 doi:10.20944/preprints202109.0264.v1

The history of *Salpingoeca rosetta* as a model for reconstructing animal origins

David S. Booth[1] and Nicole King[2]

[1]Chan Zuckerberg Biohub and Department of Biochemistry and Biophysics

University of California, San Francisco

[2]Howard Hughes Medical Institute and Department of Molecular and Cell Biology

University of California, Berkeley

 doi:10.20944/preprints202109.0264.v1

Abstract:

Choanoflagellates, the closest living relatives of animals, have the potential to reveal the genetic and cell biological foundations of complex multicellular development in animals. Here we describe the history of research on the choanoflagellate *Salpingoeca rosetta.* From its original isolation in 2000 to the establishment of CRISPR-mediated genome editing in 2020, *S. rosetta* provides an instructive case study in the establishment of a new model organism.

Keywords: choanoflagellates, multicellularity, animal origins, genome editing, electroporation

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 15 September 2021 doi:10.20944/preprints202109.0264.v1

Introduction

By virtue of their close relationship with animals (Figure 1A) and the many animal-like features of their cell physiology and genomes, choanoflagellates have great potential as experimental models for reconstructing animal origins and core mechanisms underlying animal cell biology. Interest in choanoflagellates began following their first observations in the mid-to-late nineteenth century (Leadbeater, 2015). Early microscopists documented the fascinating collar complex, a diagnostic feature of choanoflagellates in which a picket fence of actin-filled microvilli encircles an apical flagellum (Figure 1B and C; Karpov and Leadbeater, 1998). Cells in the feeding chambers of sponges (choanocytes; Figure 1D) also contain a collar complex and these similarities in cellular architecture led Elie Metchnikoff in 1886 to hypothesize that these features were homologous and that animals and choanoflagellates shared a recent common ancestor (Metchnikoff, 1886; Brunet and King, 2020). The collar complex is now recognized as a widespread feature of animal cell biology, consistent with the results from multiple independent phylogenetic and phylogenomic comparisons that firmly established choanoflagellates as the sister group of animals (Figure 1A; Wainright et al., 1993; Lang et al., 2002; Burger et al., 2003; Philippe et al., 2004; Ruiz-Trillo et al., 2008). Subsequent forays into comparative genomics and transcriptomics revealed that many mediators of intercellular signaling, adhesion, cell differentiation, extracellular matrix and immunity in animals are also encoded in and transcribed from choanoflagellate genomes (King and Carroll, 2001; King et al., 2003; Snell et al., 2006; King et al., 2008; Fairclough et al., 2013; Richter et al., 2018; López-Escardó et al., 2019). These shared cellular and genetic traits suggested that mechanistic studies in choanoflagellates could prove valuable for core features of animal cell and molecular biology.

Nonetheless, the value of choanoflagellates to molecular and cell biologists was historically limited by the absence of key molecular genetic approaches that would allow mechanistic investigation of the functions of conserved genes. Moreover, understanding the transition between unicellular and colonial (i.e. multicellular) life history stages in choanoflagellates was poorly understood because of the absence of multicellular development in laboratory cultures.

Here we describe the history of molecular genetic approaches in choanoflagellates and the establishment of the colony-forming choanoflagellate *Salpingoeca rosetta* as a simple model for investigating the ancestry of animal gene function and cell biology. We focus on rosette colony development to show how discoveries from the study of this process not only revealed the biology of *S. rosetta* but also provided crucial insights that led to the establishment of genetic tools. We emphasize that a deep appreciation of choanoflagellate biology gained through the application of (relatively) inexpensive and general tools provided foundational knowledge required to establish methods for gene delivery, transgene expression, and genome editing.

S. rosetta growth and life history in the laboratory

When one of us (NK) began to study choanoflagellates in 2000, choanoflagellates were known to exist in diverse unicellular and colonial forms in nature (Leadbeater, 1983, 2015), but there were no readily available colony-forming species in culture. Fortunately, a visit with Dr. Tom Nerad at the ATCC inspired him to isolate a colony-forming choanoflagellate from brackish water samples collected near Hog Island, Virginia. This species, initially named *Proterospongia* sp. 50818 (King et al., 2003; Ruiz-Trillo et al., 2007; Carr et al., 2008) and later renamed *Salpingoeca rosetta* (Fairclough et al., 2010), was isolated as a single rosette colony (Figure 2A) that was then propagated in cereal grass media with environmental bacteria (as food) to establish a stable culture.

Within a small number of generations in the laboratory, *S. rosetta* cultures transitioned into a stably unicellular state (Figure 2B) and the rosette colonies that were apparent in the first

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 15 September 2021 doi:10.20944/preprints202109.0264.v1

environmental samples were rarely detected, preventing their study. Diverse environmental perturbations, including changes in media content, temperature, pH, and salinity, failed to re-stimulate rosette formation. However, a serendipitous observation made during the preparation of *S. rosetta* for genome sequencing revealed how rosette development could be induced (Fairclough et al., 2013).

Like all other choanoflagellates, *S. rosetta* must be co-cultured with bacteria as a food source. The diversity and abundance of bacteria in choanoflagellate cultures can interfere with eukaryotic genome sequencing and assembly, so then-undergraduate Rick Zuzow treated the cultures with different cocktails of antibiotics (Fairclough et al., 2010). One antibiotic cocktail blocked all rosette development while another unexpectedly yielded a culture in which rosettes developed readily and reproducibly. Experiments with these cultures performed by then-post-doc Rosie Alegado and members of the laboratory of Prof. Jon Clardy (Harvard Medical School) revealed that a key factor in rosette induction was the presence of a previously-uncharacterized bacterial species, which we named *Algoriphagus machipongonensis* (Alegado et al., 2012, 2013). Addition of *A. machipongonensis* to a culture of single cells was sufficient to induce rosette development through serial rounds of cell division.

Subsequent improvements to *S. rosetta* culture conditions, including co-culture with the feeder bacterium *Echinicola pacifica* and the development of new semi-defined media formulations (Levin and King, 2013; Woznica et al., 2016), increased the efficiency of rosette induction and later aided the establishment of transfection and genome editing methods (Booth et al., 2018; Booth and King, 2020). These advances provided the first evidence of bacterial regulation of choanoflagellate cell state transitions and provided a means by which to control rosette development in the laboratory.

Multicellular development was not the only *S. rosetta* life history transition regulated by environmental cues. In a separate line of research, then-graduate student Tera Levin found that prolonged starvation (i.e. over two weeks) induced mating in *S. rosetta* (Levin and King, 2013). Four years later, then-graduate student Arielle Woznica found that treatment of *S. rosetta* cultures with the bacterium *Vibrio fischeri* induced *S. rosetta* to swarm and then mate, in this case within an hour after treatment (Woznica et al., 2017). These findings demonstrated the essentiality of bacterial cues to the regulation of choanoflagellate biology and provided two independent methods for inducing mating, which later proved critical for the establishment of genetic screens. This period of *S. rosetta* research demonstrated that simple alterations to growth conditions, including altering nutrient content or changing the bacterial composition, could reveal novel biology and provide levers for controlling life history transitions in the laboratory.

Forward genetics and the hunt for rosette defective mutants

Although rosette development in *S. rosetta* could now be readily induced by treatment with *A. machipongonensis*, nothing was known about the genes that regulated the process. The gold standard for discovering the genes underlying a biological process is to perform an unbiased mutant screen, but such screens faced multiple barriers in choanoflagellates. Methods for mutagenesis had not been established (and there were reasons to think that it might be challenging), no genetic markers had yet been identified and the cells are grown in liquid media, meaning mutants could not be isolated, propagated, and phenotyped on plates.

The same Tera Levin who previously uncovered a method for inducing mating (Levin and King, 2013; see above) systematically overcame each of these barriers, allowing the establishment of forward genetics in *S. rosetta* (Levin et al., 2014). She first developed methods for mutagenesis of *S. rosetta* with both EMS and X-rays and identified effective doses using “death curves.” Next, she established a phenotypically wild type mapping strain containing over 39,000 genetic variants distributed throughout the genome. The mapping strain was clonally propagated from an isolated cell that had been exposed to high doses of mutagen, found to

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 15 September 2021 doi:10.20944/preprints202109.0264.v1

have no obvious mutant phenotypes (e.g. growth defects or loss of rosette development), and then sequenced to identify genetic variants across the genome that could be used for mapping mutations through crosses.

These tools set up the stage for the first genetic screen in a choanoflagellate. Over 37,000 clones were isolated by serial dilution and screened by eye to identify 16 mutants that fell into seven phenotypic classes. One mutant, Rosetteless, was of particular interest; while it lacked the ability to form rosettes, all other aspects of its life history and cell biology seemed to be wild type. By crossing the Rosetteless mutant to the mapping strain, Levin identified a genetic lesion that co-segregated with the loss of rosette development. The lesion fell in a C-type lectin gene that we named "*rosetteless"* after its phenotype. Rosetteless protein is secreted into the extracellular matrix (ECM) at the basal ends of cells in rosettes (Figure 2A, C), and the mutant allele abrogates Rosetteless protein expression, resulting in cells that grow normally except for their inability to form rosettes (Figure 2D, C; (Levin et al., 2014)).

In two other rosette-defect mutants, Jumble and Couscous, single cells aggregated into large amorphous clumps and true rosettes were never observed to develop (Wetzel et al., 2018). Then-graduate student Laura Wetzel used crosses between these mutants and the mapping strain, coupled with bulk segregant analysis, to identify genetic lesions in two different glycosyltransferases that co-segregated with the clumping phenotype. Using new transgenic approaches (described in more detail below), Wetzel corroborated the mapping cross results and demonstrated that Jumble localizes to the Golgi, whereas Couscous localizes to puncta that may represent the endoplasmic reticulum.

Taken together, these studies demonstrated the efficacy of forward genetics for identifying genes required for *S. rosetta* rosette development. Moreover, the unbiased genetics approach emphasized the importance of the extracellular matrix during the development of rosettes. The criticality of the ECM for rosette development was shown independently through a recent study of biophysical constraints on rosette morphogenesis (Larson et al., 2020). Finally, the connection of genotype to phenotype in the context of the Rosetteless mutant provided a toehold against which different genome editing approaches could later be tested.

Gene delivery and genome editing in *S. rosetta*

While genomic, biochemical, and forward genetic approaches provided key early insights into rosette development, the absence of transgenic approaches in *S. rosetta* limited its experimental tractability and potential relevance as a laboratory model. From its isolation in 2000 until 2018, repeated efforts to establish transgenics failed, in large part because the field lacked methods for robustly delivering macromolecules into *S. rosetta* cells. Choanoflagellate cells are too small (~ 5 μm) for microinjection and standard electroporation methods had thus far failed, both because of the necessity of culturing *S. rosetta* in sea water and because we lacked a positive control for transfection. Key advances in our understanding of *S. rosetta's* genetics and cell physiology that arose through our basic research into rosette development proved to be critical for overcoming these barriers.

In many single-celled organisms, electroporation with nucleic acids or proteins is physically blocked by an organic layer (e.g. a cell wall or extracellular matrix) surrounding the cell (Meilhoc et al., 1990; Shimogawara et al., 1998; Thompson et al., 1998; Benatuil et al., 2010; Kawai et al., 2010). However, because choanoflagellates lack a cell wall and prior electron microscopy studies of unicellular life stages suggested that they might be "naked," we admittedly paid little heed to the extracellular matrix in our early forays into establishing transgenics. This changed with the discovery of *rosetteless* (Levin et al., 2014) and improvements in electron microscopy that revealed a layer of extracellular matrix encapsulating *S. rosetta* cells (Figure 3A; Dayel et al., 2011). Following lessons learned in other microeukaryotes, we therefore developed a method (Figure 3B) to prime *S. rosetta* for transfection by bathing single cells in a cocktail of papain (a protease) and specific salts that

gently removes extracellular proteins (dx.doi.org/10.17504/protocols.io.h68b9hw), leaving the cell membrane vulnerable to pore-forming electrical pulses delivered through electroporation, but healthy enough for the pores to seal and the cells to recover.

The next step was to identify the best buffering conditions and electrical pulses to create pores in the *S. rosetta* membrane. We chose to use nucleofection, a specialized form of electroporation, as it has proven effective for delivering DNA and protein into difficult-to-transfect cells (Gresch et al., 2004; Hamm et al., 2002). Although the exact details of nucleofection remain proprietary, an excellent review of the biophysical principles that underlie a variety of gene delivery methods can be found in (Stewart et al., 2018). A major strength of nucleofection, particularly when used in conjunction with a plate-based platform, is that the combination of pulses and diverse transfection buffers creates a large parameter space to explore for testing a wide range of nucleofection conditions. By first removing the extracellular matrix and then applying the most effective nucleofection conditions, we achieved transfection efficiencies of 1%. (An important downside for nucleofection is the expense of the machine itself and the proprietary buffers. Therefore, we have recently established a second transfection protocol for *S. rosetta* using an inexpensive, square-pulse electroporator [Figure 3C; dx.doi.org/10.17504/protocols.io.bx5vpq66]. We hope that these initial parameters will make transfection accessible to a broader community of choanoflagellate researchers and serve as the start of further optimization.)

Our second big challenge was the detection of successful gene delivery events. Until everything worked – gene delivery, gene expression, and detection – we would not know if any part of the transfection protocol was working and therefore were operating in the dark. To improve our chances of success, we cloned promoters and regulatory elements from highly-expressed *S. rosetta* genes that lacked introns and replaced their open reading frames with genes, e.g. *luciferase* or fluorescent proteins, that would report successful transfection. Although genes that confer resistance to a toxic drug can aid in selection for rare drug-resistant transformants, we were discouraged by the variable and unpredictable responses of *S. rosetta* cultures to drug treatments. Fluorescent protein expression in *S. rosetta* also presented problems because, as we later learned, green fluorescent protein (GFP) and its derivatives did not fold properly in *S. rosetta*.

Therefore, we settled on a reporter gene that encodes a small, sensitive, and stable protein, Nanoluc (Hall et al., 2012), to detect expression from transfected plasmids. In other experimental models (Vinayak et al., 2015), the use of *nanoluc* provided exceptional signal to noise, allowing clear distinction between successful and unsuccessful transfection experiments (Figure 3C). Moreover, the small size of *nanoluc* (a 519 bp open reading frame that encodes a ~19 kD protein), helped reduce the overall size of our transgenic constructs. After testing a broad array of transfection parameters with *nanoluc* constructs, we identified a small combination of nucleofection buffers and pulse programs that yielded luciferase activity. We subsequently optimized around these parameters to develop a robust and reproducible transfection protocol (dx.doi.org/10.17504/protocols.io.h68b9hw).

The establishment of this protocol almost immediately transformed the study of choanoflagellates. The expression of fluorescent markers brought the cellular architecture of *S. rosetta* into living color in a way not possible previously (Figure 3D). The ability to express and perform live imaging of fluorescently tagged proteins also helped to validate the mapping of the *jumble* and *couscous* glycosyltransferases (Wetzel et al., 2018), revealed that septins localize to the basal pole of *S. rosetta* cells (Booth et al., 2018), and revealed the cytoskeletal dynamics that accompany the transition from flagellated swimming cells to crawling amoeboid cells upon confinement (Brunet et al., 2021). In addition, this protocol served as the foundation for transfection of the choanoflagellate *Monosiga brevicollis* (Woznica et al., 2021). Finally, and most importantly for the development of *S. rosetta* as an experimental model, it opened the possibility for delivering the machinery for genome editing.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 15 September 2021 doi:10.20944/preprints202109.0264.v1

CRISPR/Cas9-mediated genome editing has become a robust method for performing reverse genetics in established and emerging model systems. An easily designed synthetic guide RNA (gRNA) associates with the Cas9 endonuclease, and together they drive cleavage at specific targeted sequences in the genome (Jinek et al., 2013, 2012; Cong et al., 2013). After the Cas9 endonuclease introduces double-stranded breaks in the genome, the incorporation of mutations ultimately relies on errors introduced during DNA repair (Rouet et al., 1994; Choulika et al., 1995; Bibikova et al., 2001; Jinek et al., 2013; Cong et al., 2013). The nature of those errors depends on the type of DNA repair mechanism implemented by an organism – either homology-directed repair (HDR) in which repair of the double stranded break incorporates mutations from a DNA template that shares homology with the region surrounding the cut site or non-homologous end-joining (NHEJ) in which repair of the double stranded break can lead to untemplated insertions or deletions as the DNA ends rejoin. Thus, a snag for implementing Cas9-mediated genome editing in emerging model systems can be the paucity of information on the DNA repair mechanisms, which can widely vary between organisms and cell types (Yeh et al., 2019). For example, *S. rosetta* preferentially incorporates mutations from HDR templates at a frequency ten-fold higher than untemplated insertions and deletions (Booth and King, 2020). In contrast, *M. brevicollis* seems to prefer the incorporation of untemplated mutations, as evidenced by the incorporation of insertions and deletions in a targeted gene despite the co-delivery of HDR templates (Woznica et al., 2021).

To establish genome editing in *S. rosetta*, we first focused on converting the ribosomal protein gene *rpl36a* to *rpl36a*^{P56Q}, a cycloheximide resistance allele (Kawai et al., 1992; Dehoux et al., 1993; Kondo et al., 1995; Kim et al., 1998; Stevens et al., 2001). This relatively simple edit allowed us to select edited cells due to their cycloheximide resistance, thereby simplifying one stage of the optimization process. By iterating around the editing of *rpl36a*, we identified key parameters that promoted genome editing success, including the need to use least 20 pmol of Cas9 ribonucleoprotein (RNP) and more than 200 nmol of a single-stranded DNA repair template that had 50 bases of homology flanking the desired mutation (Booth and King, 2020).

Afterward, we targeted *rosetteless*, both because a successful edit should yield a predictable phenotype (loss of rosette development) and, more broadly, to demonstrate the utility of genome editing for investigating gene function in *S. rosetta*. The original *rosetteless* allele was likely a hypomorph whose phenotype manifested through aberrant splicing from a transition mutation in a 5' splice site. By introducing an independent knockout of *rosetteless* by genome editing, we corroborated the necessity of *rosetteless* for rosette development (Figure 3E-H). Thus, genome editing in combination with mating crosses are essential tools for the verification of mutations identified in forward genetic screens. Most importantly, genome editing now provides a method to directly test hypotheses about gene function in *S. rosetta*.

The major barrier that remains for genome engineering in *S. rosetta* is the accurate and precise introduction of large DNA fragments. Even with small insertions (~20 bases), the efficiency of genome editing (~1%) requires screening through hundreds of clonal strains to find one that bears a genome edit. Engineering larger insertions, which will likely integrate into the genome with even lower frequencies, will require modifications to improve gene delivery, to increase integration frequencies, and/or to select for stable transgenes. A puromycin resistance gene has already been used for the selection of transgenic cassettes (Wetzel et al., 2018; Brunet et al., 2021; Woznica et al., 2021), but puromycin also affects the feeder bacteria that support choanoflagellate growth, leading to variable selection conditions. The development of more selectable markers that confer resistance to drugs that selectively target eukaryotes would help advance genetics in choanoflagellates and other heterotrophic organisms that require bacterial feeding.

The door is now open to study the molecular biology of *S. rosetta* and other relatives of animals

 doi:10.20944/preprints202109.0264.v1

The lessons learned from *S. rosetta* have aided efforts to establish genetics in other emerging microeukaryotic models. Of particular note is the recent development of molecular methods for *M. brevicollis* (Woznica et al., 2021). The establishment of gene delivery and genome editing in *M. brevicollis* helped to uncover potential ancestral functions of the innate immunity gene STING. Although the exact protocols from *S. rosetta* were not readily exported to *M. brevicollis*, variations on the general principles used during the development of *S. rosetta* transfection facilitated the construction of transgenic lines and the generation of a STING knockout with genome editing. This work clearly demonstrates the maturation of choanoflagellates as model systems to investigate core process conserved in both choanoflagellates and animals; the Woznica et al., 2021 study drew on comparative genomics (Richter et al., 2018), reported the discovery of a bacterial pathogen of *M. brevicollis*, and developed methods for molecular genetics to show that STING likely functioned as an innate immunity gene in the last common ancestor of choanoflagellates and animals. Similar advances in the development of genetic methods in other protistan relatives of choanoflagellates and animals (Suga and Ruiz-Trillo, 2013; Parra-Acero et al., 2018; Li et al., 2018; Faktorová et al., 2020; Kożyczkowska et al., 2021) now provide a menagerie of microeukaryotes to reconstruct how evolutionary changes prior to the emergence of animals laid the foundations for complex, multicellular development.

We hope that this review serves as an invitation for the ‘choano curious’ to pursue their research in these fascinating organisms. Rosette development is just one dimension of the rich biology that is now amenable to molecular investigation in *S. rosetta*; others include cell differentiation, interspecies interactions, and evolutionary cell biology. As we move forward with a new era of molecular genetics in choanoflagellates, we highlight the important roles that classical methods from biochemistry, cytology, and genetics played in developing modern tools for transforming *S. rosetta*, editing its genome, and illuminating its cell biology (Supplementary Table).

Acknowledgements

No amount of thanks can properly convey our deep gratitude to the many scientists – those from the King Lab and those in the larger community of holozoan researchers – for the passion, discoveries and insights that fostered the emergence of choanoflagellates as model systems, only a fraction of whom could be highlighted here. It really takes dedication from a community to found new model systems. We also thank the Gordon and Betty Moore Foundation Marine Microbiology and HHMI for their generous support of work that established fundamental biology and genetic methods in choanoflagellates. Finally, we thank Matt Welch and laboratory for use of the square-wave electroporator.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 15 September 2021 doi:10.20944/preprints202109.0264.v1

Figure 1

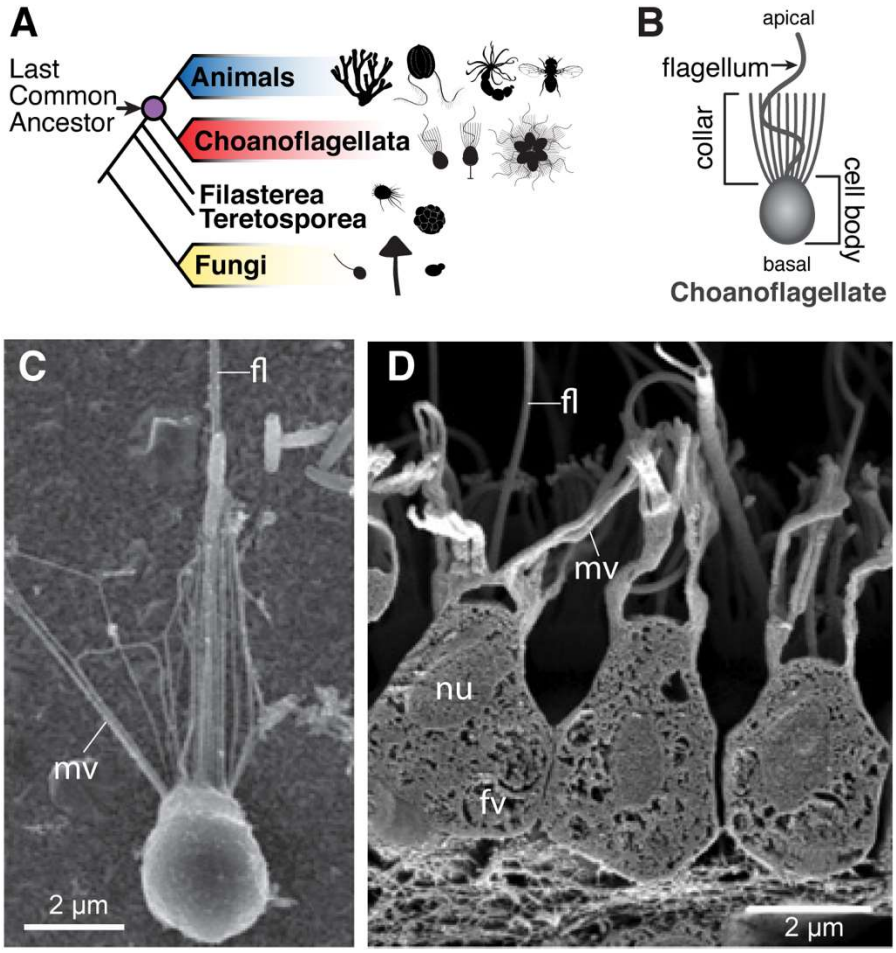
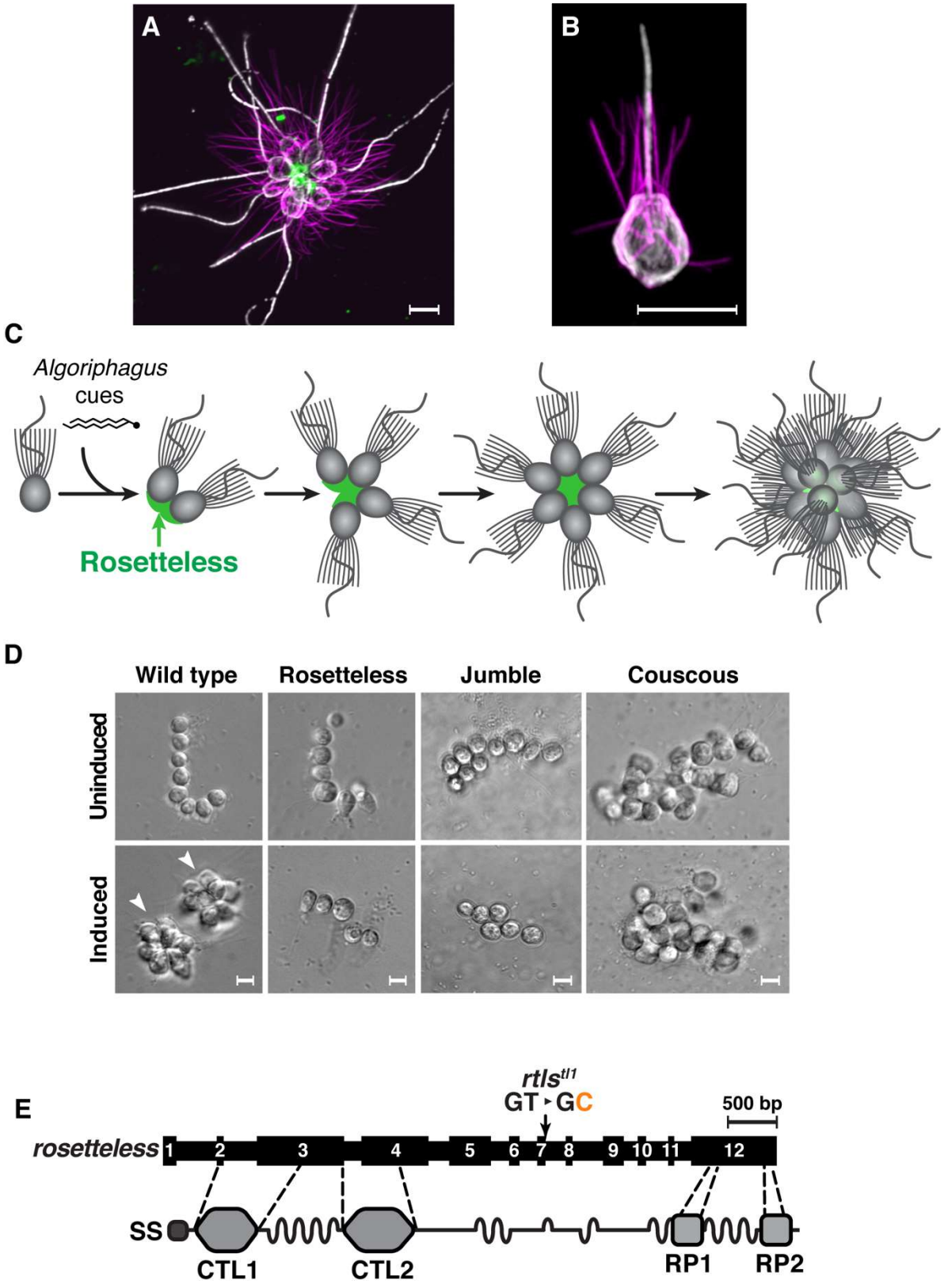


Figure 1. Choanoflagellates and their closest living relatives, the animals, possess cells with a collar complex.
(**A**) Choanoflagellates are the closest living relatives of animals and diverged from their last common ancestor over 600 million years ago (Parfrey et al., 2011). Fungi, the closest living group of organisms with complex multicellularity, last shared a common ancestor with animals and choanoflagellates ~1.5 billion years ago. (**B**) A schematic of a choanoflagellate cell highlights their distinctive trait: the collar complex. At the apical pole of the cell, the collar forms from actin-filled microvilli encircling the flagellum. (**C**) A scanning electron micrograph of *S. rosetta* (from Dayel *et al.* 2011), showing the apical flagellum (fl) surrounded by the collar of microvilli (mv). **(D)** Scanning electron micrograph of a cross-section through sponge choanocytes (Leys and Hill, 2012) showing the collar complex: the apical flagellum (fl) surrounded by microvilli (mv). Also indicated are the nucleus (nu) and food vacuole (fv).

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 15 September 2021 doi:10.20944/preprints202109.0264.v1

Figure 2



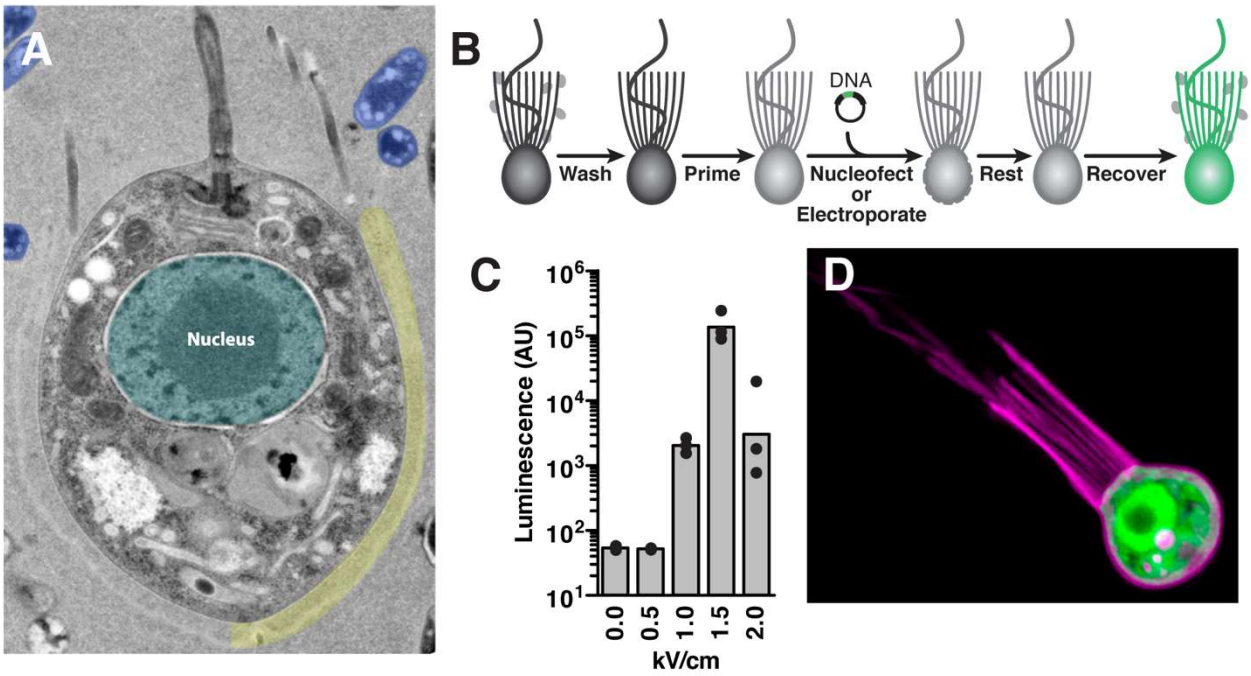
Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 15 September 2021 doi:10.20944/preprints202109.0264.v1

Figure 2. Environmental cues and a C-type lectin together regulate multicellular development in *S. rosetta*

The dynamic life history of *S. rosetta* includes a multicellular "rosette" stage **(A)** and a single celled "slow swimmer" stage **(B)** The phalloidin-stained microvilli are highlighted in magenta, microtubules in white, and the extracellular matrix protein Rosetteless in green**. (C)** The transition from the slow swimmer stage to the development of the multicellular rosette proceeds through multiple rounds of incomplete cytokinesis (Fairclough et al., 2010). Thus, every rosette is clonal. Rosette development can be induced by secreted cues from *Algoriphagus* bacteria (Alegado et al., 2012; Woznica et al., 2016) and requires the activity of the C-type lectin protein Rosetteless (green; Levin et al., 2014), that contributes to the extracellular matrix at the basal poles of cells in rosettes. **(D)** The Rosetteless mutant was uncovered, along with rosette-defective mutants Jumble and Couscous (previously called Branched), as part of a forward genetic screen. Wild type cells typically exist as single cells or simple chains in the absence of inducing factors produced by *Algoriphagus*; whereas, the presence of *Algoriphagus* cues induces the development of compact and robust rosettes (Alegado et al., 2012; Levin et al., 2014). In contrast, Rosetteless mutants never develop into rosettes, even in the presence of *Algoriphagus* conditioned medium (Levin et al., 2014). Jumble and Couscous (previously called Branched in Levin et al., 2014) mutant cells never develop into rosettes, and instead aggregate into large amorphous clumps of cells (Wetzel et al., 2018). **(E)** The 12-exon *rosetteless* gene encodes a protein with an N-terminal signal sequence (SS), two C-type lectin domains (CTL1 and 2) and a pair of repeats (RP1 and 2) near the C-terminus. The mutant allele of *rosetteless* (*rtls*^{tl1}) in the original Rosetteless mutant contained a GT to GC transition at the splice junction between exon 7 and intron 7, which disrupted splicing and protein expression. All scale bars are 5 µm. (Panels A and E adapted from Booth and King, 2020, Panel B adapted from Booth et al., 2018, and Panel D from Levin et al., 2014)

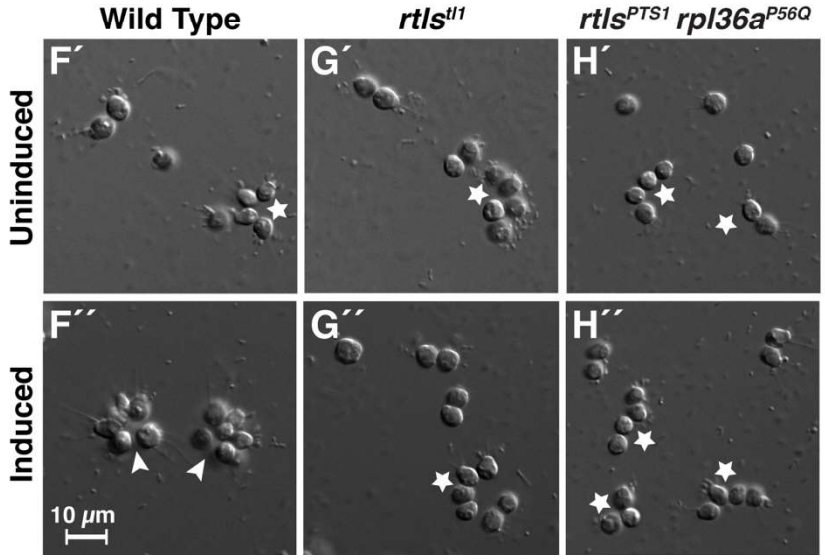
Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 15 September 2021 doi:10.20944/preprints202109.0264.v1

Figure 3



E

Strain	Locus (supercontig:nucleotide)	
	rtls exon 4 (8:429,354)	*rtls* intron 7 (8:427,804)
Wild Type	GAT----------TGACGG	TCGTGAG
rtls^{tl1}	GAT----------TGACGG	TCGCGAG
rtls^{PTS1}*rpl36a*^{P56Q}	GATTTTATTTAATTAAATAAATGACGG	TCGTGAG



Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 15 September 2021 doi:10.20944/preprints202109.0264.v1

Figure 3. Establishment of transfection and genome editing in *S. rosetta*
(A) Transmission electron microscopy reveals a layer of extracellular matrix (ECM; highlighted in yellow on one half of the cell, leaving the ECM on the other half unlabeled for easier viewing). Also labeled with pseudocolor: nucleus (teal) and representative bacteria (blue). **(B)** A step-wise procedure to deliver plasmids and other macromolecules into *S. rosetta* (Booth et al., 2018). Cells are prepared for gene delivery by washing away feeder bacteria–through centrifugation–from *S. rosetta* that has been grown to its mid-logarithmic phase. Afterwards, *S. rosetta* is primed for gene delivery by removing a proteinaceous ECM with a buffer comprised of a chaotropic cation (Lithium), a gentle chelator (Citrate), a reducing agent (Cysteine), and a gentle protease (Papain). The primed cells and cargoes (expression plasmids or gene editing materials) are then combined in a nucleofection or electroporation buffer. After applying an electrical pulse to deliver cargoes inside of *S. rosetta*, a rest buffer, which has low ionic strength and high osmolarity, is added to cells to promote the closure of holes that persist in membranes long after the application of the electrical pulse. Finally, cells are transferred to growth media to recover before being assayed for gene activity or cultured further to establish transgenic strains. An optimized protocol for transfection was developed using a nucleofector and proprietary buffers (Booth and King, 2020 and https://dx.doi.org/10.17504/protocols.io.bx58pq9w). (**C**) More recently we have established a transformation protocol using an inexpensive, square-pulse electroporator (https://dx.doi.org/10.17504/protocols.io.bx5vpq66), which can be used to deliver plasmids for exogenous gene expression in *S. rosetta*. By electroporating cells with an expression plasmid pEFL-*nanoluc* (Addgene NK606), containing the EFL promoter from *S. rosetta* fused in-frame to the *nanoluc* luciferase gene and varying the field strength (kV/cm), we found that voltages between 1 - 1.5 kV/cm transfected cells most effectively and yielded the highest nanoluc activities. Each dot represents one of three independent transfections, and the grey bar indicates the average of those experiments. Because the nanoluc activity in one electroporation reaction uses ten-fold more material to achieve a similar level of activity as one nucleofection reaction, we estimate that these electroporation conditions transfect cells with ten-fold lower efficiency than nucleofection. **(D)** Transfection of *S. rosetta* with plasmids expressing fluorescent proteins can be used to highlight cellular architecture in live cells. The collar and cell membranes here are visualized through the expression of mCherry fused to an *S. rosetta* geranyl-geranylation sequence (magenta; Addgene NK644) and the cell body, illuminated by the expression of cytoplasmic mTFP (green; Addgene NK591). **(E, F)** Genome editing of the *rosetteless* gene recapitulates the phenotype of the previously isolated rosetteless mutant. **(E)** Genotypes of *S. rosetta* strains show two distinct alleles in *rosetteless* that were either isolated from a forward genetic screen ($rtls^{tl}$; Levin et al., 2014) or engineered through CRISPR-Cas9 mediated genome editing ($rtls^{PTS1}$ $rpl36a^{P56Q}$; Booth and King, 2020). The $rtls^{tl1}$ allele is a T to C transition in the 5'-splice site of intron 7 that abrogates the expression of *rosetteless* through aberrant splicing (also see Fig. 2). The $rtls^{PTS1}$ $rpl36a^{P56Q}$ strain was engineered to bear a cycloheximide resistance allele ($rpl36^{P56Q}$) and a *rosetteless* allele with a palindromic premature termination sequence (PTS: TTTATTTAATTAAATAAA) inserted into the fourth exon that stops the transcription and translation of *rosetteless* products. **(F)** Independent mutations show the necessity of *rosetteless* for multicellular development in *S. rosetta*. In the absence of inducing cues from *Algoriphagus*, all strains (F'-H') grew as chains (stars) or single cells. Upon the addition of inducing cues, only the wild-type (F") strain formed rosettes (arrowheads) while strains with $rtls^{tl1}$ (E") or $rtls^{PTS1}$ $rpl36^{P56Q}$ (F") alleles failed to form rosettes. (Panel A courtesy of Kent McDonald and adapted from Booth et al., 2018; Panels E-H adapted from Booth and King, 2020)

References

Alegado RA, Brown LW, Cao S, Dermenjian RK, Zuzow R, Fairclough SR, Clardy J, King N. 2012. A bacterial sulfonolipid triggers multicellular development in the closest living relatives of animals. *eLife* **1**:e00013. doi:10.7554/eLife.00013

Alegado RA, Grabenstatter JD, Zuzow R, Morris A, Huang SY, Summons RE, King N. 2013. Algoriphagus machipongonensis sp. nov., co-isolated with a colonial choanoflagellate. *Int J Syst Evol Microbiol* **63**:163–168. doi:10.1099/ijs.0.038646-0

Benatuil L, Perez JM, Belk J, Hsieh C-M. 2010. An improved yeast transformation method for the generation of very large human antibody libraries. *Protein Eng Des Sel* **23**:155–159. doi:10.1093/protein/gzq002

Bibikova M, Carroll D, Segal DJ, Trautman JK, Smith J, Kim Y-G, Chandrasegaran S. 2001. Stimulation of Homologous Recombination through Targeted Cleavage by Chimeric Nucleases. *Mol Cell Biol* **21**:289–297. doi:10.1128/MCB.21.1.289-297.2001

Booth DS, King N. 2020. Genome editing enables reverse genetics of multicellular development in the choanoflagellate *Salpingoeca rosetta*. *eLife* **9**:e56193. doi:10.7554/eLife.56193

Booth DS, Szmidt-Middleton H, King N. 2018. Choanoflagellate transfection illuminates their cell biology and the ancestry of animal septins. *Mol Biol Cell* mbcE18080514. doi:10.1091/mbc.E18-08-0514

Brunet T, Albert M, Roman W, Coyle MC, Spitzer DC, King N. 2021. A flagellate-to-amoeboid switch in the closest living relatives of animals. *eLife* **10**. doi:10.7554/eLife.61037

Brunet T, King N. 2020. The Single-Celled Ancestors of Animals: A History of Hypotheses. doi:10.20944/preprints202011.0302.v1

Burger G, Forget L, Zhu Y, Gray MW, Lang BF. 2003. Unique mitochondrial genome architecture in unicellular relatives of animals. *Proc Natl Acad Sci* **100**:892–897. doi:10.1073/pnas.0336115100

Carr M, Leadbeater BSC, Hassan R, Nelson M, Baldauf SL. 2008. Molecular phylogeny of choanoflagellates, the sister group to Metazoa. *Proc Natl Acad Sci U S A* **105**:16641–16646. doi:10.1073/pnas.0801667105

Choulika A, Perrin A, Dujon B, Nicolas JF. 1995. Induction of homologous recombination in mammalian chromosomes by using the I-SceI system of *Saccharomyces cerevisiae*. *Mol Cell Biol* **15**:1968–1973.

Cong L, Ran FA, Cox D, Lin S, Barretto R, Habib N, Hsu PD, Wu X, Jiang W, Marraffini LA, Zhang F. 2013. Multiplex Genome Engineering Using CRISPR/Cas Systems. *Science* **339**:819–823. doi:10.1126/science.1231143

Dayel MJ, Alegado RA, Fairclough SR, Levin TC, Nichols SA, McDonald K, King N. 2011. Cell differentiation and morphogenesis in the colony-forming choanoflagellate *Salpingoeca rosetta*. *Dev Biol* **357**:73–82. doi:10.1016/j.ydbio.2011.06.003

Dehoux P, Davies J, Cannon M. 1993. Natural cycloheximide resistance in yeast. The role of ribosomal protein L41. *Eur J Biochem* **213**:841–848.

Fairclough SR, Chen Z, Kramer E, Zeng Q, Young S, Robertson HM, Begovic E, Richter DJ, Russ C, Westbrook MJ, Manning G, Lang BF, Haas B, Nusbaum C, King N. 2013. Premetazoan genome evolution and the regulation of cell differentiation in the choanoflagellate *Salpingoeca rosetta*. *Genome Biol* **14**:R15. doi:10.1186/gb-2013-14-2-r15

Fairclough SR, Dayel MJ, King N. 2010. Multicellular development in a choanoflagellate. *Curr Biol* **20**:R875–R876. doi:10.1016/j.cub.2010.09.014

Faktorová D, Nisbet RER, Fernández Robledo JA, Casacuberta E, Sudek L, Allen AE, Ares M, Aresté C, Balestreri C, Barbrook AC, Beardslee P, Bender S, Booth DS, Bouget F-Y,

Bowler C, Breglia SA, Brownlee C, Burger G, Cerutti H, Cesaroni R, Chiurillo MA, Clemente T, Coles DB, Collier JL, Cooney EC, Coyne K, Docampo R, Dupont CL, Edgcomb V, Einarsson E, Elustondo PA, Federici F, Freire-Beneitez V, Freyria NJ, Fukuda K, García PA, Girguis PR, Gomaa F, Gornik SG, Guo J, Hampl V, Hanawa Y, Haro-Contreras ER, Hehenberger E, Highfield A, Hirakawa Y, Hopes A, Howe CJ, Hu I, Ibañez J, Irwin NAT, Ishii Y, Janowicz NE, Jones AC, Kachale A, Fujimura-Kamada K, Kaur B, Kaye JZ, Kazana E, Keeling PJ, King N, Klobutcher LA, Lander N, Lassadi I, Li Z, Lin S, Lozano J-C, Luan F, Maruyama S, Matute T, Miceli C, Minagawa J, Moosburner M, Najle SR, Nanjappa D, Nimmo IC, Noble L, Novák Vanclová AMG, Nowacki M, Nuñez I, Pain A, Piersanti A, Pucciarelli S, Pyrih J, Rest JS, Rius M, Robertson D, Ruaud A, Ruiz-Trillo I, Sigg MA, Silver PA, Slamovits CH, Jason Smith G, Sprecher BN, Stern R, Swart EC, Tsaousis AD, Tsypin L, Turkewitz A, Turnšek J, Valach M, Vergé V, von Dassow P, von der Haar T, Waller RF, Wang L, Wen X, Wheeler G, Woods A, Zhang H, Mock T, Worden AZ, Lukeš J. 2020. Genetic tool development in marine protists: emerging model organisms for experimental cell biology. *Nat Methods* **17**:481–494. doi:10.1038/s41592-020-0796-x

Gresch O, Engel FB, Nesic D, Tran TT, England HM, Hickman ES, Körner I, Gan L, Chen S, Castro-Obregon S, Hammermann R, Wolf J, Müller-Hartmann H, Nix M, Siebenkotten G, Kraus G, Lun K. 2004. New non-viral method for gene transfer into primary cells. *Methods*, Transfection of Mammalian Cells **33**:151–163. doi:10.1016/j.ymeth.2003.11.009

Hall MP, Unch J, Binkowski BF, Valley MP, Butler BL, Wood MG, Otto P, Zimmerman K, Vidugiris G, Machleidt T, Robers MB, Benink HA, Eggers CT, Slater MR, Meisenheimer PL, Klaubert DH, Fan F, Encell LP, Wood KV. 2012. Engineered Luciferase Reporter from a Deep Sea Shrimp Utilizing a Novel Imidazopyrazinone Substrate. *ACS Chem Biol* **7**:1848–1857. doi:10.1021/cb3002478

Hamm A, Krott N, Breibach I, Blindt R, Bosserhoff AK. 2002. Efficient Transfection Method for Primary Cells. *Tissue Eng* **8**:235–245. doi:10.1089/107632702753725003

Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna JA, Charpentier E. 2012. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* **337**:816–821. doi:10.1126/science.1225829

Jinek M, East A, Cheng A, Lin S, Ma E, Doudna J. 2013. RNA-programmed genome editing in human cells. *eLife* **2**:e00471. doi:10.7554/eLife.00471

Karpov SA, Leadbeater BSC. 1998. Cytoskeleton Structure and Composition in Choanoflagellates. *J Eukaryot Microbiol* **45**:361–367. doi:https://doi.org/10.1111/j.1550-7408.1998.tb04550.x

Kawai S, Hashimoto W, Murata K. 2010. Transformation of *Saccharomyces cerevisiae* and other fungi. *Bioeng Bugs* **1**:395–403. doi:10.4161/bbug.1.6.13257

Kawai S, Murao S, Mochizuki M, Shibuya I, Yano K, Takagi M. 1992. Drastic alteration of cycloheximide sensitivity by substitution of one amino acid in the L41 ribosomal protein of yeasts. *J Bacteriol* **174**:254–262.

Kim IG, Nam SK, Sohn JH, Rhee SK, An GH, Lee SH, Choi ES. 1998. Cloning of the ribosomal protein L41 gene of *Phaffia rhodozyma* and its use a drug resistance marker for transformation. *Appl Environ Microbiol* **64**:1947–1949.

King N, Carroll SB. 2001. A receptor tyrosine kinase from choanoflagellates: Molecular insights into early animal evolution. *Proc Natl Acad Sci* **98**:15032–15037. doi:10.1073/pnas.261477698

King N, Hittinger CT, Carroll SB. 2003. Evolution of Key Cell Signaling and Adhesion Protein Families Predates Animal Origins. *Science* **301**:361–363. doi:10.1126/science.1083853

King N, Westbrook MJ, Young SL, Kuo A, Abedin M, Chapman J, Fairclough S, Hellsten U, Isogai Y, Letunic I, Marr M, Pincus D, Putnam N, Rokas A, Wright KJ, Zuzow R, Dirks

W, Good M, Goodstein D, Lemons D, Li W, Lyons JB, Morris A, Nichols S, Richter DJ, Salamov A, Sequencing JGI, Bork P, Lim WA, Manning G, Miller WT, McGinnis W, Shapiro H, Tjian R, Grigoriev IV, Rokhsar D. 2008. The genome of the choanoflagellate *Monosiga brevicollis* and the origin of metazoans. *Nature* **451**:783–788. doi:10.1038/nature06617

Kondo K, Saito T, Kajiwara S, Takagi M, Misawa N. 1995. A transformation system for the yeast *Candida utilis*: use of a modified endogenous ribosomal protein gene as a drug-resistant marker and ribosomal DNA as an integration target for vector DNA. *J Bacteriol* **177**:7171–7177.

Kożyczkowska A, Najle SR, Ocaña-Pallarès E, Aresté C, Shabardina V, Ara PS, Ruiz-Trillo I, Casacuberta E. 2021. Stable transfection in protist *Corallochytrium limacisporum* identifies novel cellular features among unicellular animals relatives. *Curr Biol*. doi:10.1016/j.cub.2021.06.061

Lang BF, O'Kelly C, Nerad T, Gray MW, Burger G. 2002. The Closest Unicellular Relatives of Animals. *Curr Biol* **12**:1773–1778. doi:10.1016/S0960-9822(02)01187-9

Larson BT, Ruiz-Herrero T, Lee S, Kumar S, Mahadevan L, King N. 2020. Biophysical principles of choanoflagellate self-organization. *Proc Natl Acad Sci* **117**:1303–1311. doi:10.1073/pnas.1909447117

Leadbeater BSC. 2015. The choanoflagellates: evolution, biology, and ecology. Cambridge, United Kingdom: Cambridge University Press.

Leadbeater BSC. 1983. Life-history and ultrastructure of a new marine species of Proterospongia (Choanoflagellida). *J Mar Biol Assoc U K* **63**:135–160. doi:10.1017/S0025315400049857

Levin TC, Greaney AJ, Wetzel L, King N. 2014. The Rosetteless gene controls development in the choanoflagellate *S. rosetta*. *eLife* **3**. doi:10.7554/eLife.04070

Levin TC, King N. 2013. Evidence for Sex and Recombination in the Choanoflagellate *Salpingoeca rosetta*. *Curr Biol* **23**:2176–2180. doi:10.1016/j.cub.2013.08.061

Leys SP, Hill A. 2012. Chapter one - The Physiology and Molecular Biology of Sponge Tissues In: Becerro MA, Uriz MJ, Maldonado M, Turon X, editors. Advances in Marine Biology, Advances in Sponge Science: Physiology, Chemical and Microbial Diversity, Biotechnology. Academic Press. pp. 1–56. doi:10.1016/B978-0-12-394283-8.00001-1

Li R, Neundorf I, Nitsche F. 2018. First Efficient Transfection in Choanoflagellates using Cell-Penetrating Peptides. *bioRxiv* 260190. doi:10.1101/260190

López-Escardó D, Grau-Bové X, Guillaumet-Adkins A, Gut M, Sieracki ME, Ruiz-Trillo I. 2019. Reconstruction of protein domain evolution using single-cell amplified genomes of uncultured choanoflagellates sheds light on the origin of animals. *Philos Trans R Soc B Biol Sci* **374**:20190088. doi:10.1098/rstb.2019.0088

Meilhoc E, Masson JM, Teissié J. 1990. High efficiency transformation of intact yeast cells by electric field pulses. *Biotechnol Nat Publ Co* **8**:223–227.

Metchnikoff E. 1886. Embryologische Studien an Medusen : Ein Beitrag Zur Genealogie Der Primitiv-Organe. Vienna: A. Hölder.

Parfrey LW, Lahr DJG, Knoll AH, Katz LA. 2011. Estimating the timing of early eukaryotic diversification with multigene molecular clocks. *Proc Natl Acad Sci U S A* **108**:13624–13629. doi:10.1073/pnas.1110633108

Parra-Acero H, Ros-Rocher N, Perez-Posada A, Kożyczkowska A, Sánchez-Pons N, Nakata A, Suga H, Najle SR, Ruiz-Trillo I. 2018. Transfection of *Capsaspora owczarzaki*, a close unicellular relative of animals. *Dev Camb Engl* **145**. doi:10.1242/dev.162107

Philippe H, Snell EA, Bapteste E, Lopez P, Holland PWH, Casane D. 2004. Phylogenomics of Eukaryotes: Impact of Missing Data on Large Alignments. *Mol Biol Evol* **21**:1740–1752. doi:10.1093/molbev/msh182

Richter DJ, Fozouni P, Eisen MB, King N. 2018. Gene family innovation, conservation and loss on the animal stem lineage. *eLife* **7**:e34226. doi:10.7554/eLife.34226

Rouet P, Smih F, Jasin M. 1994. Introduction of double-strand breaks into the genome of mouse cells by expression of a rare-cutting endonuclease. *Mol Cell Biol* **14**:8096–8106. doi:10.1128/MCB.14.12.8096

Ruiz-Trillo I, Burger G, Holland PWH, King N, Lang BF, Roger AJ, Gray MW. 2007. The origins of multicellularity: a multi-taxon genome initiative. *Trends Genet TIG* **23**:113–118. doi:10.1016/j.tig.2007.01.005

Ruiz-Trillo I, Roger AJ, Burger G, Gray MW, Lang BF. 2008. A Phylogenomic Investigation into the Origin of Metazoa. *Mol Biol Evol* **25**:664–672. doi:10.1093/molbev/msn006

Shimogawara K, Fujiwara S, Grossman A, Usuda H. 1998. High-efficiency transformation of *Chlamydomonas reinhardtii* by electroporation. *Genetics* **148**:1821–1828.

Snell EA, Brooke NM, Taylor WR, Casane D, Philippe H, Holland PWH. 2006. An unusual choanoflagellate protein released by Hedgehog autocatalytic processing. *Proc R Soc B Biol Sci* **273**:401–407. doi:10.1098/rspb.2005.3263

Stevens DR, Atteia A, Franzén LG, Purton S. 2001. Cycloheximide resistance conferred by novel mutations in ribosomal protein L41 of *Chlamydomonas reinhardtii*. *Mol Gen Genet MGG* **264**:790–795.

Stewart MP, Langer R, Jensen KF. 2018. Intracellular Delivery by Membrane Disruption: Mechanisms, Strategies, and Concepts. *Chem Rev* **118**:7409–7531. doi:10.1021/acs.chemrev.7b00678

Suga H, Ruiz-Trillo I. 2013. Development of ichthyosporeans sheds light on the origin of metazoan multicellularity. *Dev Biol* **377**:284–292. doi:10.1016/j.ydbio.2013.01.009

Thompson JR, Register E, Curotto J, Kurtz M, Kelly R. 1998. An improved protocol for the preparation of yeast cells for transformation by electroporation. *Yeast* **14**:565–571. doi:10.1002/(SICI)1097-0061(19980430)14:6<565::AID-YEA251>3.0.CO;2-B

Vinayak S, Pawlowic MC, Sateriale A, Brooks CF, Studstill CJ, Bar-Peled Y, Cipriano MJ, Striepen B. 2015. Genetic modification of the diarrhoeal pathogen *Cryptosporidium parvum*. *Nature* **523**:477–480. doi:10.1038/nature14651

Wainright PO, Hinkle G, Sogin ML, Stickel SK. 1993. Monophyletic Origins of the Metazoa: an Evolutionary Link with Fungi. *Science* **260**:340–342.

Wetzel LA, Levin TC, Hulett RE, Chan D, King GA, Aldayafleh R, Booth DS, Sigg MA, King N. 2018. Predicted glycosyltransferases promote development and prevent spurious cell clumping in the choanoflagellate *S. rosetta*. *eLife* **7**. doi:10.7554/eLife.41482

Woznica A, Cantley AM, 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. doi:10.1073/pnas.1605015113

Woznica A, Gerdt JP, Hulett RE, Clardy J, King N. 2017. Mating in the Closest Living Relatives of Animals Is Induced by a Bacterial Chondroitinase. *Cell* **170**:1175-1183.e11. doi:10.1016/j.cell.2017.08.005

Woznica A, Kumar A, Sturge CR, Xing C, King N, Pfeiffer JK. 2021. STING mediates immune responses in a unicellular choanoflagellate. *bioRxiv* 2021.05.13.443778. doi:10.1101/2021.05.13.443778

Yeh CD, Richardson CD, Corn JE. 2019. Advances in genome editing through control of DNA repair pathways. *Nat Cell Biol* **21**:1468–1478. doi:10.1038/s41556-019-0425-z