrata (2)
c-Rel
Lophotrochozoa (4)
Saccoglossus kowalevskii Spur_XP_780741
Capitella teleta Ctel_220294
Capitella teleta Ctel_181359
Lottia gigantea Lgig_106471
Aplysia californica Acal_524868441
Crassostrea gigas Cgig_10028750
Arthropoda (2)
Daphnia pulex Dpul_329057
Arthropoda (2)
Ixodes scapularis Isca_XP_002434504
Strigamia maritima Smar_011690
Helobdella robusta Hrob_77620
Salpingoecidae (9)
Acanthoecidae (6)
Stephanoecidae (2)
Stephanoecidae (2)
Acanthoecidae (2)
Stephanoecidae (2)
Diaphanoeca grandis Dgra_m.68592
Ministeria vibrans Mvib_comp15962_c0_seq1_fr6
Chromosphaera perkinsi Nk52_evm57s1073
Creolimax fragrantissima Cfra_2801T1
Sphaeroforma arctica Sarc4_g42T
Ichthyophonus hoferi Ihof_3724
Amoebidium parasiticum Apar_comp14023_c0_seq1_fr6
Pirum gemmata Pgem_c67824_fr1
Abeoforma whisleri Awhi_c11394_fr6
Capsaspora owczarzaki Cowc_CAOG_01632
Parvularia atlantis Patl_3322_fr6
Pigoraptor chileana Pigchi_Opistho2_74492
< 1/8 seq. only RHD_DBD
< 2/6 seq. no Ank, 1/6 seq. only RHD_DBD
< 2/9 seq. only RHD_DBD, 3/9 seq. only Dim
< 2/6 seq. only Dim
NFAT
Rel/NF-κB
(see legend on next page)

**Figure S3. IQ-Tree of Rel-homology DNA-binding Domain and domain architecture.** Taxa are color-coded according to taxonomic assignment (indicated in the upper right legend).


**Figure S4. Runt DNA-Binding Domain alignment of Runx proteins.** *C. owczarzaki* Runx1 (CAOG_00880) and Runx2 (CAOG_08438) were aligned with several animal and unicellular holozoan Runx orthologs. Taxa includes *Homo sapiens (Hsap), Drosophila melanogaster (Dmel), Strongylocentrotus purpuratus (Spur), Helobdella robusta (Hrob), Capitella teleta (Ctel), Nematostella vectensis (Nvec), Mnemiopsis leidyi (Mlei), Sycon ciliatum (Scil), Salpingoeca infusionum (Sinf), Salpingoeca kvevrii (Skve), Salpingoeca urceolata (Surc), Salpingoeca dolichothecata (Sdol), Codosiga hollandica (Chol), Salpingoeca punica (Spun), Salpingoeca macrocollata (Smac), Capsaspora owczarzaki (Cowc), Pigoraptor chileana (Pigchi), Ministeria vibrans (Mvib), Creolimax fragrantissima (Cfra), Sphaeroforma arctica (Sarc), Amoebidium parasiticum (Apar), Ichthyophonus hoferi (Ihof), Abeoforma whisleri (Awhi), Pirum gemmata (Pgem), Chromosphaera perkinsi (Cper), Sphaerothecum destruens (Sdes)* and *Corallochytrium limacisporum (Clim).* The alignment was trimmed to the Runt Domain predicted by PFAM. Key DNA binding aminoacids are highlighted in blue and Cys residues involved in redox binding affinity regulation are highlighted in orange. Numbers on the right represent the number of aminoacids in the alignment. Taxa are color-coded according to taxonomic assingment (indicated in the upper-left legend).

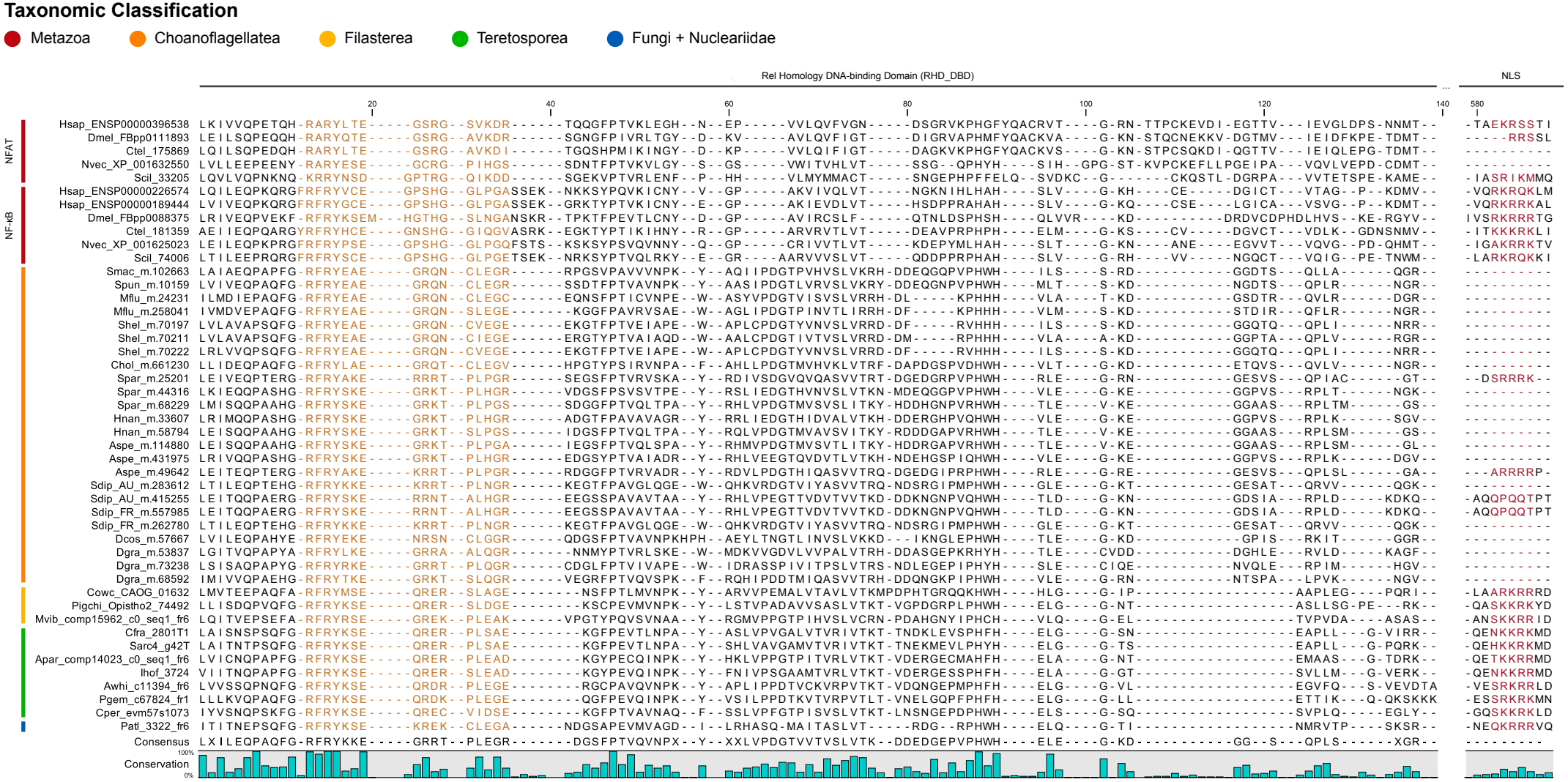



**Figure S5. Rel Homology DNA-binding Domain alignment of NF-κB and NFAT proteins.** *C. owczarzaki* NF-κB (CAOG_01632) was aligned with several opisthokonta NF-κB orthologs. Taxa includes *Homo sapiens (Hsap), Drosophila melanogaster (Dmel), Capitella teleta (Ctel), Nematostella vectensis (Nvec), Sycon ciliatum (Scil), Salpingoeca macrocollata (Smac), Salpingoeca punica (Spun), Mylnosiga fluctuans (Mflu), Salpingoeca helliandica (Shel), Codosiga hollandica (Chol), Savillea parva (Spar), Helgoeca nana (Hnan), Acanthoeca spectabilis (Aspe), Stephanoeca diplocostata isolated from Australia or France (Sdip_AU or Sdip_FR, respectively), Didymoeca costata (Dcos), Diaphanoeca grandis (Dgra), Capsaspora owczarzaki (Cowc), Pigoraptor chileana (Pigchi), Ministeria vibrans (Mvib), Creolimax fragrantissima (Cfra), Sphaeroforma arctica (Sarc), Amoebidium parasiticum (Apar), Ichthyophonus hoferi (Ihof), Abeoforma whisleri (Awhi), Pirum gemmata (Pgem), Chromosphaera perkinsi (Cper)* and *Parvularia atlantis (Parv).* The alignment was trimmed to the Rel Homology DNA-binding Domain predicted by PFAM and the Nuclear Localization Signal (NLS). Key DNA binding aminoacids (recognition loop) are highlighted in orange and NLS is highlighted in red. Taxa are color-coded according to their taxonomic assignment (indicated in the upper-left legend).

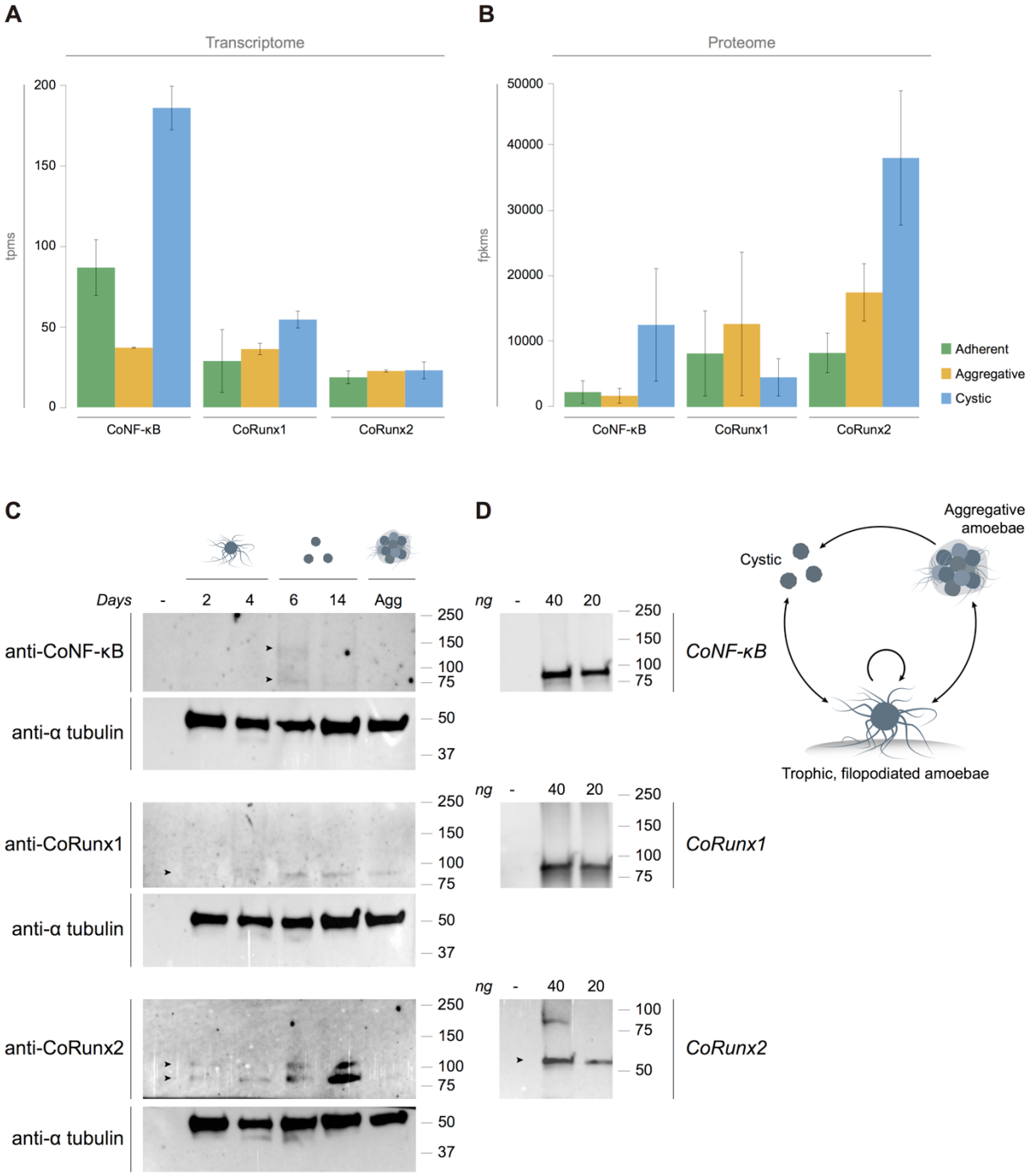



**Fig. S6: CoNF-κB, CoRunx1 and CoRunx2 expression across *Capsaspora* life stages.** (A) *CoNF-κB*, *CoRunx1* and *CoRunx2* transcriptomic expression in adherent, cystic and aggregative stages (data from Sebé-Pedrós et al., 2013a). (B) *CoNF-κB, CoRunx1* and *CoRunx2* protein abundance in adherent, cystic and aggregative stages (data from Sebé-Pedrós et al., 2016b). (C) Western Blot analysis of *Capsapora* protein extracts from adherent (2-day and 4-day), cystic (6-day and 14-day) and aggregates stages against *CoRunx1, CoRunx2* and *CoNF-κB* custom antibodies. Anti-alpha tubulin antibody was used as loading control. (D) Antibody validation against *CoNF-κB, CoRunx1* and *CoRunx2* recombinant proteins.

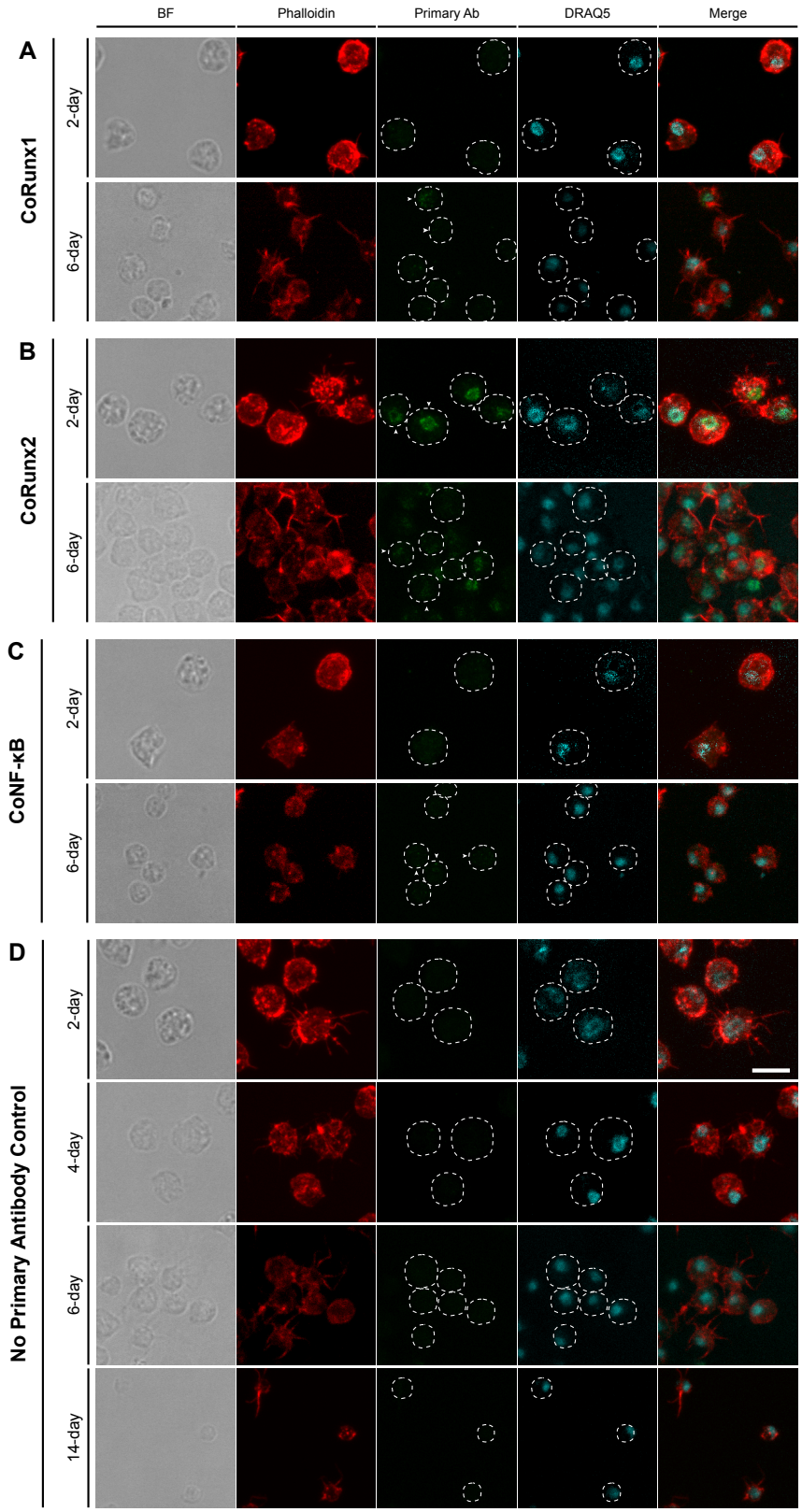



**Fig. S7. Localization of CoRunx1, CoRunx2 and CoNF-κB in *Capsaspora*.** Immunofluorescence of 2-day adherent and 6-day cystic stage *Capsaspora* cells using custom *CoRunx1* (A), *CoRunx2* (B) and *CoNF-κB* (C) antibodies. (D) No primary antibody control across timepoints (2, 4, 6 and 14 days). Dashed line indicate cell body and arrows indicate antibody signal. Scale bar represents 5 µm.

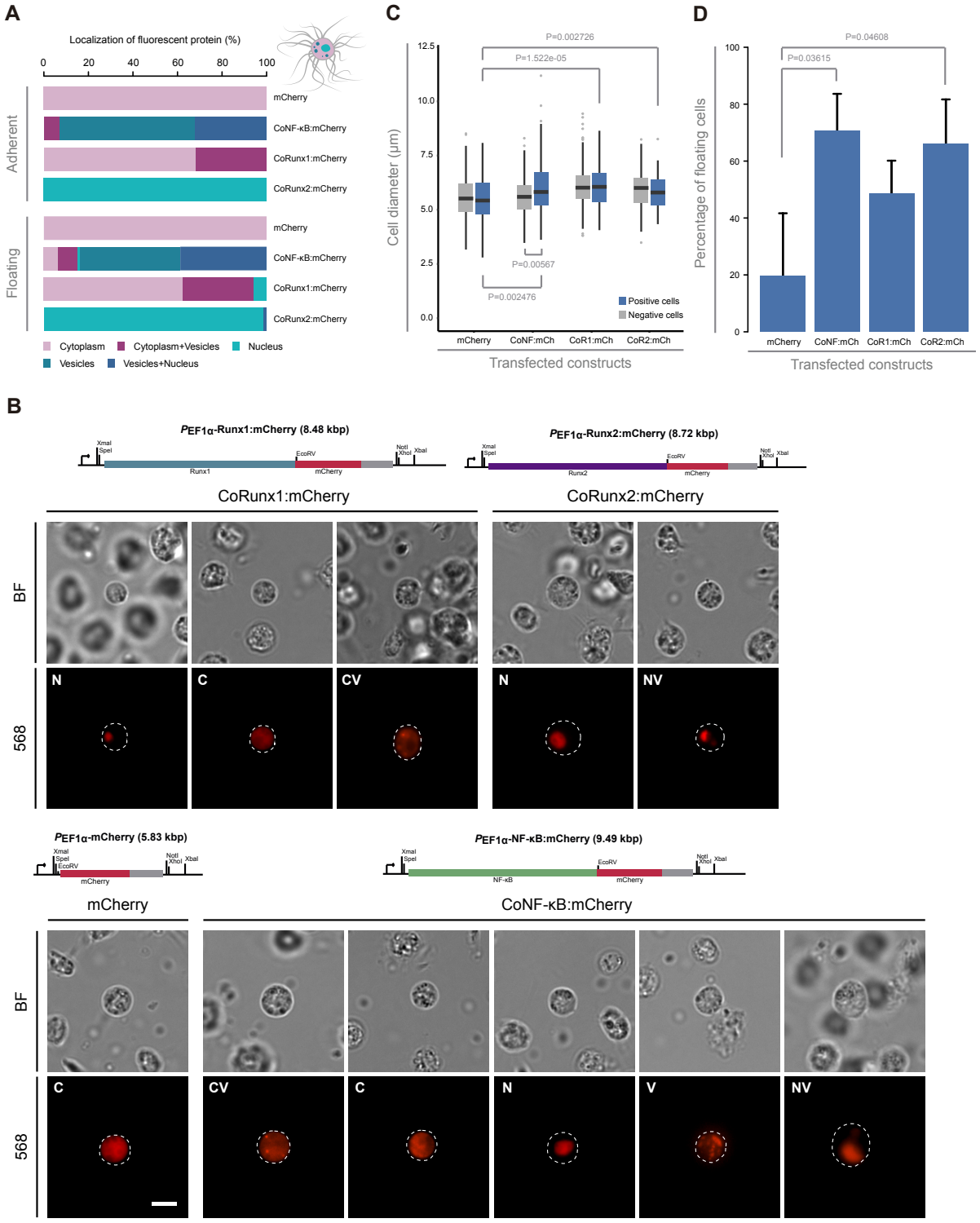



**Fig. S8. Overexpression of CoRunx1, CoRunx2 and CoNF-κB 24h post-transfection in *Capsaspora*.** (A) Relative localization of fluorescent proteins (%). (B) Live imaging of transfected cells with pONSY-mCherry, pONSY-CoNF-κB:mCherry, pONSY-CoRunx1:mCherry and pONSY-CoRunx2:mCherry constructs. Note different localization of fluorescent proteins in the cytoplasm (C), cytoplasm and vesicles (CV), nucleus (N), vesicles (V) or nucleus and vesicles (NV). (C) Cell diameter of positive and negative transfected cells with pONSY-mCherry, pONSY-CoNF-κB:mCherry, pONSY-CoRunx1:mCherry and pONSY-CoRunx2:mCherry constructs. (D) Percentage of floating cells of cells transfected with each construct. Error bars represent s.d. Statistical tests in (C) and (D) are performed using Student's T-Test. Results from 3 biological replicates (n=2.286k cells).

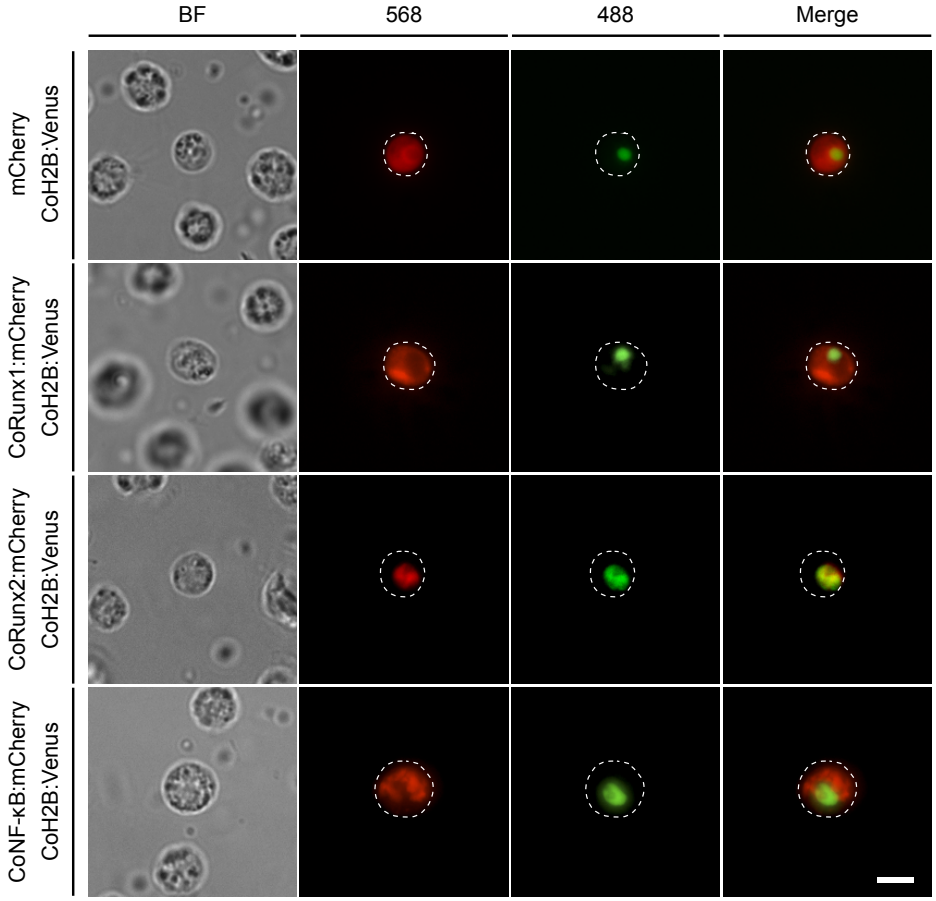



**Fig. S9: Overexpression of CoNF-κB, CoRunx1 and CoRunx2 24h post-transfection in *Capsaspora*.** Co-transfection of CoH2B:Venus with either pONSY-mCherry, pONSY-CoNF-κB:mCherry, pONSY-CoRunx1:mCherry or pONSY-CoRunx2:mCherry constructs. Note CoRunx2 localisation in the nucleus. CoRunx1 and CoNF-κB are mainly located in vesicles and in the cytoplasm.

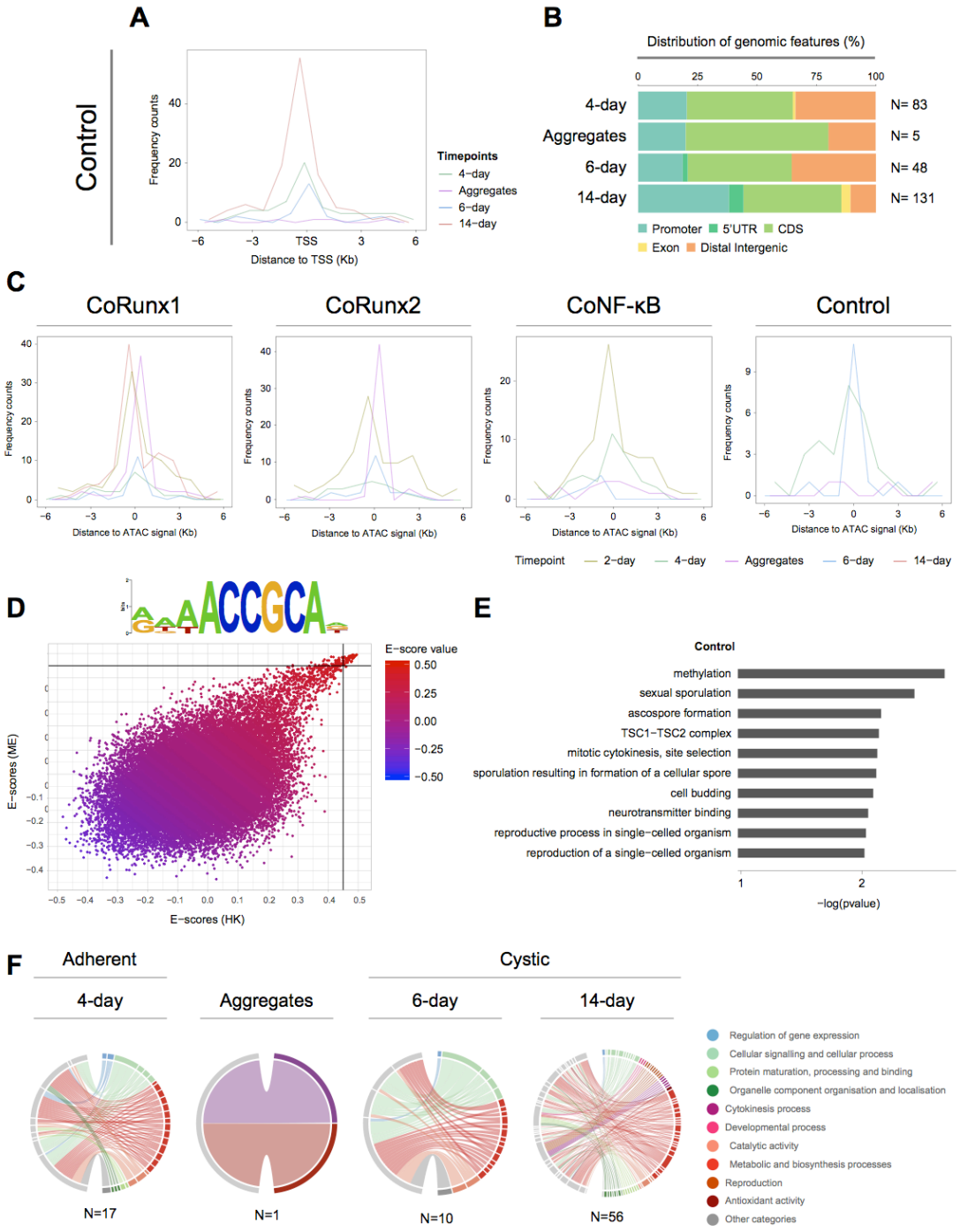



**Fig. S10. Distribution and GO enrichment of high confidence ChIPseq peaks mapping regulatory sites.** (A) Distribution of mock (control) ChIP peaks near TSS across *Capsaspora* life stages. (B) Genomic distribution of mock ChIP (control) peaks across *Capsaspora* life stages. Number of mapped peaks (N) are indicated at the right. (C) Distribution of *CoRunx1, CoRunx2, CoNF-κB* and mock (control) ChIP peaks around ATAC-defined regulatory sites across *Capsaspora* life stages, normalized to timepoints. (D) Distribution of 8-mer E-scores from HK and ME arrays in *CoRunx2* PBM and *CoRunx2* enriched motif. (E) Enriched GO terms among genes associated with mock (control) ChIP peaks mapping regulatory sites (P-value cutoff is 0.01). (F) Enriched GO terms among (N) number of genes associated with mock (control) ChIP peaks mapping regulatory sites across *Capsaspora* life stages (P-value cutoff is 0.05).

## SUPPLEMENTARY TABLES

**Table S3. List of primers used to build *Capsaspora* expression vectors with reporter genes.**
Restriction enzymes sites are underlined.

| Gene | Region | Name | Sequence 5'-3' |
|---|---|---|---|
| *CoRunx1* | Full-length | 1 | ATATCCCGGGATGAGCCTGAC |
| | | 2 | TATAGATATCAAAATGTAATTCGCCCATG |
| | DBD (±50 aa) | 3 | AGCTGGCGCGCCGAATTCATGAGCCTGACAGC |
| | | 4 | CGGTCCTGCAGGGAAATTCGGGAGCCATATAAGG |
| | DBD (±300 aa) | 5 | ATTAGGCGCGCCATGAGCCTGACAGCGACT |
| | | 6 | CACCCCTGCAGGAGCGCGACGGAACTTTGA |
| *CoRunx2* | Full-length | 7 | TATACCCGGGATGAGCATCG |
| | | 8 | TATAGATATCCGCGTCCATACGTAT |
| | DBD (±50 aa) | 9 | GAATGGCGCGCCATCGCTGCAGCATCACCG |
| | | 10 | GTGGCCTGCAGGGTTGCAAACCAAGCTGAGGA |
| | DBD (±300 aa) | 11 | ATATGGCGCGCCATGAGCATCGCTGCAGCAT |
| | | 12 | TAAACCTGCAGGTGGCTGACATTGGCGTGG |
| *CoNF-κB* | Full-length | 13 | TATACCCGGGATGGACCTCTCC |
| | | 14 | TATAGATATCGTCGACGCTGTAGAGGGC |
| | DBD (±50 aa) | 15 | GTCCGGCGCGCCTCTGCCAACACGAGTATGGT |
| | | 16 | GGTGCCTGCAGGAGTTGGCGCCAGGTACATT |
| | DBD (±300 aa) | 17 | ATAAGGCGCGCCATGGACCTCTCCGAATTGTCTGGAT |
| | | 18 | GGGCCCTGCAGGTGTGCAAGGGTGAGTCGC |

*Section 4*

# Discussion

*The transition to animal multicellularity in light of Capsaspora owczarzaki*

The origin and evolution of animal multicellularity can only be addressed by comparing animals with their closest unicellular relatives, the unicellular holozoans. In the last decades, the phylogenetic framework of Holozoa has been resolved by phylogenomic analyses and expanded thanks to the description of new species (Arkush et al., 2003; Cavalier-Smith and Chao, 2003; Glockling et al., 2013; Hehenberger et al., 2017; Knauth, 2005; Lang et al., 2002; Marshall et al., 2008; Marshall and Berbee, 2011; Raghu-kumar, 1987; Ruiz-Trillo et al., 2008; Shalchian-Tabrizi et al., 2008; Steenkamp et al., 2006; Torruella et al., 2012, 2015). Moreover, comparative analyses between unicellular holozoans and animals showed us the complex genetic repertoires as well as the diversity of lifestyles of animals' relatives, including regulated multicellular-like behaviours (Alegado et al., 2012; de Mendoza et al., 2015; Fairclough et al., 2013; Hehenberger et al., 2017; Ondracka et al., 2018; Raghu-kumar, 1987; Sebé-Pedrós et al., 2013b, 2016a; Suga and Ruiz-Trillo, 2013). This suggests that the unicellular ancestor of animals was probably capable of temporal cell differentiation, presented different regulated cell stages (some of them multicellular) and possessed a complex repertoire of genes related to multicellular functions that evolved prior the divergence of animals (Mikhailov et al., 2009; Sebé-Pedrós et al., 2017).

Although the lack of genetic tools among unicellular holozoans has limited our understanding of the function of key genes in a unicellular context, functional-based analyses can now be performed thanks to the recent development of transfection in *S. rosetta* (Booth et al., 2018), *C. fragrantissima* (Suga and Ruiz-Trillo, 2013), and *A. whisleri* and *C. limacisporum* (Kożyczkowska et al., unpublished). Altogether, comparative genomic analyses and functional analyses of genes key for animal multicellularity represent an ideal synergy to improve our knowledge of the origin of animal multicellularity.

In the following sections, I will discuss how the development of *Capsaspora owczarzaki* as an experimentally tractable system (section *3.1. Transfection of Capsaspora owczarzaki, a close unicellular relative to animals*) and the study of the function of key genes for multicellularity in *Capsaspora* (section *3.2. Evolution of Runx and NF-κB developmental Transcription Factor families and the origin of animal multicellularity)* can improve our understanding of the role of co-option in the transition to animal multicellularity.

# *Section 4.1*

# Developing genetic tools in *Capsaspora owczarzaki,* an emerging model system

In section *3.1. Transfection of Capsaspora owczarzaki, a close unicellular relative to animals* we describe a reliable transfection protocol that opens new avenues for molecular and functional-based analyses. Moreover, this protocol can be used as a starting point to achieve stable transfection in *Capsaspora* and to develop transfection in other species.

## 4.1.1. Transfection of *Capsaspora owczarzaki*

We have developed a robust method to efficiently transfect *Capsaspora,* based on the classical calcium-phosphate precipitation coupled with a glycerol shock (Gaudet et al., 2007; Graham and van der Eb, 1973; Grosjean et al., 2006; Ling et al., 2017). We optimized the transfection procedure which now overcomes the barriers that prevented efficient transfection in our prior attempts to deliver plasmid DNA in *Capsaspora*.

We achieved a mean transfection efficiency of 1.132%±0.529 (mean±s.d.) 18 hours post-transfection, being the highest efficiency of 2.083%±0.248 (which is equivalent to ~$2.5*10^2$ transformants per µg of DNA). This transfection efficiency is sufficient to screen for positive cells and perform further manipulations, such as *in vivo* fluorescence imaging, *fluorescence activating cell sorting* (FACS) of live cells and immunofluorescence assays. Moreover, this efficiency is comparable to other transfection efficiencies obtained in other unicellular species (Booth et al., 2018; Caro et al., 2012; Gaudet et al., 2007; Janse et al., 2006b, 2006a; Kawai et al., 2010; Schiestl and Gietz, 1989; Vinayak et al., 2015).

We additionally show that co-transfection of two different plasmids is relatively high in *Capsaspora,* similarly to those observed in other unicellular eukaryotes (Lerche and Hallmann, 2013, 2014; Schiedlmeier et al., 1994). Finally, we also observed that positive cells persisted for more than 10 days, indicating that transient expression of a gene of interest can be analysed during different life stages of *Capsaspora* (Sebé-Pedrós et al., 2013b; Stibbs et al., 1979). We provide a detailed step-by-step protocol available in *protocols.io*[1], an open access repository for science methods. The protocol includes reagent preparation and critical steps to be considered for efficient transfection that can be easily implemented in other laboratories and can serve as a starting point to develop transfection in other closely-related species.

This achievement situates *Capsaspora* as an ideal system to address the origin and evolution of animal multicellularity from a functional perspective for different reasons. First, *Capsaspora* grows in a rich medium axenically and cultures are easy to grow and to maintain in the lab. In fact, axenic cultures facilitate experimental manipulations in a controlled environment. Moreover, the multicellular aggregative stage of *Capsaspora* can be easily and reproducibly induced under culture conditions (Sebé-Pedrós et al., 2013b). Furthermore, *Capsaspora's* full nuclear and mitochondrial genomes are sequenced and well-annotated (Suga et al., 2013), an essential requirement to survey interesting homologs to animal genes related to multicellular functions. Third, the transcriptome, proteome, phosphoproteome and histone post-translational modifications of *Capsaspora* are available from its three life stages (Sebé-Pedrós et al., 2013b, 2016b, 2016a), allowing the comparison of phenotypical/morphological and genomic differences associated to gene function from a temporal perspective. Finally, it overcomes the limitations of previous analyses in *Capsaspora* which exclusively relied on comparative and functional genomics (Sebé-Pedrós et al., 2016a, 2013a) or on custom antibodies for subcellular localisation and functional study of a gene of interest (Sebé-Pedrós et al., 2016a).

Thus, the development of the reliable transfection protocol in *Capsaspora* described here will, hopefully, accelerate localization and in-depth studies on gene function of animal homologs that are conserved in *Capsaspora.*

[1] Protocol avaiable in *protocols.io:* dx.doi.org/10.17504/protocols.io.p4adqse

Moreover, comparisons to other experimentally-tractable unicellular holozoans (Booth et al., 2018; Kożyczkowska et al., unpublished; Suga et al., 2013) promises to yield increasingly mechanistic insights into the co-option of key genes for animal multicellularity.

### 4.1.2. Uncovering cell biological features in *Capsaspora owczarzaki*

Transfection allows the exploration of the cell biology of live cells through the expression of fluorescently tagged proteins. These proteins (or peptide sequences) contain key localization signals that localize fluorescent markers to specific organelles or subcellular structures. Moreover, *in vivo* localization of proteins overcomes distortion of native structures derived from cross-linking procedures, such as the use of aldehyde fixatives or organic solvents in fluorescence staining or antibody labelling.

Recently, a panel of fluorescently tagged cassettes allowed the visualization of the nucleus, cytoplasm, collar, filopodia, flagellum, membrane, mitochondria and endoplasmic reticulum in the choanoflagellate *S. rosetta* (Booth et al., 2018). For example, fluorescent labelling of the actin-filled collar and flagellar structures using a Lifeact:mCherry fusion in *S. rosetta* live cells revealed the parallel arrangement of straight microvilli in the collar and filopodia extending from the basal pole of the cell. Moreover, it also uncovered the existence of actin filaments originated in the cell body, and their organization at the base of the collar forming the microvilli. The mitochondrial marker encoded by the *S. cerevisiae Cox4* gene illuminated the network of mitochondria around the nucleus and a fusion of mCherry to the endogenous C-terminal *Histone H3 (H3)* gene or the N-terminal nuclear localization signal of the simian virus 40 (SV40) allowed us to specifically visualize the nucleoplasm (Booth et al., 2018).

Similarly, the nucleus of the ichthyosporean *C. fragrantissima* was labelled using the endogenous *Histone 2B2 (H2B2)* protein fused to Venus fluorescent protein (Suga and Ruiz-Trillo, 2013). The expression of this cassette allowed the tracing of nuclear movement during *C. fragrantissima* cell growth and revealed that nuclear divisions happen strictly synchronously.

To this end, we also provide a foundational set of expression vectors with signalling tags to label the nucleus, filopodia, membrane and actin cytoskeleton of *Capsaspora* cells *in vivo* **(Figure 20)**. These vectors have been deposited in *Addgene*[2], the non-profit plasmid repository. *Capsaspora* expression vectors encode tagged subcellular markers fused to Venus or mCherry fluorescent proteins under the control of an endogenous promoter and terminator of the *Capsaspora* Elongation Factor 1 alpha gene (EF1α). We showed by *in vivo* fluorescence imaging that each subcellular location is efficiently tracked. For example, we monitored the dynamics of filopodia *in vivo* using the cell membrane and filopodia marker, which allowed us to observe the retraction, breakage and foci of membrane accumulation in filopodia. Membrane and filopodia labelling allowed us to determine with unprecedented detail the three-dimensional organisation of filopodia around the *Capsaspora* cell body. Interestingly, filopodia are distributed all around the cell body, which in fact is not in direct contact with the substrate, but rather is held up by the numerous filopodia around them. Moreover, we used the actin cytoskeleton marker to observe the organisation of actin bundles around the cell body. These observations provide a valuable information for understanding *Capsaspora* cell biology and organization *in vivo.*

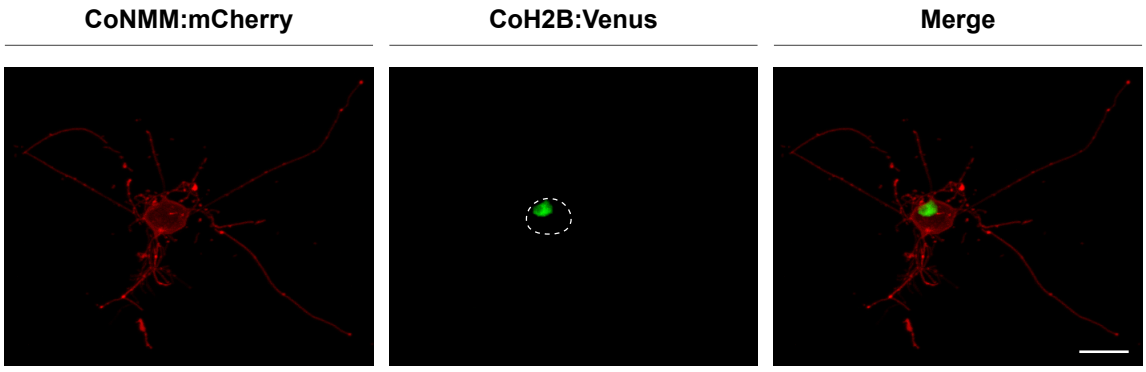



**Figure 20. Live imaging of a transfected *Capsaspora* cell.** Cell co-transfected with the membrane marker (from the endogenous *Capsaspora* Src2 protein) fused to mCherry (pONSY-CoNMM:mCherry) and the nuclear marker labelling the endogenous *Capsaspora* Histone 2B (H2B) protein fused to Venus (pONSY-CoH2B:Venus). Dashed lines indicate cell body. Scale bar: 5 μm.

[2] *Addgene:* www.addgene.org

### 4.1.3. From transient to stable transfection

In order to study the function of a particular gene of interest, it is often necessary to work with a homogeneous population of cells overexpressing the corresponding protein. This can be achieved by stable transfection, which generates stable cell lines that integrate the gene of interest. Stable transfection requires DNA integration into the host genome, which can occur randomly using plasmids, actively at random sites of the genome using transposases or viruses, or site-specifically when using genome editing tools, such as CRISPR/Cas9. Importantly, it also requires a strategy to successfully identify the positive population of transfected cells and permit the isolation of stable transfectants. For example, stable transfection was achieved in the volvocine green algae *Volvox carteri, Eudorina elegans, Gonium pectoral* and *Pandorina morum* using the same selection strategy with an antibiotic resistance gene to Paromomycin, the aminoglycoside 3'phosphotransferase VIII gene (aphVIII) of *Streptomyces rimosus* (Jakobiak et al., 2004; Lerche and Hallmann, 2014, 2013, 2009).

Thus, a key step for the optimization and development of stable cell lines is the determination of the selection conditions. This selection strategy needs to afford the positive population of transfected cells a growth advantage in culture. For example, selectable markers, such as antibiotic resistance genes, are a common strategy used for selection in mammalian cell lines. Conditioned media is also used to identify the positive population of transfected cells, especially in yeast. Thus, the implementation of either approach will depend on the transfection of an antibiotic-resistance marker into cells, along with the gene of interest. The antibiotic-resistance gene will allow the positive transfected cells to survive under the stress of the corresponding antibiotic, whereas non-transfected cells will perish after several weeks in culture.

Nevertheless, it is key to consider several steps when designing such experiments. For example, antibiotic-resistance markers work well when the gene of interest has either a neutral or stimulatory effect upon cell growth. However, if the gene of interest has a negative or toxic effect on cell growth antibiotic-resistance markers will not be useful to achieve stable expression of that gene. This is because cells lose their growth advantage when expressing a toxic gene, even if they carry the antibiotic-resistance marker. Moreover, the isolation of stable cell populations expressing a gene under antibiotic selection conditions often requires a longer period of selection (usually a month or more)

to be successfully isolated. Thus, by the time a stably transfected pool of cells is isolated, the toxic gene expression may have produced negative irreversible effects on the cell line. An alternative is to isolate positive transfectants soon after transfection, either by micromanipulation or by FACS.

Moreover, to achieve stable transfection, the selection conditions need to be optimized depending on the susceptibility of a particular cell line to the drug. Precise concentration ranges, time and duration of treatment, and toxicity will vary depending on the organism or cell line of choice.

In the present thesis, I also participated in developing a strategy to select *Capsaspora* by the use of antibiotics[3]. We tried the effect of several families of antibiotics on *Capsaspora* cell growth and viability **(Table 1)**. For example, we tested the effect of the family of *Aminoglycoside* antibiotics (*e.g.,* Neomycin, Paromomycin, Geneticin (G418), Kanamycin, Nourseothricin, Puromycin and Hygromycin), which are widely used in mammalian cell cultures (Southern and Berg, 1982). *Aminoglycoside* antibiotics function as protein synthesis inhibitors by interfering with ribosomal function. We also tested three antibiotics from the Bleomycin family (*i.e.,* Bleomycin, Phleomycin and Zeocin), a broad-spectrum antibiotic family effective against most bacteria, filamentous fungi, yeast, plant and animal cell lines. This family induces DNA double strand breaks and putatively inhibiting the incorporation of thymidine into DNA strands. Other tests were performed by assessing the effect of Pyrimethamine-sulfadoxine and WR99210 antibiotics in *Capsaspora*, two inhibitors of the dihydrofolate reductase *(DHFR)* gene involved in folic acid synthesis (Hastings and Sibley, 2002). These antibiotics act as antiparasitic and are commonly used in the treatment of uncomplicated chloroquine resistant *P. falciparum* malaria. Finally, we also tested the effect of different combinations of the previous antibiotics at similar concentration ranges. Overall, none of these antibiotics or a combination of antibiotics at the conditions tested had any remarkable effect on *Capsaspora* cells after one-week treatment. Ultimately, we tested the Carboxin family of fungicides, which includes Carboxin, Isopyrazam, Boscalin, Fuopyram and Bixafen. The Carboxin family of antibiotics primarily affects the succinate dehydrogenase (*SdhB*) gene and is generally used in sensitive yeast and other fungi.

[3] People involved: Núria Ros-Rocher, Helena Parra-Acero, Aleksandra Kożyczkowska, Sebastián R. Najle and Claudio Scazzoccio

For *Capsaspora,* we tested multiple concentration ranges and assessed their effect on cell growth and viability, yielding promising results for Carboxin, Bixafen and Isopyrazam **(Table 1).** The strongest effect was noticed 30 hours post-treatment, especially for Carboxin and Isopyrazam.

**Table 1. Antibiotics screening in *Capsaspora.***

| Family | Mechanism | Antibiotic Name | Concentration range |
|---|---|---|---|
| Aminoglycosides | Protein synthesis inhibitor | Neomycin | 200-1000 µg/mL |
| | | Paromomycin | 50-1000 µg/mL |
| | | Geneticin | 200-1000 µg/mL |
| | | Kanamycin | 200-1000 µg/mL |
| | | Hygromycin | 50-1000 µg/mL |
| | | Nourseothricin | 50-1000 µg/mL |
| | | Puromycin | 50-1000 µg/mL |
| Bleomycin | DNA double strand breaks inducer | Bleomycin | 3.5-35 µM |
| | | Phleomycin | 50-1000 µg/mL |
| | | Zeocin | 25-300 µg/mL |
| Aminonucleosides | Folic Acid synthesis inhibitor | Pyrimethamine | 50-1000 µg/mL |
| | | WR 99210 | 50-1000 µg/mL |
| Carboxin | Succinate dehydrogenase | Carboxin | 5-150 µg/mL |
| | | Isopyrazam | 5-150 µg/mL |
| | | Bixafen | 5-150 µg/mL |
| | | Boscalin | 5-150 µg/mL |
| | | Fuopyram | 5-150 µg/mL |

Overall, future work can be built on the transfection protocol here presented to establish stable transfection and genome editing in *Capsaspora.* The determination of a selection strategy and the effect of the drug in *Capsaspora* opens new avenues of research to establish stable transfection in *Capsaspora.* Efforts are underway to further test the effect of a resistance cassette to achieve stable transfection in *Capsaspora.*

### 4.1.4. From one species to another: developing transfection in new species

Previously reported and optimized transfection methods in closely-related species or species showing similar biological features can help to further develop a transfection protocol in a new species of interest. For example, transfection in the volvocine green alga *Chlamydomonas reinhardtii* was based on previously described protocols developed for plant cells using DNA-coated tungsten microprojectiles accelerated by particle gun (Klein et al., 1988a, 1988b, 1987). From the *C. reinhardtii* protocol (Kindle et al., 1989), stable transfection using particle bombardment was developed in *Volvox carteri* (Schiedlmeier et al., 1994) and *Gonium pectorale* (Lerche and Hallmann, 2009), other closely related species in the volvocine lineage. In fact, the *G. pectorale* resistance cassette was built using flanking sequences that include promoters from *C. reinhardtii* and from *V. carteri* (Lerche and Hallmann, 2009). This cassette allowed the functional expression of heterologous genes in *G. pectorale,* such as the codon-optimized luciferase gene from the marine copepod *Gaussia princeps*, which turned out to be a suitable reporter in *G. pectorale.*

A few years later, these protocols also contributed to the development of stable nuclear transfection in the volvocine green algae *Eudorina elegans* (Lerche and Hallmann, 2013) and *Pandorina morum* (Lerche and Hallmann, 2014). In both cases, stable transfection was achieved conferring resistance to the aminoglycoside antibiotic paromomycin using the same resistance cassette previously built for *V. carteri* (Lerche and Hallmann, 2009), which contains the aminoglycoside 3′-phosphotransferase (aphH) gene from *Streptomyces rimosus*.

In a similar way, the protocol here presented was informed by previous approaches developed in the slime mold *D. discoideum* (Gaudet et al., 2007; Nellen et al., 1984). Like *Capsaspora*, *D. discoideum* presents a single-celled trophic amoeboid stage adhered to the substrate which alternates with a multicellular (aggregative) dispersal stage under certain conditions. The trophic stage (vegetative cycle) is strictly unicellular and consists of an independent amoeba without a cell wall that multiplies by binary fission, an event that occurs roughly every 8-10 hours under optimal growth conditions (Escalante and Vicente, 2000), similarly to what happens in *Capsaspora.* This was the reason why we decided to try this method.

Thus, our method established in *Capsaspora* could, as well, extend to aid gene delivery in other non-model amoeboid unicellular holozoans or other microeukaryotes. Moreover, our panel of reporter cassettes for expression of fluorescent markers might also be of interest to report transfection in other closely-related species without sequenced genomes.

# *Section 4.2*

# Conserved transcription factor networks in *Capsaspora owczarzaki*

Unicellular holozoans possess and express an unexpected repertoire of developmental transcription factors previously thought to be animal-specific (de Mendoza et al., 2015; Fairclough et al., 2013; King et al., 2008; Richter et al., 2018; Sebé-Pedrós et al., 2013a, 2011), suggesting that animal evolution was not solely dependent on gene innovation. Thus, co-option, tinkering and expansion of pre-existing transcription factor networks were probably key for animal multicellularity (Arenas-Mena, 2017; King, 2004; Sebé-Pedrós et al., 2016a).

In section *3.2. Evolution of Runx and NF-κB developmental Transcription Factor families and the origin of animal multicellularity* we investigated the evolution and diversification of *Runx* and *NF-KB* transcription factor families in a taxon-rich paneukaryotic survey and assessed their downstream regulatory network in *Capsaspora* through localization and Chromatin Immunoprecipitation experiments.

## 4.2.1. Evolution of *Runx* and *NF-κB* transcription factor networks

Taxon sampling richness is of critical importance to trace a more realistic evolutionary history of gene families, including lineage-specific gene losses. Thus, recently sequenced eukaryotic genomes and transcriptomes, especially those of unicellular holozoans, can help update the phylogenetic history of key protein families. This newer lineage representation can be analysed with novel methods and provide new insights into the diversification and expansion of these families at the origin of animals.

Our taxon-rich paneukaryotic survey of *Runx* and *NF-κB* TFs-related protein domains, including Runt and Rel-homology (RHD) DNA-binding domains and other key domains

present in *Runx* and *NF-κB* interacting partners extended the number of non-metazoan taxa in which those genes have been identified. We confirmed Runt-related protein families were gained at the base of Holozoa and RHD at the base of Opisthokonta (being putatively secondarily lost in fungi), in agreement with previous studies (de Mendoza et al., 2013; Robertson et al., 2009). Our results also suggested that RHD TF families putatively underwent a duplication event of an ancestral Rel domain-like protein prior to the last metazoan common ancestor, followed by acquisitions of new domains in the paralogs leading to the emergence of modern NF-κB and NFAT proteins. Moreover, key domains for *Runx* and *NF-κB* regulation in animals and well-described interacting partners appeared to be metazoan innovations. Thus, an expansion of these TF networks, protein domain gains and interaction with new partners were key to increasing their combinatorial regulatory capabilities during the evolutionary transition to animal multicellularity and may have contributed to the increased developmental potential and cell type diversity in animals (Degnan et al., 2009).

### 4.2.2. Evolution of ancestral transcription factor networks

The origin of animal multicellularity has been associated with the genesis and expansion of the transcription factor toolkit, which enabled increased regulation of specific gene expression programs and cell differentiation (Bonner, 1998; Carroll, 2000; Larroux et al., 2008; Raff, 1996). The evolution of many TF networks, especially the regulatory components found in complex developmental programs, could have been produced by a period of genome innovation and gene duplication, prior the divergence of all major extant Metazoa phyla. In fact, some developmental TFs are present and expressed in the sponge *Amphimedon queenslandica,* suggesting that the ancestral animal developmental network was populated by many of the same regulatory components that are operating in modern metazoans (Adamska et al., 2007a, 2007b; Larroux et al., 2006).

On one side, early innovations of novel TF genes in the ancestor of animals and their closest unicellular relatives could have provided the regulatory foundation for the evolution of multicellularity and embryogenesis. These novel TFs would have extended the regulatory capacity of the genome which, followed by other innovation events during the transition to animal multicellularity, allowed new combinatorial interactions between

TFs and further expansions of their regulatory complexity **(Figure 21)** (de Mendoza et al., 2013; Larroux et al., 2008; Phillips and Luisi, 2000; Wilson and Koopman, 2002). This second period of innovation could have given rise to the full diversity of TF classes, generating lineage-specific classes and subfamilies, especially after the divergence of sponges and eumetazoans, and allowed their co-option into new roles. In fact, functional co-option of pre-existing TF networks may have been the first step towards the evolution of complex body plans and life cycles of eumetazoans (Larroux et al., 2008).

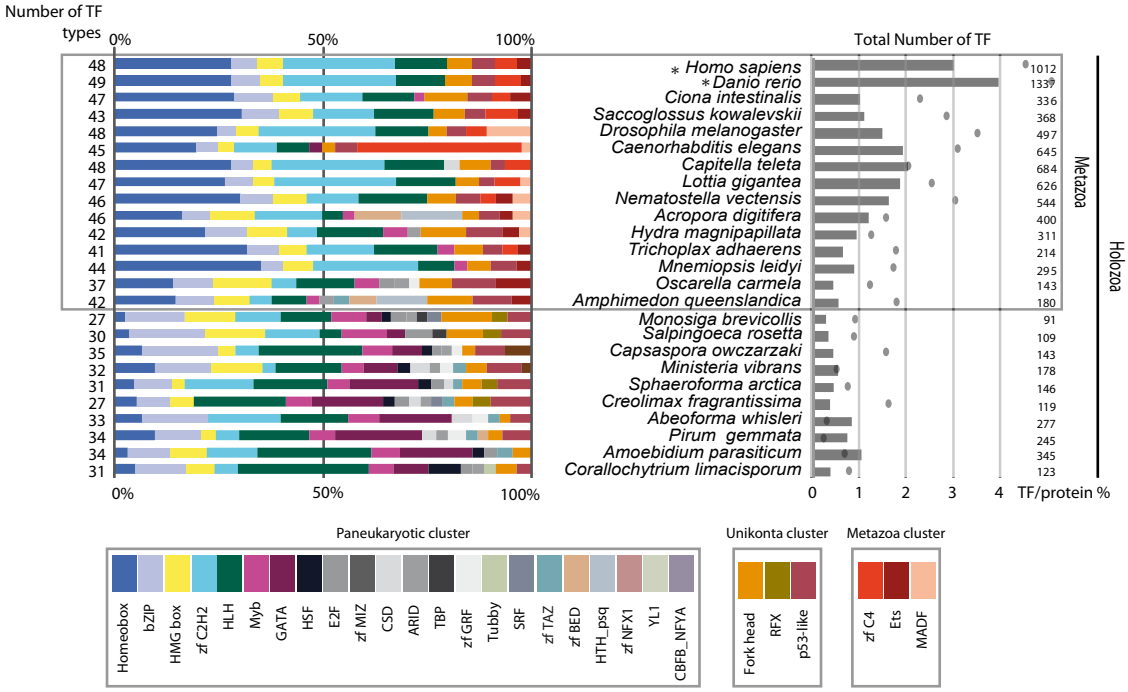



**Figure 21. Phylogenetic patterns of Transcription Factor families in Holozoa.** Transcription Factors (TFs) types that represent less than 2% of the corresponding TF repertoire (TFome) are not considered. The following TF types are summarized in higher-level categories according to structural similarities: (i) Homeobox supergroup (including Homeobox and Homeobox_KN/TALE); (ii) the bZIP supergroup (including bZIP_1, bZIP_2 and bZIP_Maf) and (iii) the p53-like supergroup (including p53, STAT, Runx, NDT80, LAG1 and RHD). To the left, total number of TF types present in each taxa and the relative abundance of each DBD type in the TFome of every species are depicted using the color code in the legend of DBDs. To the right, the bar plot indicates the total number of TFs in each species and dots indicate the percentage of total TFs/total number of proteins. The black asterisks indicate species with Whole Genome Duplications (WGDs). Modified from de Mendoza et al., 2013.

On the other side, gene duplication events in some TF families during the evolution of animals were also key for the expansion and diversification of their regulatory networks. Interestingly, gene duplication products often present asymmetric evolutionary rates (*i.e.,* one duplicate evolves faster than the other). This differential rate of evolution may produce functional constraints on the protein's original role, maintaining one copy similar to the ancestral form while the other diverges, acquiring a new function (Larroux et al., 2008; Taylor and Raes, 2004). Altogether, both sources of TF evolution contributed to a gradual increase in TF gene number in animals. This increase was putatively key for cell type differentiation and increasing the number of cell types, allowing the expansion of the developmental potential in animals, especially during early eumetazoan evolution (Degnan et al., 2009). In fact, TF numbers correlate with morphological and cell type complexity across different animal phyla **(Figure 21)** (de Mendoza et al., 2013). For example, cnidarians have a TF gene repertoire typically two to three times larger than that of sponges and placozoans (Degnan et al., 2009; Putnam et al., 2007; Srivastava et al., 2010), and morphologically simpler forms within these lineages have fewer TFs that morphologically more complex groups (de Mendoza et al., 2013).

Functional co-option of ancestral genes was also an important driving force for the emergence of animals (Sebé-Pedrós et al., 2017). Consistent with previous studies (Sebé-Pedrós et al., 2016a), our results suggested that complex gene regulatory networks (GRNs) may exist in *Capsaspora*. In particular, we showed in this thesis that core GRNs of developmental TFs, such as *Runx* and *NF-κB,* evolved long before the advent of animal multicellularity (Davidson and Erwin, 2006), probably controlling essential cell cycle and developmental control behaviours in the first animals. These core conserved TF networks were subsequently integrated into complex developmental programs during animal evolution (Peter and Davidson, 2011).

Nevertheless, the mechanistic changes in the ancestral genetic regulatory program that were necessary for the evolution of the developmental body plan at the onset of animals are still not fully resolved. Classic evolutionary theories, based on selection of small incremental changes, has sought explanations by extrapolation from observed patterns of adaptation (Davidson and Erwin, 2006). Macroevolutionary theories have invoked multilevel selection, which proposes a series of analytical tractable stages during the transition to multicellularity and the evolution of developmental programs (Damuth and Heisler, 1988; Heisler and Damuth, 1987). Depending on their functional properties,

(gradual) changes in the structure of the diverse kinds of subcircuits of GRNs will have different consequences for the outcome of the developmental process, and therefore for evolution. For example, some GRNs that perform essential upstream functions in building given body parts would remain as evolutionarily inflexible subcircuits, termed “kernels” (Davidson and Erwin, 2006). Other GRNs constituting small subcircuits, the “plug-ins” of GRNs, would be repeatedly co-opted to diverse developmental processes. Finally, other classes of GRNs would act as switches that would allow or disallow developmental subcircuits to function in a given context or would be differentiated into gene batteries that encode the detailed functional properties of the body part (Davidson, 2001; Davidson and Erwin, 2006).

Interestingly, the first class of GRNs, or *kernels,* would be mainly constituted by subcircuits of regulatory genes (*i.e.,* genes encoding TFs), which would execute developmental patterning functions required to specify the spatial domain of an embryo in which a given body part will form. Moreover, they would be exclusively dedicated to given developmental functions. This means that *kernels* would have a particular well-conserved structure, and each product of the multiple regulatory genes constituting that *kernel* are required for its proper functioning. Thus, any interference in any of its components would destroy *kernel’s* function, leading to catastrophic alternations of those GRNs and possibly lethal phenotypes during development. The result is extraordinary conservation of *kernel* architecture generally common to all members of a given phylum or superphylum (Davidson and Erwin, 2006). Examples include *kernels* involved in cell patterning specification of the nervous system or development of the immune system (Arendt and Nübler-Jung, 1999; Cornell and Von Ohlen, 2000; Hirth et al., 2003; Lichtneckert and Reichert, 2005; Lowe et al., 2003). Another interesting example of well-conserved *kernels* includes GRNs involved in endoderm specification in echinoderms, known to have existed at least since diverging at the end of the Cambrian (Hinman et al., 2003).

Similarly to *kernels,* co-option of plug-ins are usually related to GRNs subcircuits involved in many diverse developmental functions. Despite both of them being entirely regulatory and well-conserved structurally, plug-ins provide inputs to a wider variety of regulatory programs, such as signalling transduction systems. Examples include the *Wnt* and *Notch* signalling pathways or transforming growth factor-beta. In general, those GRNs are more flexible, and even homologous processes in related species may be used differently

(Tokuoka et al., 2004). Finally, gene batteries would be directed to cell-type specific functions, especially in cell differentiation programs (Davidson, 2001), being more evolutionary labile and being subjected to continuous renovation. For example, incremental changes of this type of GRNs would gradually alter gene functionality in their protein-coding sequences. New genes may be also added to them, or others may be lost. Nevertheless, they reside at the periphery of GRNs, being at the terminal part of the network. Thus, they are not regulating other genes, but instead they are expressed in the final stages of given developmental processes.

In the present thesis, I have addressed the transition to animal multicellularity from a functional perspective by developing a transfection protocol to turn the filasterean amoeba *Capsaspora owczarzaki* into an experimentally tractable system and studying the evolution and diversification of *Runx1*, *Runx2* and *NF-κB* transcription factor families. Moreover, our results suggest that complex regulatory networks of transcription factors exist in *Capsaspora,* and are temporally regulated across its different life stages. This implies that the unicellular ancestor of animals was probably capable of temporarily regulate distinct cell states (being one or more of them multicellular) through a pre-existing repertoire of TFs, an scenario that was putatively dependent on environmental stimuli, as observed in some extant unicellular holozoans. During the transition to animal multicellularity, an expansion of the TFs repertoire and the evolution of gene regulatory networks could have been key to providing additional during the spatial integration of pre-existing cell types into the first metazoans.

Altogether, the present thesis places *Capsaspora* as a good model to understand the function of key genes in a unicellular context and offers new insights into the evolution of animal multicellularity.

*Section 5*

# Conclusions

*Capsaspora owczarzaki as an emerging model*

# *Section 5*

# *Conclusions*

The main conclusions of the present work are the following:

1. *Capsaspora owczarzaki* is efficiently and reproducibly transiently transfected using a calcium-phosphate DNA precipitation protocol coupled with a glycerol shock, opening new opportunities to explore its cell biology *in vivo.*

2. *Capsaspora* mean transfection efficiency resulted in 1.132%±0.529 (mean±s.d.) 18 hours post-transfection, being the highest efficiency of 2.083%±0.248. This transfection efficiency is sufficient to screen for positive cells and perform further manipulations, such as *in vivo* fluorescence imaging, *fluorescence activating cell sorting* (FACS) of live cells and immunofluorescence assays.

3. Fluorescence of transfected cells persists for more than 10 days, indicating that it continues through multiple rounds of cell division. Thus, transient expression of a gene of interest can be analysed during different life stages of *Capsaspora*. Moreover, transfection did not irreparably harm *Capsaspora* as positive transfected cells resembled non-transfected cells in their shape, motility, and ability to propagate.

4. Co-transfection of two different plasmids is possible in *Capsaspora,* allowing visualizing two different cellular structures simultaneously in live cells. Moreover, co-transfection was relatively high as ~65% to ~83% of positive cells simultaneously expressed two fluorescent markers.

5. A panel of fluorescently-tagged proteins under the control of the endogenous *Capsaspora* EF1α promoter and terminator sequences allows visualizing the cytoplasm, membrane and filopodia, the actin cytoskeleton and the nucleus in *Capsaspora* live cells.

6. The cell membrane and filopodia marker allow monitoring the dynamics of filopodia *in vivo* and revealed with unprecedented detail the three-dimensional organization of filopodia around *Capsaspora* cell body. The actin cytoskeleton marker allows observing the organisation of actin bundles around the cell body.

7. The protocol of transfection here developed situates *Capsaspora* as a potential powerful model to analyze the origin of animals.

8. Both Runt and Rel-homology DNA binding domains were significantly enriched at the origin of Metazoa, possibly through gene duplication events and diversification of protein domain architectures.

9. *Runx* was gained at the base of Holozoa and *NF-κB* at the root of Opisthokonta, putatively being secondarily lost in Fungi.

10. *Runx* homologs among unicellular holozoans share key DNA binding amino acids and Cysteine residues involved in redox binding affinity regulation in animals. The *Runx* Inhibitor domain, the WRPY motif and the *NF-κB*-related Death domain appeared to be metazoan innovations, *RunxI* appearing in chordates and the Death domain and WRPY in early branching animals.

11. *CoRunx1* dynamically localized in small vesicles in adherent cells (4-day old culture) and in cystic stage cells (14-days old) and in the nucleus and small vesicles in early cystic stage cells (6-days old). In contrast, *CoRunx2* localized clearly in the nucleus both in adherent and cystic stages.

12. *CoNF-κB* was localized in small vesicles around the nucleus in the early cystic stage (6-day old), and more intensely in small vesicles and in the cytoplasm in the late cystic stage**.**

13. *CoRunx2*, and possibly *CoRunx1* and *CoNF-κB*, could be capable of functioning as transcription factors dynamically across *Capsaspora* life stages.

14. The most significantly overrepresented motif for *CoRunx2* corresponded to a perfect match to the animal *Runx2* consensus binding sequence in the aggregative stage. This motif was also enriched in PBM experiments, providing independent

confirmation of the ChIP-seq data quality. *CoRunx1* overrepresented motifs do not resemble the binding motifs of *Runx1* animal homologs, altogether suggesting that complex regulatory networks may exist in *Capsaspora.*

15. The downstream regulatory networks of *CoRunx1* and *CoRunx2* are enriched with genes related to regulation of cell growth and proliferation and response to stress, particularly, to oxygen response, similar to some of the pathways in which animal *Runx1* and *Runx2* homologs are known to participate.

16. The *CoNF-κB* downstream regulatory network is enriched in genes related to response to stimuli, including putative response to oxidative stress, similar to its animal homologs.

17. Altogether our results indicate that complex regulatory networks may exist in *Capsaspora* and are dynamically regulated across *Capsaspora* life stages.

# References

*Publications and communications*

# References

Aanen, D.K., Debets, A.J.M., de Visser, J.A.G.M., Hoekstra, R.H., 2008. The social evolution of somatic fusion. BioEssays 30, 1193–1203.

Abedin, M., King, N., 2010. Diverse evolutionary paths to cell adhesion. Trends Cell Biol. 20, 734–42.

Adamska, M., Degnan, S.M., Green, K.M., Adamski, M., Craigie, A., Larroux, C., Degnan, B.M., 2007a. Wnt and TGF-β expression in the sponge *Amphimedon queenslandica* and the origin of metazoan embryonic patterning. PLoS One 2.

Adamska, M., Matus, D.Q., Adamski, M., Green, K., Rokhsar, D.S., Martindale, M.Q., Degnan, B.M., 2007b. The evolutionary origin of hedgehog proteins. Curr. Biol. 17, 836–837.

Adl, S.M., Simpson, A.G.B., Farmer, M.A., Andersen, R.A., Anderson, O.R., Barta, J.R., Bowser, S.S., Brugerolle, G., Fensome, R.A., Fredericq, S., James, T.Y., Karpov, S., Kugrens, P., Krug, J., Lane, C.E., Lewis, L.A., Lodge, J., Lynn, D.H., Mann, D.G., Mccourt, R.M., Mendoza, L., Moestrup, Ø., Mozley-Standridge, S.E., Nerad, T.A., Shearer, C.A., Smirnov, A. V., Spiegel, F.W., Taylor, M.F.J.R., 2005. The new higher level classification of eukaryotes with emphasis on the taxonomy of protists. J. Eukaryot. Microbiol. 52, 399–451.

Adl, S.M., Simpson, A.G.B., Lane, C.E., Lukeš, J., Bass, D., Bowser, S.S., Brown, M.W., Burki, F., Dunthorn, M., Hampl, V., Heiss, A., Hoppenrath, M., Lara, E., Gall, L. Le, Lynn, D.H., McManus, H., Mitchell, E.A.D., Mozley-Stanridge, S.E., Parfrey, L.W., Pawlowski, J., Rueckert, S., Shadwick, L., Schoch, C.L., Smirnov, A., Spiegel, F.W., 2012. The revised classification of eukaryotes. J. Eukaryot. Microbiol. 59, 429–493.

Adli, M., 2018. The CRISPR toolkit for genome editing and beyond. Nat. Commun. 9, 1–13.

Alegado, R.A., Brown, L.W., Cao, S., Dermenjian, R.K., Zuzow, R., Fairclough, S.R., Clardy, J., King, N., 2012. A bacterial sulfonolipid triggers multicellular development in the closest living relatives of animals. Elife 2012, 1–16.

Alegado, R.A., King, N., 2014. Bacterial Influences on Animal Origins. Cold Spring Harb. Perspect. Biol. 6.

Allwood, A., Walter, M., Burch, I., Kamber, B., 2007. 3.43 billion-year-old stromatolite reef from the Pilbara Craton of Western Australia: Ecosystem-scale insights to early life on Earth. Precambrian Res. 158, 198–227.

Anderson, C.A., Eser, U., Korndorf, T., Borsuk, M.E., Skotheim, J.M., Gladfelter, A.S., 2013. Nuclear repulsion enables division autonomy in a single cytoplasm. Curr. Biol. 23, 1999–2010.

Anderson, M., 1984. The Evolution of Eusociality. Annu. Rev. Ecol. Syst. 15, 165–189.

Antcliffe, J.B., Callow, R.H.T., Brasier, M.D., 2014. Giving the early fossil record of sponges a squeeze. Biol. Rev. 89, 1–33.

Arenas-Mena, C., 2017. The origins of developmental gene regulation. Evol. Dev. 19, 96–107.

Arendt, D., Nübler-Jung, K., 1999. Comparison of early nerve cord development in insects and vertebrates. Dev. 126, 2309–2325.

Arkush, K.D., Mendoza, L., Adkinson, M.A., Hedrick, R.P., 2003. Observations on the Life Stages of *Sphaerothecurn destruens* n. g., n. sp., a Mesomycetozoean Fish Pathogen Formally Referred to as the *Rosette* Agent. J. Eukaryot. Microbiol. 50, 430–438.

Arndt, H., Dietrich, D., Auer, B., Cleven, E.-J., Gråfenhan, T., Weitere, M., Mylnikov, A.P., 2000. Functional diversity of heterotrophic flagellates in aquatic ecosystems, Flagellates: Unity, Diversity and Evolution. B.S.C. Leadbeater, J. Green (Eds.). CRC Press.

Awramik, S.M., 1971. Precambrian columnar Stromatolite diversity: Reflection of metazoan appearance. Science. 174, 825–827.

Babu, M.M., Luscombe, N.M., Aravind, L., Gerstein, M., Teichmann, S.A., 2004. Structure and evolution of transcriptional regulatory networks. Curr. Opin. Struct. Biol. 14, 283–291.

Barrett, L.E., Sul, J.Y., Takano, H., Van Bockstaele, E.J., Haydon, P.J., Eberwine, J.H., 2006. Region-directed phototransfection reveals the functional significance of a dendritically synthesized transcription factor. Nat. Methods 3, 455–460.

Bates, J.M., Mittge, E., Kuhlman, J., Baden, K.N., Cheesman, S.E., Guillemin, K., 2006. Distinct signals from the microbiota promote different aspects of zebrafish gut differentiation. Dev. Biol. 297, 374–386.

Beaumont, N.J., 2009. Modelling the transport of nutrients in early animals. Evol. Biol. 36, 256-266.

Becker, B., Marin, B., 2009. Streptophyte algae and the origin of embryophytes. Ann. Bot. 103, 999–1004.

Bell, G., Mooers, A.O., 1997. Size and complexity among multicellular organisms. Biol. J. Linn. Soc. 60, 345–363.

Benny, G.L., O'Donnell, K., 2000. *Amoebidium parasiticum* is a protozoan, not a Trichomycete. Mycologia 92, 1133–1137.

Berkner, L. V., Marshall, L.C., 1965. On the Origin and Rise of Oxygen Concentration in the Earth's Atmosphere. J. Atmos. Sci.

Bisson, M.A., Beilby, M.J., 2002. The transport systems of *Ventricaria ventricosa*: Hypotonic and hypertonic turgor regulation. J. Membr. Biol. 190, 43–56.

Bonner, J.T., 1998. The origins of multicellularity. Integr. Biol. Issues, News, Rev. 1, 27–36.

Bonner, J.T., 2000a. The Origin of Multicellularity. In: First Signals, The Evolution of Multicellular Development. Princeton University Press. pp. 19–48.

Bonner, J.T., 2000b. Size and Evolution. In: First Signals, The Evolution of Multicellular Development. Princeton University Press. pp. 49–62.

Bonner, J.T., 2004. Perspective: The size-complexity rule. Evolution (N. Y). 58, 1883–1890.

Booth, D.S., Szmidt-Middleton, H., King, N., 2018. A robust method for transfection in choanoflagellates illuminates their cell biology and the ancestry of animal septins.

Bosher, J.M., Labouesse, M., 2000. RNA interference: genetic wandand genetic watchdog. Nat. Cell Biol. 2, 31–36.

Bråte, J., Neumann, R.S., Fromm, B., Haraldsen, A.A.B., Grini, P.E., Shalchian-Tabrizi, K., *in press.* Pre-metazoan origin of animal miRNAs.

Brown, L.E., Sprecher, S.L., Keller, L.R., 1991. Introduction of exogenous DNA into *Chlamydomonas reinhardtii* by electroporation. Mol. Cell. Biol. 11, 2328–2332.

Brown, M.W., Spiegel, F.W., Silberman, J.D., 2009. Phylogeny of the 'forgotten' cellular slime mold, *Fonticula alba,* reveals a key evolutionary branch within Opisthokonta. Mol. Biol. Evol. 26, 2699–2709.

Brunet, T., King, N., 2017. The Origin of Animal Multicellularity and Cell Differentiation. Dev. Cell 43, 124–140.

Budd, G.E., Jensen, S., 2017. The origin of the animals and a 'Savannah' hypothesis for early bilaterian evolution. Biol. Rev. 92, 446–473.

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

Buss, L.W., 1987. The Evolution of Individuality. Princeton University Press.

Butterfield, N., 2000. Bangiomorpha pubescens n. gen., n. sp.: implications for the evolution of sex, multicellularity, and the Mesoproterozoic/Neoproterozoic radiation of eukaryotes. Paleobiology 26, 386–404.

Butterfield, N.J., 2009. Modes of pre-Ediacaran multicellularity. Precambrian Res. 173, 201–211.

Butterfield, N.J., 2011. Terminal developments in ediacaran embryology. Science. 334, 1655–1656.

Cafaro, M.J., 2005. Eccrinales (Trichomycetes) are not fungi, but a clade of protists at the early divergence of animals and fungi. Mol. Phylogenet. Evol. 35, 21–34.

Calcium phosphate – mediated transfection of eukaryotic cells, 2005. Nat. Methods 2, 319–320.

Cardarelli, F., Digiacomo, L., Marchini, C., Amici, A., Salomone, F., Fiume, G., Rossetta, A., Gratton, E., Pozzi, D., Caracciolo, G., 2016. The intracellular trafficking mechanism of Lipofectamine-based transfection reagents and its implication for gene delivery. Sci. Rep. 6, 1–8.

Caro, F., Miller, M.G., DeRisi, J.L., 2012. Plate-based transfection and culturing technique for genetic manipulation of Plasmodium falciparum. Malar. J. 11, 1–10.

Carr, M., Leadbeater, B.S.C., Hassan, R., Nelson, M., Baldauf, S.L., 2008. Molecular phylogeny of choanoflagellates, the sister group to Metazoa. Proc. Natl. Acad. Sci. U. S. A. 105, 16641–16646.

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

Carroll, S.B., 2000. Endless forms: The evolution of gene regulation and morphological diversity. Cell 101, 577–580.

Carroll, S.B., 2001. Chance and necessity: The evolution of morphological complexity and diversity. Nature 409, 1102–1109.

Cavalier-Smith, T., 1987. The origin of fungi and pseudofungi. In: Evolutionary Biology of the Fungi. Rayner A. D.M., Braiser C.M., Moore D. (Eds). Cambridge University Press. pp. 339-353.

Cavalier-Smith, T., 1998. A revised six-kingdom system of life. Biol. Rev. 73, 203–266.

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.

Cavalier-Smith, T., Allsopp, M.T.E.P., 1996. *Corallochytrium,* an enigmatic non-flagellate protozoan related to choanoflagellates. Eur. J. Protistol. 32, 306–310.

Cavalier-Smith, T., Chao, E.E.Y., 2003. Phylogeny of choanozoa, apusozoa, and other protozoa and early eukaryote megaevolution. J. Mol. Evol. 56, 540–563.

Chen, L., Xiao, S., Pang, K., Zhou, C., Yuan, X., 2014. Cell differentiation and germ-soma separation in Ediacaran animal embryo-like fossils. Nature 516, 238–241.

Chow, Y.T., Chen, S., Wang, R., Liu, C., Kong, C.W., Li, R.A., Cheng, S.H., Sun, D., 2016. Single cell transfection through precise microinjection with quantitatively controlled injection volumes. Sci. Rep. 6, 1–9.

Chowdhury, E.H., Sasagawa, T., Nagaoka, M., Kundu, A.K., Akaike, T., 2003. Transfecting mammalian cells by DNA/calcium phosphate precipitates: Effect of temperature and pH on precipitation. Anal. Biochem. 314, 316–318.

Cloud, P.E., 1965. Significance of the Gunflint (Precambrian) Microflora. Science. 148, 27–35.

Cloud, P.E., 1968. Atmospheric and hydrospheric evolution on the primitive earth. Science. 160, 729–736.

Cloud, P.E., 1976. Beginnings of biospheric evolution and their biogeochemical consequences. Paleobiology 2, 351–387.

Cohen, P.A., Knoll, A.H., Kodner, R.B., 2009. Large spinose microfossils in Ediacaran rocks as resting stages of early animals. Proc. Natl. Acad. Sci. U. S. A. 106, 6519–6524.

Corliss, J.O., 1984. The kingdom PROTISTA and its 45 phyla. BioSystems 17, 87–126.

Cornell, R.A., Von Ohlen, T., 2000. Vnd/nkx, ind/gsh, and msh/msx: conserved regulators of dorsoventral neural patterning? Curr. Opin. Neurobiol. 10, 63–71.

Cunningham, J.A., Liu, A.G., Bengtson, S., Donoghue, P.C.J., 2017. The origin of animals: Can molecular clocks and the fossil record be reconciled? BioEssays 39, 1–12.

Damuth, J., Heisler, L., 1988. Alternative formulations of multilevel selection. Biol. Phylosofy 3, 407–430.

Davidson, E.H., 2001. Gene Regulatory Systems. In: Development and Evolution (Academic Press, San Diego).

Davidson, E.H., Erwin, D.H., 2006. Gene regulatory networks and the evolution of animal body plans. Science. 311, 796–797.

Dawkins, R., 1976. The Selfish Gene. Oxford University Press.

Dayel, M.J., Alegado, R.A., Fairclough, S.R., Levin, T.C., Nichols, S.A., McDonald, K., King, N., 2011. Cell differentiation and morphogenesis in the colony-forming choanoflagellate *Salpingoeca rosetta.* Dev. Biol. 357, 73–82.

Dayel, M.J., King, N., 2014. Prey capture and phagocytosis in the choanoflagellate *Salpingoeca rosetta.* PLoS One 9, 1–6.

de Mendoza, A., Ruiz-Trillo, I., 2014. Forward genetics for back-in-time questions. Elife 3, 1–3.

de Mendoza, A., Sebé-Pedrós, A., Sestak, M.S., Matejcic, M., Torruella, G., Domazet-Loso, T., Ruiz-Trillo, I., 2013. Transcription factor evolution in eukaryotes and the assembly of the regulatory toolkit in multicellular lineages. Proc. Natl. Acad. Sci. U. S. A. 4858–4866.

de Mendoza, A., Suga, H., Permanyer, J., Irimia, M., Ruiz-Trillo, I., 2015. Complex transcriptional regulation and independent evolution of fungal-like traits in a relative of animals. Elife 4, 1–26.

de Vargas, C., Audic, S., Henry, N., Decelle, J., Mahé, F., Logares, R., Lara, E., Berney, C., Bescot, N., Engelen, S., Hingamp, P., Sieracki, M., Probert, I., Carmichael, M., Poulain, J., Romac, S., 2015. Eukaryotic plankton diversity in the sunlit ocean. Science (80-. ). 348, 1261605–1/11.

Debroas, D., Hugoni, M., Domaizon, I., 2015. Evidence for an active rare biosphere within freshwater protists community. Mol. Ecol. 24, 1236–1247.

Degnan, B.M., Vervoort, M., Larroux, C., Richards, G.S., 2009. Early evolution of metazoan transcription factors. Curr. Opin. Genet. Dev. 19, 591–599.

del Campo, J., Mallo, D., Massana, R., de Vargas, C., Richards, T.A., Ruiz-Trillo, I., 2015. Diversity and distribution of unicellular opisthokonts along the European coast analysed using high-throughput sequencing. Environ. Microbiol. 17, 3195–3207.

del Campo, J., Massana, R., 2011. Emerging diversity within Chrysophytes, Choanoflagellates and Bicosoecids based on molecular surveys. Protist 162, 435–448.

del Campo, J., Ruiz-Trillo, I., 2013. Environmental survey meta-analysis reveals hidden diversity among unicellular opisthokonts. Mol. Biol. Evol. 30, 802–805.

del Campo, J., Sieracki, M.E., Molestina, R., Keeling, P., Massana, R., Ruiz-Trillo, I., 2014. The others: our biased perspective of eukaryotic genomes. Trends Ecol. Evol. 29, 252–259.

Desrosiers, C., Leflaive, J., Eulin, A., Ten-Hage, L., 2013. Bioindicators in marine waters: Benthic diatoms as a tool to assess water quality from eutrophic to oligotrophic coastal ecosystems. Ecol. Indic. 32, 25–34.

Dìez, B., Pedrós-alió, C., Massana, R., 2001. Study of Genetic Diversity of Eukaryotic Picoplankton in Different Oceanic Regions by Small-Subunit rRNA Gene Cloning and Sequencing Study of Genetic Diversity of Eukaryotic Picoplankton in Different Oceanic Regions by Small-Subunit rRNA Gene Cloning and. Appl. Environ. Microbiol. 67, 2932–2941.

Dobson, J., 2006. Gene therapy progress and prospects: Magnetic nanoparticle-based gene delivery. Gene Ther. 13, 283–287.

dos Reis, M., Thawornwattana, Y., Angelis, K., Telford, M.J., Donoghue, P.C.J., Yang, Z., 2015. Uncertainty in the Timing of Origin of Animals and the Limits of Precision in Molecular Timescales. Curr. Biol. 25, 2939–2950.

Du, Q., Kawabe, Y., Schilde, C., Chen, Z.H., Schaap, P., 2015. The Evolution of Aggregative Multicellularity and Cell-Cell Communication in the Dictyostelia. J. Mol. Biol. 427, 3722–3733.

Duffy, J.E., Morrison, C.L., Rios, R., 2000. Multiple origins of eusociality among sponge-dwelling shrimps (Synalpheus). Evolution (N. Y). 54, 503–516.

Dunn, F.S., Liu, A.G., Donoghue, P.C.J., 2018. Ediacaran developmental biology. Biol. Rev. 93, 914–932.

Eme, L., Sharpe, S.C., Brown, M.W., Roger, A.J., 2014. On the Age of Eukaryotes: Evaluating Evidence from Fossils and Molecular Clocks. Cold Spring Harb. Perspect. Biol. 6, 1–16.

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

Erwin, D.H., Laflamme, M., Tweedt, S.M., Sperling, E.A., Pisani, D., Peterson, K.J., 2011. The Cambrian conundrum: Early divergence and later ecological success in the early history of animals. Science. 334, 1091–1097.

Escalante, R., Vicente, J.J., 2000. *Dictyostelium discoideum:* A model system for differentiation and patterning. Int. J. Dev. Biol. 44, 819–835.

Fairclough, S.R., Chen, Z., Kramer, E., Zeng, Q., Young, S., Robertson, H.M., Begovic, E., Richter, D.J., Russ, C., Westbrook, M.J., Manning, G., Lang, B.F., Haas, B., Nusbaum, C., King, N., 2013. Premetazoan genome evolution and the regulation of cell differentiation in the choanoflagellate *Salpingoeca rosetta.*

Fairclough, S.R., Dayel, M.J., King, N., 2010. Multicellular development in a choanoflagellate. Curr. Biol. 20, R875-876.

Fedonkin, M.A., Gehling, J.G., Grey, K., Narbonne, G.M., Vickers-Rich, P., 2007. The rise of animals: Evolution and diversification of the Kingdom Animalia.

Feldman, S.H., Wimsatt, J.H., Green, D.E., 2005. Phylogenetic Classification of the Frog Pathogen *Amphibiothecum* (Dermosporidium) Penneri Based on Small Ribosomal Subunit Sequencing. J. Wildl. Dis. 41, 701–706.

Felgner, P.L., 1990. Particulate systems and polymers for in vitro and in vivo delivery of polynucleotides. Adv. Drug Deliv. Rev. 5, 163–187.

Felgner, P.L., Gadek, T.R., Holm, M., Roman, R., Chan, H.W., Wenz, M., Northrop, J.P., Ringold, G.M., Danielsen, M., 1987. Lipofection: A highly efficient, lipid-mediated DNA-transfection procedure 84, 7413–7417.

Ferrier, D.E.K., 2015. The origin of the *Hox/ParaHox* genes, the Ghost Locus hypothesis and the complexity of the first animal. Brief. Funct. Genomics 15, 333–341.

Ford Doolittle, W., 1998. You are what you eat: A gene transfer ratchet could account for bacterial genes in eukaryotic nuclear genomes. Trends Genet. 14, 307–311.

Garrett, P., 1970. Phanerozoic Stromatolites: Noncompetitive ecologic restriction by grazing and burrowing animals. Science. 169, 171–173.

Gaudet, P., Pilcher, K.E., Fey, P., Chisholm, R.L., 2007. Transformation of *Dictyostelium discoideum* with plasmid DNA. Nat. Protoc. 2, 1317–24.

Gavelis, G.S., Wakeman, K.C., Tillmann, U., Ripken, C., Mitarai, S., Herranz, M., Özbek, S., Holstein, T., Keeling, P.J., Leander, B.S., 2017. Microbial arms race: Ballistic "nematocysts" in dinoflagellates represent a new extreme in organelle complexity. Sci. Adv. 3.

Gerhart, J., Kirscher, M., 1997. Cells, Embryos and Evolution. Blackwell Science.

Gladfelter, A.S., Hungerbuehler, A.K., Philippsen, P., 2006. Asynchronous nuclear division cycles in multinucleated cells. J. Cell Biol. 172, 347–362.

Glockling, S.L., Marshall, W.L., Gleason, F.H., 2013. Phylogenetic interpretations and ecological potentials of the Mesomycetozoea (Ichthyosporea). Fungal Ecol. 6, 237–247.

Goldstein, B., King, N., 2016. The Future of Cell Biology: Emerging Model Organisms. Trends Cell Biol. 26, 818–824.

Gong, J.H., Nicholls, E.F., Elliott, M.R., Brown, K.L., Hokamp, K., Roche, F.M., Cheung, C.Y.K., Falsafi, R., Brinkman, F.S.L., Bowdish, D.M.E., Hancock, R.E.W., 2010. G-protein-coupled receptor independent, immunomodulatory properties of chemokine CXCL9. Cell. Immunol. 261, 105–113.

González-Hernández, M., Denoël, M., Duffus, A.J.L., Garner, T.W.J., Cunningham, A.A., Acevedo-Whitehouse, K., 2010. Dermocystid infection and associated skin lesions in free-living palmate newts *(Lissotriton helveticus)* from Southern France. Parasitol. Int. 59, 344–350.

Gooday, a J., Nott, J. a, 1982. Intracellular Barite Crystals in 2 Xenophyophores, Aschemonella-Ramuliformis and Galatheammina Sp (Protozoa, Rhizopoda) with Comments on the Taxonomy of Aschemonella-Ramuliformis. J. Mar. Biol. Assoc. United Kingdom 62, 595–605.

Graham, F.L., 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., Ruiz-Trillo, I., 2017. Dynamics of genomic innovation in the unicellular ancestry of animals. Elife 6, 1–35.

Grazhdankin, D., 2014. Patterns of Evolution of the Ediacaran Soft-Bodied Biota. J. Paleontol. 88, 269–283.

Grosberg, R.K., Strathmann, R.R., 2007. The Evolution of Multicellularity: A Minor Major Transition? Annu. Rev. Ecol. Evol. Syst. 38, 621–654.

Grosjean, F., Bertschinger, Æ.M., Hacker, D.L., Wurm, Æ.F.M., 2006. Multiple glycerol shocks increase the calcium phosphate transfection of non-synchronized CHO cells 1827–1833.

Grossman, A.R., Harris, E.E., Hauser, C., Lefebvre, P.A., Martinez, D., Rokhsar, D., Shrager, J., Sil, C.D., Stern, D., Vallon, O., Zhang, Z., 2003. *Chlamydomonas reinhardtii* at the Crossroads of Genomics. Eukaryot. Cell 2, 1137–1150.

Grotzinger, J., Knoll, A., 1999. Stromatolites in Precambrian carbonates: evolutionary mileposts or environmental dipsticks? Annu. Rev. Earth Planet. Sci. 27, 313–358.

Haeckel, E., 1874. Die Gastraea Theorie, Die Phylogenetische Klassifikation Des Thierreichs Und Die Homologie Der Keimblätter. Jenaische Z. Naturwiss 8, 1–55.

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.

Hamdi, B., Brasier, M.D., Zhiwen, J., 1989. Earliest skeletal fossils from Precambrian-Cambrian boundary strata, Elburz Mountains, Iran. Geol. Mag. 126, 283–289.

Harris, E.H., 2001. *Chlamydomonas* as a Model Organism. Annu. Rev. Physiol. Plant Mol. Biol. 52, 363–406.

Hassett, B.T., López, J.A., Gradinger, R., 2015. Two New Species of Marine Saprotrophic Sphaeroformids in the Mesomycetozoea Isolated From the Sub-Arctic Bering Sea. Protist 166, 310–322.

Hastings, M.D., Sibley, C.H., 2002. Pyrimethamine and WR99210 exert opposing selection on dihydrofolate reductase from *Plasmodium vivax.* Proc. Natl. Acad. Sci. U. S. A. 99, 13137–

13141.

Hehenberger, E., Tikhonenkov, D. V., Kolisko, M., del Campo, J., Esaulov, A.S., Mylnikov, A.P., Keeling, P.J., 2017. Novel Predators Reshape Holozoan Phylogeny and Reveal the Presence of a Two-Component Signaling System in the Ancestor of Animals. Curr. Biol. 27, 2043–2050.e6.

Heisler, L., Damuth, J., 1987. A Method for Analyzing Selection in Hierarchically Structured Populations. Am. Nat. 130, 582–602.

Herr, R.A., Ajello, L., Taylor, J.W., Arseculeratne, S.N., Mendoza, L., 1999. Phylogenetic analysis of Rhinosporidium seeberi's 18S small-subunit ribosomal DNA groups this pathogen among members of the protoctistan Mesomycetozoa clade. J. Clin. Microbiol. 37, 2750–2754.

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 Mesomycetozoea. Int. J. Parasitol. 32, 1183–91.

Hinman, V.F., Nguyen, A.T., Cameron, R.A., Davidson, E.H., 2003. Developmental gene regulatory network architecture across 500 million years of echinoderm evolution. Proc. Natl. Acad. Sci. U. S. A. 100, 13356–61.

Hirth, F., Kammermeier, L., Frei, E., Walldorf, U., Noll, M., Reichert, H., 2003. An urbilaterian origin of the tripartite brain: developmental genetic insights from Drosophila. Development 130, 2365–2373.

Hoffmeyer, T.T., Burkhardt, P., 2016. Choanoflagellate models — *Monosiga brevicollis* and *Salpingoeca rosetta.* Curr. Opin. Genet. Dev. 39, 42–47.

Holmen, S.L., Vanbrocklin, M.W., Eversole, R.R., Stapleton, S.R., Ginsberg, L.C., 1995. Efficient lipid-mediated transfection of DNA into primary rat hepatocytes. Vitr. Cell Dev. Biol. 30, 347–351.

Holter, W., Fordis, M.C., Howard, B.H., 1989. Efficient Gene Transfer by Sequential Treatment of Mammalian Cells with DEAE-Dextran and Deoxyribonucleic Acid. Exp. Cell Res. 184, 546–551.

Hyman, L.H., 1940. The invertebrates: protozoa through Ctenophora (1st Ed.), McGraw-Hill, New

York.

Ispolatov, I., Ackermann, M., Doebeli, M., 2012. Division of labour and the evolution of multicellularity. Proc. R. Soc. B Biol. Sci. 279, 1768–1776.

Ivics, Z., Hiripi, L., Hoffmann, O.I., Mátés, L., Yau, T.Y., Bashir, S., Zidek, V., Landa, V., Geurts, A., Pravenec, M., Rülicke, T., Bösze, Z., Izsvák, Z., 2014. Germline transgenesis in rabbits by pronuclear microinjection of Sleeping Beauty transposons. Nat. Protoc. 9, 794–809.

Jakobiak, T., Mages, W., Scharf, B., Babinger, P., Stark, K., Schmitt, R., 2004. The bacterial paromomycin resistance gene, aphH, as a dominant selectable marker in *Volvox carteri.* Protist 155, 381–393.

James-Clark, H., 1866. Conclusive proofs on the Animality of the ciliat Sponges, and of their affinities with the Infusioria Flagellata. Am. J. Sci. 2, 320–325.

Janse, C.J., Franke-Fayard, B., Mair, G.R., Ramesar, J., Thiel, C., Engelmann, S., Matuschewski, K., Gemert, G.J. Van, Sauerwein, R.W., Waters, A.P., 2006a. High efficiency transfection of *Plasmodium berghei* facilitates novel selection procedures. Mol. Biochem. Parasitol. 145, 60–70.

Janse, C.J., Ramesar, J., Waters, A.P., 2006b. High-efficiency transfection and drug selection of genetically transformed blood stages of the rodent malaria parasite *Plasmodium berghei.* Nat. Protoc. 1, 346–356.

Janssen, B.D., Chen, Y.P., Molgora, B.M., Wang, S.E., Simoes-Barbosa, A., Johnson, P.J., 2018. CRISPR/Cas9-mediated gene modification and gene knock out in the human-infective parasite *Trichomonas vaginalis.* Sci. Rep. 8, 1–14.

Jordan, M., Schallhorn, A., Wurm, F.M., Francisco, S.S., 1996. Transfecting mammalian cells : optimization of critical parameters affecting calcium-phosphate precipitate formation 24, 596–601.

Jordan, M., Wurm, F., 2004. Transfection of adherent and suspended cells by calcium phosphate 33, 136–143.

Kaiser, D., 2001. Building a Multicellular Organism. Annu. Rev. Genet. 35, 103–123.

Kasting, J.F., Siefert, J.L., 2002. Life and the evolution of Earth's atmosphere. Science. 296, 1066–1068.

Kawai, S., Hashimoto, W., Murata, K., 2010. Transformation of *Saccharomyces cerevisiae* and other fungi: methods and possible underlying mechanism. Bioeng. Bugs 1, 395–403.

Keeling, P.J., Burger, G., Durnford, D.G., Lang, B.F., Lee, R.W., Pearlman, R.E., Roger, A.J., Gray, M.W., 2005. The tree of eukaryotes. Trends Ecol. Evol. 20, 670–676.

Kiel, J.A.K.W., 2010. Autophagy in unicellular eukaryotes. Philos. Trans. R. Soc. B Biol. Sci. 365, 819–830.

Kim, H.J., Greenleaf, J.F., Kinnick, R.R., Bronk, J.T., Bolander, M.E., 1996. Ultrasound-Mediated Transfection of Mammalian cells. Hum. Gene Ther. 7, 1339–1346.

Kim, T.K., Eberwine, J.H., 2010. Mammalian cell transfection: The present and the future. Anal. Bioanal. Chem. 397, 3173–3178.

Kindle, K.L., Schnell, R.A., Fernández, E., Lefebvre, P.A., 1989. Stable nuclear tranformation of *Chlamydomonas* using the *Chlamydomonas* gene for Nitrate Reductase. J. Cell Biol. 109, 2589–2601.

King, N., 2004. The unicellular ancestry of animal development. Dev. Cell 7, 313–25.

King, N., 2005. Choanoflagellates. Curr. Biol. 15, 113–114.

King, N., Hittinger, C.T., Carroll, S.B., 2003. Evolution of key cell signaling and adhesion protein families predates animal origins. Science. 301, 361–363.

King, N., Westbrook, M.J., Young, S.L., Kuo, A., Abedin, M., Chapman, J., Fairclough, S., Hellsten, U., Isogai, Y., Letunic, I., Marr, M., Pincus, D., Putnam, N., Rokas, A., Wright, K.J., Zuzow, R., Dirks, W., Good, M., Goodstein, D., Lemons, D., Li, W., Lyons, J.B., Morris, A., Nichols, S., Richter, D.J., Salamov, A., Bork, P., Lim, W.A., Manning, G., Miller, W.T., McGinnis, W., Shapiro, H., Tjian, R., Grigoriev, I. V., Rokhsar, D., 2008. The genome of the choanoflagellate *Monosiga brevicollis* and the origin of metazoans. Nature 451, 783–788.

Kirk, D.L., 2003. Seeking the Ultimate and Proximate Causes of Volvox Multicellularity and Cellular Differentiation. Integr. Comp. Biol. 43, 247–253.

Klein, T.M., Fromm, M., Weissinger, A., Tomes, D., Schaaf, S., Sletten, M., Sanford, J.C., 1988a. Transfer of foreign genes into intact maize cells with high-velocity microprojectiles. Proc. Natl. Acad. Sci. U. S. A. 85, 4305–4309.

Klein, T.M., Harper, E.C., Svab, Z., Sanford, J.C., Fromm, M.E., Maliga, P., 1988b. Stable genetic transformation of intact Nicotiana cells by the particle bombardment process. Proc. Natl. Acad. Sci. U. S. A. 85, 8502–8505.

Klein, T.M., Sanford, J.C., Wolf, E.D., Wu, R., 1987. High-velocity microprojectiles for delivering nucleic acids into living cells. Nature 327, 70–73.

Knauth, L.P., 2005. Temperature and salinity history of the Precambrian ocean: implications for the course of microbial evolution. Palaeogeogr. Palaeoclimatol. Palaeoecol. 219, 53–69.

Knoll, A.H., 1992. Biological and biogeochemical preludes to the ediacaran radiation. In: Origin and Early Evolution of the Metazoa. New York Plenum Press. Lipps JH, Signor PW (Eds).

Knoll, A.H., 2011. The Multiple Origins of Complex Multicellularity.

Knoll, A.H., 2014. Paleobiological perspectives on early microbial evolution. Cold Spring Harb. Perspect. Biol. 7, 1–17.

Knoll, A.H., Carroll, S.B., 1999. Early animal evolution: Emerging views from comparative biology and geology. Science. 284, 2129–2137.

Knoll, A.H., Hewitt, D., 2011. Phylogenetic, Functional, and Geological Perspectives on Complex Multicellularity. In: The Major Transitions in Evolution Revisited (Eds. Brett Calcott and Kim Sterenly). pp. 251–270.

Konijn, T.M., Raper, 1961. Cell Aggregation in Dictyostelium discoideum. Public Health 756, 725–756.

Kumar, P., Nagarajan, A., Uchil, P.D., 2018. DEAE-Dextran transfection. Cold Spring Harb. Protoc. 2018, 498–500.

Lane, N., Martin, W., 2010. The energetics of genome complexity. Nature 467, 929–34.

Lang, B.F., O'Kelly, C., Nerad, T., Gray, M.W., Burger, G., 2002. The closest unicellular relatives of animals. Curr. Biol. 12, 1773–1778.

Lankester, E.R., 1873. On the primitive cell layers of the embryo as the basis of the genealogical classification of animals, and on the origin of the vascular and lymph systems. Ann. Mag. Nat. Hist. 4, 321–328.

Lankester, E.R., 1877. Notes on the embryology and classification of the animal kingdom: comprising a revision of speculations relative to the origin and significance of germ layers. Q. J. Microsc. Soc. 17, 399–454.

Larroux, C., Fahey, B., Liubicich, D., Hinman, V.F., Gauthier, M., Gongora, M., Green, K., Worheide, G., Leys, S.P., Degnan, B.M., 2006. Developmental expression of transcription factor genes in a demosponge: insights into the origin of metazoan multicellularity. Evol Dev 8, 150–173.

Larroux, C., Luke, G.N., Koopman, P., Rokhsar, D.S., Shimeld, S.M., Degnan, B.M., 2008. Genesis and expansion of metazoan transcription factor gene classes. Mol. Biol. Evol. 25, 980–996.

Leadbeater, B.S.C., 2015. The choanoflagellates: Evolution, biology and ecology.

Lerche, K., Hallmann, A., 2009. Stable nuclear transformation of *Gonium pectorale.* BMC Biotechnol. 9, 64.

Lerche, K., Hallmann, A., 2013. Stable nuclear transformation of *Eudorina elegans.* BMC Biotechnol. 13, 11.

Lerche, K., Hallmann, A., 2014. Stable nuclear transformation of *Pandorina morum.* BMC Biotechnol. 14, 65.

Levin, M., Anavy, L., Cole, A.G., Winter, E., Mostov, N., Khair, S., Senderovich, N., Kovalev, E., Silver, D.H., Feder, M., Fernandez-valverde, S.L., Nakanishi, N., Simmons, D., Simakov, O., Larsson, T., Liu, S., Jerafi-vider, A., Yaniv, K., Ryan, J.F., Martindale, M.Q., Rink, J.C., Arendt, D., Degnan, S.M., Degnan, B.M., Hashimshony, T., Yanai, I., 2016. The mid-developmental transition and the evolution of animal body plans. Nature.

Levin, T.C., Greaney, A.J., Wetzel, L., King, N., 2014. The rosetteless gene controls development in the choanoflagellate *S. rosetta.* Elife 3, 1–23.

Levin, T.C., King, N., 2013. Report Evidence for Sex and Recombination in the Choanoflagellate *Salpingoeca rosetta.* Curr. Biol. 23, 2176–2180.

Lewandowska, K., Choi, H.U., Rosenberg, L.C., Sasse, J., Neame, P.J., Culp, L.A., 1991. Extracellular matrix adhesion-promoting activities of a dermatan sulfate proteoglycan-associated protein (22K) from bovine fetal skin. J. Cell Sci. 99, 657–668.

Lichtneckert, R., Reichert, H., 2005. Insights into the urbilaterian brain: conserved genetic patterning mechanisms in insect and vertebrate brain development. Heredity (Edinb). 94, 465–477.

Lieber, M.R., Hesse, J.E., Nickol, J.M., Felsenfeld, G., 1987. The mechanism of osmotic transfection of avian embryonic erythrocytes: Analysis of a system for studying developmental gene expression. J. Cell Biol. 105, 1055–1065.

Ling, G., Liyang, W., Ronghua, Y., Rui, F., Zhongguang, L., Nishi, M., Qi, Z., Isaacs, W., Ma, J., Xuehong, X., 2017. Optimizing conditions for calcium phosphate mediated transient transfection. Saudi J. Biol. Sci. 24, 622–629.

Liu, Y., Steenkamp, E.T., Brinkmann, H., Forget, L., Philippe, H., 2009. Phylogenomic analyses predict sistergroup relationship of nucleariids and fungi and paraphyly of zygomycetes with significant support. BMC Evol. Biol. 9, 272.

Lohr, J.N., Laforsch, C., Koerner, H., Wolinska, J., 2010. A daphnia parasite *(Caullerya mesnili)* constitutes a new member of the ichthyosporea, a group of protists near the animal-fungi divergence. J. Eukaryot. Microbiol. 57, 328–336.

López-García, P., Rodríguez-Valera, F., Pedrós-Alió, C., Moreira, D., 2001. Unexpected diversity of small eukaryotes in deep-sea Antarctic plankton. Nature 409, 603.

Lord, J.C., Hartzer, K.L., Kambhampati, S., 2012. A nuptially transmitted ichthyosporean symbiont of *Tenebrio molitor* (Coleoptera: Tenebrionidae). J. Eukaryot. Microbiol. 59, 246–250.

Love, G.D., Grosjean, E., Stalvies, C., Fike, D.A., Grotzinger, J.P., Bradley, A.S., Kelly, A.E., Bhatia, M., Meredith, W., Snape, C.E., Bowring, S.A., Condon, D.J., Summons, R.E., 2009. Fossil steroids record the appearance of Demospongiae during the Cryogenian period. Nature 457, 718.

Lowe, C.J., Wu, M., Salic, A., Evans, L., Lander, E., Stange-Thomann, N., Gruber, C.E., Gerhart, J., Kirschner, M., 2003. Anteroposterior patterning in hemichordates and the origins of the chordate nervous system. Cell 113, 853–865.

Lynch, M., Conery, J.S., 2003. The Origins of Genome Complexity. Science. 302, 1401–1404.

Lyons, T.W., Reinhard, C.T., Planavsky, N.J., 2014a. The rise of oxygen in Earth's early ocean and atmosphere. Nature 506, 307–315.

Lyons, T.W., Reinhard, C.T., Planavsky, N.J., 2014b. Evolution: A fixed-nitrogen fix in the early ocean? Curr. Biol. 24, R276–R278.

Maloof, A.C., Rose, C. V., Beach, R., Samuels, B.M., Calmet, C.C., Erwin, D.H., Poirier, G.R., Yao, N., Simons, F.J., 2010. Possible animal-body fossils in pre-Marinoan limestones from South Australia. Nat. Geosci. 3, 653–659.

Marshall, W.L., Berbee, M.L., 2011. Facing Unknowns: Living cultures (*Pirum gemmata* gen. nov., sp. nov., and *Abeoforma whisleri*, gen. nov., sp. nov.) from invertebrate digestive tracts represent an undescribed clade within the unicellular opisthokont lineage ichthyosporea (Mesomycetozoea). Protist 162, 33–57.

Marshall, W.L., Celio, G., McLaughlin, D.J., 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, 415–33.

Matt, G., Umen, J., 2016. *Volvox*: a simple algal model for embryogenesis, morphogenesis and cellular differentiation. Dev. Biol. 419, 99–113.

Mazmanian, S.K., Cui, H.L., Tzianabos, A.O., Kasper, D.L., 2005. An Immunomodulatory molecule of symbiotic bacteria directs maturation of the Host Immune System. Cell 122, 107–118.

McCutchan, J.H., Pagano, J.S., 1968. Enchancement of the infectivity of simian virus 40 deoxyribonucleic acid with diethylaminoethyl-dextran. J. Natl. Cancer Inst. 41, 351–357.

McFall-Ngai, M.J., Ruby, E.G., 2000. Developmental biology in marine invertebrate symbioses. Curr. Opin. Microbiol. 3, 603–607.

Medina, M., Collins, A.G., Taylor, J.W., Valentine, J.W., Lipps, J.H., Amaral-Zettler, L., Sogin, M.L., 2003. Phylogeny of Opisthokonta and the evolution of multicellularity and complexity in Fungi and Metazoa. Int. J. Astrobiol. 2, 203–211.

Medlin, L., Elwood, H.J., Stickel, S., Sogin, M.L., 1988. The characterization of enzymatically amplified eukaryotic 16S-like rRNA-coding regions. Gene 71, 491–499.

Mendoza, L., Ajello, L., Taylor, J.W., 2001. The taxonomic status of *Lacazia loboi* and *Rhinosporidium seeberi* has been finally resolved with the use of molecular tools. Rev. Iberoam. Micol. organo la Asoc. Esp. Espec. en Micol. 18, 95–98.

Mendoza, L., Taylor, J.W., Ajello, L., 2002. The Class Mesomycetozoea: A Heterogeneous Group of Microorganisms at the Animal-Fungal Boundary. Annu. Rev. Microbiol. 56, 315–344.

Metchnikoff, E., 1886. Embryologische Studien an Medusen: Ein Beitrag zur Genealogie der Primitivorgane, Wien, A. Holder.

Michod, R.E., 2006. The group covariance effect and fitness trade-offs during evolutionary transitions in individuality. Proc. Natl. Acad. Sci. U. S. A. 103, 9113–9117.

Michod, R.E., 2007. Evolution of individuality during the transition from unicellular to multicellular life. Proc. Natl. Acad. Sci. U. S. A. 104, 8613–8618.

Michod, R.E., Herron, M.D., 2006. Cooperation and conflict during evoluitonary transitions in individuality. J. Evol. Biol. 19, 1406-1409.

Michod, R.E., Viossat, Y., Solari, C.A., Hurand, M., Nedelcu, A.M., 2005. Life-history evolution and the origin of multicellularity. J. Theor. Biol. 239, 257–272.

Mikhailov, K. V., Konstantinova, A. V., Nikitin, M.A., Troshin, P. V., Rusin, L.Y., Lyubetsky, V.A., Panchin, Y. V., Mylnikov, A.P., Moroz, L.L., Kumar, S., Aleoshin, V. V., 2009. The origin of Metazoa: A transition from temporal to spatial cell differentiation. BioEssays 31, 758–768.

Mills, D.B., Canfield, D.E., 2014. Oxygen and animal evolution: Did a rise of atmospheric oxygen ‘trigger’ the origin of animals? BioEssays 36, 1145–1155.

Mills, D.B., Ward, L.M., Jones, C., Sweeten, B., Forth, M., Treusch, A.H., Canfield, D.E., 2014. Oxygen requirements of the earliest animals. Proc. Natl. Acad. Sci. U. S. A. 111, 4168–

4172.

Mojzsis, S.J., Arrhenius, G., McKeegan, K.D., Harrison, T.M., Nutman, A.P., Friend, C.R.L., 1996. Evidence for life on Earth before 3,800 million years ago. Nature 384, 55–59.

Montgomery, M.K., McFall-Ngai, M., 1994. Bacterial symbionts induce host organ morphogenesis during early postembryonic development of the squid *Euprymna scolopes.* Development 120, 1719–1729.

Nair, D.R., D'Ausilio, C.A., Occhipinti, P., Borsuk, M.E., Gladfelter, A.S., 2010. A conserved G1 regulatory circuit promotes asynchronous behavior of nuclei sharing a common cytoplasm. Cell Cycle 9, 3771–3779.

Narbonne, G.M., 2005. The Ediacara Biota: Neoproterozoic Origin of Animals and Their Ecosystems. Annu. Rev. Earth Planet. Sci. 33, 421–442.

Narbonne, G.M., Gehling, J.G., 2003. Life after snowball: The oldest complex Ediacaran fossils. Geology 31, 27–30.

Nedelcu, A.M., 2009. Environmentally induced responses co-opted for reproductive altruism. Biol. Lett. 5, 805–808.

Nellen, W., Silan, C., Firtel, R.A., 1984. DNA-Mediated Transformation in *Dictyostelium discoideum*: Regulated Expression of an Actin Gene Fusion 4, 2890–2898.

Neumann, E., Schaefer-Ridder, M., Wang, Y., Hofschneider, P.H., 1982. Gene transfer into mouse lyoma cells by electroporation in high electric fields. EMBO J. 1, 841–845.

Newman, S.A., 2012. Physico-genetic determinants in the evolution of development. Science. 338, 217–219.

Nielsen, C., 2008. Six major steps in animal evolution: Are we derived sponge larvae? Evol. Dev. 10, 241–257.

Nitsche, F., Carr, M., Arndt, H., Leadbeater, B.S.C., 2011. Higher level taxonomy and molecular phylogenetics of the Choanoflagellatea. J. Eukaryot. Microbiol. 58, 452–462.

Nitsche, F., Weitere, M., Scheckenbach, F., Hausman, K., Wylezich, C., Arndt, H., 2007. Deep

Sea Records of Choanoflagellates with a Description of Two New Species. Acta Protozool. 46, 99–106.

Nursall, J.R., 1959. Oxygen as a prerequisite to the origin of the Metazoa. Nature 183, 1170–1172.

Olive, L.S., Blanton, R.L., 1980. Aerial Sorocarp Development by the Aggregative Ciliate, *Sorogena stoianovitchae.* J. Protozool. 27, 293–299.

Oliverio, A.M., Power, J.F., Washburne, A., Cary, S.C., Stott, M.B., Fierer, N., 2018. The ecology and diversity of microbial eukaryotes in geothermal springs. ISME J. 12, 1918–1928.

Ondracka, A., Dudin, O., Ruiz-Trillo, I., 2018. Decoupling of Nuclear Division Cycles and Cell Size during the Coenocytic Growth of the Ichthyosporean *Sphaeroforma arctica.* Curr. Biol. 28, 1964–1969.e2.

Otsuka, J., 2008. A theoretical approach to the large-scale evolution of multicellularity and cell differentiation. J. Theor. Biol. 255, 129–136.

Ou, Q., Xiao, S., Han, J., Sun, G., Zhang, F., Zhang, Z., Shu, D., 2015. A vanished history of skeletonization in Cambrian comb jellies. Sci. Adv. 1, 1–9.

Owczarzak, A., Stibbs, H.H., Bayne, C.J., 1980. The destruction of *Schistosoma mansoni* mother sporocysts in vitro by amoebae isolated from *Biomphalaria glabrata:* an ultrastructural study. J. Invertebr. Pathol. 35, 26–33.

Paps, J., Medina-Chacón, L. a, 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.

Patterson, D.J., Nygaard, K., Steinberg, G., Turley, C.M., 1993. Heterotrophic flagellates and other protists associated with oceanic detritus throughout the water column in the mid north atlantic. J. Mar. Biol. Assoc. United Kingdom 73, 67–95.

Pawlowski, J., Audic, S., Adl, S., Bass, D., Belbahri, L., Berney, C., Bowser, S.S., Cepicka, I., Decelle, J., Dunthorn, M., Fiore-Donno, A.M., Gile, G.H., Holzmann, M., Jahn, R., Jirků, M., Keeling, P.J., Kostka, M., Kudryavtsev, A., Lara, E., Lukeš, J., Mann, D.G., Mitchell, E.A.D.,

Nitsche, F., Romeralo, M., Saunders, G.W., Simpson, A.G.B., Smirnov, A. V., Spouge, J.L., Stern, R.F., Stoeck, T., Zimmermann, J., Schindel, D., de Vargas, C., 2012. CBOL Protist Working Group: Barcoding Eukaryotic Richness beyond the Animal, Plant, and Fungal Kingdoms. PLoS Biol. 10.

Pawlowski, J., Lejzerowicz, F., Apotheloz-Perret-Gentil, L., Visco, J., Esling, P., 2016. Protist metabarcoding and environmental biomonitoring: Time for change, European Journal of Protistology. Elsevier GmbH.

Pereira, C.N., Rosa, I. Di, Fagotti, A., Simoncelli, F., Pascolini, R., Mendoza, L., 2005. The Pathogen of Frogs *Amphibiocystidium ranae* Is a Member of the Order Dermocystida in the Class Mesomycetozoea. J. Clin. Microbiol. 43, 192–198.

Peter, I.S., Davidson, E.H., 2011. Evolution of gene regulatory networks controlling body plan development. Cell 144, 970–985.

Pfeifer, A., Verma, I.M., 2001. Gene therapy: Promises and Problems. Annu. Rev. Genomics Hum. Genet. 2, 177–211.

Phillips, K., Luisi, B., 2000. The virtuoso of versatility: POU proteins that flex to fit. J. Mol. Biol. 302, 1023–1039.

Plank, C., Schillinger, U., Scherer, F., Bergemann, C., Rémy, J.-S., Krötz, F., Anton, M., Lausier, J., Rosenecker, J., 2003. The magnetofection method: using magnetic force to enhance gene delivery. Biol. Chem. 384, 737–747.

Potter, H., Heller, R., 2010. Transfection by electroporation. Curr. Protoc. Mol. Biol. 9.3.1-9.3.10.

Potter, H., Weir, L., Leder, P., 1984. Enhancer-dependent expression of human kappa immunoglobulin genes introduced into mouse pre-B lymphocytes by electroporation. Proc. Natl. Acad. Sci. U. S. A. 81, 7161–7165.

Prelich, G., 2012. Gene overexpression: Uses, mechanisms, and interpretation. Genetics 190, 841–854.

Prochnik, S.E., Umen, J., Nedelcu, A.M., Hallmann, A., Miller, S.M., Nishii, I., Ferris, P., Kuo, A., Mitros, T., Fritz-Laylin, L.K., Hellsten, U., Chapman, J., Simakov, O., Rensing, S.A., Terry, A., Pangilinan, J., Kapitonov, V., Jurka, J., Salamov, A., Shapiro, H., Schmutz, J.,

Grimwood, J., Lindquist, E., Lucas, S., Grigoriev, I. V., Schmitt, R., Kirk, D., Rokhsar, D.S., 2010. Genomic analysis of organismal complexity in the multicellular green alga *Volvox carteri.* Science. 329, 223–226.

Putnam, N.H., Srivastava, M., Hellsten, U., Dirks, B., Chapman, J., Salamov, A., Terry, A., Shapiro, H., Lindquist, E., Kapitonov, V. V., Jurka, J., Genikhovich, G., Grigoriev, I. V., Lucas, S.M., Steele, R.E., Finnerty, J.R., Technau, U., Martindale, M.Q., Rokhsar, D.S., 2007. Sea anemone genome reveals ancestral eumetazoan gene repertoire and genomic organization. Science. 317, 86–94.

Raff, M.C., 1996. Size Control: The Regulation of Cell Numbers in Animal Development. Cell 86, 173–175.

Ragan, M.A., Goggin, C.L., Cawthorn, R.J., Cerenius, L., Jamieson, A. V, Plourde, S.M., Rand, T.G., Söderhäll, K., Gutell, R.R., 1996. A novel clade of protistan parasites near the animal-fungal divergence. Proc. Natl. Acad. Sci. U. S. A. 93, 11907–12.

Raghu-kumar, S., 1987. Occurrence of the Thraustochytrid, Corallochytrium limacisporum gen. et sp. nov. in the Coral Reef Lagoons of the Lakshadweep Islands in the Arabian Sea. Bot. Mar. 30, 83–90.

Ran, F.A., Hsu, P.D., Wright, J., Agarwala, V., Scott, D.A., Zhang, F., 2013. Genome engineering using the CRISPR-Cas9 system. Nat. Protoc. 8, 2281–2308.

Rasala, B.A., Mayfield, S.P., 2011. The microalga *Chlamydomonas reinhardtii* as a platform for the production of human protein therapeutics. Bioeng. Bugs 2, 50–54.

Raz, A.A., Srivastava, M., Salvamoser, R., Reddien, P.W., 2017. Acoel regeneration mechanisms indicate an ancient role for muscle in regenerative patterning. Nat. Commun. 8, 1–8.

Recillas-Targa, F., 2006. Multiple strategies for gene transfer, expression, knockdown, and chromatin influence in mammalian cell lines and transgenic animals. Mol. Biotechnol. 34, 337–354.

Reitsema, R.E., Meire, P., Schoelynck, J., 2018. The Future of Freshwater Macrophytes in a Changing World: Dissolved Organic Carbon Quantity and Quality and Its Interactions With Macrophytes. Front. Plant Sci. 9, 1–15.

Richards, G.S., Degnan, B.M., 2009. The dawn of developmental signaling in the metazoa. Cold

Spring Harb. Symp. Quant. Biol. 74, 81–90.

Richter, D.J., Fozouni, P., Eisen, M.B., King, N., 2018. Gene family innovation, conservation and loss on the animal stem lineage. Elife 7, 1–43.

Richter, D.J., King, N., 2013. The Genomic and Cellular Foundations of Animal Origins. Annu. Rev. Genet. 47, 509–537.

Robertson, A.J., Larroux, C., Degnan, B.M., Coffman, J. a, 2009. The evolution of *Runx* genes II. The C-terminal *Groucho* recruitment motif is present in both eumetazoans and homoscleromorphs but absent in a haplosclerid demosponge. BMC Res. Notes 2, 59.

Rokas, A., 2008a. The molecular origins of multicellular transitions.

Rokas, A., 2008b. The Origins of Multicellularity and the Early History of the Genetic Toolkit For Animal Development. Annu. Rev. Genet. 42, 235–251.

Rosing, M.T., 1999. 13C-Depleted Carbon microparticles in >3700-Ma Sea-Floor sedimentary rocks from West Greenland. Science. 283, 674–676.

Rossetti, V., Schirrmeister, B.E., Bernasconi, M. V., Bagheri, H.C., 2010. The evolutionary path to terminal differentiation and division of labor in cyanobacteria. J. Theor. Biol. 262, 23–34.

Ruiz-Trillo, I., Inagaki, Y., Davis, L.A., Landfald, B., Roger, A.J., 2004. *Capsaspora owczarzaki* is an independent opisthokont lineage. Curr. Biol. 14, 946–947.

Ruiz-Trillo, I., Lane, C.E., Archibald, J.M., Roger, A.J., 2006. Insights into the evolutionary origin and genome architecture of the unicellular opisthokonts *Capsaspora owczarzaki* and *Sphaeroforma arctica.* J. Eukaryot. Microbiol. 53, 379–84.

Ruiz-Trillo, I., Roger, A.J., Burger, G., Gray, M.W., Lang, B.F., 2008. A phylogenomic investigation into the origin of Metazoa. Mol. Biol. Evol. 25, 664–672.

Runnegar, B., 1991. Precambrian oxygen levels estimated from the biochemistry and physiology of early eukaryotes. Palaeogeogr. Palaeoclimatol. Palaeoecol. 97, 97–111.

Ryan, J.F., Pang, K., Schnitzler, C.E., Nguyen, A.D., Moreland, R.T., Simmons, D.K., Koch, B.J., Francis, W.R., Havlak, P., Smith, S.A., Putnam, N.H., Haddock, S.H.D., Dunn, C.W.,

Wolfsberg, T.G., C.Mullikin, J., Martindale, M.Q., Baxevanis, A.D., 2013. The genome of the ctenophore *Mnemiopsis leidyi* and its implications for cell type evolution. Science. 342, 1-8.

Ryser, C., Wang, J., Mimietz, S., Zimmermann, U., 1999. Determination of the Individual Electrical and Transport Properties of the Plasmalemma and the Tonoplast of the Giant Marine Alga *Ventricaria ventricosa* by Means of the Integrated Perfusion/Charge-Pulse Technique: Evidence for a Multifolded Tonoplast. J. Membr. Biol. 168, 183–197.

Ryser, H.J., 1967. Studies on protein uptake by isolated tumor cells. Apparent stimulations due to pH, hypertonicity, polycations, or dehydration and their relation to the enhanced penetration of infectious nucleic acids. J. Cell Biol. 32, 737–750.

Savin, M.C., Martin, J.L., LeGresley, M., Giewat, M., Rooney-Varga, J., 2004. Plankton diversity in the bay of fundy as measured by morphological and molecular methods. Microb. Ecol. 48, 51–65.

Schaap, P., 2011. Evolutionary crossroads in developmental biology: *Dictyostelium discoideum.* Development 138, 387–396.

Schenborn, Goiffon, 2000. DEAE-dextran transfection of mammalian cultured cells. Methods Mol. Biol. (Clifton, NJ) 130, 147–153.

Scherer, F., Anton, M., Schillinger, U., Henke, J., Bergemann, C., Krüger, A., Gänsbacher, B., Plank, C., 2002. Magnetofection: enhancing and targeting gene delivery by magnetic force in vitro and in vivo. Gene Ther. 9, 102–109.

Schiedlmeier, B., Schmitt, R., Moller, W., Kirkt, M.M., Gruber, H., Mages, W., Kirkt, D.L., 1994. Nuclear transformation of *Volvox carteri.* Proc. Natl. Acad. Sci. U. S. A. 91, 5080–5084.

Schiestl, R.H.R., Gietz, R.D.R., 1989. High efficiency transformation of intact yeast cells using single stranded nucleic acids as a carrier. Curr. Genet. 16, 339–346.

Schirrmeister, B.E., Gugger, M., Donoghue, P.C.J., 2015. Cyanobacteria and the Great Oxidation Event: Evidence from genes and fossils. Palaeontology 58, 769–785.

Schlichting, C.D., 2003. Origins of differentiation via phenotypic plasticity. Evol. Dev. 5, 98–105.

Schmitz, J.F., Zimmer, F., Bornberg-Bauer, E., 2016. Mechanisms of transcription factor evolution

in Metazoa. Nucleic Acids Res. 44, 6287–6297.

Schneckenburger, H., Hendinger, A., Sailer, R., Strauss, W.S.L., Schmitt, M., 2002. Laser-assisted optoporation of single cells. J. Biomed. Opt. 7, 410–416.

Schopf, J.W., 1993. Microfossils of the Early Archean Apex Chert: New Evidence of the of the Antiquity of Life. Science. 260, 640–646.

Schwartzman, D., 2002. Life, temperature, and the Earth: the self-organizing biosphere. Columbia University Press.

Sebé-Pedrós, A., Ariza-Cosano, A., Weirauch, M.T., Leininger, S., Yang, A., Torruella, G., 2013a. Early evolution of the *T-box* transcription factor family. Proc. Natl. Acad. Sci. U. S. A. 110, 16050–16055.

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., 2016a. The Dynamic Regulatory Genome of Capsaspora and the Origin of Animal Multicellularity. Cell 165, 1224–1237.

Sebé-Pedrós, A., Chomsky, E., Pang, K., Lara-Astiaso, D., Gaiti, F., Mukamel, Z., Amit, I., Hejnol, A., Degnan, B.M., Tanay, A., 2018. Early metazoan cell type diversity and the evolution of multicellular gene regulation. Nat. Ecol. Evol. 2, 1176–1188.

Sebé-Pedrós, A., de Mendoza, A., Lang, B.F., Degnan, B.M., Ruiz-Trillo, I., 2011. Unexpected repertoire of metazoan transcription factors in the unicellular holozoan *Capsaspora owczarzaki.* Mol. Biol. Evol. 28, 1241–54.

Sebé-Pedrós, A., Degnan, B.M., Ruiz-Trillo, I., 2017. The origin of Metazoa: a unicellular perspective. Nat Rev Genet 18, 498–512.

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

Sebé-Pedrós, A., Peña, M.I., Capella-Gutiérrez, S., Antó, M., Gabaldón, T., Ruiz-Trillo, I., Sabidó, E., 2016b. High-Throughput Proteomics Reveals the Unicellular Roots of Animal Phosphosignaling and Cell Differentiation. Dev. Cell 39, 186–197.

Sebé-Pedrós, A., Roger, A.J., Lang, F.B., King, N., Ruiz-trillo, I., 2010. Ancient origin of the integrin-mediated adhesion and signaling machinery. PNAS 107, 10142–10147.

Sebé-Pedrós, A., Ruiz-Trillo, I., 2010. Integrin-mediated adhesion complex. Commun. Integr. Biol. 3, 475–477.

Shalchian-Tabrizi, K., Minge, M.A., Espelund, M., Orr, R., Ruden, T., Jakobsen, K.S., Cavalier-Smith, T., 2008. Multigene phylogeny of Choanozoa and the origin of animals. PLoS One 3.

Shirahata, Y., Ohkohchi, N., Itagak, H., Satomi, S., 2001. New Technique for Gene Transfection Using Laser Irradiation. J. Investig. Med. 49, 184–190.

Southern, P., Berg, P., 1982. Transformation of mammalian cells to antibiotic resistance with a bacterial gene under control of the SV40 early region promoter. J. Mol. Appl. Genet. 1, 327–341.

Sperling, E.A., Frieder, C.A., Raman, A. V., Girguis, P.R., Levin, L.A., Knoll, A.H., 2013. Oxygen, ecology, and the Cambrian radiation of animals. Proc. Natl. Acad. Sci. U. S. A. 110, 13446–13451.

Sperling, E.A., Peterson, K.J., Pisani, D., 2009. Phylogenetic-signal dissection of nuclear housekeeping genes supports the paraphyly of sponges and the monophyly of Eumetazoa. Mol. Biol. Evol. 26, 2261–2274.

Sperling, E.A., Robinson, J.M., Pisani, D., Peterson, K.J., 2010. Where's the glass? Biomarkers, molecular clocks, and microRNAs suggest a 200-Myr missing Precambrian fossil record of siliceous sponge spicules. Geobiology 8, 24–36.

Srivastava, M., Begovic, E., Chapman, J., Putnam, N.H., Hellsten, U., Kawashima, T., Kuo, A., Mitros, T., Salamov, A., Carpenter, M.L., Signorovitch, A.Y., Moreno, M.A., Kamm, K., Grimwood, J., Schmutz, J., Shapiro, H., Grigoriev, I. V., Buss, L.W., Schierwater, B., Dellaporta, S.L., Rokhsar, D.S., 2008. The Trichoplax genome and the nature of placozoans. Nature 454, 955–960.

Srivastava, M., Mazza-curll, K.L., Wolfswinkel, J.C. Van, Reddien, P.W., 2014. Whole-Body Acoel Regeneration Is Controlled by *Wnt* and *Bmp-Admp* Signaling. Curr. Biol. 24, 1107–1113.

Srivastava, M., Simakov, O., Chapman, J., Fahey, B., Gauthier, M.E.A., Mitros, T., Richards, G.S., Conaco, C., Dacre, M., Hellsten, U., Larroux, C., Putnam, N.H., Stanke, M., Adamska,

M., Darling, A., Degnan, S.M., Oakley, T.H., Plachetzki, D.C., Zhai, Y., Adamski, M., Calcino, A., Cummins, S.F., Goodstein, D.M., Harris, C., Jackson, D.J., Leys, S.P., Shu, S., Woodcroft, B.J., Vervoort, M., Kosik, K.S., Manning, G., Degnan, B.M., Rokhsar, D.S., 2010. The *Amphimedon queenslandica* genome and the evolution of animal complexity. Nature 466, 720–726.

Stanley, S.M., 1973. An Ecological Theory for the Sudden Origin of Multicellular Life in the Late Precambrian. Proc. Natl. Acad. Sci. U. S. A. 70, 1486–1489.

Steenkamp, E.T., Wright, J., Baldauf, S.L., 2006. The protistan origins of animals and fungi. Mol. Biol. Evol. 23, 93–106.

Stibbs, H.H., Owczarzak, A., Bayne, C.J., DeWan, P., 1979. Schistosome sporocyst-killing amoebae isolated from *Biomphalaria glabrata.* J. Invertebr. Pathol. 33, 159–170.

Stougaard, P., Jørgensen, F., Johnsen, M.G., Hansen, O.C., 2002. Microbial diversity in ikaite tufa columns: An alkaline, cold ecological niche in Greenland. Environ. Microbiol. 4, 487–493.

Suga, H., Chen, Z., de Mendoza, A., Sebé-Pedrós, A., Brown, M.W., Kramer, E., Carr, M., Kerner, P., Vervoort, M., Sánchez-Pons, N., Torruella, G., Derelle, R., Manning, G., Lang, B.F., Russ, C., Haas, B.J., Roger, A.J., Nusbaum, C., Ruiz-Trillo, I., 2013. The *Capsaspora* genome reveals a complex unicellular prehistory of animals. Nat. Commun. 4, 2325.

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. Sci. Signal. 5, ra35.

Suga, H., Miller, W.T., 2018. *Src* signaling in a low-complexity unicellular kinome. Sci. Rep. 8, 1–11.

Suga, H., Ruiz-Trillo, I., 2013. Development of ichthyosporeans sheds light on the origin of metazoan multicellularity. Dev. Biol. 377, 284–92.

Sumathi, J.C., Raghukumar, S., Kasbekar, D.P., Raghukumar, C., 2006. Molecular Evidence of Fungal Signatures in the Marine Protist *Corallochytrium limacisporum* and its Implications in the Evolution of Animals and Fungi. Protist 157, 363–376.

Szathmáry, E., 2015. Toward major evolutionary transitions theory 2.0: Table 1. Proc. Natl. Acad. Sci. U. S. A. 112, 10104–10111.

Szathmáry, E., Smith, J.M., 1995. The major evolutionary transitions. Nature.

Szathmáry, E., Smith, J.M., 1997. From replicators to reproducers: The first major transitions leading to life. J. Theor. Biol. 187, 555–571.

Takishita, K., Miyake, H., Kawato, M., Maruyama, T., 2005. Genetic diversity of microbial eukaryotes in anoxic sediment around fumaroles on a submarine caldera floor based on the small-subunit rDNA phylogeny. Extremophiles 9, 185–196.

Takishita, K., Tsuchiya, M., Kawato, M., Oguri, K., Kitazato, H., Maruyama, T., 2007. Genetic Diversity of Microbial Eukaryotes in Anoxic Sediment of the Saline Meromictic Lake Namako-ike (Japan): On the Detection of Anaerobic or Anoxic-tolerant Lineages of Eukaryotes. Protist 158, 51–64.

Takishita, K., Tsuchiya, M., Reimer, J.D., Maruyama, T., 2006. Molecular evidence demonstrating the basidiomycetous fungus *Cryptococcus curvatus* is the dominant microbial eukaryote in sediment at the Kuroshima Knoll methane seep. Extremophiles 10, 165–169.

Tang, F., Bengtson, S., Wang, Y., Wang, X.L., Yina, C.Y., 2011. Eoandromeda and the origin of Ctenophora. Evol. Dev. 13, 408–414.

Taylor, J.S., Raes, J., 2004. Duplication and Divergence: The Evolution of New Genes and Old Ideas. Annu. Rev. Genet. 38, 615–643.

Thorne, H. V, Evans, J., Warden, D., 1968. Detection of Biologically Defective Molecules in Component I of Polyoma Virus DNA. Nature 219, 728–730.

Tokuoka, M., Imai, K.S., Satou, Y., Satoh, N., 2004. Three distinct lineages of mesenchymal cells in *Ciona intestinalis* embryos demonstrated by specific gene expression. Dev. Biol. 274, 211–224.

Tomitani, A., Knoll, A.H., Cavanaugh, C.M., Ohno, T., 2006. The evolutionary diversification of cyanobacteria: Molecular-phylogenetic and paleontological perspectives. Proc. Natl. Acad. Sci. U. S. A. 103, 5442–5447.

Tomizawa, M., Shinozaki, F., Motoyoshi, Y., Sugiyama, T., Yamamoto, S., Sueishi, M., 2013.

Sonoporation: Gene transfer using ultrasound. World J. Methodol. 3, 39.

Tong, S.M., 1997. Heterotrophic flagellates and other protists from southampton water, u.k. Ophelia 47, 71–131.

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., Paley, R., Roger, A.J., Sitjà-Bobadilla, A., Donachie, S., Ruiz-Trillo, I., 2015. Phylogenomics Reveals Convergent Evolution of Lifestyles in Close Relatives of Animals and Fungi. Curr. Biol. 25, 2404–2410.

Torruella, G., Derelle, R., Paps, J., Lang, B.F., Roger, A.J., Shalchian-Tabrizi, K., Ruiz-Trillo, I., 2012. Phylogenetic relationships within the Opisthokonta based on phylogenomic analyses of conserved single-copy protein domains. Mol. Biol. Evol. 29, 531–544.

Towe, K.M., 1970. Oxygen-Collagen Priority and the Early Metazoan Fossil Record. Proc. Natl. Acad. Sci. U. S. A. 65, 781–788.

Umen, J.G., 2014. Green Algae and the Origins of Multicellularity in the Plant Kingdom. Cold Spring Harb. Perspect. Biol. 6, 1–27.

Vaheri, A., Pagano, J.S., 1965. Infectious poliovirus RNA: a sensitive method of assay. Virology 27, 434–436.

Valentine, J.W., Tiffney, B.H., Sepkoski, J.J.J., 1991. Evolutionary dynamics of plants and animals: a comparative approach. Palaios 6, 81–88.

Velicer, G.J., Vos, M., 2009. Sociobiology of the Myxobacteria. Annu. Rev. Microbiol. 63, 599–623.

Venter, P.C., Nitsche, F., Arndt, H., 2018. The Hidden Diversity of Flagellated Protists in Soil. Protist 169, 432–449.

Vinayak, S., Pawlowic, M.C., Sateriale, A., Brooks, C.F., Studstill, C.J., Bar-Peled, Y., Cipriano, M.J., Striepen, B., 2015. Genetic modification of the diarrhoeal pathogen *Cryptosporidium parvum.* Nature 523, 477–480.

Wainright, P.O., Hinkle, G., Sogin, M.L., Stickel, S.K., 1993. Monophyletic origins of the metazoa: An evolutionary link with fungi. Science. 260, 340–342.

Walter, M.R., Buick, R., Dunlop, J.S.R., 1980. Stromatolites 3,400–3,500 Myr old from the North-Pole area, Western Australia. Nature 284, 443–445.

Walter, M.R., Heys, G.R., 1985. Links between the rise of the Metazoa and the decline of Stromatolites. Precambrian Res. 29, 149–174.

Washbourne, P., McAllister, A.K., 2002. Techniques for gene transfer into neurons. Curr. Opin. Neurobiol. 12, 566–573.

Webster, N.S., Smith, L.D., Heyward, A.J., Watts, E.M., Webb, R.I., Blackall, L.L., Negri, P., Watts, J.E.M., Negri, A.P., 2004. Metamorphosis of a Scleractinian Coral in Response to Microbial Biofilms. Appl. Environ. Microbiol. 70, 1213–1221.

Weis, V.M., Davy, S.K., Hoegh-Guldberg, O., Rodriguez-Lanetty, M., Pringle, J.R., 2008. Cell biology in model systems as the key to understanding corals. Trends Ecol. Evol. 23, 369–376.

Weisse, T., Anderson, R., Arndt, H., Calbet, A., Hansen, P.J., Montagnes, D.J.S., 2016. Functional ecology of aquatic phagotrophic protists – Concepts, limitations, and perspectives. Eur. J. Protistol. 55, 50–74.

Wilcox, J.J.S., Hollocher, H., 2018. Unprecedented Symbiont Eukaryote Diversity Is Governed by Internal Trophic Webs in a Wild Non-Human Primate. Protist 169, 307–320.

Willmer, P., 1990. The origin of the Metazoa. In: Invertebrate Relationships: Patterns in Animal Evolution. Cambridge University Press. pp. 163–198.

Wilson, M., Koopman, P., 2002. Matching SOX: Partner proteins and co-factors of the SOX family of transcriptional regulators. Curr. Opin. Genet. Dev. 12, 441–446.

Woese, C.R., 1996. Phylogenetic trees: Whither microbiology? Curr. Biol. 6, 1060–1063.

Woese, C.R., Fox, G.E., 1977. Phylogenetic structure of the prokaryotic domain: The primary kingdoms. Proc. Natl. Acad. Sci. U. S. A. 74, 5088–5090.

Wolpert, L., Szathmáry, E., 2002. Multicellularity: Evolution and the egg. Nature 420, 745.

Woollacott, R., Hadfield, M., 1996. Induction of metamorphosis in larvae of a sponge. Invertebr. Biol. 115, 257–262.

Woznica, A., Cantley, A.M., Beemelmanns, C., Freinkman, E., Clardy, J., King, N., 2016. Bacterial lipids activate, synergize, and inhibit a developmental switch in choanoflagellates. Proc. Natl. Acad. Sci. U. S. A. 113, 7894–7899.

Woznica, A., Gerdt, J.P., Hulett, R.E., Clardy, J., King, N., 2017. Mating in the Closest Living Relatives of Animals Is Induced by a Bacterial Chondroitinase. Cell 170, 1–9.

Woznica, A., King, N., 2018. Lessons from simple marine models on the bacterial regulation of eukaryotic development. Curr. Opin. Microbiol. 43, 108–116.

Wylezich, C., Karpov, S.A., Mylnikov, A.P., Anderson, R., Jürgens, K., 2012. Ecologically relevant choanoflagellates collected from hypoxic water masses of the Baltic Sea have untypical mitochondrial cristae. BMC Microbiol. 12, 1–13.

Xiao, S., Knoll, A.H., Schiffbauer, J.D., Zhou, C., Yuan, X., 2012. Comment on '"Fossilized Nuclei and Germination Structures Identify Ediacaran 'Animal Embryos' as Encysting Protists"'. Science. 335, 1696–1699.

Yang, Y., Yang, J., 1997. Studies of DEAE-dextran-mediated gene transfer. Biotechnol. Appl. Biochem. 47–51.

Yao, C.P., Zhang, Z.X., Rahmanzadeh, R., Huettmann, G., 2008. Laser-based gene transfection and gene therapy. IEEE Trans. Nanobioscience 7, 111–119.

Yin, Z., Zhu, M., Davidson, E.H., Bottjer, D.J., Zhao, F., Tafforeau, P., 2015. Sponge grade body fossil with cellular resolution dating 60 Myr before the Cambrian. Proc. Natl. Acad. Sci. U. S. A. 112, E1453–E1460.

Young, S.L., Diolaiti, D., Conacci-Sorrell, M., Ruiz-Trillo, I., Eisenman, R.N., King, N., 2011. Premetazoan ancestry of the Myc-Max network. Mol. Biol. Evol. 28, 2961–71.

Zettler, L.A.A., Nerad, T.A., O'Kelly, C.J., Sogin, M.L., 2001. The nucleariid amoebae: More protists at the animal-fungal boundary. J. Eukaryot. Microbiol. 48, 293–297.

# Appendix

*Antibody validation*

# *Appendix*

## Antibody validation

Antibodies, especially custom antibodies, are among the most critical research reagents used in a lab, both for their performance and for their cost of production (both at the economic and time-consuming level). Still, most publications using both non-commercial and commercial antibodies do not include complete information about their validation, challenging the reproducibility of antibodies-related research, which thus faces a replication crisis (Weller, 2018). It is not only essential to prove that they are able to bind to the right target (specificity) and not to other non-specific proteins (selectivity), but it is also important to make sure they are ultimately suitable for the desired applications (functionality). Thus, validated antibodies must be specific, selective, and reproducible in the context for which they are to be used (Bordeaux et al., 2010).

In the present thesis, I designed custom polyclonal antibodies using synthetic peptides and protein fragments against the endogenous *Runx1*, *Runx2* and *NF-κB* proteins of *Capsaspora owczarzaki*. This work took almost three years, including antigen design, antigen production, animal immunization, antibody validation and purification; without counting the actual experiments. Thus, I here present an overview of the know-how generated in the lab from hands-on experience and training by the Custom Antibody Service Unit (CAbS, CIBER-BBN, IQAC-CSIC) from the ICTS "NANBIOSIS and the European Monoclonal Antibodies Network "EuroMabNet" (Roncador et al., 2015).

When planning an antibody production, it is key to consider many factors including the conjugation method (if necessary), the immunization protocol and the antibody purification procedure. In this appendix, I will briefly summarize the steps for designing and producing antibodies against synthetic peptides and protein fragments and I will expose the reasoning of each decision we made to validate the antibodies.

# *Section A.1*

## Candidate proteins for functional studies in *Capsaspora owczarzaki*

For our purpose, we surveyed *Runx* and the *NF-κB* animal orthologs in *Capsaspora. Runx* and *NF-κB* are two well-described families of developmental TFs with myriad roles in development, cell fate determination, cell differentiation and stress responses in animals (Coffman, 2003; Hayden and Ghosh, 2004; Macian, 2005; Robertson et al., 2009). Both *Runx* and *NF-κB* were identified in *Capsaspora owczarzaki* and interestingly they both presented different transcriptional levels and protein abundance across its different life stages (Sebé-Pedrós et al., 2016b, 2016a, 2013, 2011). Moreover, they present different phosphorylation profiles across life stages (Sebé-Pedrós et al., 2016b).

*Capsaspora* has two homologs to animal *Runx* genes. Both *Runx* genes in *Capsaspora* (*CoRunx1* and *CoRunx2*) possess a conserved Runt DNA-binding Domain (DBD) with key DNA contacting amino acids **(Figure 22A-B).** In particular, *CoRunx1* possesses two Cysteine residues involved in redox regulation (shared with other animal *Runx* genes) and Ankyrin repeats (Ank) involved in protein-protein interactions. Moreover, both *Capsaspora Runx* genes lack the C-terminal WRPY motif. The *CoRunx1* transcript has a predicted 2661 bp coding sequence, corresponding to a 886 amino acid predicted protein (94.9 kDa). The *CoRunx2* transcript has a predicted 2901 bp coding sequence, corresponding to a 988 amino acid predicted protein (103.3 kDa).

*Capsaspora* has one homolog of the animal *NF-κB* TF family. Remarkably, *Capsaspora* ortholog (*CoNFκB*) possesses a conserved Rel-homology DNA-binding and dimerization Domain (RHD) with a highly conserved and specific recognition loop (RL) and other key amino acids for DNA-binding specificities **(Figure 22C)**. Moreover, it contains a basic nuclear localization signal (NLS), a Glycine-Serine rich region (GS) and Ankyrin repeats (Ank).

The *CoNF-κB* transcript has a predicted 3672 bp coding sequence, corresponding to a 1223 amino acids protein (129.2 kDa). It shares with its animal homologs key domains for protein-protein interactions and DNA-binding specificity:

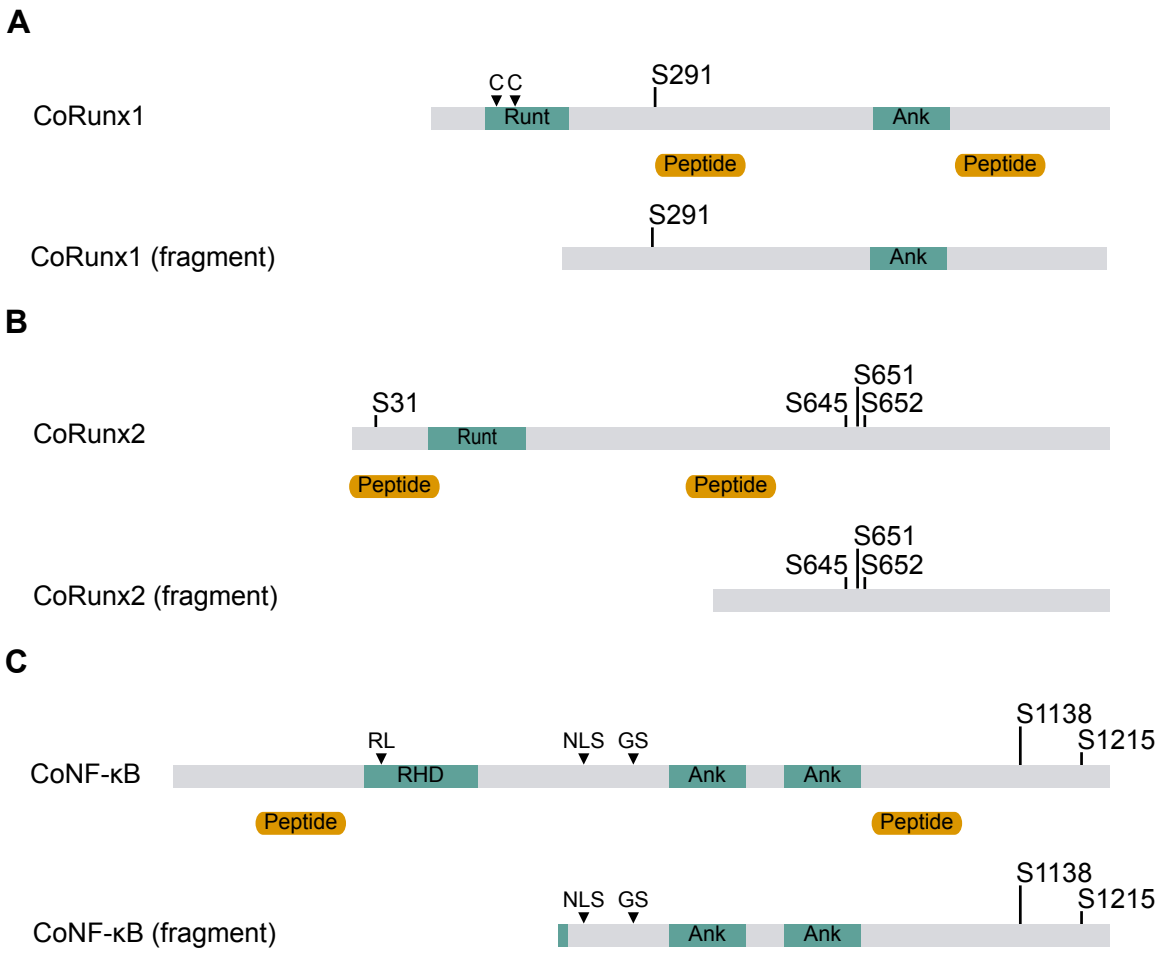



**Figure 22. Schematic representation of *Capsaspora* CoRunx1, CoRunx2 and CoNF-κB candidates.** (A) CoRunx1 protein architecture representation, depicting the Runt DNA-binding Domain (Runt), key Cysteine residues (C) within Runt, residue and position of key phosphorylation sites (S291), and Ankyrin repeats (Ank). (B) Same as (A) for CoRunx2. (C). CoNF-κB protein architecture representation, depicting the Rel-homology DNA-binding and dimerization domain (RHD), a specific recognition loop (RL) within the RHD, the nuclear localisation signal (NLS), a Glycine-Serine rich region (GS), Ankyrin repeats (Ank) and residue and position of key phosphorylation sites (example S1138). Key domains and regions are depicted in the green boxes. Protein fragments used to raise especific antibodies against each candidate are illustrated below as (fragment). Orange boxes indicate peptide regions used to raise especific antibodies against each candidate.

# *Section A.2*

# Generating antibodies against synthetic peptides

Raising antibodies against synthetic peptides has gained popularity in the recent years, especially among commercial companies. Despite these peptides being poor immunogens *per se*[1], using them can overcome cross-reactivity issues and increase specificity.

Choosing the protein region for the synthetic peptide design is critical. In fact, the 3D-structure of the candidate protein and other unpredictable factors will affect the productivity and specificity of antibodies generated during animal immunization. Thus, it is usually recommended to use more than one synthetic peptide (two or three peptides) derived from the same protein sequence to raise the chances that they will be functional.

In our case, we designed two peptides for each of our candidate genes in *Capsaspora (CoRunx1, CoRunx2* and *CoNFκB)* to generate hyperimmune antisera in two rabbits per peptide during an immunization period of 8 months.

## A.2.1. Synthetic peptides design

As mentioned before, deciding which region would be more adequate for raising antibodies against is critical, especially considering that synthetic peptide sequences range between 12-15 amino acids long. Nevertheless, recommended peptides are not guaranteed to successfully produce antibodies.

[1] Synthetic peptides are usually covalently conjugated to highly immunogenic carrier proteins (*e.g.,* Bovine Serum Albumin (BSA) and Keyhole limpet haemocyanin (KLH)) to raise effective anti-peptide antibodies and for antibody titration during the immunization period.

Our criteria for their design was based in three key points:

- Protein architecture and peptide uniqueness.
- Immunogenicity.
- Molecular technique to be used.

### A.2.1.1. Protein architecture and peptide uniqueness

Peptide design over particular regions and uniqueness will be ultimately related to the specificity and selectivity of the peptide for antibody recognition, respectively. We evaluated protein architecture by assessing protein domain architecture, hydrophilicity profiles of key regions and predicting secondary and tertiary structures of the protein.

**a) Protein domain architecture:**

We used PfamScan v.1.5 (Mistry et al., 2007) for protein domains prediction. Despite PfamScan being based on Hidden Markov Models built mainly from animal sequences, conserved regions are also detected in *Capsaspora* proteins. Thus, our criteria consisted in avoiding ubiquitous protein domains conserved between various protein families to minimize cross reactivity. These included, for example, Ankyrin repeats or Immunoglobulin-like folds. Moreover, depending on the user's particular purpose, it might be critical to also avoid Cysteine-rich regions, Glycine-Serine rich regions or Phosphorylation sites.

**b) Hydrophilicity:**

Assessment of the hydrophilicity profiles provides information about the putative accessibility for antibody recognition. Thus, it is critical to avoid complex and inaccessible regions (generally with higher hydrophobic profiles) and aim for accessible regions which are surface-exposed and hydrophilic. To this end, we evaluated hydrophilicity profiles of key regions using Kyte & Doolittle Hydropathy analysis[2] (Gasteiger et al., 2005) from ExPASy Bioinformatics Resource Portal (Artimo et al., 2012).

[2] ExPASy ProtScale Server: https://web.expasy.org/protscale/

**c) Secondary and tertiary predicted structures:**

Finally, another essential aspect of peptide region selection is to avoid disrupting predicted secondary and tertiary structures. For secondary structure prediction, we used the Geneious[3] 7.1.9 predict structure tool, which predicts protein secondary structure (*e.g.,* alpha helix and beta strand) using the original Garnier Osguthorpe Robson algorithm (GOR I) provided by the EMBOSS suite[4]. We additionally used the Phyre[2] (Protein Homology/analogY Recognition Engine v.2.0)[5] protein fold recognition Server (Kelley et al., 2015). This tool uses the alignment of hidden Markov models via HHsearch (Söding, 2005) to significantly improve the accuracy of alignment and detection rate and *ab initio* folding simulation, called "Poing", to model regions with no detectable homology to known structures. Another useful tool to compare crystallized structures and evaluate the accessibility of the peptide region for antibody recognition is UCSF Chimera (Pettersen et al., 2004), that allows the user to model and compare crystallized structures (*e.g.,* deposited in the Protein Data Bank (PDB); (Berman et al., 2000).

**d) Peptide uniqueness:**

To decrease putative cross-reactivity with other non-target proteins, it is essential to select unique peptides, *i.e.,* peptide regions that are only and exclusively present in the candidate protein. A recommendation is to BLAST selected peptides against the proteome of the target species to make sure they are exclusive for the target protein.

### A.2.1.2. Immunogenicity

We predicted the immunogenicity in each protein comparing available information on known epitopes and an immunogenicity prediction online tool. Even though protein homology between the three *Capsaspora* genes and animal homologs is very low, we decided to perform a detailed study based on the epitopes used to raise antibodies in commercially available antibodies.

[3] Geneious. See: https://www.geneious.com

[4] EMBOSS 6.5.7 website. See: http://emboss.sourceforge.ne

[5] Phyre[2] Server. See: http://www.sbg.bio.ic.ac.uk/phyre2/html/page.cgi?id=index

However, in most of the cases the main providers used either larger peptides (around 50 amino acids) or recombinant fragments that could not be compared at the protein sequence level with *Capsaspora* proteins or did not specify the sequence. Nevertheless, this knowledge was used to at least select putative similar regions in our candidate proteins. We performed protein alignments using Mafft v7.299b (Katoh et al., 2005, 2002; Katoh and Standley, 2013) to assess the homology of selected regions in animals.

Moreover, we used the AbDesigner online tool[6] (Pisitkun et al., 2012) to analyse and predict optimal immunizing regions of a given protein for antibody production. This tool recognizes trade-off between immunogenicity, specificity, animal species targets and post-translational modifications and provides each region a score correlating with immunogenicity.

### A.2.1.3. Molecular technique to be used

Finally, it is key to consider for what purpose the antibodies produced will be for. In our case, we want to detect native proteins by immunofluorescence and Chromatin Immunoprecipitation followed by DNA sequencing (ChIPseq). Thus, we should consider avoiding additional domains that are directly correlated with DNA binding activities of our candidate proteins.

## A.2.2. Synthetic peptides production

Finally, 10-12 mg of the following candidate synthetic peptides with >90% purity were produced at the Applied Molecular Receptor Group (AMRg) - CIBER of Bioengineering, Biomaterials and Nanomedicine IQAC-CSIC (Spanish Council for Scientific Research).

P1-RUNX1: Cys-PSSPRASSNNNDS (14aa)
P2-RUNX1: Cys-LNFGGLSQNTSPP (14 aa)
P1-RUNX2: GRAKRAARPASGR-Cys (14 aa)

[6] AbDesigner online tool. See: https://hpcwebapps.cit.nih.gov/AbDesigner/index.jsp.

P2-RUNX2: Cys-PQSQPHHPQQPG (13 aa)
P1-NFκB: Cys-PPRRGHNQSSSSD (14 aa)
P2-NFκB: Cys-DELKKDKDDPKKE (14 aa)

### A.2.3. Antibody production protocol against synthetic peptides

The final antibody production protocol against synthetic peptides included the conjugation of each peptide to both KLH, for rabbit immunization, and BSA, for ELISA titration assays and antibody affinity purification. 12 polyclonal antibodies were raised in rabbits by the Custom Antibody Service Unit (CAbS, CIBER-BBN, IQAC-CSIC) from the ICTS "NANBIOSIS". Rabbits were immunized monthly (6 times) using 100 µg/rabbit per immunization of *CoRunx1*, *CoRunx2* or *CoNF-κB* synthetic peptides (conjugated to KLH) and affinity-purified from the final antisera using *CoRunx1*, *CoRunx2* or *CoNF-κB* synthetic peptides (conjugated with BSA).

# *Section A.3*

# Generating antibodies against protein fragments

## A.3.1. Protein fragment design

Antibodies generated against full-length or partial fragments of native proteins lead to antibodies recognizing multiple epitopes, increasing the chances of detecting the target protein. For our purpose, we ensured the antibodies are suitable for Chromatin immunoprecipitation experiments and immunofluorescence assays by selecting the most exposed regions of each candidate and excluding regions that might interfere the technical performance of both approaches (*i.e.,* DNA contacting amino acids).

We sought for the largest fragment possible of each candidate protein excluding the N-terminal region and the DNA-binding Domains (DBDs) of *CoRunx1, CoRunx2* and *CoNF-κB* **(Figure 22).** We additionally generated recombinant proteins under native conditions, as both Chromatin Immunoprecipitation experiments and Immunofluorescence assays are performed with native proteins.

## A.3.2. Recombinant proteins production

Protein fragments of *CoRunx1* (711 aa, 75.8 kDa), *CoRunx2* (518 aa, 54.2 kDa) and *CoNF-κB* (721 aa, 75.6 kDa) were cloned into pETM14 with a His-tag, expressed in *E. coli* BL21(DE3) strain and purified by Nickel affinity chromatography under native conditions with an Elution Buffer containing 50 mM Tris-HCl pH=7.4, 300 mM NaCl and 10% Glycerol. *CoRunx1* and *CoNF-κB* pellets were additionally re-solubilized with GuHCl, purified by Nickel affinity chromatography in denaturing conditions and refolded in a second Nickel affinity chromatography in native conditions to avoid protein degradation products. Recombinant proteins were produced by the Biomolecular Screening & Protein Technologies Unit from the Centre for Genomic Regulation (CRG).

Recombinant proteins were used for antibody production and as controls for western blot analysis.

### A.3.3. Antibody production protocol against protein fragments

Polyclonal antibodies were raised in rabbits by the Custom Antibody Service Unit (CAbS, CIBER-BBN, IQAC-CSIC) from the ICTS "NANBIOSIS". Rabbits were immunized monthly (6 times) using 100 µg/rabbit per immunization of *CoRunx1, CoRunx2* or *CoNF-κB* recombinant proteins (conjugated with KLH) and affinity-purified from the final antiserums using *CoRunx1, CoRunx2* or *CoNF-κB* recombinant proteins. The final concentrations of affinity-purified antibodies were 170 µg/mL, 350 µg/mL and 50 µg/mL.

# *Section A.4*

# Antibody validation

Antibody validation corresponds to the experimental proof(s) and documentation that a specific antibody is suitable for an intended application or purpose, and specifically recognizes what it is intended to recognize (Bordeaux et al., 2010).

In general, it is recommended to follow more than one strategy to completely assess antibody specificity, selectivity and functionality for the desired technique to be used.

## A.4.1. Target specificity validation

To ensure that the antibody will bind to the correct target, it is recommended to validate the antibody's proper functionality by at least two independent methods. There are several strategies to address antibody validation; some examples include:

- Immunoprecipitation coupled to Tandem Mass Spectrometry (IP-MS/MS)
- Genetic modification:
    - Negative controls: testing antibodies on Knockout cell lines (*e.g.,* using CRISPR-Cas9) or Knockdown cell lines (*e.g.,* using RNAi).
    - Positive controls: testing antibodies on cell lines expressing target protein with or without a tag.
    - Independent antibody verification (IAV): measurement of target expression using two differentially raised antibodies recognizing the same target protein.
- Biological verification:
    - Cell treatment: detecting downstream events following cell treatment.
    - Relative expression: using naturally occurring variable expression across cell stages or life cycle to confirm antibody specificity.

- Neutralization: functional blocking of protein activity by antibody binding (*e.g.,* Peptide Competition Assays).
- Peptide array: using arrays to test reactivity against known protein modifications.

### A.4.2. Functional application validation

To ensure the antibody is suitable for a particular application of interest, it is critical to assess the antibody's functionality and optimize protocol conditions. Especially when raising multiple antibodies against the same target protein, it is also essential to test which one generates acceptable or the best results in a specific application to confirm antibody performance and decide which one/s to further purify and use for subsequent experiments, that may include (among others):

- Western blot analysis
- Immunofluorescence assays
- Immunoprecipitation experiments
- Chromatin Immunoprecipitation

#### A.4.2.1. Validation of target antigen recognition

Assessing the antibody reactivity against the immunizing agent by high-throughput screening techniques is usually the starting point for antibody validation.

Western blotting is widely used to determine an antibody's specificity and in general is the first validation step to ensure the antibody recognizes the denatured antigen. The first indication that the antibody is specific for the selected target would be detecting the immunizing agent (*e.g.,* recombinant protein). Another good indicator is observing a single band at the (approximately) known molecular weight for the target in whole cell protein extract. Presence of multiple bands or bands too far from the proper molecular weight might indicate nonspecific recognition for another undesired protein.

Preimmune sera (*a.k.a.,* animal sera before immunization) are also worth testing by dot blot screens and western blotting. Preimmune sera can also be used as negative controls and to "preclear" protein extracts when testing partial bleedings before antibody purification.

Nevertheless, especially when antibodies are generated from proteins in native conformations, western blotting will not be sufficient for antibody validation.

### A.4.2.2. Immunoprecipitation coupled to Tandem Mass Spectrometry (IP-MS/MS)

Immunoprecipitation in combination with tandem Mass Spectrometry (IP-MS/MS) allows the user to verify that an antibody interacts specifically with an intended target. IP-MS/MS directly identifies peptide sequences from the proteins in the sample.

Generally, the top 20 peptides detected by IP-MS/MS are considered (or the ones that show a score higher than 90%). However, at times some proteins identified might be marked as doubtful because its identification is supported by only one or few confident peptides. This does not necessary mean that the protein identification is wrong, but it should, logically, be considered with care.

One way to double check IP-MS/MS results is to examine *in silico* trypsin digestion prediction of the target protein to verify the number and features of the identifiable peptides by Mass Spectrometry, as not all the peptides are equally detectable by MS/MS.

- Peptide size: small peptides (generally shorter than 8 amino acids) or large peptides (generally longer than 25 amino acids) will not be detected by MS/MS.
- Technical limitations: the necessary condition of a MS detector to be able to "see" a peptide is the peptide's ionization. However, there is always a percentage of uncharged peptides that will not be detected.
- Unique peptides: a very important feature to assess the specificity of the assignment/prediction of target protein recognition.

### A.4.2.3. Functional blocking by antigen competition assays

Neutralization, or functional blocking by competition assays, of endogenous proteins in a cell lysate by antibodies is another strategy to validate antibody specificity. Basically, when an antibody exhibits functional blocking characteristics, it is strong evidence that the antibody is binding the intended target, or at least, a target specifically (the actual identity of the target can only be determined through IP-MS/MS).

The immunizing peptide or recombinant protein can be used as a method for antibody specificity verification by blocking the working concentration of the desired antibody with different dilutions of the immunizing agent (peptides or recombinant proteins) in western blotting applications.

# Appendix References

Artimo, P., Jonnalagedda, M., Arnold, K., Baratin, D., Csardi, G., De Castro, E., Duvaud, S., Flegel, V., Fortier, A., Gasteiger, E., Grosdidier, A., Hernandez, C., Ioannidis, V., Kuznetsov, D., Liechti, R., Moretti, S., Mostaguir, K., Redaschi, N., Rossier, G., Xenarios, I., Stockinger, H., 2012. ExPASy: SIB bioinformatics resource portal. Nucleic Acids Res. 40, 597–603.

Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N., Bourne, P.E., 2000. The protein data bank. Nucleic Acids Res. 28, 235–242.

Bordeaux, J., Welsh, A.W., Agarwal, S., Killiam, E., Baquero, M.T., Hanna, J.A., Anagnostou, V.K., Rimm, D.L., 2010. Antibody validation. Biotechniques 48, 197–209.

Coffman, J.A., 2003. *Runx* transcription factors and the developmental balance between cell proliferation and differentiation. Cell Biol. Int. 27, 315–324.

Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S., Wilkins, M.R., Appel, R.D., Bairoch, A., 2005. Protein Identification and Analysis Tools on the ExPASy Server. Proteomics Protoc. Handb. 571–607.

Hayden, M.S., Ghosh, S., 2004. Signaling to *NF-kB.* Genes Dev. 18, 2195–2224.

Katoh, K., Kuma, K.I., Toh, H., Miyata, T., 2005. MAFFT version 5: Improvement in accuracy of multiple sequence alignment. Nucleic Acids Res. 33, 511–518.

Katoh, K., Misawa, K., Kuma, K., Miyata, T., 2002. MAFFT : a novel method for rapid multiple sequence alignment based on fast Fourier transform 30, 3059–3066.

Katoh, K., Standley, D.M., 2013. MAFFT Multiple Sequence Alignment Software Version 7 : Improvements in Performance and Usability Article Fast Track 30, 772–780.

Kelley, L.A., Mezulis, S., Yates, C., Wass, M., Sternberg, M., 2015. The Phyre2 web portal for protein modelling, prediction, and analysis. Nat. Protoc. 10, 845–858.

Macian, F., 2005. NFAT proteins: Key regulators of T-cell development and function. Nat. Rev. Immunol. 5, 472–484.

Mistry, J., Bateman, A., Finn, R.D., 2007. Predicting active site residue annotations in the Pfam database. BMC Bioinformatics 8, 1–14.

Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C., Ferrin, T.E., 2004. UCSF Chimera - A visualization system for exploratory research and analysis. J. Comput. Chem. 25, 1605–1612.

Pisitkun, T., Hoffert, J.D., Saeed, F., Knepper, M.A., 2012. NHLBI-AbDesigner: an online tool for design of peptide-directed antibodies. AJP Cell Physiol. 302, C154–C164.

Robertson, A.J., Larroux, C., Degnan, B.M., Coffman, J. a, 2009. The evolution of *Runx* genes II. The C-terminal Groucho recruitment motif is present in both eumetazoans and homoscleromorphs but absent in a haplosclerid demosponge. BMC Res. Notes 2, 59.

Roncador, G., Engel, P., Maestre, L., Anderson, A.P., Cordell, J.L., Cragg, M.S., Šerbec, V., Jones, M., Lisnic, V.J., Kremer, L., Li, D., Koch-Nolte, F., Pascual, N., Rodríguez-Barbosa, J.I., Torensma, R., Turley, H., Pulford, K., Banham, A.H., 2015. The European antibody network's practical guide to finding and validating suitable antibodies for research. MAbs 8, 27–36.

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., 2016a. The Dynamic Regulatory Genome of *Capsaspora* and the Origin of Animal Multicellularity. Cell 165, 1224–1237.

Sebé-Pedrós, A., de Mendoza, A., Lang, B.F., Degnan, B.M., Ruiz-Trillo, I., 2011. Unexpected repertoire of metazoan transcription factors in the unicellular holozoan *Capsaspora owczarzaki.* Mol. Biol. Evol. 28, 1241–54.

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.

Sebé-Pedrós, A., Peña, M.I., Capella-Gutiérrez, S., Antó, M., Gabaldón, T., Ruiz-Trillo, I., Sabidó, E., 2016b. High-Throughput Proteomics Reveals the Unicellular Roots of Animal Phosphosignaling and Cell Differentiation. Dev. Cell 39, 186–197.

Söding, J., 2005. Protein homology detection by HMM-HMM comparison. Bioinformatics 21, 951–960.

Weller, M.G., 2018. Ten basic rules of antibody validation. Anal. Chem. Insights 13.